Opole University of Technology

Faculty of Electrical Engineering, Automatic Control and Informatics

Automatic Diagnosis of the Patient's Knee Joint Using Selected Methods of Vibroarthrographic Signal Analysis

Author: mgr inż. Adam Łysiak

Supervisor: Dr hab. inż. Mirosław Szmajda
Assistant supervisor: Dr hab. Dawid Bączkowicz

Opole, September 2023

Streszczenie

Wpływ przeciążeń oraz siedzącego trybu życia na stan stawów kolanowych jest związany z zależnością struktury stawu od sił działających na jego powierzchnie oraz od regularnego ruchu. Zarówno nadmierne obciążanie powierzchni stawowych, jak i brak aktywności fizycznej może prowadzić do degradacji chrząstki i redukcji ilości i jakości płynu stawowego. Małe uszkodzenia niepowodujące bólu mogą z czasem przyczyniać się do poważniejszych zaburzeń stawu kolanowego. Wibroartografia (VAG) jest metodą obrazowania jakości funkcjonowania stawu poprzez pomiar wibracji generowanych podczas ruchu, tworząc sygnał VAG. Pomimo znacznej liczby badań dotyczących przetwarzania sygnału VAG, w literaturze nadal brakuje zgody co do niektórych metod przetwarzania i cech ekstrahowanych z sygnału VAG.

W niniejszej dysertacji przedstawione zostały badania sygnałów VAG zarejestrowanych ze stawów kolanowych należących do pięciu klas: trzech stopni chondromalacji rzepki, choroby zwyrodnieniowej oraz grupy kontrolnej zawierającej zdrowe stawy kolanowe. W tezie pracy zawarto, że metody cyfrowego przetwarzania sygnałów w kontekście analizy sygnałów VAG, szczególnie w dziedzinie czasu, częstotliwości oraz czasowo-częstotliwościowej, umożliwią ekstrakcję cech prowadzących do wyższej niż obecnie (0.69) dokładności klasyfikacji. Teza została udowodniona poprzez szczegółową analizę wyodrębnionych cech sygnału, optymalizację cech parametrycznych oraz wybór najbardziej informatywnego zestawu cech. Najdokładniejszym klasyfikatorem okazała się Liniowa Maszyna Wektorów Nośnych uczona z wykorzystaniem 110 cech, osiągając dokładność klasyfikacji na poziomie 0.80.

Ograniczenia badań obejmują zastosowanie stosunkowo prostych metod przetwarzania wstępnego, potencjalnie ograniczając informatywność sygnału. Ponadto, informatywne cechy zostały wybrane na podstawie ich *średniej* możliwości różnicowania par klas. Wykorzystanie informacji o konkretnych parach klas mogłoby poprawić wyniki klasyfikacji. Ograniczeniem jest także ekstrakcja cech czasowo-częstotliwościowych przeprowadzona na całym spektrum. Ograniczenie analiz tylko do informatywnych zakresów częstotliwości mogłoby poprawić wyniki. Wreszcie, brak informacji o fazie ruchu stawu kolanowego zsynchronizowanych z sygnałem VAG może być uznany za ograniczenie, ponieważ w poprzednich badaniach, segmentacja sygnału prowadziła do podniesienia informatywności ekstrahowanych cech.

Z przeprowadzonych badań wyciągnięte zostały następujące wnioski. Cechy wyekstrahowane z dziedziny czasu sygnału filtrowanego, zazwyczaj są bardziej informatywne niż te obliczone na sygnale surowym lub pochodnej sygnału. Najbardziej informatywny zakres częstotliwości to ok. 10 Hz do 350 Hz. Algorytm Maximum Relevance Minimum Redundancy dostarcza bardziej informatywnego zestawu cech, w porównaniu z Analizą Głównych Składowych. Postępująca degeneracja stawu kolanowego zwykle prowadzi do wzrostu mocy, zmienności i złożoności sygnału VAG, oraz spadku cech mierzących efekt długiej pamięci. W przyszłych badaniach wartościowe byłoby rozważenie niższych częstotliwości próbkowania, rozszerzenie sygnału informacjami o fazie ruchu i przeprowadzania analizy czasowo-częstotliwościowej w ściśle określonym zakresie częstotliwości.

Słowa kluczowe: wibroartrogerafia, VAG, przetwarzanie sygnału, ekstrakcja cech, analiza czasowo-częstotliwościowa, system ekspercki, klasyfikacja, automatyczna diagnostyka.

Abstract

The impact of overload and a sedentary lifestyle on the health of knee joints can be linked to the joint structure's reliance on the forces acting on its surface and on regular movement. Both excessive loading of the joint and absence of physical activity can lead to the cartilage degradation and reduction in the amount and quality of synovial fluid. Minor damages to the cartilage that may not cause pain, can contribute to more severe knee joint conditions. Vibroarthrography (VAG), an imaging method, assesses joint function quality by measuring vibrations generated during motion, producing a VAG signal. However, despite significant body of VAG-related studies, there is still a lack of consensus regarding some of the specific methods and features extracted from the VAG signal.

This dissertation presents an extensive examination of VAG signals derived from five classes of knee joint conditions. These include three Chondromalacia Patellae stages, one Osteoarthritis class, and a control group consisting of healthy knee joints. The thesis of this work proposes that the digital signal processing methods, in the context of VAG signal analysis, specifically in the time domain, frequency domain, and time-frequency domain, will enable extraction of features that lead to a classification accuracy higher than the current state-of-the-art (0.69). It was achieved through a thorough analysis of extracted signal features, optimization of parametric features, and selection of the most informative feature set. Ultimately, the most accurate classifier proved to be Linear Support Vector Machine trained on 110 features, achieving an accuracy of 0.80.

The study's limitations include the use of relatively simple preprocessing methods, which might have limited the extracted signal's information. Moreover, informative features were selected based on their *average* class-pairs differentiation potential. Utilizing information about specific class pairs could improve the classification results. Another limitation is that the time-frequency feature selection was conducted on the whole spectrum. Focusing the analyzes on the informative frequency ranges only could potentially further improve the results. Finally, the lack of explicit information about the knee joint movement phase synchronized with the VAG signal can be considered a limitation, as segmentation of the signal has been shown to result in more informative features in previous studies.

Following conclusions could be derived from the conducted research. Time domain features, calculated on a filtered signal, generally are more informative than the ones calculated on raw or derivative signals. In the frequency domain, the most informative range ranges from 10 Hz to 350 Hz. The Maximum Relevance Minimum Redundancy algorithm provides more informative feature set, as compared to the Principal Component Analysis. The progression of knee joint conditions typically results in an increase in the VAG signal's power, variability, and complexity measures, along with a decline in features measuring long-range dependencies. For future studies, it would be beneficial to consider lower sampling frequencies, augmenting the signal with movement phase information, and conducting time-frequency analysis on a specific frequency range.

Keywords: vibrorathrography, VAG, signal processing, feature extraction, time-frequency analysis, expert system, classification, automatic diagnosis.

Contents

Streszczenie iii

Abstract iv

List of Tables ix

List of Figures x

List of Acronyms xiv

I Introduction 1

1 Introduction 2
1.1 Research motivation and justification . . . 2
1.2 Thesis and goals of the work . . . 3
1.3 Structure of the work . . . 3

2 Description of the vibroarthrogram 6
2.1 Knee anatomy and physiology . . . 6
2.1.1 Overview of the Knee Joint . . . 6
2.1.2 Forces Within the Knee Joint . . . 7
2.2 Knee joint conditions classified in this work . . . 7
2.2.1 Chondromalacia Patellae . . . 8
2.2.2 Osteoarthritis . . . 8
2.2.3 Comparison of Chondromalacia Patellae and Osteoarthritis . . . 8
2.3 Diagnosis methods and imaging techniques . . . 9
2.4 Vibroarthrography . . . 10
2.5 Vibroarthrograms studied in this work . . . 10

3 Review of the vibroarthrography processing chain 12
3.1 Signal processing steps used in automatic diagnostics . . . 12
3.2 Introduction to vibroarthrography . . . 14
3.2.1 Knee joint conditions studied in the literature . . . 15
3.3 Signal acquisition . . . 19
3.3.1 Sensor types and frequency ranges . . . 19
3.3.2 Sensor placement on the knee joint . . . 21
3.3.3 Experimental procedures and kinetic chain configurations . . . 23
3.4 Signal preprocessing . . . 26
3.4.1 Filters . . . 26
3.4.2 Reference denoising . . . 29

3.4.3 Decomposition methods . . . 29
3.4.4 Normalization . . . 31
3.4.5 Segmentation . . . 31
3.4.6 Windowing . . . 32
3.5 Feature extraction . . . 33
3.6 Feature selection and dimensionality reduction . . . 34
3.7 Classification . . . 36

II Methods 40

4 Preprocessing 41
4.1 Introduction . . . 41
4.2 Normality and stationarity of a signal . . . 41
4.2.1 Normality . . . 41
4.2.2 Stationarity . . . 43
4.3 Description of the preprocessing methods . . . 44
4.3.1 Butterworth filter . . . 44
4.3.2 Differentiation . . . 46
4.4 Comparison of the preprocessing methods . . . 46

5 Signal analysis in time domain 48
5.1 Statistical features . . . 48
5.1.1 Basic statistical features . . . 48
5.1.2 Complex statistical features . . . 54
5.2 Rolling features . . . 54
5.3 Difference and frequency features . . . 56
5.4 Self-similarity features . . . 57
5.4.1 Autocorrelation function features . . . 58
5.4.2 Detrended Fluctuation Analysis . . . 59
5.5 Phase Space Reconstruction features . . . 62
5.5.1 Phase space complexity measures . . . 64
5.5.2 Recurrence plots . . . 64
5.5.3 Entropy measures . . . 66

6 Signal analysis in frequency domain 69
6.1 Power Spectral Density estimation . . . 69
6.2 Frequency Range Maps . . . 72
6.3 Scalar features used to calculate FRMs . . . 73
6.4 Other frequency features . . . 77

7 Time-frequency analysis 79
7.1 The Spectrogram . . . 79
7.2 Spectral Fluctuation Signals . . . 81
7.3 Time analysis of the Spectral Fluctuation Signals . . . 82

8 Dimensionality reduction and classification 85
8.1 Feature evaluation . . . 85
8.2 Dimensionality reduction . . . 87
8.2.1 Summary of all extracted features and general dimensionality reduction approach . . . 87
8.2.2 The Maximum Relevance Minimum Redundancy algorithm . . . 87
8.2.3 The Principal Component Analysis method . . . 89
8.3 Classification algorithms . . . 90

III Results, Discussion and Conclusions 93

9 Preprocessing 94
9.1 Comparison of the preprocessing methods . . . 94
9.2 Summary . . . 97
9.2.1 Preprocessing methods limitations . . . 97
9.2.2 Summary and conclusions . . . 97

10 Time domain features 98
10.1 Statistical features . . . 98
10.2 Rolling features . . . 107
10.3 Difference and frequency features . . . 108
10.4 Self-similarity features . . . 111
10.5 Phase Space Reconstruction features . . . 115
10.6 Correlation of the features . . . 121
10.7 Summary . . . 130
10.8 Conclusions . . . 132

11 Frequency domain features 133
11.1 Estimation of the Power Spectral Density . . . 133
11.2 Frequency Range Maps . . . 134
11.3 Other frequency features . . . 150
11.4 Correlations of the features . . . 153
11.5 Summary . . . 155
11.6 Conclusions . . . 157

12 Time-frequency domain features 158
12.1 Frequency perspective—Spectral Fluctuation Signals . . . 158
12.2 Temporal perspective—Time features . . . 163
12.3 Spectrogram parameters . . . 168
12.4 Summary . . . 169
12.5 Conclusions . . . 171

13 Dimensionality reduction and classification 172
13.1 Dimensionality reduction . . . 172
13.2 Classification . . . 174

13.3 Summary and conclusions 176

14 Summary and conclusions **178**
14.1 Summary 178
14.2 Limitations 179
14.3 Conclusions 180
14.3.1 General analysis of the vibroarthrogram 181
14.3.2 Changes in the signal with progressing knee joint conditions 181
14.3.3 Directions for the future VAG-related studies 182
14.3.4 Novelty of the research 182

A Detailed results of the PSD estimation comparison **183**

B List of the most informative features **186**

List of Tables

2.1 Condition classes analyzed in the research. 10

3.1 Summary of knee joint conditions studied in the literature. 19
3.2 Summary of sensors used to acquire vibroarthrographic signals. 20
3.3 Summary of sampling frequencies used to acquire vibroarthrographic signals. 21
3.4 Summary of the sensor localizations used in the literature. 24
3.5 Summary of the movement patterns used in the literature. 25
3.6 Summary of the preprocessing methods used in vibroarthrography literature. 26
3.7 Summary of the filter frequencies used in the literature. 27
3.8 Summary of the feature extraction methods used in the literature. 33
3.9 Summary of the dimensionality reduction approaches used in the literature. . 35
3.10 Summary of the classification algorithms used in the literature. 38

5.1 Summary of the central tendency and power statistical features. 49
5.2 Summary of the statistical spread features. 51
5.3 Summary of the simple statistical shape and entropy features. 53
5.4 Summary of the complex statistical shape features. 55
5.5 Summary of the rolling features. 56
5.6 Summary of the difference and frequency features. 58
5.7 Summary of the ACF-based features. 60
5.8 Summary of the DFA-based features. 62
5.9 Summary of the RPS-based features. 65
5.10 Summary of the RQA-based features. 66
5.11 Summary of the entropy-based features. 68

6.1 Summary of the PSD distribution features. 75
6.2 Summary of the PSD power features. 76
6.3 Summary of the PSD shape features. 77
6.4 Summary of the frequency features not following the FRM methodology. . . 78

7.1 A review of window sizes used in the literature. 81
7.2 Summary of the frequency features used for SFS calculation. 83
7.3 Summary of the time features extracted from each SFS. 84

11.1 Average rankings of the PSD estimation methods. 133

13.1 Summary of the first dimensionality reduction step. 173

A.1 Detailed results of the PSD estimation method for the raw signal. 183
A.2 Detailed results of the PSD estimation method for the filtered signal. 184
A.3 Detailed results of the PSD estimation method for the derivative signal. . . . 185

B.1 List of the most informative features selected by the MRMR algorithm. . . . 186

List of Figures

2.1 Exemplary sensor placement. 11
2.2 Exemplary VAG signals. 11

3.1 General pipeline used in most signal classification contexts. 12
3.2 Sensors placement used across the VAG literature. 22
3.3 Filtering bands used in the literature. 28

4.1 Frequency response of the Butterworth filter. 45
4.2 Frequency response of the difference operator. 46
4.3 Exemplary vibroarthrogram after preprocessing. 47

5.1 Exemplary percentile ranges and singular percentiles plots. 50
5.2 Exemplary reconstructions of the phase space. 63
5.3 Exemplary distance matrices and corresponding Recurrence Plots. 65

6.1 Periodogram of the exemplary vibroarthrogram. 70
6.2 Comparison of the Welch's method to the periodogram. 71
6.3 Comparison of the Multitaper method to the periodogram. 71
6.4 Exemplary Frequency Range Map. 72

7.1 Exemplary spectrograms with varying window sizes. 81
7.2 Exemplary time-frequency feature extraction. 82

8.1 Visualization of the $coef_B$ 86
8.2 Visualization of the two-stage dimensionality reduction approach. 88
8.3 Visualization of the exemplary MRMR score. 89

9.1 QQ plots and histograms for each preprocessing. 95
9.2 Boxplots of the VAG signals' kurtosis and skewness. 95
9.3 Exemplary normalized moving average value and variance. 96
9.4 Boxplots of the CVNMA and CVNMV coefficients. 97

10.1 Central measures informativeness. 98
10.2 Boxplot of the RAV feature calculated for the filtered signal. 99
10.3 Spread measures informativeness. 99
10.4 Boxplot of the MeanAD feature. 100
10.5 Detailed informativeness results of the percentiles. 101
10.6 Boxplot of the 2-nd percentile feature. 101
10.7 Boxplot of the 98-th percentile feature. 101
10.8 Boxplot of the 10-th to 40-th Interpercentile Range feature. 102
10.9 Simple statistical shape measures' informativeness. 103
10.10 Detailed informativeness results of the Spiky Indices. 103
10.11 Detailed informativeness results of the entropy estimation features. 104

10.12 Boxplot of the k-nn Entropy feature. 105
10.13 Complex statistical shape measures' informativeness. 106
10.14 Rolling signal measures' informativeness. 107
10.15 Detailed informativeness results of the rolling features. 108
10.16 Difference and time-domain frequency measures' informativeness. 109
10.17 Boxplot of the MoAD feature calculated for the filtered signal. 109
10.18 Detailed results of the ZCR informativeness. 110
10.19 Detailed results of the TCR informativeness. 110
10.20 ACF measures' informativeness. 111
10.21 Multifractal DFA measures' informativeness. 112
10.22 Detailed results of the Multifractal DFA informativeness. 113
10.23 Boxplot of the Hurst Exponent feature. 115
10.24 RPS measures' informativeness. 115
10.25 Boxplot of the LLE feature. 116
10.26 RQA measures' informativeness. 117
10.27 Boxplot of the Maximum Distance in the RPS feature. 118
10.28 Entropy measures' informativeness. 118
10.29 Detailed informativeness results of the ApEn and the SampEn with fixed r. 119
10.30 Spread and standard deviation of the signals after each preprocessing. . . 119
10.31 Informativeness results of the ApEn and the SampEn with varying r. . . . 120
10.32 Boxplot of the ApEn feature. 121
10.33 Time domain features' correlation for the raw signal. 122
10.34 Time domain features' correlation for the filtered signal. 123
10.35 Time domain features' correlation for the derivative signal. 124
10.36 Time domain features' correlation for the raw and filtered signals. 127
10.37 Time domain features' correlation for the raw and derivative signals. . . . 129
10.38 Time domain features' correlation for the derivative and filtered signals. . 130

11.1 Informativeness of the FRM distribution features. 134
11.2 FRM of the F_{max} feature. 135
11.3 FRM of the PVn feature. 135
11.4 FRM of the PVr feature. 135
11.5 FRM of the PVs (Crest) feature. 136
11.6 Boxplot of the Peak Value (non-normalized) (PVn) PSD feature. 136
11.7 FRM of the Spectral Median feature. 137
11.8 FRM of the Spectral Centroid feature. 137
11.9 Boxplot of the Spectral Median feature. 137
11.10 Boxplot of the Spectral Centroid feature. 138
11.11 FRM of the Spectral Spread feature. 138
11.12 Boxplot of the Spectral Spread feature. 138
11.13 FRM of the Spectral Skewness feature. 139
11.14 Boxplot of the Spectral Skewness feature. 139
11.15 FRM of the Spectral Kurtosis feature. 140
11.16 Informativeness of the Frequency Range Map power features. 140
11.17 FRM of the Average Power (non-normalized) feature. 141

11.18 Boxplot of the Pn PSD feature. . . . 141
11.19 FRM of the Pr feature. . . . 142
11.20 FRM of the Gn feature. . . . 142
11.21 FRM of the Gr feature. . . . 143
11.22 FRM of the Gs (Flatness) feature. . . . 143
11.23 Boxplot of the Gn PSD feature. . . . 144
11.24 FRM of the RMSn feature. . . . 144
11.25 FRM of the RMSr feature. . . . 145
11.26 FRM of the RMSs feature. . . . 145
11.27 Informativeness of the FRM shape features. . . . 146
11.28 FRM of the Spectral Slope feature. . . . 146
11.29 Boxplot of the Spectral Slope feature. . . . 147
11.30 FRM of the Spectral Decrease feature. . . . 147
11.31 FRM of the Hr feature. . . . 147
11.32 Boxplot of the Hr feature. . . . 148
11.33 FRM of the Hs feature. . . . 148
11.34 Boxplot of the Hs feature. . . . 149
11.35 FRM of the II feature. . . . 149
11.36 Boxplot of the Irregularity Index feature. . . . 150
11.37 Informativeness of the other frequency features. . . . 150
11.38 Detailed results of the frequency ratios' informativeness. . . . 151
11.39 Boxplot of the High-to-Low ratio feature. . . . 151
11.40 Detailed results of the frequency Roll-off Points informativeness. . . . 151
11.41 Boxplot of the Roll-off Point feature calculated for the raw signal. . . . 152
11.42 Boxplot of the Roll-off Point feature calculated for the filtered signal. . . . 152
11.43 Frequency domain features' correlation for the raw signal. . . . 153
11.44 Average informativeness of each frequency bin. . . . 156
11.45 Number of best FRM points for each frequency bin. . . . 157

12.1 Informativeness of time features of the PSD distribution SFSs. . . . 158
12.2 Informativeness of time features of the PVn SFS. . . . 159
12.3 Informativeness of the PVn SFS for varying spectrogram parameters. . . . 159
12.4 Boxplot of the Median feature extracted from the PVn SFS. . . . 160
12.5 Boxplot of the MoAD feature extracted from the PVn SFS. . . . 160
12.6 Informativeness of time features of the PSD power SFSs. . . . 161
12.7 Informativeness of time features extracted from the Pn SFS. . . . 161
12.8 Informativeness of the Pn SFS, for varying spectrogram parameters. . . . 162
12.9 Boxplot of the Median feature extracted from the Pn SFS. . . . 162
12.10 Boxplot of the MoAD feature extracted from the Pn SFS. . . . 163
12.11 Informativeness of time features of the PSD shape SFSs. . . . 163
12.12 Informativeness of time features of the Spectral Slope SFS. . . . 164
12.13 Informativeness of Spectral Slope SFS, for varying spectrogram parameters. 164
12.14 Boxplot of the Median feature extracted from the Spectral Slope SFS. . . . 164
12.15 Boxplot of the MoAD feature extracted from the Spectral Slope SFS. . . . 165
12.16 Informativeness of time domain distribution features extracted from all SFSs. 165

12.17 Informativeness of the median feature extracted from specific SFSs. 166
12.18 Informativeness of time domain shape features extracted from all SFSs. . 166
12.19 Informativeness of time domain temporal features extracted from all SFSs. 167
12.20 Informativeness of the MoAD feature extracted from specific SFSs. 167
12.21 Informativeness of the MoAD feature for varying spectrogram parameters. 168
12.22 Average informativeness of SFSs, for varying spectrogram parameters. . . 169
12.23 Average informativeness of time features, for varying spectrogram parameters. 170

13.1 Informativeness of features obtained from the MRMR and the PCA. . . . 174
13.2 Average absolute correlation of each feature selected by the MRMR algorithm. 174
13.3 Classification results for each classifier. 175

List of Acronyms

μ_f	Spectral Centroid
σ_f	Spectral Spread
$decrease_f$	Spectral Decrease
F_{max}	Frequency of the most prominent power
$kurt_f$	Spectral Kurtosis
$skew_f$	Spectral Skewness
$slope_f$	Spectral Slope
$coef_B$	Binary Bhattacharyya coefficient
$coef_{BMC}$	Multiclass Bhattacharyya coefficient
AACF	Average Autocorrelation Function
ACF	Autocorrelation Function
ACFTCR	Autocorrelation Function Turns Count Rate
ACFZCR	Autocorrelation Function Zero Crossing Rate
ACL	Anterior Cruciate Ligament
ADC	Analog to Digital Converter
ADL	Average Diagonal Line
ApEn	Approximate Entropy
ASP	Average Signal Power
AUC	Area Under the ROC Curve
CD	Correlation Dimension
CF	Crest Factor
CKC	Closed Kinetic Chain
CMP	Chondromalacia Pattellae
cmp1	Stage I Chondromalacia Patellae
cmp2	Stage II Chondromalacia Patellae
cmp3	Stage III Chondromalacia Patellae
CT	Computed Tomography
ctrl	Control Group
CV	Coefficient of Variation
CVNMA	Coefficient of Variation of the Normalized Moving Average
CVNMV	Coefficient of Variation of the Normalized Moving Variance
DET	Determinism
DFA	Detrended Fluctuation Analysis
DFT	Discrete Fourier Transform
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
FF	Form Factor
FRM	Frequency Range Map
FZCT	First Zero Crossing Time
Hr	Spectral Entropy (relatively-normalized)

(Continued)

Hs	Spectral Entropy (self-normalized)
IF	Impulse Factor
II	Irregularity Index
IMF	Intrinsic Mode Function
IPerR	Interpercentile Range
KNN	k-Nearest Neighbors
LAM	Laminarity
LCL	Lateral Collateral Ligament
LDA	Linear Discriminant Analysis
LEE	Log Energy Entropy
LLE	Largest Lyapunov Exponent
MCL	Medial Collateral Ligament
MCV	Modified Coefficient of Variation
MeanAD	Mean Absolute Deviation
MedF	Median Frequency
MedianAD	Median Absolute Deviation
MF	Margin Factor
MoAD	Mean of Absolute Differences
MRI	Magnetic Resonance Imaging
MRMR	Maximum Relevance Minimum Redundancy
MSACF	Mean Squared Autocorrelation Function
NN	Neural Network
oa	Osteoarthritis
OKC	Open Kinetic Chain
PCA	Principal Component Analysis
PCL	Posterior Cruciate Ligament
PSD	Power Spectral Density
PVn	Peak Value (non-normalized)
PVr	Peak Value (relatively-normalized)
PVs	Peak Value (self-normalized)
QQ	Quantile-Quantile
RAACF	Rectified Average Autocorrelation Function
RAV	Rectified Average
RCV	Rectified Coefficient of Variation
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
RP	Recurrence Plot
RPS	Reconstructed Phase Space
RQA	Recurrence Quantification Analysis
RR	Recurrence Rate
RSI	Rectified Spiky Index
SampEn	Sample Entropy
SFS	Spectral Fluctuation Signal

(Continued)

SI	Spiky Index
std	Standard Deviation
SVM	Support Vector Machine
TCR	Turns Count Rate
VAG	Vibroarthrography
VMD	Variational Mode Decomposition
VMS	Variance of the Mean Squared
ZCR	Zero Crossing Rate

Part I

Introduction

Chapter 1

Introduction

In this chapter, a broad introduction to the whole dissertation will be included, with the motivation and justification of the research, specific goals and thesis of this work, as well as its structure.

1.1 Research motivation and justification

A sedentary lifestyle, being one of the most prevalent issues in the current times, significantly impacts the human knee joint health [1]. This is caused by the joint's structure and reliance on regular movement. The knee, a predominantly hinge joint, consists of hyaline cartilage and bone, largely dependent on a regular load for nutrient diffusion and metabolite removal. The absence of physical activity due to a sedentary lifestyle leads to a lack of regular load on the joint, leading to cartilage degradation. Additionally, inactivity reduces the production and circulation of synovial fluid, a critical joint lubricant, causing joint stiffening [2].

One unique characteristic of the hyaline cartilage is its lack of nerves [2]. Due to this, minor injuries or wear-and-tear to the cartilage usually do not cause any immediate pain or discomfort, making them easy to overlook in their initial stages. Over time, however, these minor injuries can compound, leading to cartilage thinning or wear, which can trigger more significant knee joint conditions.

Traditional imaging techniques, like the x-rays [3] or magnetic resonance imaging [4] enable assessment of the joint's *structure*, providing an image of the static joint. On the other hand, Vibroarthrography (VAG) is an imaging technique, that quantifies quality of the joint's *function*, by registering vibrations generated by the knee joint in motion. The examination provides a signal of those vibrations, called the vibroarthrogram. There is a significant body of work focusing on VAG processing methods, usually in the screening context, i.e., differentiation between healthy and (somehow) damaged knee joint. However, there is still a lack of consensus in the literature about specific processing techniques and exact features extracted from the VAG signal.

Therefore, in the current work, a plethora of VAG signal features were extracted. Moreover, their informativeness was thoroughly analyzed and interpreted. Those analyzes led to selection of specific features that allowed classification results better than current state-of-the-art. A number of important conclusions were derived, including ones related to the general analysis of the VAG signal, changes in the signal with progressing degeneration of knee joint, and potentially fruitful directions for the future VAG-related research.

1.2 Thesis and goals of the work

The thesis of the work is as follows. The application of digital signal processing methods in vibroarthrographic signal analysis, specifically in the time domain, frequency domain, and time-frequency domain, will enable extraction of features allowing the classification of knee joint conditions with accuracy higher than the current state-of-the-art (0.69).

Relating to the thesis, the following main goals were set.

1. Analysis of signal features used in VAG field and identification of parametric features, which parameters can be optimized in terms of informativeness.
2. Feature extraction from the signal in the time domain.
3. Feature extraction from the signal in the frequency domain.
4. Feature extraction from the signal in the time-frequency domain.
5. Selection of the feature informativeness' measure.
6. Selection of the most informative feature set.
7. Selection of a machine learning model allowing for classification of the extracted features with accuracy higher than 0.69, achieved by Kręcisz and Bączkowicz [5].

In addition, the following minor goals were defined.

1. Comparison of selected preprocessing methods in the context of feature informativeness.
2. Analysis and interpretation of the obtained features' informativeness.
3. For parametric features, analysis of the impact of feature parameters on their informativeness.
4. Comparative analysis of spectrum estimation methods in the context of feature informativeness.
5. Comparative analysis of spectrogram parameters in the context of feature informativeness.
6. Comparative analysis of selected dimensionality reduction methods in the context of classification accuracy.
7. Comparative analysis of machine learning models in the context of classification accuracy.

1.3 Structure of the work

The work is divided into three parts. The first part constitutes an introduction to the study and vibroarthrography. The first chapter, i.e., the current one, contains an introduction to the work. Chapter two describes a vibroarthrographic signal, or a vibroarthrogram. The chapter begins with a brief description of the knee joint. Next, the conditions analyzed in this study are described, along with the potential for diagnosis using traditional methods and vibroarthrography. The chapter ends with a brief description of the specific signal base used in this study.

The third chapter is a review of methods used in the VAG field, at every stage of a signal processing. The chapter begins with a brief description of the typical signal processing pipeline. Then, an introduction to vibroarthrography is presented, in a literature context. Following sections focus on signal acquisition, description of various sensors used, methods

of preprocessing, feature extraction, dimensionality reduction, and, finally, classification. The section on feature extraction is written rather concisely, because major part of this work is focused on a feature extraction, and specific methods are described in more detail in respective chapters. The third chapter concludes the first part of the work, i.e., the introduction.

The second part of the study includes a detailed description of the methods used in the work. It begins with Chapter 4, devoted to preprocessing. The chapter contains a brief description of normality and stationarity, followed by a description and justification of the methods used, namely the Butterworth filter and differentiation. The chapter concludes with a comparison of selected methods in the context of the work.

Chapter 5 contains a description of the signal features extracted from the time domain. It begins with a section devoted to basic statistical methods. Next, rolling, differentiation and time domain frequency features are described. In the next section, self-similarity measures, i.e., ones based on the Autocorrelation Function and the Detrended Fluctuation Analysis, are described. The last section delves into the Phase Space Reconstruction and measures defined on its basis, such as the complexity features, Recurrence Quantification Analysis, and selected parametric entropy measures, i.e., the Approximate and Sample Entropy. For many parametric features, an in-depth analysis of the impact of parameters on feature informativeness in the context of classification is conducted.

The sixth chapter contains a description of frequency analysis. The first section is an introduction to Power Spectral Density estimation, a description of several most popular methods of the estimation, and the methodology implemented to compare these methods in the context of feature extraction. The next section introduces the Frequency Range Map, which is an intuitive way of visualizing informativeness of individual frequency ranges, measured with various features. The last section describes frequency features not based on the Frequency Range Maps, and also includes a description of the parameters of these features.

Chapter 7 contains a description of the time-frequency analysis. The first section describes the spectrogram, its parameters, and the methodology for comparing these parameters in the context of feature extraction. The next chapter describes the Spectral Fluctuation Signals, which are generated by reducing one dimension of the spectrogram. The chapter ends with a description of time features, which are calculated from each Spectral Fluctuation Signal.

In chapter eight, methods for reducing dimensionality of the features obtained in previous chapters and classifying signals based on reduced-dimensionality features are discussed. The first section introduces the method of feature quality evaluation, or, in other words, method of quantifying its informativeness in the context of classification. The second section provides a summary of all features obtained in previous chapters, as well as two methods for reducing dimensionality of these features: the Maximum Relevance Minimum Redundancy algorithm and the Principal Component Analysis. The last section is dedicated to the description of classification algorithms used in this study. Chapter eight is the last chapter of the second part, concerning methods.

In the third part of the dissertation, the results of analyzes corresponding to the chapters of the second part are presented and discussed. In Chapter 9, results of the preprocessing methodology are discussed. Chapter 10 includes the results of signal analysis in the time

domain. The informativeness of all features is thoroughly discussed. Parametric features are additionally analyzed in terms of the influence of parameters on the informativeness. Additionally, the most informative features are analyzed in terms of their ability to distinguish between individual class pairs analyzed in the dissertation. Next, the results of the correlation analysis of time domain features are presented and discussed. The chapter ends with a summary and specific conclusions that have been drawn from the conducted analyzes.

In Chapter 11, results of the frequency analysis are presented. In the first section, results of a comparison of selected spectrum estimation methods are included and discussed. In the next section, Frequency Range Maps are visualized and analyzed in terms of informativeness of selected frequency ranges. Then, informativeness of frequency features not following the Frequency Range Map methodology is presented. Similar to the previous chapter, the most informative features are analyzed in terms of distinguishing individual class pairs. The next section includes the results of a correlation analysis of frequency features and their interpretation. In the next section, the summary, the author attempted to answer the question of which frequency range is the most informative, and why. The final section contains specific conclusions derived from the conducted research.

In Chapter 12, the results of the time-frequency analysis are presented. The chapter begins with a section on the results of research from a frequency perspective. The informativeness of selected features is analyzed, and as in previous chapters, the most informative features are analyzed for their capability to differentiate specific class pairs. The next section presents the results from the perspective of a time domain. In the following section, the influence of spectrogram parameters on the informativeness of the features is discussed. The last two sections constitute a summary of time-frequency analyzes and conclusions derived from them.

Chapter 13 includes the results of dimensionality reduction and classification. The first section is devoted to dimensionality reduction, comparison of the Maximum Relevance Minimum Redundancy and the Principal Component Analysis methods, and the results of dimensionality reduction in this study. In the next section, the results of VAG signal classification using a reduced set of features are described. The final section contains a summary and conclusions.

The last, fourteenth chapter, includes a summary of all the conducted studies and conclusions drawn from them. The contribution of this work to the field of vibroarthrography is summarized, and the limitations of the conducted studies are discussed. Finally, potentially fruitful directions for future studies are suggested.

Chapter 2

Description of the vibroarthrogram

In this chapter, a general description of the vibroarthrogram will be provided. The chapter will begin with the anatomy and physiology of the knee. Then, the specific joint conditions analyzed in this work will be described, followed by a description of the standard methods used for their diagnosis. Finally, vibroarthrography will be briefly described, along with the specific signals analyzed in this work.

2.1 Knee anatomy and physiology

In this section, an overview of the knee joint will be provided, as well as the forces acting within it.

2.1.1 Overview of the Knee Joint

The knee joint, classified as a hinge joint, is among the most complex and vital joints in the human body, playing a significant role in weight-bearing activities, such as walking, running, and jumping [2]. Primarily composed of three bones, the femur (thighbone), the tibia (shinbone), and the patella (kneecap), the knee joint allows for flexion and extension movements, as well as minimal rotation in the flexed position [6]. The ends of the femur and tibia, along with the posterior surface of the patella, are covered with a layer of articular cartilage that helps to reduce friction and absorb shock during motion [6].

Four major ligaments stabilize and connect the bones in the knee joint: the Anterior Cruciate Ligament (ACL), the Posterior Cruciate Ligament (PCL), the Medial Collateral Ligament (MCL), and the Lateral Collateral Ligament (LCL) [2]. The ACL and PCL prevent excessive anterior and posterior translation of the tibia relative to the femur, while the MCL and LCL provide stability against forces that would push the knee inward or outward, respectively [2]. Working together, these ligaments ensure the stability and integrity of the knee joint during various movements and activities [6].

Additionally, the knee joint contains two C-shaped cartilage-like structures called menisci, which are the medial and lateral menisci [2]. These menisci provide stability to the joint and help in distributing the load across the articular surfaces of the tibia and femur, effectively reducing the contact pressure and minimizing wear on the articular cartilage. Furthermore, the menisci serve as secondary shock absorbers, providing an additional layer of protection for the joint [2].

A noteworthy characteristic of hyaline cartilage, including the articular cartilage in the knee joint, is its near-complete absence of nerves [2]. This feature has significant implications for the detection and perception of injuries in the cartilage. Due to the lack of nerve endings,

minor damage or micro-injuries to the cartilage often do not cause pain, making them largely unnoticed by the individual. This absence of pain sensations can lead to the continuous use and loading of the affected joint, potentially increasing the damage and causing the cartilage to degrade over time. As a result, knee joint issues like chondromalacia patellae and osteoarthritis may develop slowly while unnoticed, becoming apparent only when the damage has reached a more serious stage [2].

2.1.2 Forces Within the Knee Joint

The knee joint experiences various forces during movement, including compressive, tensile, and shear forces, as well as friction [2]. Compressive forces occur when the joint surfaces are pushed together, while tensile forces result from the joint surfaces being pulled apart. Shear forces, on the other hand, occur when the joint surfaces slide against each other [2]. The ligaments, muscles, and other soft tissues surrounding the knee joint play a crucial role in managing these forces by providing stability and distributing the load across the joint [6, 7].

Friction in the knee joint is minimized by the presence of cartilage and synovial fluid, which provide a smooth, low-friction surface for movement [2]. Factors affecting friction within the knee joint include joint alignment, cartilage wear, and the properties of synovial fluid, such as viscosity and lubrication [2]. Increased friction can have several consequences, including joint pain, inflammation, and degeneration, which may contribute to the development of knee joint conditions such as osteoarthritis [2, 8].

During knee joint motion, friction between sliding joint surfaces and surrounding tissues generates vibrations. These vibrations can be influenced by various factors, such as the health of the cartilage, joint biomechanics, and muscle activity. For instance, changes in cartilage properties due to wear or degradation may result in altered vibration patterns. Similarly, variations in joint biomechanics and muscle activation can also impact the generation and propagation of vibrations within the knee joint [9].

The study of these vibrations is called vibroarthrography, and it aims to noninvasively evaluate knee joint health and function by analyzing the vibration signals measured on the skin surface [9]. The measured vibration signal is a composite, created by various components such as joint surface vibrations, muscle contractions, and soft tissue vibrations [10]. Joint surface vibrations originate from the interaction between the articulating surfaces of the knee joint, while muscle contractions produce vibrations as the muscles activate and generate force. Soft tissue vibrations result from the transmission and reflection of the vibrations through the surrounding tissues, such as tendons, ligaments, and skin. By examining the characteristics of the composite vibration signal, specific signal features can be identified and associated with healthy and unhealthy knee joints [5, 10].

2.2 Knee joint conditions classified in this work

In this section, two specific conditions will be described, namely the Chondromalacia Patellae and the Osteoarthritis, as well as their respective grades and, finally, their comparison.

2.2.1 Chondromalacia Patellae

Chondromalacia Pattellae (CMP) is a condition characterized by the softening and deterioration of the cartilage on the underside of the patella [11]. The pathophysiology of this condition involves the breakdown of the hyaline cartilage, leading to increased friction and altered joint mechanics. Some factors contributing to the development of CMP include muscle weakness, abnormal wear and tear of the cartilage and patellar tracking, i.e., kneecap moving out of place in bending or straightening motion [11]. The deterioration of the patellar cartilage in CMP leads to increased joint friction due to abnormal patellofemoral joint mechanics and increased joint stress and pressure [11]. As a result, the vibrations generated by the joint are affected by changes in contact between the joint surfaces [5].

The severity of the CMP is graded the Outerbridge Classification System [12], consisting of five stages. Stage 0 denotes normal cartilage. Stage I is characterized by an intact but soft, swollen, and inflamed articular surface. In the second stage, there is a presence of fractures and fragmentation of the articular surface, with an area measuring half an inch or less in diameter. Stage III refers to localized, partial thickness cartilage damage, with an area measuring more than half an inch in diameter, and Stage IV indicates a full-thickness defect that extends down to the subchondral bone [11]. While arthroscopic evaluation of the knee is the gold standard for assessing the grade of articular cartilage wear, Magnetic Resonance Imaging (MRI) or radiography can also be used to classify the degree of cartilage wear. However, those methods are not as accurate as the visual assessment performed during arthroscopy.

2.2.2 Osteoarthritis

oa is a complex joint disease characterized by the progressive loss of articular cartilage, which can be caused by a range of factors including obesity, aging, and joint injuries. Although cartilage loss is a critical component of oa, the disease is not limited to the cartilage and involves changes in the entire joint [13].

The pathophysiology of the oa involves the degradation of articular cartilage, as well as changes in the biomechanics of the joint and inflammation. As oa progresses, joint space narrowing can occur, and bone remodeling may take place as the joint attempts to adapt to the changing conditions [13]. The impact of oa on joint friction involves changes in the properties of synovial fluid, which can become thinner and less effective at lubricating the joint. This can result in increased friction, which can lead to pain, stiffness, and reduced joint motion. Consequently, the vibrations generated by the joint are affected by altered joint loading patterns and the effects of joint instability and misalignment [10, 14, 15].

2.2.3 Comparison of Chondromalacia Patellae and Osteoarthritis

Both CMP and oa have similar impacts on joint friction, as both conditions result in cartilage degradation and increased joint stress and pressure [11, 13]. However, chondromalacia patellae primarily affects the patellar cartilage, while osteoarthritis affects articular cartilage throughout the joint [13].

The impact of both conditions on vibrations generated by the joint is also similar, as both cause changes in contact between joint surfaces and altered joint mechanics due to cartilage loss and joint instability. However, chondromalacia patellae may produce more localized vibrations due to patellofemoral joint involvement, while osteoarthritis may generate more widespread vibrations due to the involvement of multiple joint surfaces.

2.3 Diagnosis methods and imaging techniques

During the clinical examination of a patient with suspected knee joint condition, such as chondromalacia patellae or osteoarthritis, a physical examination is initially performed, involving inspection, palpation, and assessment of joint mobility [16]. Furthermore, the range of motion of the knee joint is evaluated, which provides valuable information about joint function and possible restrictions [17]. Special tests designed for knee joint evaluation can be utilized to assess the integrity of ligaments, menisci, and other joint structures, contributing to a comprehensive assessment of the joint's condition [17].

Upon completing the clinical examination, healthcare professionals may decide to employ imaging techniques to obtain a more in-depth understanding of the knee joint's structure. One such technique is X-ray imaging [3], a widely used diagnostic method for assessing bone and joint abnormalities. While it is limited in its ability to visualize soft tissues, such as cartilage, it can reveal joint space narrowing, bone spurs (osteophytes), and other bony changes indicative of osteoarthritis [18]. However, X-ray imaging is less effective in detecting chondromalacia patellae, as it mainly affects the cartilage rather than the bone.

Another imaging technique is the MRI [4]. It is an imaging technique that offers detailed visualization of soft tissues, including cartilage, ligaments, and menisci. It is particularly useful in detecting CMP, as it can reveal cartilage deterioration and other abnormalities related to this condition [18]. MRI also enables the assessment of oa by showing changes in articular cartilage and joint space narrowing, among other features [18]. The advantages of MRI include its high-resolution images and non-invasive nature.

Computed Tomography (CT) is another tool widely used in knee joint imaging [19]. The CT scans provide cross-sectional images of the knee joint, allowing the visualization of bone and some soft tissue structures. While CT scans can show bone changes related to osteoarthritis, such as joint space narrowing and osteophytes, they are less effective in visualizing cartilage and other soft tissues [18]. As a result, CT scans have limited utility in diagnosing chondromalacia patellae.

Another method used for imaging the knee joint is the Ultrasound Imaging [20]. It uses high-frequency sound waves to produce images of soft tissues, such as cartilage, ligaments, and tendons. While it can provide some information about the knee joint structures, its ability to visualize deeper tissues, such as cartilage, is limited in comparison to the MRI [18]. Ultrasound may help to detect joint effusion, synovial thickening, and other abnormalities related to oa, but its effectiveness in diagnosing CMP is restricted due to its limited visualization of cartilage.

It is worth noting that the mentioned imaging methods, while effective in providing valuable information about the knee joint's structure, can be time-consuming, expensive, and potentially uncomfortable for the patient. Clinical examinations and imaging techniques,

such as X-rays, MRI, CT scans, and ultrasound, require specialized equipment and professional interpretation, which can result in significant healthcare costs. Additionally, patients may experience discomfort during the examination process, particularly when undergoing MRI or CT scans, which involve enclosed spaces and potentially long examination times.

2.4 Vibroarthrography

Imaging techniques, such as X-ray, MRI, CT, and ultrasound, play a vital role in the diagnosis and assessment of knee joint conditions, as they allow for the visualization of joint structures in a static, non-moving state. These qualities enable the examination of bone and soft tissue morphology, joint space, and other structural features that can be indicative of conditions such as chondromalacia patellae and osteoarthritis. However, their inherent nature as static assessments means that they can only provide limited information about the dynamic behavior of the knee joint during movement.

In contrast, VAG is a dynamic technique that captures information during joint motion, offering insights into the real-time interactions between joint components and their biomechanical function [10]. VAG focuses on detecting and analyzing vibrations generated by the knee joint as it moves, which can provide valuable information about quality of the joint's function. The dynamic nature of VAG allows it to capture aspects of joint function that may not be clearly visible using static imaging methods.

2.5 Vibroarthrograms studied in this work

Vibroarthrograms used in this work were obtained and firstly analyzed by Kręcisz and Bączkowicz [5]. Signals were acquired from a total of 184 knee joints. Division into specific classes was included in Table 2.1. Diagnoses were done by a single radiologist blinded to the patients' symptoms, using X-ray imaging.

Table 2.1: Condition classes analyzed in the research.

signal class	signal count	description
ctrl	66	Control Group
cmp1	26	Stage I Chondromalacia Patellae
cmp2	30	Stage II Chondromalacia Patellae
cmp3	36	Stage III Chondromalacia Patellae
oa	26	Osteoarthritis

All subjects were subjected to standard medical interviews and physical examinations. Classification of patients with stage I–III CMP was carried out based on the Outerbridge Classification System [12]. Concurrently, patients diagnosed with oa exhibited mild to moderate knee oa, corresponding to grade II and III in the Kellgren-Lawrence grading system [21], with a condition duration exceeding 2 years.

In order to prevent signal artifacts stemming from conditions other than chondral lesions, the study excluded individuals with a prior history of knee surgeries, fractures or significant

instabilities, meniscal tears, or patellar maltracking. Furthermore, participants with pronounced muscle weakness, significantly swollen knees or restricted range of motion in the knee joint (within 0° to 100°), were also excluded from the study due to the methodology of the VAG assessment.

Throughout the assessment, participants were seated and directed to perform a total of four complete cycles of knee joint extension and flexion alternation, with a motion range spanning from 90° to 0° and back to 90°, all cycles lasting 6.144 seconds total. A metronome set at 82 beats per minute was employed to facilitate consistent and repeatable movements. To capture the vibration signals produced by the knee joint, an acceleration sensor was positioned 1 cm above the apex of the patella. Exemplary acceleration attachment, seen from the front and the side, was shown in Figure 2.1.

Figure 2.1: Exemplary sensor placement.

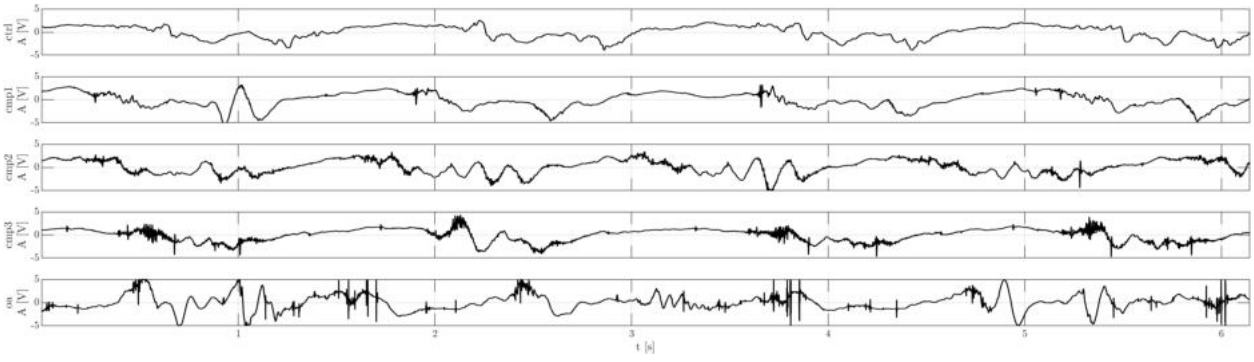

Figure 2.2: Exemplary VAG signals.

The signals analyzed in this research were obtained using the Brüel & Kjær 4513B-002 accelerometer and Nexus 2692-C signal amplifier connected to a computer using an Analog to Digital Converter (ADC). Each signal was sampled at 10 kHz and lasted exactly 6.144 seconds. Exemplary signals from all condition classes were shown in Figure 2.2.

Chapter 3

Review of the vibroarthrography processing chain

In this chapter, a VAG literature review will be provided. The first section will include a short description of a typical signal processing pipeline. In the next section, an introduction to vibroarthrography will be provided, in the literature context. Following sections will include methods used in VAG-related studies, relating to the signal acquisition, preprocessing, feature extraction, dimensionality reduction and classification.

3.1 Signal processing steps used in automatic diagnostics

The signal processing pipeline for classification tasks usually involves several separate steps, as illustrated in Figure 3.1. The first step, *signal acquisition*, includes capturing raw signals using appropriate sensors and sampling techniques. The objective of this stage is the transformation of the analog signal into a digital one, i.e., to convert continuous physical quantities into a discrete representation. This stage is vital for ensuring the accuracy and reliability of the acquired data, which forms the foundation for subsequent analysis.

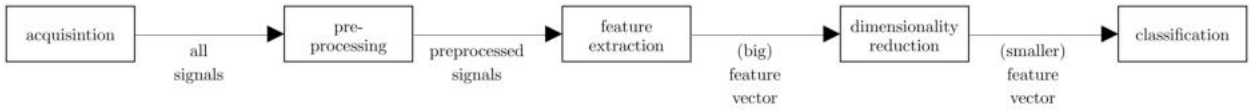

Figure 3.1: General pipeline used in most signal classification contexts.

Once the signals are acquired, *preprocessing* is carried out to eliminate noise, artifacts, and unwanted signal components. This step enhances the signal quality and enables further processing. It may include various filtering or decomposition techniques, normalization, or artifacts removal. In general, this step produces signal (or signals) in the same domain as the input signal (for example, the time domain).

Following preprocessing, *feature extraction* is used to transform the processed signals into meaningful and representative features. These features can be used to characterize and distinguish different classes in the classification task, providing insight into the underlying patterns and relationships within the data. This step can be thought of as the quantification of the intuition about the signal, providing specific values for the intuitive concepts, such as frequency content, signal energy, or statistical properties, which contribute to a more comprehensive understanding of the signal's behavior and characteristics.

For example, the average value of a signal quantifies intuition about the baseline behavior, or central tendency of a quantity in time. Standard deviation, or variance, provide

direct measure of the variability, or dispersion of signal's values around the average value. Low standard deviation value indicates that the signal values are consistently close to the average, which could be interpreted as a stable and predictable signal. Conversely, a high standard deviation might imply greater variability, indicating a signal that is more dynamic or changeable over time.

Since feature extraction step can provide virtually infinite number of more or less informative features, *dimensionality reduction* techniques, in a form of feature selection or transformation, are subsequently applied to reduce the complexity of the feature space. For instance, in a scenario where height is a feature, it can be recorded in both meters and inches. These two features are perfectly correlated, because they represent the same characteristic, but recorded in different units. In this case, keeping both features would not add new information but would unnecessarily increase the dimensionality of the feature space. Therefore, one of them could be removed, reducing the complexity of the feature set, while keeping its information content. The removed features do not need to be perfectly correlated to be removed, and the general intuition about the dimensionality reduction is to keep as much information as possible, while reducing the dimensionality of the feature space. This step aims to retain the most information in the least number of features for classification purposes, thereby improving the efficiency and performance of the subsequent classification algorithms.

The difference between feature selection and transformation is that feature selection involves selecting a subset of the original features, while feature transformation transforms the original features into a new, lower-dimensional feature space, typically using statistical or machine learning methods. The biggest advantage of the feature selection is interpretability. Informative features selected from the original feature set are not transformed in any way, which allows them to keep their interpretation, or rationale. Dimensionality reduction algorithms, on the other hand, usually deprive the features of their original interpretations. It is worth mentioning, that feature selection and transformation algorithms can also be used sequentially, i.e., first selecting the most informative features, and then transforming them into even lower dimensional representation.

Finally, the *classification* step utilizes machine learning algorithms to assign the extracted features to distinct classes. By learning from training data, these algorithms enable accurate classification or diagnosis. This step can be considered the *intelligence* of the whole signal processing pipeline, since it directly draw inferences about a signal, based on the values of its specific features. In general, quality of the classification is highly dependent on the quality of the features. However, additional factor greatly influencing the classification accuracy is the selection of the classification algorithm that can utilize specific distribution of the features.

Deep learning techniques, particularly when provided with a large number of training samples, have the potential to replace virtually all explicit steps in the signal processing pipeline following the signal acquisition. Despite their ability to effectively learn features and perform classification tasks, there are two main disadvantages associated with deep learning approaches: the need for a large number of training samples and the poor interpretability caused by the complex, nonlinear structures of the algorithms. Within these structures, the transformations and calculations are often hidden, rendering the process opaque. This lack of transparency in the model's operation frequently results in difficulty in understanding how specific transformations are made, a phenomenon commonly referred to as the "black box"

problem [22].

The structure of the following literature review mirrors the organization of the signal processing pipeline. After the next, VAG introductory section, each section of the review will focus on one of the pipeline components, discussing the relevant methods, techniques, and advancements applied in the vibroarthrography field. Since this dissertation is focused mostly on the feature extraction step, the corresponding part of the review will be considerably less in-depth than parts corresponding to the remaining steps. Specific feature extraction methods will be described in more detail in a following chapters.

3.2 Introduction to vibroarthrography

The first publication on vibroarthrography can be dated back to 1902, to "Auscultation of the Knee Joint" by Dr. William Ernest Blodgett [23]. To listen to the sounds, he used stethoscope placed on the patella. In the experimental procedure, he flexed (up to the maximum flexion) and extended (up to the maximum extension) subject's leg, without the heel touching the ground. He moved the leg of the patient himself, in order to minimize patient's use of muscles, which could interfere with the joint sounds. He recorded the data by hand, using dots, dashes and vertical lines, indicating various sounds, such as snapping, grating or squeaking.

Similar descriptive approaches were used in VAG research until 1970's, when digital signal processing started to be implemented in the field, for example, by Chu et al. [24]. Analyzed signals were acquired by microphones and the research focused mostly on the sources of knee generated noises. Microphone was usually placed at a short distance to the knee [25]. However, signal generated by the knee joint was damped by the skin and highly distorted by the ambient noises. Therefore, as a direct sensors of the vibrations generated by the knee, accelerometers were proposed by Mollan et al. [26]. They remain the main VAG acquisition device used to date.

Up to 1990's, the digital signal processing methods were implemented in a limited way. Some studies, for example Kernohan et al. [27, 28] used explicitly extracted features (such as the peak frequency or duration of the signal), however, automated signal analysis and classification was not used. Steps of the signal processing pipeline started to be automated in the 1990's, with signal preprocessing, such as noise removal in a work by Liu et al. [29], classification in a study by Zhang et al. [30] or feature extraction in a study by Tavathia et al. [31]. Those automated steps remain the focus of the literature to date.

In the literature, methods using sounds or vibrations generated by the joints can be found under a variety of names, such as joint auscultation (especially in the early years of VAG development) [23], phono arthrography [32], vibroarthrography [33], vibration arthrometry [25] or joint acoustic emissions [34]. Those naming conventions mirror historical trends, with a gradual change of microphones to accelerometers as main acquisition devices. For more in-depth analysis of VAG domain with great insights on historical developments, see critical review by Abbott and Cole [25].

3.2.1 Knee joint conditions studied in the literature

Vibroarthrography in the healthy knees joints

Some references use VAG for healthy subjects only, studying dependencies of the vibroarthrogram on some external factors. For example, Bączkowicz et al. [35] investigated the impact of knee immobilization and subsequent re-mobilization on the quality of arthrokinematics within the patellofemoral joint. They found significant differences in VAG power features between immobilized patients and controls. Furthermore, a strong correlation was observed between these features and the duration of immobilization. Although a 2-week rehabilitation program yielded some improvement, it did not fully restore the VAG power features to the level of controls, underlining the profound and lasting effects of immobilization on joint function.

In some other publications, the effects of muscle contraction and additional load on the VAG signals were described. Zhang et al. [36] found that muscle contractions result in an increased VAG signal power. Andersen et al. [37] and Kalo et al. [38] studied external load in open kinetic chain, while Bolus et al. [39] and Ołowiana et al. [40] analyzed it in a closed kinetic chain. These findings collectively suggest a consistent increase in VAG signal power with external load and muscle activity, regardless of the kinetic chain context.

The impact of age on the VAG signal features has also been a focus of investigation, with studies dividing participants into junior and senior groups [41], or further into 5 age groups spanning the 3rd to 7th decade of life [42, 43]. The findings consistently indicate a positive correlation between age and the power of the VAG signal, and a negative correlation between age and signal complexity. Specific features measuring power and complexity will be later described in more detail.

Vibroarthrography as a screening tool

Vibroarthrography was often used as a screening tool to classify knee joints into normal or abnormal classes, without specifying the exact type of abnormality or condition. An example of this approach is the dataset created by Professor Rangayyan's team at the University of Calgary [44]. This dataset comprises VAG signals from both healthy and unhealthy knee joints, with the latter group containing various conditions such as different grades of chondromalacia patellae, anterior cruciate ligament injuries, meniscal tear, and tibial chondromalacia. This dataset has made a significant impact on the VAG field, being used in a number of publications [44–71].

In another study, Hersek et al. [72] also used a binary classification system, distinguishing between healthy and unhealthy (injured) knee joints using VAG signals. They focused on knees affected by acute injuries, which included anterior cruciate ligament tears, lateral meniscus tears, and medial collateral ligament sprains. Taking a similar approach, Ozmen et al. [73] utilized another dataset, classifying VAG signals into healthy and injured groups, with the injured group containing knee joints with a nine different conditions.

Similarly, Zheng et al. [74] conducted a study distinguishing healthy and injured classes, where the injured class included patients with meniscus injury or four various stages of osteoarthritis. Further, Befrui et al. [75] also employed a binary classification approach, distinguishing between healthy knee joints and non-healthy ones. The non-healthy class

included knee joints affected by osteoarthritis and three stages of chondromalacia patellae (second, third, and fourth stage).

Vibroarthrography in Osteoarthritis

This binary classification approach, which doesn't specify a particular disease, implies that various knee diseases might share similar characteristic features, or lack of those, in their VAG signals. These shared features may allow for differentiation from healthy knee signals. Typically, these distinguishing features involve measurements related to the power or spectral parameters of the signal, which will be described in more detail later. Majority of studies, however, use more homogeneous abnormal groups, with the osteoarthritis being one of the most popular.

The severity of oa can be categorized into four grades according to the Kellgren-Lawrence grading system [21]. While many studies examine either a single OA grade or combine several without making distinctions, a study by Tanaka and Hoshiyama diverge from this trend [76]. They considered three groups: healthy knees, knees with oa grades I and II, and knees with oa grades III and IV. They identified statistical differences in low-frequency VAG power features, observing an increase in power corresponding with the progression of oa severity. In a subsequent study [77], they classified participants into three groups: an oa class including Kellgren-Lawrence grades 1 to 3, a healthy senior group age-matched to the oa class, and a healthy junior group. They found that the VAG signals of the healthy junior group were less variable compared to the healthy senior group, with the OA group exhibiting the most variable signals. Statistically significant differences were found between the OA group and both healthy groups.

However, within the context of osteoarthritis, the primary application of VAG has typically been to differentiate oa knees from healthy ones [14, 15, 78–88]. In many instances, this involves categorizing oa by a specific grade or, more frequently, combining several grades to create a unified group without specific grade distinction. Publications implementing this type of approach usually differ from each other mostly in implemented signal processing algorithms. However, general consensus is reached that oa knees, in comparison to healthy ones, have more power in lower frequency bands, making spectral features highly informative [89]. Another common observation is that the oa knee joints generally produce more acoustic emissions (according to Mascaro et al. [80], even up to 10 times more) than healthy knee joints, which is indicated in the VAG signal by singular amplitude bursts or spikes.

Therapy monitoring in Osteoarthritis

In a couple of studies, VAG method was used as a tool to assess the effects of viscosupplementation in patients with knee osteoarthritis. For instance, Falkowski et al. [90] employed VAG to investigate changes in the arthrokinematics of a knee joint after a single injection of hyaluronate in patients with moderate knee osteoarthritis. The findings revealed a significant reduction in VAG power features, suggesting an improvement in arthrokinematics. The study concluded that viscosupplementation could potentially reduce joint motion-related vibrations, improving the qualitative aspects of arthrokinematics in patients with moderate knee osteoarthritis. Similarly, Bączkowicz et al. [91] utilized VAG to evaluate the influence

of viscosupplementation on knee arthrokinematics in patients with grade II oa. The patients were analyzed at three different time points: before the injection, two weeks post-injection, and four weeks post-injection. The study showed an initial decrease in the VAG power two weeks after the injection, indicating an improvement in arthrokinematics. However, at four weeks post-injection, VAG power increased again, reaching levels similar to those before the injection. These findings suggest that while viscosupplementation can lead to short-term improvements in joint motion, the effects may be temporary.

In another study, Kalo et al. [92] investigated the immediate effects of a neuromuscular exercise intervention on knee joint sounds in individuals with osteoarthritis. The study involved a neuromuscular exercise intervention, designed to enhance sensorimotor control of the trunk and lower limbs, as well as improve the quality of movement. In addition, a placebo treatment, using laser needle acupuncture, was also administered. Vibroarthrography was used to measure knee joint acoustic emissions before and after both interventions, during various daily activities. The results indicated that the exercise intervention indeed had an impact on knee joint sounds. This study, along with the viscosupplementation studies previously discussed, illustrate the potential of VAG as a robust and descriptive tool, showing its capability to effectively evaluate the quality of knee joint function.

Distinguishing Osteoarthritis and Chondromalacia

Certain studies have gone beyond distinguishing oa from healthy knee joints, exploring the capability of differentiating oa from other diseases using VAG signals. For instance, Bączkowicz and Majorczyk [93] conducted a comparative analysis of VAG signals from oa, chondromalacia, lateral patellar compression syndrome, and healthy subjects. They first extracted features from the VAG signals, specifically focusing on measures of variability and spectral power across two frequency bands. Their analysis revealed that all groups with patellofemoral joint disorders exhibited higher levels of both variability and spectral power, compared to the healthy controls. More interestingly, the study identified differences in all examined features between osteoarthritis and chondromalacia. Moreover, certain features were also able to differentiate lateral patellar compression syndrome from osteoarthritis, and lateral patellar compression syndrome from chondromalacia.

In the study conducted by Kręcisz and Bączkowicz [5], five distinct classes were identified: oa, three stages of chondromalacia (first, second and third), and healthy knees. This dataset, used for the first time in their research, was also utilized in this dissertation. The authors employed a plethora of signal features to quantify power, various frequency components, and the complexity of the signal. When classifying signals, they achieved an accuracy of 69% for the 5-class classification. For binary classification, where all non-healthy signals were combined into one group, the accuracy increased to 90%. These findings suggest that VAG signals may serve as a valuable tool, not only for distinguishing healthy knee joints from those with specific disorders, but for differentiating between various knee disorders as well.

Vibroarthrography in Chondromalacia

Wu et al. [94] leveraged vibroarthrographic (VAG) signals to distinguish between healthy knees and knees with chondromalacia patellae, with the latter group containing various

stages of the condition. The results indicated that vibroarthrograms of knees with chondromalacia were characterized by a higher degree of variability than those from healthy knees, as indicated by a greater number of signal turns.

In a comparative analysis of open and closed kinetic chains, Bączkowicz et al. [95] undertook binary differentiation of healthy and chondromalacia-affected knees. The chondromalacia group included stages 2 and 3. Results showed that patients with chondromalacia exhibited heightened vibroacoustic emission levels in both open and closed kinetic chains. That is, their VAG signals were more variable and demonstrated higher power within specific frequency bands.

Similar approach was implemented by Jonak et al. [96], who also sought to distinguish between healthy knees and those affected by chondromalacia. Also employing both open and closed kinetic chains, they utilized Recurrence Quantification Analysis to measure signal complexity. Results indicated that signals obtained from chondromalacia-affected knees exhibited greater Recurrence Rate, Entropy, and Trapping Time features, in both open and closed kinetic chains. This suggests a higher complexity and variability in chondromalacia signals compared to those from healthy knees.

Using a different approach, Bączkowicz and Majorczyk [97] differentiated among the three consecutive stages of chondromalacia, rather than treating them as a unified group. Each stage was found to generate unique vibroacoustic signals, with a clear relationship observed between signal variability, power in specific frequency bands, and the degree of cartilage damage (i.e., the chondromalacia stage). Similar approach was achievement in the previously mentioned research by Kręcisz and Bączkowicz [5], where three stages of chondromalacia were also distinguished. Both works show the potential for vibroarthrographic analysis to differentiate among the stages of chondromalacia patellae.

Vibroarthrography in Juvenile Idiopathic Arthritis and other diseases

Several publications have utilized vibroarthrography to distinguish healthy knee joints from joints affected by juvenile idiopathic arthritis. In a pilot study by Semiz et al. [98], signals from 8 subjects were used. In following studies by Whittingslow et al. [99] and Gharehbaghi et al. [100], the number of study participants was increased. All mentioned studies classified knee joints as either healthy or affected by juvenile idiopathic arthritis. Their signal processing approach was different from previously referenced works, with additional signal segmentation step before the feature extraction. It will be discussed in more detail later. For now, the most significant conclusion which can be derived from those studies is that the signals obtained from the knee joints affected by the juvenile idiopathic arthritis can be effectively distinguished from the healthy-generated ones.

Vibroarthrography has also been employed for distinguishing between various other health conditions. Reddy et al. [101] used it to distinguish between two forms of arthritis: spondyloarthropathy and rheumatoid arthritis. Their research demonstrated that the mean power of acceleration signals within a specific frequency range varied significantly between the two patient groups. In a later research, Faria et al. [102] conducted a study focusing on the distinction between normal and rheumatoid arthritis-affected knee joints. They found a statistically significant difference in knee joint vibration power within a lower frequency range. In a different study, Richardson et al. [103] applied vibroarthrography to separate healthy

knees from those afflicted with a meniscus tear. They extracted a set of features, which allow them to distinguish healthy knee joints from knees with meniscus tear.

Table 3.1: Summary of knee joint conditions studied in the literature.

references	studied conditions	groups	details
[44,47–75,104–114]	healthy, non-healthy	2	various conditions
[14,15,78–88]	healthy, oa	2	various oa grades combined
[76]	healthy, oa	3	oa grades I and II, and oa grades III and IV
[77]	healthy, oa	3	oa grades I to III, healthy senior, and junior
[90,91]	healthy, oa	3	oa before and after viscosupplementation
[92]	oa	4	before and after neuromuscular exercise interventions and placebo
[93]	healthy, oa, CMP, LPCS*	4	CMP combined stages II and III
[5,115,116]	healthy, oa, CMP	5	CMP separate stages I, II and III
[94–96]	healthy, CMP	2	various CMP stages combined
[97]	healthy, CMP	4	CMP separate stages I, II and III
[98–100]	healthy, JIA†	2	
[101]	SPA‡ and RA§	2	
[102]	healthy, RA§	2	
[103]	healthy, MT¶	2	
[35]	healthy	3	controls, immobilized before rehabilitation, and immobilized after rehabilitation
[36]	healthy	3	min., medium and max. contraction levels
[37]	healthy	4	0, 1.25 kg, 2.5 kg and 5 kg additional load
[117]	healthy	3	0 kg, 10% and 40% of one repetition max
[39]	healthy	3	0, 50%, 100% body weight
[40]	healthy	3	0 kg, 10 kg, 20 kg
[41]	healthy	2	healhty junior, healthy senior
[42,43]	healthy	5	3rd to 7th decade of life

*LPCS: Lateral Patellar Compression Syndrome †JIA: Juvenile Idiopathic Arthritis ‡SPA: Spondyloarthropathy §RA: Rheumatoid Arthritis ¶MT: Meniscus Tear

In the Table 3.1, various conditions of the knee joint, with the corresponding references, were summarized.

3.3 Signal acquisition

In this section, details of signal acquisition will be discussed, broadly divided into sensor types and their frequency ranges, placement of the sensor on the studied joint and experimental procedures used during the recording of the vibroarthrogram.

3.3.1 Sensor types and frequency ranges

As mentioned earlier, early developments in VAG signal analysis were focused on using microphones, which were later changed for accelerometers. Usage of microphones is still to be found in recent works [92], for the advantage of sensor not requiring direct contact with the skin.

Specific accelerometer sensors currently used to measure vibroarthrographic signals are of frequency band up to 10 kHz. Microphones of different types on the other hand, have considerably wider frequency range, even up to 200 kHz. Summary of various sensors used in literature, with specific works which used those sensors, was included in table 3.2.

Table 3.2: Summary of sensors used to acquire vibroarthrographic signals.

references	sensor type	brand	model	frequency range
[36,44,45,47–71,94, 105–111,118,119]	ACC*	Dytran	3315A	5 Hz–10 kHz
[39,73,98–100, 103,120–122]	ACC*	Dytran	3225F7	2 Hz–10 kHz
[5,35,40,42,43,90, 91,93,95,97,115, 116,123]	ACC*	Brüel&Kjær	4513B-002	1 Hz–10 kHz
[34,38,92,124]	ACC*	Knowles	SPU0414HR5H-SB	100 Hz–10 kHz
[125]	ACC*	Analog Devices	ADXL325	0 Hz–550 Hz
[102]	ACC*	Analog Devices	ADXL335	0 Hz–550 Hz
[75]	ACC*	PCB Piezotronics	352A24	1 Hz–8 kHz
[112]	ACC*	PCB Piezotronics	353-B33	1 Hz–4 kHz
[104,126]	ACC*	Vibro-Meter Corp	BBN501M601	not specified
[101]	ACC*	Entran Devices	EGA	0 Hz–500 Hz
[76,77]	ACC*	Kistler	8690B5	not specified
[37,84]	ACC*	STMicroelectronics	LIS344ALH	0 Hz–1.8 kHz
[114]	ACC*	PLUX Wireless Biosignals	xyzPlux	0 Hz–500 Hz
[83]	ACC*	InvenSense	MPU6050	not specified
[75]	PE† disk	Elektrotechnik Karl-Heinz Mauz GmbH	EPZ-27MS44F	4.4 kHz‡
[14,15,86–88,96]	PE† contact microphone	TE Connectivity	CM01b	8 Hz–2.2 kHz
[72,113,127]	electret microphone	Sanken Microphone	COS-11D	50 Hz–20 kHz
[78,79]	electro-stethoscope (PE† microphone)	MediTech	SP-S1	5 Hz–2 kHz
[82]	microphone	not specified	not specified	50 Hz–16 kHz
[113,127]	PE† contact microphone	Measurement Specialties	SDT	100 Hz–10 kHz
[113,127]	MEMS based microphone	STMicroelectronics	MP33AB01H	100 Hz–10 kHz
[80]	PE† sensor	Physical Acoustics	S9204	50 Hz–200 kHz
[41]	PE† sensor	Murata	7BB-20-6L0	6.3 kHz‡

*ACC: Accelerometer †PE: Piezoelectric ‡Frequency band information not available, resonant frequency provided instead.

The frequency bands of the sensors used indicate specific ranges within the total spectrum of measured frequencies, where the sensors have a flat frequency response. In these bands, from the lower to the upper frequency limits, the sensors maintain a consistent output level. However, another major factor determining acquired signal is the sampling frequency, i.e.,

the frequency in which the data is acquired. It determines the time resolution of a signal and is especially important in relation to the Nyquist frequency. The Nyquist frequency is half of the sampling frequency. According to the Nyquist-Shannon theorem [10], to accurately capture a signal without aliasing, the sampling frequency must be at least twice the highest frequency component of the input signal. In other words, frequencies up to half of the sampling frequency will be correctly captured in the sampled (discrete) signal.

In the VAG signal acquisition, most studies use sampling frequencies up to 10 kHz, with a few exceptions where much higher frequency is used (even up to 1 MHz). However, most studies also use some band-pass filtering or other preprocessing algorithms to filter out uninformative signal components. More in-depth preprocessing review will be provided in the following section. Summary of the acquisition frequencies was provided in Table 3.3.

Table 3.3: Summary of sampling frequencies used to acquire vibroarthrographic signals.

references	sampling frequency [kHz]
[37, 96, 114, 125]	1
[14, 15, 86–88]	1.4
[83]	1.48
[41, 44, 45, 47–71, 74, 84, 94, 106–111, 118, 119]	2
[76, 77, 101, 102]	3
[36]	4
[112]	5
[38]	5.512
[78, 79]	6
[5, 35, 40, 42, 43, 90, 91, 93, 95, 97, 115, 116, 123]	10
[85]	10.24
[105]	15
[34, 75, 92, 124]	16
[121]	25
[113]	40
[72, 81, 127]	44.1
[39, 82, 122, 127]	50
[73, 99, 100, 103]	100
[120]	102.4
[98]	108
[80]	1000

3.3.2 Sensor placement on the knee joint

Vibrations generated by the knee joint can be easily damped by soft tissues, such as muscles or fat [121] and distorted by muscle activity [25]. Therefore, sensor placement on the knee is crucial. Various locations used in literature to acquire vibroarthrograms were visualized in Figure 3.2. In the rest of this subsection, numbers referenced in the text correspond to dots in this figure.

Generally, various locations were used in the literature, sometimes depending on specific condition of the knee to be diagnosed, or type of the sensor. However, the most common localization for VAG signal acquisition is central part of the patella (dot 1). It was used in the context of healthy knee joints to analyze the impact of muscle interference [36, 119] or

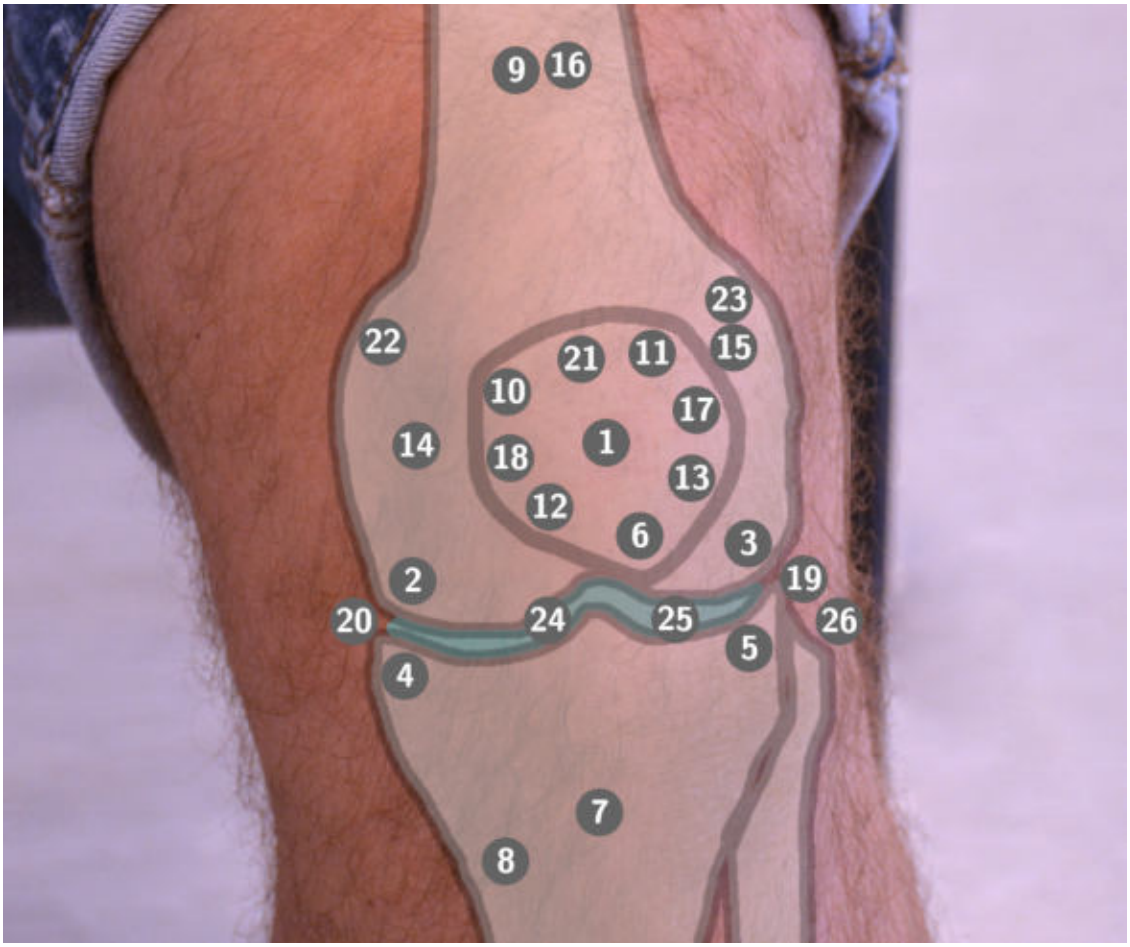

Figure 3.2: Sensors placement used across the VAG literature.

external load [38, 40] on the vibroarthrogram, to compare movement patterns [125], measure the relation of age and the quality of joint's motion [42, 43] or determine repeatability of the VAG signal in general [34, 124]. Nevertheless, differencing between normal and *somehow* abnormal knee joints is the usual goal of VAG signal analysis. Mid-patella location (dot 1) was used for knee joint screening [71] and more specific classification, with conditions such as chondromalacia [94], osteoarthritis [82], rheumatoid arthritis [102] or meniscal pathologies [128].

In some studies, meniscal pathologies were classified using VAG signals obtained from femoral condyles. For example, Bączkowicz and Kręcisz [123] used medial condyle of the femur (dot 2). Both medial (dot 2) and lateral (dot 3) condyles were earlier used by Kernohan et al. [28, 128].

Medial and lateral condyles of the tibia (dots 4 and 5, respectively), are also quite common sensor locations in literature. The medial tibial condyle (dot 4) was used by Kim et al. [78,79] for osteoarthritis detection. In the same context, similar sensor placement was used by Mascaro et al. [80], aided by the sensor placed at the upper part of tibia (dot 8). In some of the works by Kalo et al. [34, 38, 92, 124], medial tibial condyle (dot 4) was used along the mid-patella (dot 1) location, in both osteoarthritis and asymptomatic only studies. Lateral

tibial condyle (dot 5) was used for knee joint screening [112]. Tanaka and Hoshiyama [76] used the same location to differentiate between different stages of osteoarthritis, and later, in the combined contexts of osteoarthritis and age [77]. Finally, Befrui et al. [75] used both condyles of the tibia, as well as mid-patella location (dot 1) to differentiate between healthy and non-healthy knee joints.

Another popular location is the tibial tuberosity below the patella (dot 7). It was mostly used in the healthy knee joints context [36, 37, 119], however, Madeleine et al. used it for osteoarthritis detection [84].

Some studies utilized two accelerometers, placed medially and laterally to the patellar tendon (dot 24 and 25, respectively). This configuration was used by Ozmen et al. [73] for screening purposes. In the studies by Whittingslow et al. [99] and Gharehbaghi et al. [100], the same location approach was applied for juvenile idiopathic arthritis detection. Richardson et al. [103] used it for meniscus tear classification, while in the study by Bolus et al. [39], it was used in the healthy-only external load context.

In most works regarding VAG signal analysis, vibroarthrograms were acquired by one sensor. However, as mentioned earlier, some studies used multiple sensors, with a goal to remove interferences [118], illustrate spacial distribution of some signal features [37], localize the pathology within the joint [105], or determine the best localization of the sensor. For example, Befrui et al. [75] used three locations: mid-patella, medial and lateral tibial condyles (dots 1, 4 and 5, respectively). According to their results, the patellar location (dot 1) provided the most informative signal (in terms of specific frequency components), followed by medial (dot 4) and lateral (dot 5) tibial condyles. In the reliability study of the vibroarthrograms, Kalo et al. [124] similarly compared mid-patella position (dot 1) to the medial tibial plateau (dot 4). Their study indicated slightly better repeatability of the VAG signals acquired from the latter location, but only in the intrasession context, i.e., with the vibroarthrograms acquired the same day. For the interday repeatability, some features of the VAG signals recorded from the patellar location (dot 1) proved to be repeatable, which was not the case for the medial tibial plateau (dot 4). The same locations were also compared in the other study by Kalo et al. [34]. Their results indicated that in the context of two specific VAG features, patellar location (dot 1) proved to be more sensitive to both between-subject differences and movement conditions.

Based on those studies, the patella (dot 1) seem be the best sensor location for VAG acquisition. However, comparison of the VAG acquisition locations, in terms of either informativeness or repeatability, can only be done using some specific VAG features. Therefore, different features could potentially yield different results of the comparison.

There are also other sensor placement locations used in the VAG literature less commonly. Their summary was provided in Table 3.4. Location numbers provided in this table correspond to dots in Figure 3.2.

3.3.3 Experimental procedures and kinetic chain configurations

Movement patterns can be broadly categorized into Open Kinetic Chain (OKC) and Closed Kinetic Chain (CKC). In the OKC, movement of the end of the leg is unrestricted, like for example during a seated leg extension. On the other hand, CKC includes movements, where the end of the limb is fixed or in contact with a stationary object. An example of a CKC

Table 3.4: Summary of the sensor localizations used in the literature. Note, that localization column corresponds to numbering in Figure 3.2.

references	localization	description
[5, 28, 34–36, 38, 40, 42–45, 47–71, 75, 82, 83, 85, 87, 91–95, 97, 102, 104–111, 114–116, 118, 119, 123–126, 128]	1	mid patella
[28, 105, 123, 128]	2	medial femoral condyle
[28, 105]	3	lateral femoral condyle
[34, 38, 75, 78–80, 92, 124]	4	medial tibial condyle (plateau)
[75–77, 112]	5	lateral tibial condyle
[15]	6	lower pole of the patella
[36, 37, 84, 105, 119]	7	tibial tuberosity below the patella
[80]	8	upper medial part of tibia
[36, 45, 47, 118, 119]	9	rectus femoris muscle
[37, 84]	10-13	square configuraion on the patella
[37, 84]	14	medial side of the knee 1-2 cm from the medial condyle of femur towards the patella
[37, 84]	15	lateral side of the knee 1-2 cm from the lateral condyle of femur towards the patella
[37, 84]	16	above the knee over the quadriceps tendon
[120, 127]	17	lateral side of the patella
[72, 120, 127]	18	medial side of the patella
[14, 86, 88]	19	lateral joint fissure
[14, 86, 88]	20	medial joint fissure
[114]	21	proximal part of the patella
[82]	22	medial epicondyle
[82]	23	lateral epicondyle
[39, 73, 98–100, 103, 121]	24	medial to the patellar tendon
[39, 73, 99, 100, 103, 121]	25	lateral to the patellar tendon
[81]	26	lateral aspect of the knee with the axis of rotation at the joint line
[36, 41, 82, 119]	-	mid tibia

movement could be a squat exercise. In the vibroarthrography filed, both OKC and CKC are widely used, each utilized through a variety of experimental procedures. Summary of movement patterns and experimental procedures used in the vibroarthrography was provided in Table 3.5.

The OKC movements usually incorporate extending and flexing the free-hanging leg in a seated position, with range of motion ranging usually from 90 to 135 degrees. In most experimental protocols, the leg moves without any additional load. Some studies, however, implement such loadings. For example, Ladly et al. [118] repeated VAG acquisition for unloaded knee, as well as three- and six-pound weights attached to the ankle joint. Similarly, Andersen et al. [37] used four loading settings, i.e., 0 kg (no additional load), 1.25 kg, 2.5 kg and 5 kg weights, attached to the arch of the foot. Slightly different approach was used by Kalo et al. [38], which incorporated leg extension device. Loading settings in this study were normalized by the strength of the participant, and were equal to 0% (no additional load), 10% and 40% of the load the participant was able to lift once (i.e., the one repetition max). Results of those studies consistently point out, that energy of the vibrations is increased

Table 3.5: Summary of the movement patterns used in the literature to measure vibroarthrographic signals.

references	chain	additional load	details
[5, 14, 15, 35, 37, 38, 42, 43, 72, 73, 75, 78, 79, 81, 90, 91, 93, 95–102, 104, 105, 114–116, 120, 122, 123, 125, 128]	OKC	none	ROM* 90°
[44, 45, 47–71, 74, 94, 105–111]	OKC	none	ROM* 135°
[14, 15, 34, 36, 38, 41, 76, 77, 80, 82–88, 92, 95, 96, 100, 104, 112, 113, 119–121, 124, 125, 127]	CKC	none	ROM*90°
[40]	CKC	0, 10, 20 kg	ROM* 90°, squat exercise
[118]	OKC	0, 3, 6 pound	ROM* 90°
[39]	CKC	0%, 50%, 100% BW†	vertical leg press
[121]	CKC	-10.8%, 0, 19.3% BW†	walking
[37]	OKC	0, 1.25, 2.5, 5 kg	ROM* 90°
[38]	OKC	0%, 10%, 40% of 1RM‡	ROM* 90°

*ROM: Range Of Motion †BW: Body Weight ‡1RM: One-Repetition Maximum

with additional load.

Movement in CKC-based protocols is more diverse. Similarly to OKC, additional loading is usually not implemented, with few exceptions. For example, Ołowiana et al. [40] implemented squats in their study, with the additional weight being 0 kg (no additional load), 10 kg, and 20 kg. Bolus et al. [39] applied additional load using vertical leg press, with three load settings: 0% (no additional load), 50%, and 100% of participant's body weight. In the study by Scherpereel et al. [121], experimental procedure consisted of walking on 10-degree inclined or declined treadmill, with body-weight harness, no additional loading and an additional weighted vest. Harness removed on average 10.8% of participants body weight, while vest added on average 19.3%. Similarly to OKC, additional weight in CKC results in greater power of the vibroarthrogram. Nevertheless, most studies acquire VAG signals without additional load. The unloaded movement usually includes standing from a seated position with 90-degree flexion in the knee joint, walking [34, 92, 121], going upstairs [84] or downstairs [34, 38, 84, 92], squatting [40, 85, 112, 120], or forward lunging [34].

Some studies compared informativeness of vibroarthrogram for movements in OKC and CKC. Results of Machrowska et al. [125] indicated that both movement chains are important. According to their results, OKC analysis, because of reduced high-frequency muscle interference, offers better insights into the unloaded limb function, particularly cartilage structures. Closed chain analysis, on the other hand, provides data on the function of a loaded limb under real conditions. Bączkowicz et al. [95] compared CKC and OKC in the context of chondromalacia diagnosis. The results indicated that VAG signals from both CKC and OKC can provide features, which show statistically significant differences between healthy and chondromalacia knee joints. Moreover, both power and variability of the VAG signals proved to be greater in CKC conditions, which can be linked to differing stress and friction in the patellofemoral joint. Since informativeness of the signals is similar for both movement patterns, OKC can be preferred option for VAG acquisition in patients.

3.4 Signal preprocessing

Various methods of preprocessing are present in the VAG literature. For the purpose of this review, they will be divided into filtering, reference denoising, decomposition, normalization, segmentation and windowing. Note, that specific references might used various preprocessing methods selectively or in combination. Also, some references did not use any preprocessing, i.e., analyzed raw VAG signal. The summary of various preprocessing strategies with their corresponding references was provided in Table 3.6. More detailed description of specific methods will be provided in following subsections.

Table 3.6: Summary of the preprocessing methods used in vibroarthrography literature.

references	method	details
[87, 115, 116]	raw	
[5, 34–40, 42–45, 47, 49–59, 61, 63–67, 71–74, 76–79, 82, 84, 90–95, 97–103, 105–108, 110, 113, 121–124, 127]	frequency-selective filtering	details in Table 3.7
[85]	non-frequency-selective filtering	moving median
[62, 64]	non-frequency-selective filtering	Savitzky-Golay
[59, 62, 64, 67, 70, 71, 74, 109, 111, 114]	non-frequency-selective filtering	cascade moving average
[116]	non-frequency-selective filtering	differencing
[41, 45, 114, 118, 119]	reference denoising	
[86, 94, 112]	decomposition	empirical mode decomposition
[14, 15, 62, 63, 66, 74, 88, 96, 110, 114, 125]	decomposition	ensemble empirical mode decomposition
[64]	decomposition	variational mode decomposition
[49, 55, 57, 85, 106]	decomposition	wavelet decomposition
[50, 67, 70]	decomposition	wavelet packet decomposition
[52, 53, 55–58, 61, 68]	normalization	0 to 1
[94]	normalization	-1 to 1
[72]	normalization	zero mean, unity variance
[68]	normalization	3-bit encoding
[39, 41, 44, 45, 47, 51–55, 57, 58, 75, 78–81, 99–103, 118, 121, 122, 124]	segmentation	specific movement patterns
[14, 15, 88]	segmentation	removal of insignificant signal fragments
[5, 35, 37, 40, 42, 49, 53, 54, 57, 58, 60, 71, 73, 78, 79, 84, 86, 90, 91, 93, 95, 97, 99, 100, 103, 107–109, 114, 121, 123]	windowing	various time frames

3.4.1 Filters

Filter methods can be further divided into frequency-selective and non-frequency-selective ones. The former types, with an example of the Butterworth filter, are designed to pass

specific frequency band. In this band, their frequency response is relatively flat, while frequencies outside of it get attenuated. Those types of filters are used with an assumption, that the important information of the signal is inside their frequency range and removing other frequencies improves signal's informativeness. Summary of frequency bands used in literature, with corresponding references, was provided in Table 3.7. For better comparison of different frequency ranges, they were also visualized in Figure 3.3.

Table 3.7: Summary of the frequency-selective filter pass-band frequencies used in VAG preprocessing. Note, that while most references used band-pass filtering, low-pass or high-pass filtering were also present. Values in the *number* column correspond to visualization provided in Figure 3.3.

references	low freq.	high freq.	number
[113]	LP†	10000	1
[102]	2	HP*	2
[37]	10	250	3
[84]	10	500	4
[39]	10	800	5
[36,44,45,47,49–59,61,63–67,71,74,94,106–108,110]	10	1000	6
[76]	25	350	7
[77]	25	500	8
[101]	25	HP*	9
[5,34,35,40,42,43,91–93,95,97,123]	50	1000	10
[73]	50	5000	11
[90]	50	HP*	12
[38,124]	100	300	13
[121]	100	1000	14
[82]	100	2000	15
[122]	150	20000	16
[78,79]	220	2000	17
[103]	250	8000	18
[99,100]	250	10000	19
[98]	250	20000	20
[105]	300	800	21
[72]	1000	15000	22
[127]	7000	16000	23

*HP: High-pass filtration †LP: Low-pass filtration

As can be seen in Figure 3.3, there seem to be no hard consensus on the filtering frequency ranges in the VAG literature. The discrepancies are mostly caused by different measurement setups, i.e., types of the sensor, its placement on the knee joint and experimental procedure. Some authors used as narrow bands as 200 Hz [38,124], while other considered wider ranges. Also, in some studies, only relatively high frequencies, i.e., above 1 kHz, were considered [72, 127]. Nevertheless, most of the studies used frequency bands between 10 Hz and 1 kHz.

In an earlier study conducted by the author of this dissertation, along with other researchers [116], Frequency Range Maps were developed and implemented to identify the most informative frequency ranges in a classification context. More detailed description of this methodology will be provided in the Chapter 6.

Another type of filtering approach do not assume specific frequency bands. That is,

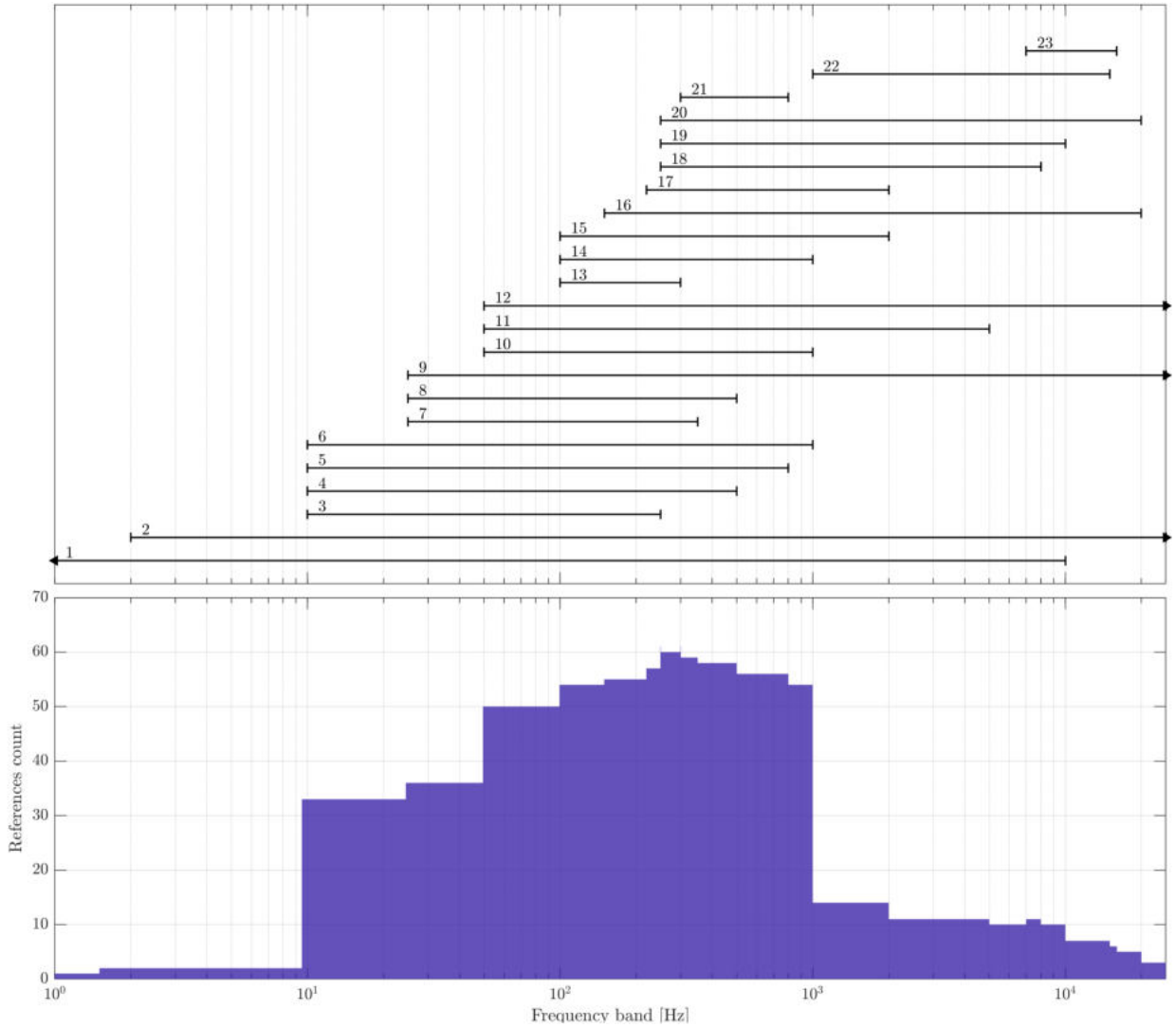

Figure 3.3: Filtering bands used in the literature. Upper plot shows specific frequency ranges used in various studies. Numbers above lines correspond to numbers in Table 3.7. Lines ended with arrows indicate low- or high-pass filtering. Lower plot shows count of references that used specific frequency range.

non-frequency-selective filters do not have flat frequency response. Their implementation usually consist of dividing signal into small windows of fixed duration, then calculating some feature(s) of those windows. Signals obtained from such windowing are usually called *moving* or *rolling*. For example, Wang et al. [85] used moving median filtration to remove high-amplitude pulses from the vibroarthrogram. Another example could be Savitzky-Golay filter, which consists of fitting low-degree polynomial to the datapoints in a given window. This method was also implemented in the VAG preprocessing context, in the studies of Sundar et al. [62,64]. In both studies, it was used at the final step of multi-step preprocessing.

One of the most popular preprocessing approaches based on non-frequency-selective filtering uses cascade moving average to remove baseline wander. It was proposed by Cai et al. [59] and can be described as having two layers. The first one consists of two overlapping

moving average operators. From their results, i.e., two smoothed signals, a linear piecewise trend is obtained, which is further smoothed in the second layer. The resulting signal can be considered the baseline wander estimation. Baseline-wander-free vibroarthrogram can be obtained by simply removing the estimation from the raw signal.

There are also high-pass non-frequency-selective filters, which can be used to attenuate lower frequencies, while preserving or emphasizing high-frequency components. An example of this filter type would be a differencing operation, i.e., subtracting adjacent signal samples. This method was applied to VAG signals in previous study [116], and and will also be used in this dissertation. Therefore, more detailed description of this method will be provided in Chapter 4.

3.4.2 Reference denoising

In some studies, the vibroarthrogram was denoised using some reference signal. For example, acquiring additional vibroarthrogram via the sensor placed on the rectus femoris muscle, Zhang and Rangayyan [119] used Widrow-Hoff LMS algorithm to remove its interference in the VAG signal. Wu et al. [114] obtained trend reference from the proximal-patella sensor location to remove motion artifacts from the mid-patella VAG signal. Both sensors used were accelerometer-based. Another usage of additional sensor to improve the quality of the vibroarthrogram was implemented by Gong et al. [41]. They used angle signal to calculate the frequency of drifting baseline, which was further used to remove the baseline wander.

3.4.3 Decomposition methods

Decomposition methods consist of breaking down signal into less complex sub-signals, which can be considered components of the original signal. That is, summing up sub-signals results in the original signal. Specific sub-signals can contain most information important in a given context. Therefore, the usual preprocessing based on decomposition, consists of firstly decomposing the signal, then determining which sub-signals are informative, optionally further filtering them, and finally summing up selected sub-signals. In some studies, however, each sub-signal was further used for feature extraction.

One of the most popular decomposition methods is the Empirical Mode Decomposition (EMD) [129]. It consist of a series of steps to obtain individual Intrinsic Mode Functions. The first step involves identifying local extremes of the signal. Then, upper and lower envelops are constructed by interpolating between local maxima and minima, respectively. In the next step, the average value of the envelopes is subtracted from the original signal, providing a candidate Intrinsic Mode Function (IMF). The IMF is a oscillatory component signal, that satisfy two conditions: its number of extremes and zero-crossings can differ by at most one, and local average value of envelopes is zero. If those conditions are not met for the obtained candidate IMF, the process is repeated on the resultant signal, as long, as the conditions are not fulfilled. After meeting the conditions, the first IMF is subtracted from the original signal, and the process is repeated on the remaining part of the original signal. Resulting IMFs are oscillatory signals with frequency bands successively decreasing with the level of the decomposition.

As mentioned earlier, the signal preprocessing using EMD usually consists of summing up specific informative IMFs. For example, Karpiński et al. [86] summed up the IMFs with normally distributed values. Opposite approach was used by Wu et al. [94]. They reconstructed the VAG signal using only IMFs with values distribution significantly different from Gaussian. Additionally, they removed baseline wander from the low-frequency IMFs. Instead of selecting IMFs based on some statistical criterion, Chen et al. [112] argued, that the highest frequency IMF captures noise, while lower frequency IMFs correspond to muscle contraction. Actual vibroarthrographic signals are captured by specific IMFs in the medium frequency range.

One of the biggest disadvantages of the EMD algorithm is so-called mode mixing problem [130]. It involves blending of different IMFs, making them less specific and informative. This phenomenon is usually attributed to non-stationarity of signals, which renders the IMFs separation difficult. One of the methods developed to overcome this problem is the Ensemble Empirical Mode Decomposition (EEMD) [131]. It consists of creating multiple copies of the original signal and contaminating them with random noise. Each signal is then decomposed into IMFs using standard EMD. Finally, the resulting IMFs of the same level are averaged across the signals, removing the contamination and yielding more robust IMF set.

The EEMD is probably the most widely used decomposition method in the vibroarthrography filed, with the preprocessing approaches similar to basic EMD. In most studies, however, informativeness of various IMFs is quantified differently. For example, Wu et al. [110] used Detrended Fluctiation Analysis to compute Fractal Scaling Index. This approach allowed them to measure anti-correlations and long-range dependencies of each IMF. After the decomposition, the vibroarthrogram was reconstructed from IMFs that did not exhibit anti-correlations. More detailed description of the Detrended Fluctuation Analysis and Fractal Scaling Index will be provided in Chapter 5.

Another approach developed to overcome potential problems of the EMD is the Variational Mode Decomposition (VMD) [132]. This method is not recursive and the specific sub-signals, called modes, are determined concurrently by solving an optimization problem. It is defined such that the total variation of the modes is minimized, while enforcing orthogonality between them. That leads to more stable decomposition. However, the VMD is conducted with a regularization parameter, which needs to be tuned. It controls the tradeoff between total variation of modes and accuracy of representing the original signal. Also, as an optimization algorithm, it needs to be initiated with initial center-frequencies of modes. Those features make the VMD less straight-forward to use, however, some studies in the VAG domain have implemented it. For example, Sundar et al. [64] compared it to some other VAG denoising strategies and find it superior. Their whole de-nosing algorithm consisted of decomposing the signal using VMD, thresholding the modes using Wiener Entropy, and filtering the modes of entropy greater than the threshold using low-pass 300 Hz filter. The first, high-frequency mode was filtered using band-pass filter of 250-300 Hz. Signal reconstructed from the filtered modes was additionally smoothed using the Savitzkty-Golay filter.

Another popular approach to decompose the vibroarthrogram employes Wavelet Decomposition. It involves breaking down the signal into different frequency components at multiple resolutions, using a predefined set of wavelet functions. Wavelet Decomposition follows a hierarchical structure, iteratively decomposing signal into two sub-signals. They capture

low- and high-frequency components, and are called approximation and detail, respectively. At each step, approximation is further decomposed into approximation and detail and the process continues for the predefined number of iterations. A similar approach, called Wavelet Packet Decomposition, consists of decomposing at each iteration not only the approximation, but also the detail into sub-signals. Both approaches were used in VAG preprocessing to either reconstruct denoised vibroarthrogram [70], or further analyze individual sub-signals for feature extraction [50].

3.4.4 Normalization

Normalization methods are used to change the signal amplitude, while keeping its shape, which can minimize the impact of variation in scales or ranges between different measurements or participants. This process often involves scaling signal samples from different ranges to a pre-defined range, typically from zero to one. Since this scaling is done on the amplitude range of a given signal, not all signals in a database, differences in statistical parameters will be transformed nonlineary between specific vibroarthrograms, potentially increasing informativeness of specific features. For example, Rangayyan and Wu [52] normalized signals to the range of zero to one, before extracting features related to the probability density function of a given signal's values. Wu et al. [94] used different values to normalize the signal's amplitudes, i.e., from negative to positive one. They conducted the normalization before decomposing the signal using EMD.

Another normalizing method is the standardization of values, i.e., transforming them to have zero mean and a unite variance. It was used by Hersek et al. [72] after initial filtering, to fix potential discrepancies between measurement's conditions across the knee joints.

One additional preprocessing method that directly changes signal's values was implemented by Athavale and Krishnan [68]. It consists of "encoding" the signal's values into 3 bits, i.e., compressing them into lower dimension. This approach is qualitatively different from the normalization approaches mentioned earlier, since it directly changes the shape of the VAG signal. Despite reducing signal's size by as much as 88%, classification results after such encoding proved to be better than without it.

3.4.5 Segmentation

Segmentation used in VAG signal preprocessing consist of dividing the signal into phases, usually involving flexion and extension. Since forces in the analyzed knee joint are different in both phases, vibroarthrograms generated by them have different characteristics. Analyzing those parts of VAG signal separately can potentially provide more information about the knee joint. The biggest disadvantage of this method is the need for additional information about the knee joint flexion, or position, in time.

There are various methods to segment the vibroarthrogram into flexion and extension phases. For example, Rangayyan and Wu [54] divided specific cycles of the signal into two halves, corresponding to extension and flexion. They extracted features for the whole cycle, as well as for flexion and extension phases separately. They found that, according to p-value of the t-test, features extracted from the flexion segment were more descriptive than those from extension segment or the whole cycle. However, in terms of area under the receiver

operating characteristic curve, both the extension segment and the whole, undivided cycle was more informative than the flexion segment alone.

The VAG signal can also be segmented manually, using only low-frequency components of the VAG signal. This approach was used by Faria et al. [102] to divide the whole signal into single standing up and sitting down movement cycles. Note, that further splitting into flexion and extension was not carried out.

The most robust information about the knee joint flexion in time, however, can be extracted from additional sensors, like electrogoniometers [75], inertial measurement units [100] or motion-capture systems [37]. For example, Befrui et al. [75] used a potentiometer to track the knee joint angle along with a VAG signal. Using obtained signal, they were able to not only segment the vibroarthrogram into flexion and extension phases, but also remove resting state between those phases. Their results indicate that frequency components are generally more informative for the flexion movement phase, especially in low frequency ranges.

Removal of irrelevant signal fragments to improve overall quality of the extracted features can also be conducted without additional segmenting. For example, it was employed by Karpiński et al. [14] in a semi-automatic way, using information provided by the encoder.

Segmentation can also be conducted to subsequently average the obtained segments. This approach was utilized by Kalo et al. [124]. Their experiment protocol consisted of standing up and sitting down. They also analyzed repeatability of two VAG features, namely amplitude and median power frequency. Their results showed, that the median power frequency feature proved to be more repeatable in sitting down motion, while the amplitude in standing up.

It seems that informativeness of specific motion phases depends on specific VAG features being extracted, as well, as kinetic chain, or more broadly, the whole experimental protocol. While VAG segmentation seem to allow for more refined information extraction, there seem to be no consensus regarding superiority of one motion phase over the other. This conclusion is also in line with the kinetic chain comparison.

3.4.6 Windowing

Some VAG-related studies used fixed-duration signal segmentation, or windowing. In this approach, the vibroarthrogram is divided into windows of fixed duration (for example, 200 ms), independently from the movement phase. Then, for each window, the features are extracted, providing information about the *dynamics* of the feature, or, in other words, changes of some VAG signal characteristics in time. Instead of providing one feature for the entire signal, those methods provide feature vectors or *feature signals*, which can be further analyzed. This approach is most widely used to conduct time-frequency analysis, i.e., to track the changes of signal's frequency distribution in time. Usually, dimensionality of the feature signals, is reduced by calculating some summary statistics. That approach was used by Shidore et al. [71]. They extracted a number of time-varying frequency features, which were further summarized by average and standard deviation values.

While time-frequency analysis is the most popular use of the widowing, there are infinitely many features, which can be obtained following this methodology. One of the most popular features extracted in this way is the variance of mean squares, defined by Rangayyan and Wu [53]. It is computed by firstly obtaining the mean squared values in a 5-ms windows, and then taking the variance of the obtained signal.

In some works, a greater number of features was extracted in the windowed manner. For example, Gharehbaghi et al. [100] extracted 72 features per window, including frequency, as well as power and cepstrum features. Each *feature signal* was further reduced to three values, namely its mean, median and standard deviation. In a work by Semiz et al. [98], authors also obtained a set if 50 windowed features, however, did not reduce its dimensionality. The feature matrix was provided as an input to a classifier.

A *feature signal* characterize the dynamics of some specific feature of the vibroarthrogram. Therefore, approach of describing this signal using some statistical features, and therefore losing the temporal information, seem to limit its informativeness. On the other hand, considering all of the time points of the *feature signal*, and treating them as separate features, seem to dilute their characteristics and may lead to redundancy of information. Therefore, an approach of describing each *feature signal* using some features, that can capture dynamics of a signal, seem to be of potentially great value. This concept will be more broadly described in Chapter 7.

3.5 Feature extraction

Most studies in the VAG domain focus on searching, or defining, the most informative features. For the sake of this review, they will be broadly divided into four categories, or families:

1. statistical and distribution features,
2. entropy, complexity and shape features,
3. frequency features,
4. time-frequency and windowed features.

Since this dissertation is also focused in a great part on the feature extraction, various VAG features will be only mentioned here, and described in more detail in later chapters. Summary of the feature categories used across literature was provided in Table 3.8.

Table 3.8: Summary of the feature families used across VAG literature.

references	feature family
[5, 14, 15, 35, 37, 39–42, 49, 51, 52, 56–58, 60, 68, 71, 74, 76, 77, 81, 84, 86–88, 90–93, 95, 97, 121, 123, 124, 126, 128]	statistical and distribution
[5, 14, 15, 37, 41, 43–45, 47, 51–55, 57, 58, 60, 61, 63, 65, 68, 71–74, 84, 86–88, 94, 96, 98, 100, 103, 109, 111, 114, 121, 122]	entropy, complexity and shape
[34, 36, 38, 41, 44, 63, 64, 72, 75–77, 81, 82, 92, 98, 100–103, 116–118, 121, 124, 128]	frequency
[5, 35, 37, 40, 42, 49, 53, 54, 57, 58, 60, 66, 69, 71–74, 78, 79, 84, 86, 90, 91, 93, 95, 97–100, 103, 106–108, 111, 115, 121–123]	time-frequency and windowing

Statistical and distribution features are measures that describe distribution of signal's values. They include features such as average, rectified average, median, standard deviation and variance, coefficient of variation (standard deviation divided by the mean), skewness, kurtosis, higher central moments, peak (maximum) and peak-to-peak (maximum minus minimum) values, root mean square or Shannon entropy.

Entropy, complexity and shape features are other features defined in the time domain, that describe dynamics of the signal in various ways. They include measures like zero crossing rate, Hjorth's parameters, turns count, Recurrence Quantification Analysis parameters, fractal dimension, fractal scaling index, various entropies, such as Tsallis, Permutation, Approximate, Sample, Symbolic or Fuzzy Entropy and various shape factors, such as crest, impact, rapid change or margin factor. Finally, autoregressive features are also included in this category.

Frequency features are generally defined in the spectral domain. They mostly include features describing power in specific frequency bands or their ratios. There are also some features describing the shape of the spectrum, such as median or peak frequency, spectral or Wiener entropy. Finally, some authors use cepstral coefficients as VAG features.

Time-frequency and windowed features are defined on signal divided into short windows, as described in previous section. Those features include mostly spectral fluctuations, i.e., changes of the spectrum in time, as described by various spectral features. Those include spectral centroid, spread, skewness, kurtosis, entropy and other statistical measures calculated for the normalized spectrum. However, different features defined on the windowed signal, such as variance of mean squared signal, are also in this category, as well as features defined on time-frequency distributions, such as spectrogram. Finally, features involving envelope calculation are classified as windowed features.

3.6 Feature selection and dimensionality reduction

After the feature extraction, and before the classification, the dimensionality of the feature space can be reduced. This step lowers the complexity of the classification, while potentially increasing generalization capabilities by minimizing the risk of overfitting. There are two general approaches of reducing the feature space. The first one, feature selection, consists of finding a subset of features that are the most informative and, at the same time, possibly the least redundant. The other approach, called in the rest of this review feature transformation, consists of transforming the features using some statistical or machine learning approaches, that can use information from all the features, potentially providing more informative features that the feature selection algorithms. However, transforming the features renders them difficult to interpret.

In the vibroarthrography filed, the first approach is much more popular. It can be further divided into all-relevant and minimal-optimal feature selection. The all-relevant approach consist of selecting features on their information content only, without taking into account their redundancy. For example, n best features can be selected, like in work by Rangayyan and Wu [54]. They selected two, four and six best features, according to their informativeness measures. In the previous study analyzing the same signal base as in this dissertation, Łysiak et al. [116] used one best feature describing each class pair. Another all-relevant approach selects all features above some informativeness measure threshold. Karpiński [88], for example, selected all features that showed statistically significant differences between groups.

Minimal-optimal feature selection approaches select the features iteratively. They can be further divided into backward and forward feature selection. The former consists of

iteratively removing singular least-informative features from the selected features set, starting with all the features. That approach was used, for example, by Mascarenhas et al. [70], who discarded the least important features describing sub-signals obtained by the Tunable Q Wavelet Transform. In the forward feature selection approach, the singular, most informative features is selected and iteratively aided by additional ones. That approach seem to be more common in the VAG literature. Gharehbaghi et al. [100], for example, used it to reduce the number of features used for classification from 216 to 44.

There are also another, less popular approaches. For example, Gong et al. [41] clustered the features base on their correlation. In a two-part study by Karpiński et al. [14, 15], the neighbor component analysis was used to select the features. Some studies, for example by Kręcisz and Bączkowicz [5], utilized the genetic algorithm approach to select the optimal feature subset. Finally, small number of features in the original feature set enables exhaustive searches for the subsets. Al-Timemy [65] extracted three features, which enabled evaluation of all seven feature subset combinations.

Note, that the feature selection approaches always consist of somehow evaluating the feature's informativeness. There is a plethora of methods providing some kind of informativeness measure in the literature. Comparison of selected methods in the classification context was conducted in an earlier study [133], and method that proved to be the best will be described in more detail in Section 8.1.

Dimensionality reduction algorithms consist of transforming the original feature set into a set of smaller number of features, which are composites of various features of the original feature set. For example, Principal Component Analysis works by identifying the orthogonal directions, in which the feature values vary the most. Then, the original feature values are projected onto these directions, resulting in a transformed dataset of lower dimensionality. In the VAG literature, it was used by Ozmen et al. [73]. Another example of the dimensionality reduction approach was utilized by Hersek et al. [72] and later by Richardson et al. [103]. They used nearest neighbors graph to compute one composite feature.

Summary of the feature dimensionality reduction approaches was provided in Table 3.9.

Table 3.9: Summary of the dimensionality reduction approaches used in the VAG literature.

references	approach	details
[54, 56, 57, 67, 71, 88, 116]	feature selection	thresholding
[61, 70]	feature selection	backward feature selection
[61, 99, 100, 121]	feature selection	forward feature selection
[41]	feature selection	correlation clustering
[14, 15]	feature selection	neighbour component analysis
[5, 51, 52]	feature selection	genetic algorithm
[65]	feature selection	all possible feature combinations
[73]	dimensionality reduction	principal component analysis
[72, 103]	dimensionality reduction	graph-based feature

3.7 Classification

All classification algorithms consist of predicting the class label, given some information about the signal, quantified by the features. Various classifiers were used in the vibroarthrography filed, and their summary was provided in Table 3.10. Short description of each model will be provided in this section.

One of the most popular models used for classification in the VAG literature is a logistic regression. Given a set of features, it calculates the probability of a certain binary event occurrence, by passing the output of the linear regression function through a sigmoid function. That constraints the prediction of the class to the range from zero to one, which can be interpreted as a probability of the given class. In the vibroarthrography field, it was used in both early references, like Krishnan et al. [45] (1997), and more recent ones, like Ozmen et al. [73] (2021).

Another popular classification algorithm is the Linear Discriminant Analysis. It consists of finding a linear combination of features that maximizes the separation between classes, by calculating a decision boundary. This boundary maximizes the distance between mean values of classes and minimizes variance of the values within those classes. This algorithm was also widely used in the VAG domain. For example, Rangayyan and Wu [52] used it for screening purposes. The Linear Discriminant Analysis algorithm assumes that all classes have the same covariance matrices, i.e., values of the features have similar variability across the classes. If this assumption cannot be made, the Quadratic Discriminant Analysis algorithm can be utilized. In the VAG domain, it was used by Wu et al. [114].

There are also classification algorithms based on the Beyes' theorem used in the vibroarthrography literature. Maximal Posterior Probability algorithm classifies the unknown sample to the class with the highest posterior probability, i.e., the class that appears most likely to have produced this sample, given the probabilities of the classes of known samples. This approach can be used when all misclassification costs are equal and numbers of observations within the classes are approximately the same. It was used by Wu et al. [109] for screening purposes. More general approach, the Bayesian Decision Rule, can be used to additionally take into account costs associated with different classifications. It assigns a sample to the class that minimizes the expected cost, which is calculated as the sum of the costs of decisions weighted by their probabilities. It was used by Yang et al. [111] on the same signal dataset, providing better screening accuracy (86.67% in [109] and 88.00% in [111]).

One of the most intuitive classification algorithms is the k-nearest neighbors algorithm. It consists of determining the distance between the feature vector to be classified and the already classified feature vectors. Then, considering k closest feature vectors (that is, nearest neighbors), algorithm classifies the feature vector to the most common class among the chosen neighbors. There are two crucial parameters, that need to be chosen in this algorithm. The first one is k, the number of neighbors to be considered. Smaller values make the model more sensitive to nuanced differences between classes, but also more sensitive to the noise. Greater values, on the other hand, make the finer class distinctions difficult, but in general, make the classification more robust to noise. Another crucial parameter is the distance calculated between the feature vectors, with the most usual choice being the

Euclidean distance. The k-nearest neighbors algorithm was used in a number of VAG studies. For example, Gong et al. [41] used it to classify knee joints into junior and senior classes. Additionally, they compared it to the linear regression classifier, proving the superiority of the k-nearest neighbors in their setup.

More complex machine learning classifier commonly used in the VAG domain is the Support Vector Machine. In summary, it constructs an optimal hyperplane separating classes in the feature space. It may seem similar to the Discriminant Analysis, however, instead of maximizing the distance between means of classes and minimizing variance within classes, it maximizes maximum marginal distances between the classes values. Moreover, the Discriminant Analysis approaches assume, that the feature's values follow Gaussian distribution. This assumption is not necessary for the Support Vector Machine. It was used by Befrui et al. [75] for knee joint screening purposes. If the data is not linearly separable, the nonlinear Support Vector Machine can be implemented, which transforms the original feature space to higher, linearly separable, dimension. It was used, for example, by Zheng et al. [74]

In comparisons to the Support Vector Machine, Two-Surface Proximal Classifier also attempts to construct separating hyperplanes for classification. However, the process consists in construction of not one, but two hyperplanes that are designed to enclose the data of each class separately. These hyperplanes are positioned such that they are closer to their own class feature vectors than to the vectors of the other class. Then, a classification rule is to assign an unknown feature vector to the class of the closest hyperplane. This approach was used by Mu et al. [51] for VAG screening purposes.

Neural networks use slightly more complex architecture to achieve more flexible, nonlinear classification boundaries. There is a plethora of neural network architectures used in literature as classifiers. In general, they are composed of stacked layers of neurons. Each neuron is characterized by at least one input and at least one output. Weighted sum of the inputs is then transformed by so-called activation function, which output becomes the output of the neuron.

The most basic neural network structure is called Multilayer Perceptron and consists of at least three layers of neurons. The first layer is called the input layer and it receives the feature vectors. It is followed by one or more hidden layers, after which the output layer is positioned, usually delivering the class labels. The role of the hidden layers is to transform the data into a space, where the classes can be separated more easily. It can be done utilizing some (commonly nonlinear) activation functions. Usually, these activation functions are monotonic, such as the hyperbolic tangent. Such structured neural network was used by Karpiński et al. [15]. The activation function does not need to be monotonic, however. It can also use the radial basis function, i.e., a function, which value depends only on the distance between the input and some fixed point. An example could be the Gaussian function, which output decreases with distance from the center. Multilayer Perceptron utilizing Radial Basis Function as an activation layer is usually called Radial Basis Function Neural Network. It is one of the most widely used neural network structures used in the VAG literature, with an example usage for the screening purposes by the Rangayyan and Wu [52]. Another type of architecture is a covolutional neural network. Those networks usually take raw (or somehow preprocessed) data as an input and perform feature extraction implicitly, through the use of convolutional layers. They are usually used in the image processing context. Kraft and Bieber [69] used them for VAG screening, feeding them spectrogram pictures of the VAG

signals.

Described classifiers all constitute singular learner strategies. Random Forest, on the other hand, provides an ensemble methodology. It consists of multiple simple classifiers, called *weak learners*, which singular classifications are later aggregated. The primary concept behind Random Forests is to aggregate predictions of multiple decision trees. A decision tree is a simple structure, organized in a hierarchical manner. Each node of the tree corresponds to a feature and each branch represents a decision rule. Those structures, however, are prone to overfitting, especially with high-dimensional features. To mitigate this problem, Random Forests employ a number of decision trees, each generated from a subset of the training data. Singular classifications are then aggregated, for example by taking the average classification label or taking a label with the greatest number of votes. Random Forests were used in VAG literature quite commonly, for example, by Nalband et al. [63].

Another ensemble classification methods are called Boosting algorithms. In a contrast to Random Forests, instead of building weak learners independently from one another, Boosting algorithms build them sequentially. Each learner is trained to correct the errors made by the previous learner, which enables more precise classification. In the VAG literature, the XGBoost is the most popular boosting algorithm. In XGBoost, each new tree is fit on the residual errors of the previous tree to gradually minimize the loss function. The process is guided by the principle of gradient descent, where at each step, the algorithm moves in the direction of steepest descent in the error surface. It was used, for example, by Semiz et al. [98] for Juvenile Idiopathic Arthritis diagnosis.

Table 3.10: Summary of the classification algorithms used in the VAG literature.

references	classifier
[5, 41, 44, 45, 47, 49, 73, 85, 99, 100, 114]	logistic regression
[50, 52–54, 56, 57, 61, 68, 109, 116, 121]	linear discriminant analysis
[114]	quadratic discriminant analysis
[109]	maximal posterior probability
[111]	bayesian decision rule
[41, 85, 116]	k-nearest neighbors
[5, 55, 57, 58, 60, 66–68, 70, 74, 75, 85, 109, 111, 114, 116, 121]	support vector machine
[51]	strict 2-surface proximal classifier
[5, 14, 15, 52–54, 56, 61, 69, 78, 79, 85, 88, 116, 121]	neural network
[5, 63, 70, 71, 116]	random forest
[65, 98, 116, 121]	boosting algorithms

Even though there are multiple VAG-related works studying more than two groups in the literature (see Table 3.1), most studies regarding classification implement binary approaches. That is, the classifier differentiates two classes, either as a screening tool—distinguishing healthy knee joint from *somehow* unhealthy or injured one—or as more specialized one, distinguishing some specific disease. There are some exception to that, however. For example, Krishan et al. [45] employed discriminant analysis to classify signals into 6 classes: normal silent, normal noisy, arthroscopically normal, arthroscopically abnormal, historically normal

silenty and historically normal noisy. In this setup, accuracy of the classifier reached 46.94%.

Another multi-class classification example is a study by Kręcisz and Bączkowicz [5], signals from which are also analyzed in this dissertation. They implemented multiple classifiers, namely logistic regression, neural network, support vector machine and random forest. Accuracy of the 5-class classification reached up to 69%, for both logistic regression and neural network. In another study utilizing the same signal base, Łysiak et al. [116] achieved up to 67% accuracy using support vector machine.

Regardless of whether the task is binary or multi-class classification, the challenge of overfitting persists. Overfitting refers to a model fitting to the training data too closely, learning the noise along with the desired information. The overfitted model is unable to generalize to new feature vectors, poorly predicting the actual class. To reduce this problem, cross-validation strategies are typically employed. They consist of partitioning the original samples into a set of *folds*. The model is then trained on all but one fold and validated, or tested, on the remaining one. This process is iterated, with each fold serving as the test set once. The average classification performance across all iterations is then reported as a predictive capability of the model. A specific case of cross-validation, the leave-one-out strategy, is especially relevant in scenarios where the dataset is small. In this strategy, the model is trained on all samples but one, which is left out for validation. The process is then repeated for each sample in the dataset, providing a comprehensive assessment of the model's predictive power. This strategy was implemented in the overwhelming majority of VAG studies regarding classification, and as so, will also be utilized in the rest of this dissertation.

Part II

Methods

Chapter 4

Preprocessing

In this chapter, the preprocessing methodology will be described. The first section will be devoted to the general introduction to the preprocessing. Then, two key characteristics of signals will be shortly discussed, i.e., the normality and stationarity. In the following section, a description of the Butterworth filter and the differentiation operation will be included. Finally, the last section will be devoted to the comparison of the chosen preprocessing methods.

4.1 Introduction

Preprocessing is an essential step in the signal classification pipeline, which, in general, consists of transforming a signal into another signal of improved quality. The main goal of the preprocessing is to remove noise, i.e., to remove such changes in the signal, that do not correspond to the classified condition. There is an additional goal, which does not necessarily takes the classification context into consideration. This goal is to transform the raw signal into a preprocessed signal that fulfills some conditions needed to be met before further feature extraction. For example, some features, in order to be interpretable, require values of a signal to follow a normal distribution. Another requirement can be stationarity, which is a property assuring the statistical properties of a signal remain constant over time. As features are extracted on the preprocessed signals, the influence of the specific preprocessing method on a signal's information content is of great importance.

There are various preprocessing methods used in the VAG literature, as listed in the Chapter 3. In this dissertation, two specific methods will be utilized, namely the Butterworth filter and the differentiation operation. Additionally, the features will be extracted on the raw signal, i.e., the signal without preprocessing. This chapter will include details of the chosen methods and their influence on the signals' properties. In the following section, the normality and stationarity properties will be described, followed by more detailed description of the utilized preprocessing methods.

4.2 Normality and stationarity of a signal

4.2.1 Normality

In the vibroarthrogram classification context, the distribution of signal values can be a highly informative descriptor of the underlying knee-joint condition. These values, depending on their occurrences, form specific patterns that can be visualized as a distribution curve. Visual inspection of this curve often reveals the general variation of values within the signal, like

clustering of values around a central point, symmetry, or skewness. In a normally distributed signal, the majority of the values gravitate around a central mean, and progressively fewer values are found as the distance from the mean increases, resulting in the characteristic bell shape. This distribution pattern is of great importance, since any deviations from it can be intuitively explained. For example, signals generated by a highly damaged, "squeaky" knee joint would contain high-amplitude spikes making the distribution of values heavy-tailed. On the other hand, healthy, "silent", knee joints would produce signals with less high amplitude spikes, leading to more normal-like distribution.

Popular statistical tests for normality, such as Kolmogorov–Smirnov test or Shapiro-Wilk test, are very useful tools for normality assessment in moderate sample sizes. However, with larger sample sizes ($n \geq 300$), they can be over-sensitive [134]. That is, even very small deviation of the sample from normal distribution, causes them to reject the null hypothesis of normality. This can lead to an overemphasis on minor irregularities that may not significantly affect subsequent analysis. Since each vibroarthrogram analyzed in this dissertation is over 60-thousand sample long, using normality tests to compare signals after each preprocessing would not be of great value. Instead, visual methods of quantile-quantile plots and histogram will be used, along with skewness and kurtosis measures.

Quantile-Quantile (QQ) plots consists of plotting the sorted values of the sample against the expected values of the corresponding quantiles from a normal distribution. In these plots, if the data points largely follow the 45-degree reference line, this suggests that the sample data has a distribution similar to the normal one [135]. Histograms, on the other hand, provide a visual representation of data distribution by grouping a range of values into bins and showing the frequency of observations in each bin. A normal distribution can be additionally overlaid on the histogram, allowing more direct comparison of the sample and the normal distribution shapes [135]. Finally, summary statistics such as skewness and kurtosis offer quantitative measures of a distribution's shape. Skewness measures the asymmetry of a distribution around its mean. Values close to zero indicate symmetric distribution, while negative and positive values indicate left- and right-skewed distribution [135]. Kurtosis quantifies the heaviness of the tails of a distribution, with a value close to zero suggesting a distribution similar to the normal one in terms of its tail behavior. Positive values of kurtosis indicate a distribution with heavier tails and a higher peak compared to a normal distribution, implying that the distribution has more extreme values or outliers than a normal distribution. On the other hand, negative values of kurtosis indicate a distribution with lighter tails and a flatter peak compared to a normal distribution, suggesting that the distribution has fewer extreme values or outliers than a normal distribution [135].

In the context of signal analysis, the normality can be a very useful concept. However, there are some major drawbacks to exclusively using measures describing the distribution. A signal represent change of some value in time. Focusing solely on distribution results in loss of information regarding the dynamics of a signal. For example, a signal composed of randomly generated normally distributed values would have the same normal distribution as a signal composed of the same values, but sorted. Their temporal ordering, or dynamics, however, would be completely different. When the distribution of a signal is examined as a whole, it does not specify which part or period within the signal contributes most significantly to this distribution. A more comprehensive understanding of the signal may be achieved by assuming that the signal's statistical properties, such as mean and variance,

remain consistent over time, across different segments of the signal. If such assumption is correct, the signal is considered to be stationary [136].

4.2.2 Stationarity

Stationarity is a fundamental property in signal analysis, referring to a signal maintaining its statistical characteristics, such as mean and variance, over time [10]. As mentioned in previous section, this property is of great importance due to the inherently temporal aspect of signals. Ensuring the statistical properties observed in one segment of the signal are likely to remain consistent in another segment allows more reliable interpretation of extracted features. In stationary signals, the statistical parameters of the signal are representative of the entire signal, despite the time-dependent nature of the data. The vibroarthrographic signals are non-stationary by nature [10]. However, making them *more* stationary, could potentially yield more informative features.

There is a plethora of strict stationarity tests used in the literature, with the most widely used being Augmented Dickey-Fuller, Kwiatkowski-Phillips-Schmidt-Shin or Phillips-Perron [137]. However, they are mostly concerned about unit-root non-stationarity (i.e., resemblance of a signal to a random walk) or trend-stationarity (i.e., presence of a trend in a signal) [136]. These forms of non-stationarity may not adequately capture or address the types of time-dependent structure that are of interest in biomedical signal analysis fields, such as vibroarthrography, where changes in physiological states can result in complex, nonlinear dynamics [10].

Therefore, stationarity will be quantified differently in the remaining part of this chapter. First, to quantify changes of summary statistics, the signal will be windowed, i.e., split into 5-ms, or 50-sample segments[1]. In those windows, summary statistics of the mean value and variance will be calculated. Finally, variability of the obtained signals will be quantified by the coefficient of variation [138], i.e., standard deviation divided by the average value. This coefficient is used mostly for non-negative samples, since average values close or equal to zero could render it extremely large or undefined [10]. And since vibroarthrograms are roughly symmetrical, before calculating the coefficient, signals obtained by the rolling statistics are normalized to scale of zero to one.

Therefore, specific values used to quantify stationarity of a signal will be called Coefficient of Variation of the Normalized Moving Average (CVNMA) and Coefficient of Variation of the Normalized Moving Variance (CVNMV). Higher values of the CVNMA indicate more changes in the average value across the signal, while higher values of the CVNMV indicate more changes in the variability across the signal. Importantly, the CVNMV value does not provide information about the general variability of a signal, but rather *consistency* of variability across the signal.

One would argue, that the fundamental limitation of using the CVNMA and CVNMV measures is similar to normality limitation described in the previous section. That is, coefficient of variation does not depend on temporal ordering of samples, and therefore cannot be used to describe *dynamics* of a signal. However, segmenting signals, and subsequently

[1]This window size was chosen since 5-ms window was consistently used in the VAG field [53], in various measuring setups, including signals analyzed in this dissertation [5].

extracting rolling features, can provide information about the dynamic characteristics of the signal that are crucial for assessing stationarity. In other words, in the context of stationarity, the critical aspect is not the specific way in which a given statistic changes over time, but the overall degree of variation it exhibits throughout the entire signal's length.

4.3 Description of the preprocessing methods

In this section, specific preprocessing methods will be described, focusing on the Butterworth filter, and the differentiation operator. Both preprocessing methods were chosen following evaluation of their properties, interpretability, and successful applications in similar contexts. Additionally, in the following sections, the no-preprocessing will also be treated as a preprocessing method, to keep in line with the signal analysis pipeline described in the Chapter 3. Unprocessed, or raw signals, though seemingly less informative, present an important baseline to which the effectiveness of other preprocessing methods will be compared. The following subsections will focus on the Butterworth filter and the differentiation operation. Their definition in time will be included as well as frequency response in form of the magnitude response and the phase response.

The magnitude response is a measure of how an input signal's amplitude is affected by the preprocessing method at each frequency component. It provides insights into how various frequency components in a signal are attenuated or amplified after the processing. For example, the 10-sample moving average filter would not greatly affect sine wave with a period of a hundred samples. However, a sine wave with a period of five samples would be greatly attenuated by the same filter.

The phase response represents the phase shift introduced by the preprocessing method at each frequency component. It shows how this method affect the timing of different frequency components in the signal, indicating how the phase of the signal is changed. For example, a step signal, transitioning from zero to one, could be processed through a 10-sample moving average filter. The signal after the processing would not instantaneously reach one, but gradually increase over subsequent samples until all samples in a window would equal to the step change. Analysis of the phase response is of great importance, since it can reveal temporal distortions introduced by the preprocessing method, potentially influencing the interpretability of subsequently extracted features.

4.3.1 Butterworth filter

The Butterworth filter is a frequency selective filter, that is, it is designed to pass or attenuate specific frequency components of the signal. Compared to different types of such filters, like Chebyshev or Elliptic, which all have a characteristic ripple in a pass-band, the Butterworth filter has by maximally flat pass-band [10]. However, compared to those other filter types, the Butterworth has relatively slow roll-off, or low steepness, around the cutoff frequency. Increasing the order of the filter can make the roll-off more steep. However, it also increases the phase distortion. The general disadvantage of the Butterworth filter is its nonlinear phase response. This issue can be easily mitigated, however, by using zero-phase filtering scheme. It consists of successively filtering the signal two times using the same filter, but

for the second filtering, the signal is reversed in time [10]. Notably, zero-phase procedure cannot be used in real time, but this is not an issue for the off-line approach took in this dissertation.

The design of a digital Butterworth filter is usually done in the continuous space. Then, using bilinear transformation, the transfer function in the s-domain is transformed to the z-domain. Finally, the filter coefficients b_k and a_k can be obtained from the transfer function, allowing to represent filter in the time domain [10]:

$$y(n) = \sum_{k=0}^{N} b_k \cdot x(n-k) - \sum_{k=1}^{N} a_k \cdot y(n-k), \tag{4.1}$$

where y is the filtered version of a signal x, and N is the order of the filter. Here the summations represent the fact that the filter's output at any time point n is a linear combination of the present and past input values, as well as past output values, weighted by the filter's coefficients.

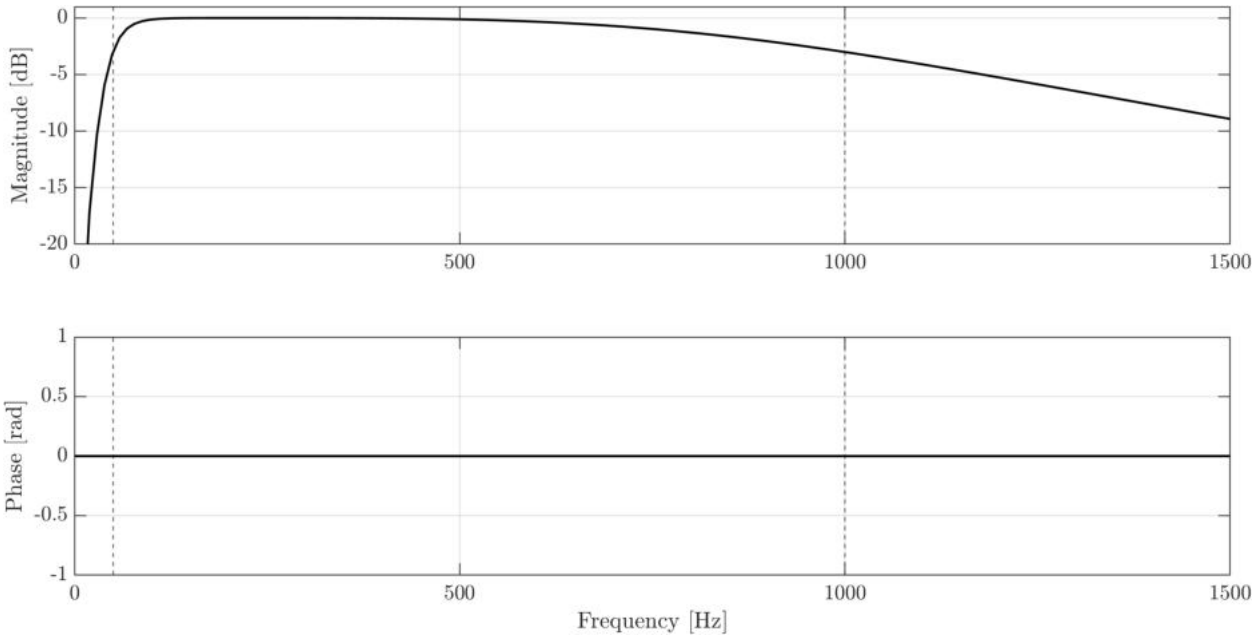

Figure 4.1: Frequency response of the 4-th order zero-phase Butterworth filter: the magnitude response (upper plot), and the phase response (lower plot). Dotted lines correspond to the filter pass-band: 50 Hz and 1000 Hz.

A specific implementation of the Butterworth filter used in this dissertation will be 4th-order and zero-phase, i.e., two 2nd-order filters cascaded in series, ensuring that the overall filter has a zero-phase response. The filter was designed to pass frequencies between 50 Hz and 1000 Hz, as in previous studies utilizing the same base of vibroarthrograms [5]. The magnitude and phase response of this filter were plotted in Figure 4.1.

4.3.2 Differentiation

Difference operator, in contrast to the Butterworth filter, is not frequency-selective. Instead, it can be generally understood as high-pass filter, with the magnitude response increasing from zero for the constant component (frequency of zero) to two for the Nyquist frequency. The difference operation is widely used in the literature to filter out low-frequency noise, such as the baseline drift [10], however, it can potentially make high-frequency noise more prominent in the filtered signal. In the time domain, it is simply defined as [10]:

$$y(n) = x(n) - x(n-1), \tag{4.2}$$

where y is the filtered version of a signal x.

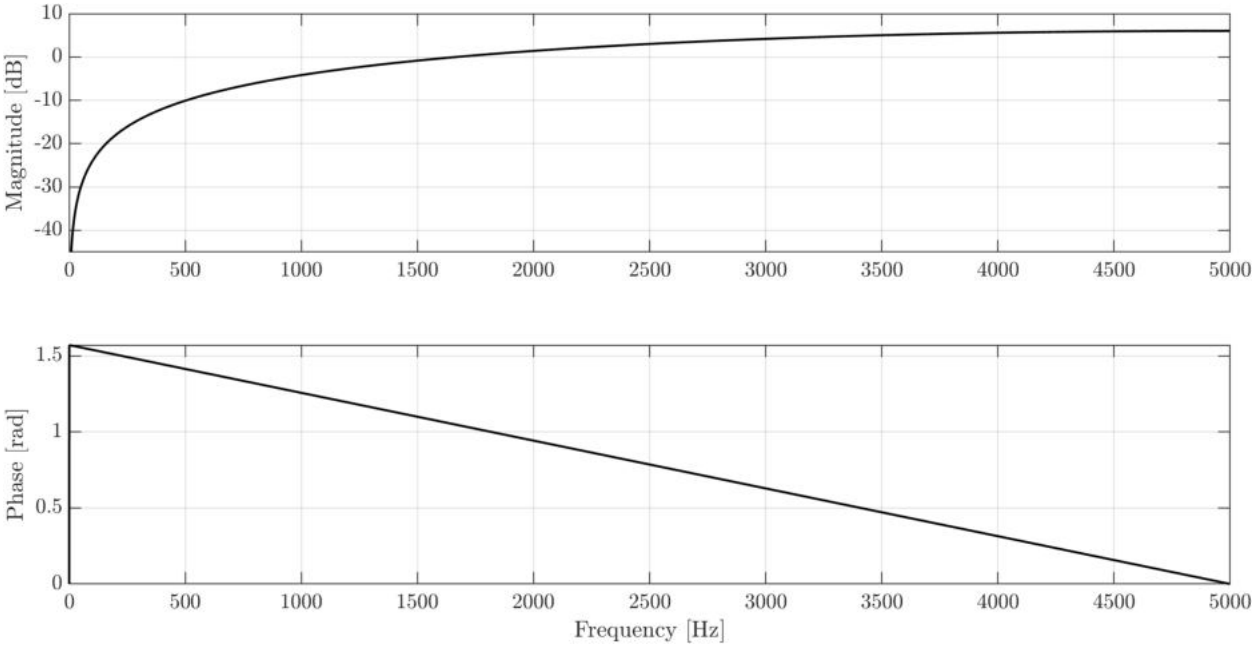

Figure 4.2: Frequency response of the difference operator: the magnitude response (upper plot), and the phase response (lower plot).

The frequency response of the difference operator was included in the Figure 4.2. Note, that the phase response is not constant like in the zero-phase Butterworth filter. It decreases linearly from $\pi/2$ to zero, indicating that the lowest frequencies are delayed in the filtered signal by about $1/4$ of their period. Consecutive frequencies are less delayed, up to the Nyquist frequency, which is not delayed at all.

4.4 Comparison of the preprocessing methods

Exemplary raw, Butterworth filtered and derivative signals, with their corresponding power spectral densities were included in Figure 4.3. To estimate the power spectral density, periodogram was used. More detailed description of this approach will be provided in Chapter 6.

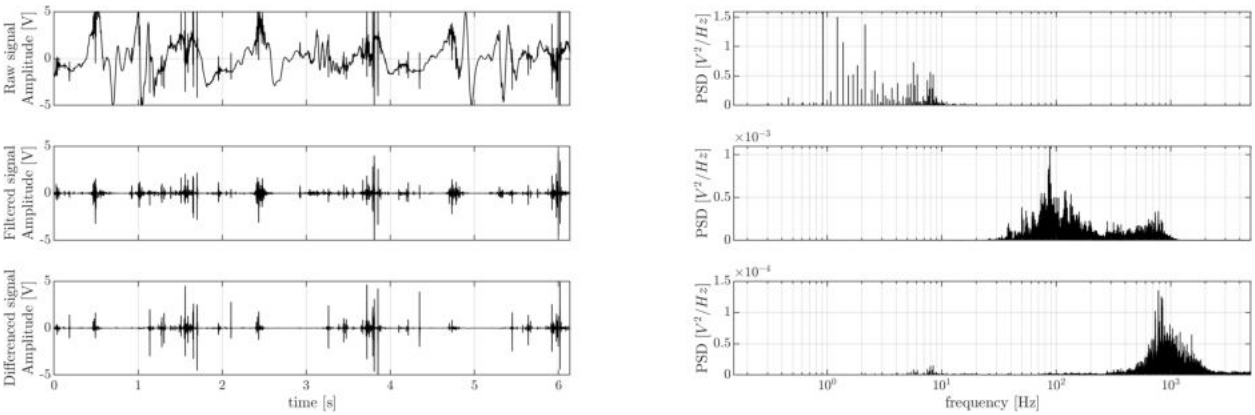

Figure 4.3: Exemplary vibroarthrogram of the osteoarthritis group (upper row) after filtering with the Butterworth filter (middle row) and after the difference operation (lower row) in the time domain (left column). Corresponding periodograms were plotted in the right column.

The most notable difference in the time domain plots is perhaps the lack of low-frequency baseline drift in both filtered signals, compared to the raw signal. Comparison of both preprocessing methods reveals generally more strict filtration of the low and medium frequencies by the difference operator, as well as filtration of very high frequencies by the Butterworth filter. See, for example, a big spike at a time of about half a second. Firstly, in the Butterworth filtered signal it is more spread out in time, it appears "fatter". In other words, it is composed of more medium-frequency components. Also, see that in the derivative signal, this spike is proceeded by a spike of similar amplitude, but shorter duration. Note, that it was almost completely attenuated by the Butterworth filter. On the other hand, a spike at around one-second mark, which is clearly visible in the Butterworth filtered signal, is almost completely lost in the derivative signal. It can be explained by its low-frequency components, which were attenuated by the difference operator. See also, that spikes which can be seen on both signals, such as the spike just after one-second mark, have clear differences in amplitudes. This difference corresponds to the duration of a spike. The longer it lasts, the lower its frequency, and hence, the more it gets attenuated by the difference operator. On the other hand, short spikes, or high-frequency spikes, get attenuated by the Butterworth filter and amplified by the difference operator. Similar conclusions can be derived from visual inspection of the remaining part of the signals.

Frequency domain plots reveal similar conclusions. Lower frequencies, up to about 10 Hz, dominate the raw signal. The Butterworth filtered signal, according to its design, occupies frequency range of 50 Hz to 1000 Hz. Frequencies of about 500 Hz to 2000 Hz are mostly visible in the derivative signal. Note, however, that while Butterworth attenuates all frequencies but the pass-band, the difference operator passes some of the more pronounced low-frequency components, such as those just below 10 Hz.

Chapter 5

Signal analysis in time domain

In this chapter, features extracted from the signal in the time domain will be described, broadly divided into five categories:

- statistical, including features describing the signal without taking into consideration temporal ordering of samples,
- rolling, including features calculated on the basis of some *feature signals*,
- difference and frequency, including features based on temporal fluctuation of the signal's values,
- self-similarity, including features based on Autocorrelation Function and Multifractal Detrended Fluctuation Analysis,
- phase space, including features based on the embedded signal.

This division of features is introduced primarily for the sake of clarity. Although it is not completely arbitrary, it should not be considered as fixed.

5.1 Statistical features

Features in this category do not extract from signal any information about the temporal ordering of its samples. In other words, the signal is treated as a collection of unrelated values, disregarding any potential patterns that may exist among them. For example, a signal will have the same standard deviation as its sorted version, as well as its randomly permuted version. Nevertheless, in the context of VAG classification, statistical measures are widely used and, in many cases, prove to be quite informative. For the sake of clarity, statistical features will be further divided into basic features, which are calculated directly on the signal's samples, and shape features, which are defined using the basic features.

5.1.1 Basic statistical features

Basic statistical features include measures of central tendency, such as mean value, mode and median. Since distributions of the vibroarthrograms analyzed in this work are slightly skewed (see Chapter 9), including multiple measures of central tendency could potentially reveal different information. The mean value (μ) indicates the baseline around which the signal fluctuates. Median, on the other hand, is a value, that separates high signal values from low signal values. Since VAG signals have quite a lot of extreme values, median can

indicate typical values more robustly than mean. Mode, in contrast to mean or median, indicates which one specific value appears most frequently in a signal. VAG signals analyzed in this dissertation were digitized by the 12-bit ADC, and therefore can take only 4096 values. And because they last for over 60 thousand samples, some discrete values have to occur multiple times. Therefore, even though the samples can generally be interpreted as real values, their mode can be determined.

Vibroarthrograms are generally symmetrical, therefore central tendency measures are not often used in VAG literature. Nevertheless, some studies successfully used them. For example, Karpiński et al. [14, 15] used the average value of the signal samples, while Scherpereel et al. [121] used both mean and median. Rangayyan and Wu [56] used mean value of the estimated probability density functions. In all those studies, however, central tendency measures were used along some other features.

Features measuring power of the signal are defined similarly to central tendencies, but are calculated on the *magnitude* of the values, i.e., discarding information about the sample's sign. Three such features will be used: the Average Signal Power (ASP), the Root Mean Square (RMS) and the Rectified Average (RAV). The ASP is defined as the average value of the squared signal values. As the name suggests, it provides information about the *power* of the signal. Its square root can be interpreted as the average magnitude of the signal and is called the RMS. Finally, the RAV feature is calculated as the average value of the signal's absolute values. It is similar in interpretation to the RMS, but since the RMS include squaring the values, it is more influenced by the extreme samples. Therefore, RAV is generally smaller than the RMS. This difference in sensitivity to extreme values means that RMS and RAV provide complementary information about the vibroarthrogram. The RMS can distinguish signals with large peaks or spikes, while the RAV can provide a more general measure of the signal's overall activity.

Calculation of power measures in VAG domain is much more popular, with the RMS feature being the most common choice [14, 15, 68, 88, 100, 121], followed by the RAV [14, 15, 88, 121]. The ASP feature was used less often [100, 121]. Both the central tendency and the power features, with their explicit definitions, were summarized in Table 5.1.

Table 5.1: Summary of the central tendency and power statistical features.

name	definition	remarks
Mean	$\mu = \frac{1}{n}\sum_{i=1}^{n} x(i)$	$x(i)$ is the i-th sample of the n-sample signal x.
Mode	$m = \max_i \{f_X(i)\}$	f_X is a frequency of values in the signal.
Median	$M = \frac{(x)(\frac{n}{2})+(x)(\frac{n}{2}+1)}{2})$	(x) is the sorted set of signal values.
Average Signal Power	$ASP = \frac{1}{n}\sum_{i=1}^{n} x(i)^2$	
Root Mean Square	$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x(i)^2}$	
Rectified Average	$RAV = \frac{1}{n}\sum_{i=1}^{n} \lvert x(i)\rvert$	

Central tendency and power features indicate some general levels of a signal, disregarding information about the dispersion of the sample values around those general levels. In order to measure this dispersion, spread features can be used. Perhaps the most straightforward measure of dispersion is the variance. First, each signal sample is subtracted from the its

mean value. Then, to measure both positive and negative discrepancies, each difference is squared. The variance is calculated as the average of those differences. Since units of variance are units of the original signal squared, the more intuitive measure is its square root, i.e., the Standard Deviation (std) or σ. Both variance [14, 15, 68, 81, 88] and std [56, 68, 121] are popular measures used in VAG domain. Measure similar to the std is the Mean Absolute Deviation (MeanAD). It is calculated similarly to the variance, but instead of squaring the differences, their absolute value is taken to calculate the mean. Since the MeanAD does not include squaring operation, it is less influenced by extreme values. The difference between std and MeanAD is the same as between RMS and RAV. The MeanAD is generally smaller than the std. Other spread measure, even more robust to extreme values is the Median Absolute Deviation (MedianAD). It is calculated similarly to the MeanAD, but both means are replaced by medians in its definition.

Another approach to measuring the spread of values is measuring the literal spread, i.e., the difference between the maximal and the minimal value. It was used in a number of VAG-related studies [14, 15, 88, 121]. Since it is defined on just two most extreme values, it can be considered the least robust measure to outliers. This robustness can be slightly improved, however, by replacing the lowest and the highest values by the average of the four lowest and highest values, respectively [5]. Similar approach of measuring the difference between *some* high and low value is employed in the definition of the Interquartile Range, i.e., the difference between the 75-th and the 25-th percentile. Percentile is a specific value in a distribution, below which a given percentage of the samples falls. For example, 100-th percentile is a maximum value, while a 50-th percentile represents the median. Interquartile range is therefore a "width" of the central part of a distribution containing half of the samples. However, various percentages different than 25-th or 75-th could potentially be used. In this work, therefore, various percentile ranges were used as features, in a range from 0-th to 100-th percentile with a step of ten. As a result, 55, that is, $\binom{11}{2}$, Interpercentile Range (IPerR) features were defined. To make their interpretation easier, a simple visualization was provided in the Figure 5.1 (left plot).

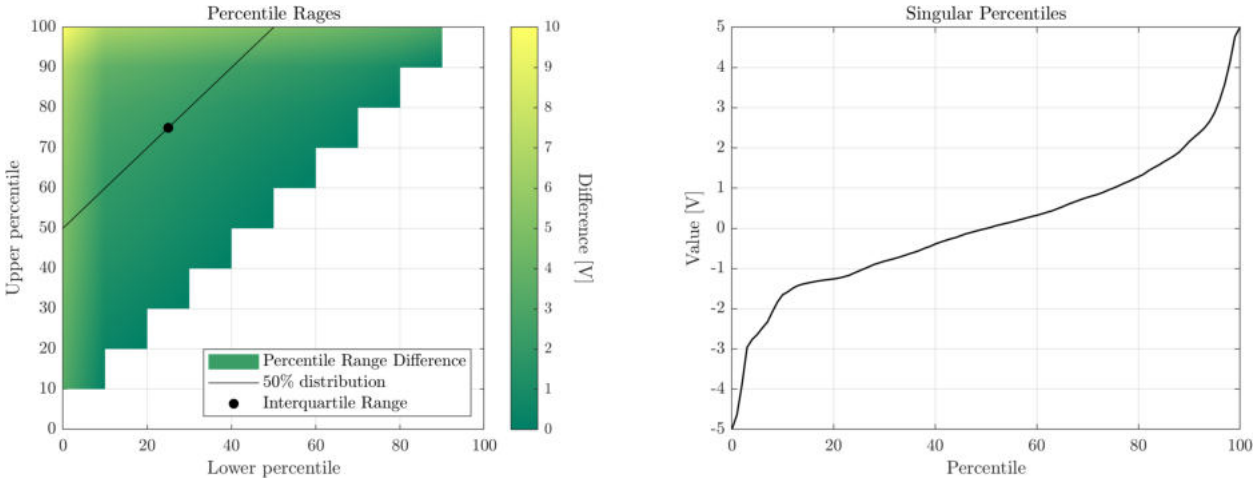

Figure 5.1: Exemplary percentile ranges (left) and singular percentiles plots (right).

The x- and y-axes in the left plot of Figure 5.1 correspond to the lower and upper range

values, respectively. The color axis indicates the specific value of the difference between the percentiles in the y- and x-axis. Interquartile Range was indicated by the black dot, and the line passing through it corresponds to various different ranges encompassing half of the samples. One could argue, that similarly to the median value, other percentiles could also be used as features independently, i.e., not in a difference with other percentiles. For example, within the VAG domain, the 0-th and 100-th percentiles, i.e., the minimal and maximal values, were used by Scherpereel et al. [121]. Therefore, in this work, 101 additional percentile values, from the 0-th to the 100-th percentile with a step of one were extracted. To make their interpretation easier, they were also plotted in the right plot of Figure 5.1, where the x-axis indicates specific percentile and the y-axis specific numerical value. Note that both plots were obtained for an exemplary signal (the same signal as in previous examples), and they contain only the information about this specific signal. In the results chapter, both the color axis in the percentile ranges plot and the y-axis in the singular percentiles plot will include informativeness of the feature instead of the specific signal's value.

Summary of the spread measures, with their respective definitions, was provided in Table 5.2.

Table 5.2: Summary of the statistical spread features.

name	definition	remarks
Variance	$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x(i)-\mu)^2$	
Standard Deviation	$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x(i)-\mu)^2}$	
Mean Absolute Deviation	$MeanAD = \frac{1}{n}\sum_{i=1}^{n}\lvert x(i)-\mu\rvert$	
Median Absolute Deviation	$MedianAD = \text{median}\left(\lvert x(i)-M\rvert\right) \forall i = 1, 2, ..., n$	
Spread	$s = \max(x) - \min(x)$	max and min denote maximum and minimum value of the signal x, respectively.
R4	$R4 = \frac{1}{4}\sum_{i=n-3}^{n}(x)(i) - \frac{1}{4}\sum_{i=1}^{4}(x)(i)$	(x) is the sorted set of signal values.
Interpercentile Range	$IPerR = QP(u) - QP(l)$	u is the upper percentile and l is the lower percentile.
Percentile	$QP(p) = (x)(\lceil \frac{np}{100} \rceil)$	(x) is the sorted set of signal values, np is the product of the total number of data points and the quantile level.

Spread measures provide general information about the dispersion of samples around some central measure. However, they do not convey any information about the shape of the distribution. Shape can be interpreted in many ways and there are many features quantifying various shape characteristics. Previously mentioned skewness, for example, measure symmetry of the distribution. Negative values reveal that the tail of the distribution is skewed to the left, i.e., that the majority of samples are concentrated on the right side of the distribution, with a few extreme values pulling the tail towards the left. Positive values indicate the contrary, while skewness equal to zero characterize distributions with equal "weight" on both

sides[1]. Kurtosis measures heaviness of the tails of a distribution, in comparison to the normal distribution, which has the kurtosis of zero. Positive and negative kurtosis values indicate a distribution with heavier and lighter tails, respectively, or a higher and flatter peak, respectively. Both skewness [41,51,52,56–58,71,74,81] and kurtosis [14,15,41,51,52,56,58,71,81,88] are popular features extracted from VAG signals.

Another characteristic of a signals shape can be how "spiky" it is, i.e., how many high-amplitude spikes it has. This can by quantified by the Spiky Index used in VAG analysis by Athavale and Krishnan [68]. It was defined as the number of prominent peaks divided by the recording time. In this work, however, number of samples will be used instead of recording time, to make the measure easily interpretable. In the original paper [68], peaks that are prominent were defined as samples of amplitude at least half of the maximal amplitude. However, choosing different threshold could potentially make to Spiky Index (SI) more informative. Therefore, in this work, multiple thresholds will be tested, from zero to one with increments of one percent. That will result in 101 distinct features. Moreover, since VAG signals are rather symmetrical, additional feature will also be tested, called the Rectified Spiky Index (RSI). It is defined similarly to the traditional SI, however, on absolute signal values. That is, it will measure *magnitude* of the signal instead of the *amplitude*. The RSI will be calculated for the same threshold range as the SI.

Entropy is another measure of the distribution's shape. It is defined for discrete distributions, as the negative sum of the specific probabilities in the distribution multiplied by their logarithm. It measures the uncertainty, and increases with the "flatness" of the distribution, achieving the highest value for the uniform distributions. Minimal entropy value is achieved for the distribution with only one value, representing a set of identical values. Signal of minimal entropy would have a constant amplitude. In such case, the entropy is equal to zero. For continuous distributions, the entropy can be estimated using, for example, the binning method, i.e., splitting the distribution into specific number of bins, and treating each real value in a specific bin as the discrete value representing this bin. Such approach was utilized in a number of VAG-related studies [5, 51, 52, 56–58, 74]. However, choosing number of bins is not a straightforward procedure and has a great influence on the estimated entropy value [139]. In the VAG literature, a hundred bins is usually used [52, 58]. However, work by Athavale and Krishnan [68] showed that encoding a signal into just three bits seem to not only keep its informativeness, but even enhance it. Authors did not extracted entropy, but their work indicated that bin number much smaller than a hundred could potentially be informative. Therefore, a range of bins was used for estimation, giving a hundred entropy estimation features for each signal: from two to a hundred bins and one additional bin of 4096 values, representing resolution of the used 12-bit ADC.

There is also an entropy measure defined for continuous distributions, called the differential entropy [140]. It is defined similarly to the discrete version, however, instead of the sum, the integral is taken. That makes it more difficult to interpret, since it is not bounded by zero and can take values from negative to positive infinity. However, in the context of feature extraction that is not a problem, since absolute entropy values are of less interest than entropy values relative to other signals. There is another issue, however. Since the

[1]Note, that distributions with the same or balanced "weight" on both sides, but different shapes would also have zero skewness.

probability distribution is not explicitly known, differential entropy needs to be estimated somehow. A popular method of estimation is using the Kozachenko-Leonenko estimator, based on the concept of the k-nearest neighbors [141].

The general idea is that the local probability density can be estimated using the reciprocal of distance between neighboring samples[2]. Choosing the number of samples, however, is not trivial. Too small values result in high statistical errors of the estimate, while too large values increase the systematic errors [141]. That is, small values make the estimation less stable, while high values tend to systematically underestimate the entropy, since they weaken the distinction between nearby and distant samples. This means that high values cause the estimation to lose sensitivity to local variations in the data, potentially missing important fine-grained patterns. Therefore, in this work, estimates based on five distinct k values were used: k equal to 1 as an extreme and simple case, k equal to 2, 3, and 4, as recommended by [141], and k equal to the rounded down square root of the number of samples, as a general rule of thumb used in the context of k-nearest neighbor approaches [142].

There is also another measure of entropy, which was used in the VAG literature, calculated directly on the samples. Log Energy Entropy [143] is defined as the sum of the logarithms of squared signal samples. In this work its normalized version will be used, i.e., using the average value instead of the sum. Since it is not calculated on the probability, its interpretation is different from the entropy used in information theory. Here, squaring the signal values creates a representation of the energy at each sample. Taking the logarithm transforms each value into a logarithmic scale, and averaging those values gives some central measure. Therefore, even though Log Energy Entropy (LEE) is called entropy, it does not measure uncertainty in a traditional sense.

Table 5.3: Summary of the simple statistical shape and entropy features.

name	definition	remarks
Skewness	$skew = \frac{\frac{1}{n}\sum_{i=1}^{n}(x(i)-\mu)^3}{s^3}$	
Kurtosis	$kurt = \frac{\frac{1}{n}\sum_{i=1}^{n}(x(i)-\mu)^4}{s^4} - 3$	
Spiky Index	$SI_\alpha = \frac{1}{n}\sum_{i=1}^{n}$ $\mathbb{I}(x_i > \alpha \cdot \max(x))$	α is the threshold, $\mathbb{I}$ is the indicator function, which equals to 1 if the condition is true, and to 0 otherwise.
Rectified Spiky Index	$RSI_\alpha = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(\lvert x(i)\rvert > \alpha \cdot \max(\lvert x\rvert))$	
Entropy (binning)	$H_b = -\sum_{i=1}^{b} p(x(i)) \ln p(x(i))$	b is the number of histogram bins, $p(x(i))$ is the probability of the i-th bin.
Entropy (k-nn)	$h_k = -\psi(k) + \psi(n) +$ $\ln(2) + \frac{1}{n}\sum_{i=1}^{n}\ln(2 \cdot r_i)$	ψ is the digamma function, k is the number of neighbors, r_i is the distance from $x(i)$ to its k-th neighbor.
Log Energy Entropy	$LEE = \frac{1}{n}\sum_{i=1}^{n}\ln(x(i)^2)$	

[2]As mentioned previously, since the signals are digitized, some of their samples have to have the same amplitude. Therefore, some of the distances will be equal to zero. To make the calculations possible, for the sake of entropy estimation, signal was additionally distorted with noise of low amplitude (10^{-10}), as recommended by [141].

Described shape and entropy measures, with their corresponding definitions, were summarized in Table 5.3.

5.1.2 Complex statistical features

Some of the described features were more complex than others, but most of them were defined on specific sample values or their probabilities. There are some statistical features used in literature, however, that are defined on already obtained summary statistics. They will be called complex statistical shape features in the remaining part of this work. One of the most popular of such features is the Coefficient of Variation (CV) [138], used on the normalized VAG signals by Rangayyan and Wu [56]. It is defined as the standard deviation divided by the average value. It measures variability of samples relative to their mean and is useful in comparing variability across signals with different scales, but can be skewed or undefined if the mean is close to zero. As most VAG signals have close-to-zero mean, two modified versions of the CV will be additionally used. The Modified Coefficient of Variation (MCV) is defined as the standard deviation divided by the mean of samples' absolute values, i.e., the RAV. It does not account for the variability of the rectified values, however, therefore one additional measure will be used, the Rectified Coefficient of Variation (RCV). It is very similar to the traditional CV, however, it is defined on the absolute values only, that is, both the standard deviation and the average are calculated on the absolute values of the signal. The RCV measures the variability in the *magnitude* of the data relative to the mean of the *magnitudes*.

There are also some shape measures used frequently in vibration signal analysis context, called factors. The Form Factor (FF) [144], defined as the RMS divided by the RAV measures how much greater is the average power of the signal compared to its average magnitude. Values close to one indicate similarity of the signal to the constant signal, while progressively smaller values indicate more fluctuations. Another measure, called the Crest Factor (CF) is defined as ratio of the maximal signal value to its RMS value [145, 146]. It quantifies closeness of extreme values to the average signal power. Similar measure, called the Impulse Factor (IF), uses RAV instead of the RMS [145, 146]. That is, it measures closeness of extreme values to the average signal magnitude. Both the CF and the IF were used in the VAG domain by Karpiński et al. [14, 15, 88], as well as Athavale and Krishnan [68]. Another factor, called the Margin Factor (MF) is defined as the maximal magnitude of the signal divided by the squared average square root of the rectified signal. Those three features, the CF, the IF and the MF, provide information about peak behavior within the signal. They differ the most in the "baseline" they use to compare peaks. It is worth noting, that their usage of maximal value make them relatively sensitive to outliers.

All of the complex statistical shape features, with their respective definitions, were summarized in Table 5.4.

5.2 Rolling features

Rolling features are the first category of time-domain measures, that actually incorporate information about temporal ordering of samples into extracted feature. They are defined as

Table 5.4: Summary of the complex statistical shape features.

name	definition
Coefficient of Variation	$CV = \frac{\sigma}{\mu}$
Modified Coefficient of Variation	$MCV = \frac{\sigma}{RAV}$
Rectified Coefficient of Variation	$RCV = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\lvert x(i)\rvert - RAV)^2}}{RAV}$
Form Factor	$FF = \frac{RMS}{RAV}$
Crest Factor	$CF = \frac{\max(x)}{RMS}$
Impulse Factor	$IF = \frac{\max(x)}{RAV}$
Margin Factor	$MF = \frac{\max(\lvert x\rvert)}{(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert x(i)\rvert})^2}$

some statistical features of the moving window signal. First, the signal is split into windows of fixed length, with one-sample shorter overlap. Then, for each window, some summary statistic is calculated. This *rolling* statistic is then treated as a *feature signal*, which can be used to extract features. Those extracted features can themselves be just summary statistics, not incorporating temporal ordering into the extracted feature. However, the first step of splitting the original signal into windows ensures, that the temporal ordering of the original signal samples is preserved. Therefore, all rolling features inherently incorporate temporal ordering information.

The first rolling feature is the Variance of the Mean Squared (VMS), defined by Rangayyan and Wu [53] and then used in a number of VAG-related studies [5,54,57,58,60,74,109]. It is defined as the variance of the mean-square rolling statistic. It represents variability of the average signal power across its duration. Another two features, the CVNMA and the CVNMV were described in previous chapter. However, in that context, they were calculated for all the signals after some preprocessing. In the current context, they will be used to differentiate specific VAG classes. In summary, they are calculated as the coefficient of variation of the moving mean and moving variance. Note, that before the calculation of CV, both rolling signals are normalized to the 0-1 range. They can be interpreted as the measures of signal stationarity.

The VMS feature was originally defined for 5-ms window [53] and, with the same widow size, used in later publications, including the work by Kręcisz and Bączkowicz [5], which originally analyzed signals used in the current work. However, to potentially gain more insights into the informativeness of the vibroarthrogram's variability, all rolling features will be calculated for windows sizes of powers of two from 0 to 12 with a step of 0.5[3], yielding windows sizes from 0.1 ms to 409.6 ms. Additionally, window size of 5 ms will be calculated, as currently used in the literature.

The rolling features, with their respective definitions, were summarized in Table 5.5.

[3]Non-integer number of samples will be rounded up.

Table 5.5: Summary of the rolling features.

name	definition	remarks
Variance of the Mean Squared	$VMS = \frac{1}{n}\sum_{i=1}^{n}(x^{MS}(i) - \mu^{MS})^2$	x^{MS} is the moving mean squared signal, and μ^{MS} is its mean value.
Coefficient of Variation of the Normalized Moving Average	$CVNMA = \frac{\sigma^{NMA}}{\mu^{NMA}}$	σ^{NMA} is the standard deviation of the normalized moving average signal, and μ^{NMA} is its mean value.
Coefficient of Variation of the Normalized Moving Variance	$CVNMV = \frac{\sigma^{NMV}}{\mu^{NMV}}$	σ^{NMV} is the standard deviation of the normalized moving variance signal, and μ^{NMV} is its mean value.

5.3 Difference and frequency features

Strict frequency analysis, with explicit transformation of the signal into the frequency domain will be discussed in the following chapter. Nevertheless, there are several features that provide insights into the signal frequency, that can be directly extracted from the signal in time domain. Those features are usually defined using two concepts: applying the difference operation to the signal, i.e., analyzing variations of its slope, or somehow counting signal turns. It is important to clarify, however, that those are not direct measures of frequency and while they can provide some general information about the rate of changes in a signal, they do not differentiate between various frequency components in a complex signal.

Sum of the absolute differences of the signal, or, more intuitively, its "length" was used in VAG domain by Al-Timemy [65]. In this work, however, it will be normalized by the number of samples, providing information about the average value of the absolute slopes. The feature will be called the Mean of Absolute Differences (MoAD). Since high values of the MoAD indicate, on average, higher absolute slopes of the signal, this feature can be interpreted as a variability measure. Note, however, that compared to the spread measures described earlier (see Table 5.2), this feature incorporate information about the temporal ordering of the samples.

Another features defined using the difference are the Hjorth parameters [147]. They were originally defined in the electroencephalogram analysis domain and proved to be informative also in the vibroarthrogram context [68]. The first feature, called the *mobility* is a ratio of the standard deviation of the derivative signal, or the slope, to the standard deviation of the raw signal. It can be interpreted as the average frequency measure of the signal. The second feature, *complexity*, is the ratio of the mobility of the derivative of the signal to the mobility of the signal itself. It measures average rate of change of the frequency. Hjorth's parameters were extensively used in the VAG literature, with slightly altered naming convention. That is, Hjorth's complexity in the VAG domain is usually called the Form Factor [5, 51, 52, 57, 58, 60, 74, 109]. Note, that in this work, the FF is defined differently (see Table 5.4).

Information about the variability, or the average frequency of the signal, could also be extracted by counting the number of zero-crossings. The Zero Crossing Rate (ZCR) is a

measure of the rate at which a signal changes its sign, i.e., goes from positive to negative values or vice versa. To make it easily interpretable, the number of crossings is divided by the length of a signal, providing direct information about the average rate of sign changes in the signal per sample. This gives an insight into the dominant frequency components of the signal. A feature similar to the ZCR is the Turns Count Rate (TCR). A "turn" in a signal represents a local minimum or maximum. Therefore, the TCR is effectively measuring how many significant peaks occur in the signal. If the signal changes direction frequently, it indicates a higher complexity or variability in the signal. Both features, however, could be sensitive to noise, and therefore they are usually implemented with some threshold.

Both ZCR [65,73,98,100,121] and TCR [54,55,60,65] were used quite extensively in VAG literature. For example, Al-Timemy [65] used the ZCR with additional threshold α. That is, only values of magnitude greater than α were considered in the counting. Yet, specific value of α was not provided. The ZCR was also used by Scherpereel et al. [121], however, the thresholding was not mentioned in their work. Rangayyan and Wu [53], on the other hand, used the TCR. They applied a threshold, that was dependent on the standard deviation of the signal. That approach led to a quite counter-intuitive result: larger values of the ZCR for the signals generated by the healthy knee. Because of that issue, in a later work, Wu and Krishnan [55] used a threshold dependent on the *range* of the values. First, they normalized signal to a zero to one range, and then applied a threshold of 0.2. Various approaches to setting the threshold make its choice quite difficult. Values too small could make the feature more noisy, and therefore less informative. Too high values, on the other hand, could make the feature not sensitive enough to extract valuable information at all. Therefore, in this work, both the std and the range approach will be used, with threshold levels α and β, respectively. Specific levels will be set in a range from zero (i.e., no thresholding) to one, with an increments of one percent. That will provide four feature vectors, namely ZCR_α and TCR_α, normalized by the standard deviation, as well as ZCR_β and TCR_β, normalized by the range. Since all threshold ranges are the same, each feature vector will have exactly 101 elements long.

Difference and frequency features calculated on signal in the time domain, with their respective definitions, were summarized in Table 5.6.

5.4 Self-similarity features

Another category of features focuses on measuring self-similarity of a signal. There are two main approaches used in this work: autocorrelation and fractality. The first approach involves the calculation of various Autocorrelation Function features to quantify the degree to which a signal repeats its structure over time. On the other hand, the second approach, fractality, employs Multifractal Detrended Fluctuation Analysis as a robust tool for quantifying the complex, scale-invariant properties of a signal. These two approaches, serving complementary roles, offer a comprehensive perspective on the self-similarity of the signal.

Table 5.6: Summary of the difference and frequency features.

name	definition	remarks
Mean of Absolute Differences	$MoAD = \frac{1}{n-1}\sum_{i=1}^{n-1} \lvert x(i+1) - x(i)\rvert)$	
Mobility	$mobility = \frac{\sigma_{x'}}{\sigma_x}$	σ_x denotes standard deviation of the signal, and $\sigma_{x'}$ denotes standard deviation of the derivative of the signal.
Complexity	$complexity = \frac{\sigma_{x''}}{\sigma'_x} / \frac{\sigma_{x'}}{\sigma_x}$	$\sigma_{x''}$ denotes standard deviation of the double-derivative of the signal.
Zero Crossing Rate α	$ZCR_\alpha = \frac{1}{n-1}\sum_{i=1}^{n-1} \mathbb{I}\left(x(i+1)\cdot x(i) < -(\alpha\cdot\sigma)^2\right)$	α is the threshold, $\mathbb{I}$ is the indicator function, which equals to 1 if the condition is true, and to 0 otherwise, and σ is the standard deviation of the signal x.
Zero Crossing Rate β	$ZCR_\beta = \frac{1}{n-1}\sum_{i=1}^{n-1} \mathbb{I}\left(x(i+1)\cdot x(i) < -(\beta\cdot(spread(x)))^2\right)$	β is the threshold, and σ is the standard deviation of the signal x.
Turns Count Rate α	$TCR_\alpha = \frac{1}{n-2}\sum_{i=1}^{n-2} \mathbb{I}\left((x(i+2) - x(i+1))\cdot(x(i+1) - x(i)) < -(\alpha\cdot\sigma)^2\right)$	
Turns Count Rate β	$TCR_\beta = \frac{1}{n-2}\sum_{i=1}^{n-2} \mathbb{I}\left((x(i+2) - x(i+1))\cdot(x(i+1) - x(i)) < -(\beta\cdot(spread(x)))^2\right)$	

5.4.1 Autocorrelation function features

Autocorrelation is a concept used to determine the similarity of a signal to a delayed version of itself over a discrete period of time. It can be very fruitful tool for characterization of biomedical signals in general [10]. The Autocorrelation Function (ACF) is defined as [148]:

$$ACF(k) = \frac{\frac{1}{n}\sum_{i=1}^{n-k}(x(i) - \mu)\cdot(x(i+k) - \mu)}{\sigma^2}, \tag{5.1}$$

where k is the lag, x is the n-sample signal, μ is the average value of this signal, and σ^2 is its variance. Note, that the sum is calculated from the first to $(n-k)$-th sample, and therefore values of the function for increasing k get calculated on progressively less samples, making them progressively less reliable. Nevertheless, in this work, for the ACF related features, function will be calculated for the biggest possible number of lags, i.e., from $k = 1$ to $k = n - 1$. A number of informative measures can be extracted from the ACF.

For example, the Average Autocorrelation Function (AACF) is the average value of the function across all lags. Values of the AACF close to zero suggest that, on average, the signal does not correlate much with delayed versions of itself, indicating that it is less structured and resembles white noise. If the AACF is far from zero, it implies that there is a high degree of autocorrelation overall, suggesting the presence of dominant frequency components or periodicities. Similar measure, i.e., the Rectified Average Autocorrelation Function (RAACF) is defined similarly, but on the absolute values of the ACF. Taking absolute value negates the effect of canceling positive and negative correlations, focusing instead on the magnitude of the ACF at each lag, regardless of its direction. It provides a measure of how strong the magnitude of autocorrelation is on average. This information can be useful to distinguish

oscillatory parts of signals, where the ACF fluctuates around zero. Similar information can be extracted using the Mean Squared Autocorrelation Function (MSACF), which squares the ACF values instead of taking their absolute value. It therefore puts an additional emphasis on larger values.

Another very important ACF measure is the First Zero Crossing Time (FZCT). It is the time delay, at which the ACF first intersects the zero line. In a pure sine wave, it can indicate quarter of the period of the wave. At multiples of its period, the ACF will be equal to one, while shifting the phase half the period will result in ACF equal to negative one. At a quarter period, the sine wave is perfectly out of phase with a version of itself that is shifted by that amount, leading to a zero crossing in the autocorrelation function. In more complex signals, it can indicate shortest typical timescale over which signal's properties change significantly. Lower values of the the FZCT might suggest rapid fluctuations in the signal, while higher values indicate slower signal evolution. This will be further discussed in the following section, devoted to the Phase Space Reconstruction. The the FZCT will be defined as the number of sample at which the ACF changes its sign divided by the number of all samples in the signal.

Another measure is the *number* of times, that ACF crosses zero, i.e., the Autocorrelation Function Zero Crossing Rate (ACFZCR). A zero-crossing in the ACF can be interpreted as indicative of a change in the signal's pattern. That is, when a zero-crossing occurs, it means the signal transitions from a state of positive autocorrelation to negative autocorrelation or vice versa. In other words, the signal at one point in time is either in phase or out of phase with a future or past version of itself. The ACFZCR, therefore, provides information about the dynamics of structural changes of the original signal. Specifically, it can reveal the frequency at which these changes occur. It will be defined as the actual number of zero-crossings divided by the number of signal samples.

The Autocorrelation Function Turns Count Rate (ACFTCR) is a similar feature, but with slightly different interpretation. Its definition is the same TCR, but without the threshold and defined on the ACF function. In complex signals, such as vibroarthrograms, the number of turns might indicate the presence of multiple overlapping frequencies or non-stationary components. A high number of turns could suggest a multi-component signal or a signal with rapidly changing structure. Similarly to the number of zero-crossings, to obtain the ACFTCR, the actual number turns count will be divided by the number of samples in the signal.

Features defined on the ACF, with their respective definitions, were summarized in Table 5.7.

5.4.2 Detrended Fluctuation Analysis

The Detrended Fluctuation Analysis (DFA) is another measure of self-similarity, or memory, in time series [149]. While the autocorrelation measures the degree of similarity between a signal's values at different time points, the DFA quantifies how the overall structure of a signal's fluctuations is similar at different time scales. This is done by creating the integral of the original signal (i.e., the cumulative sum), segmenting the obtained signal into non-overlapping windows of various sizes, and removing a local trend in each window. Trend is obtained using polynomial fitted to the signal in a window. After removing the trend, the

Table 5.7: Summary of the ACF-based features.

name	definition	remarks
Average Autocorrelation Function	$AACF = \frac{1}{n}\sum_{i=1}^{n} ACF(i)$	ACF is the Autocorrelation Function of n lags, calculated using Equation 5.1.
Rectified Average Autocorrelation Function	$RAACF = \frac{1}{n}\sum_{i=1}^{n} \lvert ACF(i)\rvert$	
Mean Squared Autocorrelation Function	$MSACF = \frac{1}{n}\sum_{i=1}^{n} ACF(i)^2$	
First Zero Crossing Time	$FZCT = \frac{1}{n}\min_k \{ACF(k)\cdot ACF(k+1) < 0\}$	
Autocorrelation Function Zero Crossing Rate	$ACFZCR = \frac{1}{n-1}\sum_{i=1}^{n-1} \mathbb{I}\left(ACF(i+1)\cdot ACF(i) < 0\right)$	$\mathbb{I}$ is the indicator function, which equals to 1 if the condition is true, and to 0 otherwise.
Autocorrelation Function Turns Count Rate	$ACFTCR = \frac{1}{n-2}\sum_{i=1}^{n-2} \mathbb{I}\left((ACF(i+2) - ACF(i+1))\cdot(ACF(i+1) - ACF(i)) < 0\right)$	

RMS value of each detrended window is calculated, providing a fluctuation measure. Averaging fluctuations over all windows of the same size provides a value, which is a function $F(s)$ of the window size s. Specific windows size choice will be discussed later. The function $F(s)$ allows to obtain the Hurst exponent, H, which is a direct measure of self-similarity. It is calculated as a slope of the line fitted to the logarithm of $F(s)$ as a function of logarithm of s. Values of the Hurst exponent smaller than 0.5 indicate anti-correlated structure of the signal, while greater than 0.5, but smaller than 1.0 indicate correlated, long-range dependent structure. Values greater than one suggest non-stationary signals, resembling random walks [150]. The Hurst exponent was previously used in the VAG feature extraction by Kręcisz and Bączkowicz [5], on the signal base the same as in this dissertation. Differences in implementation between the [5] and this work will be discussed later. The DFA was also used in the preprocessing steps in some other VAG-related studies, to determine which sub-signals should be summed up after the decomposition [63, 66, 114].

Two parameters are of great importance in the Hurst exponent calculation: window sizes and order of the polynomial used to obtained local trend. Too small window sizes could result in overfitting of the polynomial to signal samples inside the window, while too big would render the number of windows in the whole signal too small to obtain reliable average fluctuation. Ihlen [150] suggests non-smaller window size than 10 and not greater than tenth of the signal length. Therefore, since signals in this work have around 60 thousand samples, window sizes of 16 to 4096 will be used, with consecutive powers of two. Polynomial order have a direct influence on the shape of removed local trend and therefore fluctuation calculation. Too large orders could result in overfitting of local trend, resulting in underestimate of fluctuations. In smallest windows size of 10 to 20, Ihlen [150] suggest using polynomial order not greater than 3. Therefore, two orders will be used, i.e., one, creating linear trend,

and two, creating quadratic trend.

The DFA assumes that self-similarity of a signal is consistent across all amplitude scales and therefore, it is sometimes called *mono*fractal. On the other hand, *multi*fractal DFA allows for varying scaling behavior, providing a more nuanced understanding of complex signals that exhibit different degrees of self-similarity at different amplitude scales. The fluctuation in each window of the DFA is calculated as the RMS of samples in that window:

$$F(s) = \sqrt{\frac{1}{n_s}\sum_{i=1}^{n_s} x_s(i)^2} = \left(\frac{1}{n_s}\sum_{i=1}^{n_s} x_s(i)^2\right)^{\frac{1}{2}}, \tag{5.2}$$

where x_s is n_s-length window of the signal. Because of the squaring operation, the RMS is influenced much more by the greater values. Changing the squaring operation to other orders q:

$$F_q(s) = \left(\frac{1}{n_s}\sum_{i=1}^{n_s} \left(x_s(i)^2\right)^{\frac{q}{2}}\right)^{\frac{1}{q}}, \tag{5.3}$$

yields Multifractal DFA [150]. Parameter q equal to zero makes the definition of fluctuation a geometric mean[4]. Order q smaller than zero make the fluctuation more sensitive to smaller variation, since small values raised to negative power get larger. Positive q orders make the fluctuation more sensitive to larger variations, with special case $q = 2$, which is used in traditional DFA. Increasing magnitudes of the q make the Hurst exponent calculation less precise, and Ihlen [150] considers range from negative to positive five sufficient in most biomedical signal analysis cases. Therefore, this range will be used in the current work, with an increment of one.

Varying the q order allows for analysis of various multifractality measures [150]. The first such measure used in this work is the mass exponent τ_q. It quantifies the scaling behavior of the moments of the probability distribution of a signal. The linearity or nonlinearity of τ_q with respect to q provides an indication of whether the signal exhibits monofractal or multifractal behavior, respectively. It is directly related to the generalized Hurst exponent H_q through a simple relation [151]: $\tau_q = q \cdot H_q - 1$, linking the long-term memory property, captured by the Hurst exponent, with the scaling behavior of different moments of the data distribution, captured by the mass exponent.

Another Multifractal DFA feature is the Singularity Exponent α_q. It characterizes the local regularity or irregularity of a signal. It is derived from the mass exponent τ_q, as its first difference with respect to q. It indicates, therefore, how the mass exponent changes with q, or how the scaling behavior of the signal changes. Positive q values prioritize larger fluctuations, and high α_q for positive q indicates rapid changes in these large fluctuations, while a lower α_q represents slower changes. On the other hand, negative q values focus on smaller fluctuations. High α_q for negative q indicates that these small fluctuations change quickly, while a lower α_q suggests more gradual changes. Besides specific values of α_q at various q, range of its values can also be extracted. This width of the multifractal spectrum is a measure of the overall degree of multifractality in a signal. A wider and narrower spectra

[4]To avoid division by zero, special case of $q = 0$ is calculated as $F_{q=0}(s) = e^{\frac{1}{2}\ln\left(x_{s_i}^2\right)}$ [150].

indicate a higher or lower degree of multifractality, respectively. That is, signals with wider spectra show a variety of scaling behaviors with a significant amount of irregularities. On the other hand, narrower spectra indicate signals with more uniform scaling behaviors and fewer irregularities.

The last feature defined on Multifractal DFA is the Singularity Dimension $f(\alpha_q)$. It is defined using singularity exponent, mass exponent and the moment q [151]: $f(\alpha_q) = q \cdot \alpha_q - \tau(q)$. The Singularity Dimension measures how often local irregularities occur in the signal. High $f(\alpha_q)$ indicates that singularities of strength α_q are dense and occupy a large portion of the signal, while low $f(\alpha_q)$ indicates that such singularities are more sparse.

Table 5.8: Summary of the DFA-based features.

name	definition	remarks
Hurst Exponent	$H(q) = \mathbb{S}\left(F_q(s), s\right)$	$\mathbb{S}$ is a slope of a line fitted to the first argument with respect to the second argument, $F_q(s)$ is the q-th order moment fluctuation, and s is the window size.
Mass Exponent	$\tau_q = q \cdot H(q) - 1$	
Singularity Exponent	$\alpha_q = \tau_{q+1} - \tau_q$	
Width of the Multifractal Spectrum	$width = \max(\alpha) - \min(\alpha)$	
Singularity Dimension	$f(\alpha_q) = q \cdot \alpha_q - \tau(q)$	

In summary, the Multifractal DFA analysis will result in 86 features: 11 Hurst Exponents, 11 Mass Exponents, 10 Singularity Exponents, 1 witdth of the spectrum and 10 Singularity Dimensions, for q order from negative to positive five with increment of 1. All those features will be extracted for detrending with linear and quadratic trend. Summary of their definitions was provided in Table 5.8.

5.5 Phase Space Reconstruction features

Another approach to signal analysis is the reconstruction of the phase space of the dynamical system that produced the signal. Since it presents a multi-dimensional perspective on the system's dynamics, it has the potential to provide more comprehensive information, potentially allowing extraction of more informative features. The process used to reconstruct the phase space is known as signal embedding [152]. It is a simple transformation that maps a univariate time series into a multivariate one, called the Reconstructed Phase Space (RPS). Each dimension in this space corresponds to a time-delayed version of the original signal [152]:

$$PSR = (x(i - (m-1)\,\tau), x(i - (m-2)\,\tau), \ldots, x(i-\tau), x(i)), \tag{5.4}$$

where x is the signal, i is the sample number, m is embedding dimension and τ is the delay. By employing a suitable delay and selecting an appropriate number of embedding dimensions, a faithful representation of the underlying dynamical system can be created [152]. Choice of the the delay time τ and number of dimensions m is not a trivial task, however. Figure 5.2

shows exemplary reconstructed phase spaces using signal embedding method with various time delays. The signal is the same as in previous examples.

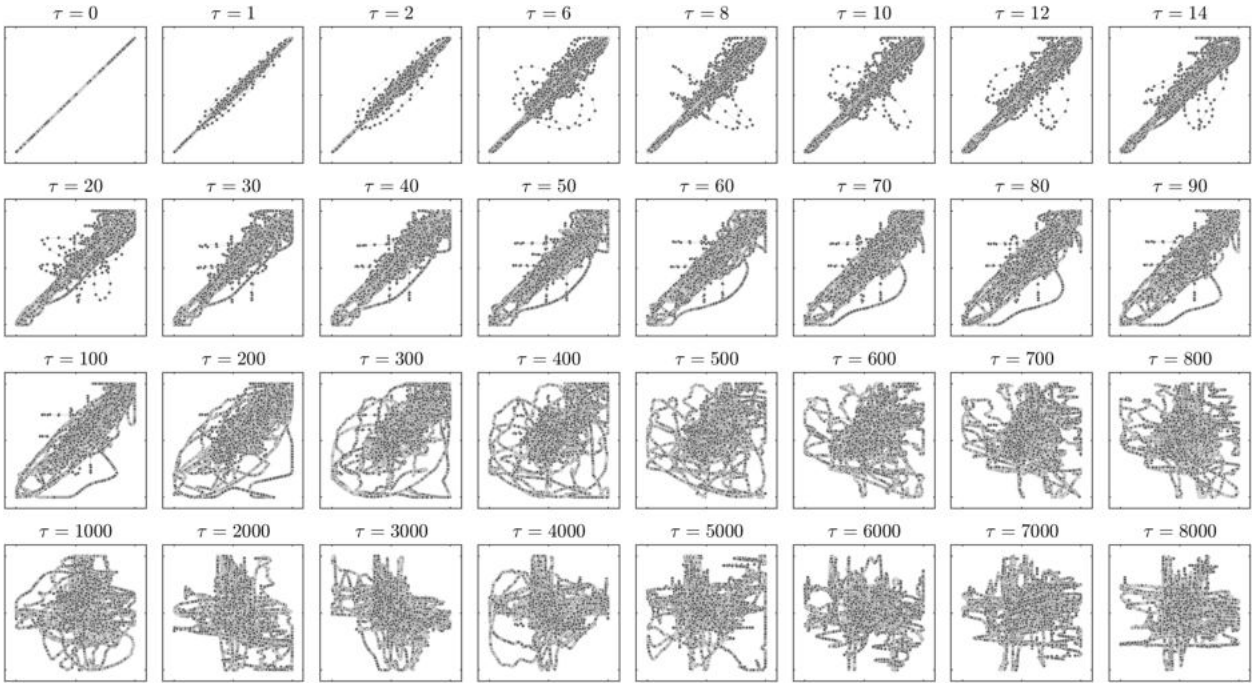

Figure 5.2: Exemplary reconstructions of the phase space. For all examples, number of dimensions (m) is equal to 2, while time delay (τ) varies. In all plots, x- and y-axes indicate specific time samples and their delayed versions, respectively.

Time delay which is too short results in highly correlated dimension values, clustered around the diagonal (see the first two rows in Figure 5.2). Too large values, on the other hand, make the informative structures spread out excessively, rendering them obscured or potentially lost within the phase space, thereby reducing the effectiveness of the embedding. A popular rule-of-thumb to obtain the τ parameter is the first zero-crossing of the ACF [152]. This point is the first time when the delayed signal becomes uncorrelated with the original signal, indicating that there is no significant information shared between the original and the delayed versions. This choice, therefore, allows for the reconstruction of the phase space that potentially allows for the extraction of features providing new information. Another method, also based on the ACF, assumes the optimal choice of the delay to be the sample, where the ACF is equal to $1 - \frac{1}{e}$ [153]. For the example from Figure 5.2, the first zero crossing in the ACF is at the 1619-th sample, while τ obtained by the second method equals to 284. In the results chapter, those methods will be indicated by the subscripts $_{zc}$ and $_{e}$, respectively.

Embedding dimension, i.e., the m parameter, can be chosen using the False Nearest Neighbors algorithm [152]. In the Figure 5.2, all phase spaces are reconstructed in two dimensions. Note, that some "loops" of the trajectory can be spotted, when the trajectory crosses itself. See, for example, lower-right part of plots for τ equal to 100, 200 and 300. This is the result of a projection of higher dimensional space into only two dimensions. Some points of those loops are "false neighbors" meaning that their proximity is a result of the projection, and in higher-dimensional embedding, their distance would be significantly

different. The algorithm of False Nearest Neighbors consists of increasing the dimensionality of embedding space as long, as the ratio of false nearest neighbors to true nearest neighbors is close to zero.

The time delays τ_{zc} and τ_e, as well as the embedding dimensions m_{zc} and m_e, will constitute four first features defined in the RPS. Note, that the τ_{zc} is a linear transformation of the FZCT feature, and therefore informativeness of those two features will be exactly the same.

5.5.1 Phase space complexity measures

The RPS is an extremely rich source of information, and there are various features measuring its complexity. For example, the Largest Lyapunov Exponent (LLE) measures the average divergence of points close to each other in the RPS [153]. Consider two proximate points in this phase space, with distance between them being δ. They symbolize marginally different initial states of the dynamical system. These points change their position over time, creating two distinct trajectories in the phase space. The LLE measures the average speed of this change in the distance between the two trajectories. In scenarios where the LLE is positive, an exponential divergence of the trajectories over time is indicated, signaling the presence of a chaotic system. Conversely, an LLE value of zero implies neither convergence nor divergence of the trajectories, reflective of a periodic system. A negative LLE value signifies the trajectories' convergence, indicative of a stable, non-chaotic system. Therefore, it provides information whether the system described by the RPS is chaotic, or, more specifically, how chaotic it is [153].

Another complexity measure is the Correlation Dimension (CD) [154]. It quantifies self-similarity of a signal in the RPS. It is calculated using a correlation sum $C(\epsilon)$ for a given distance ϵ. The correlation sum is the fraction of all possible pairs of points in the phase space that are within a distance ϵ of each other. In a m-dimensional space, for relatively small ϵ, $C(\epsilon)$ should be proportional to ϵ^m. In other words, the number of close pairs of points should increase as the m-th power of ϵ. It is because in a m-dimensional space, the volume of a hypersphere scales with the m-th power of its radius. To compute the CD, a line is fitted to the log-log plot of the $C(\epsilon)$ as a function of ϵ. The slope of this line is a value of the CD. However, for some systems, this plot may not be exactly linear, or its slope may change for different ϵ. This is where the CD reveals the fractal nature of the dynamical systems: if the CD is not an integer, it indicates that the system's dynamics have a fractal structure. Increasing values of the CD indicate higher complexity [154]. Features based on the RPS, with their corresponding definitions were summarized in Table 5.9.

5.5.2 Recurrence plots

Recurrence plot is a visualization tool allowing easier identification of patterns in the RPS [155]. To construct such plot, a matrix of distances between each point pair in the RPS is created. Those distances can be arbitrary and in this work, a Chebychev (maxiumum) distance will be used[5]. Then, a recurrence threshold r is applied, to create binary version of the distance

[5]This distance is assumed by an approximation method proposed by Webber et al. [156] and used in this work.

Table 5.9: Summary of the RPS-based features.

name	definition	remarks
Time delay (the first method)	$\tau_{zc} = \min_k \{ACF_k \cdot ACF_{k+1} < 0\}$	
Time delay (the second method)	$\tau_e = \min_k \left\{\left(ACF_k - 1 + \frac{1}{e}\right) \approx 0\right\}$	
Embedding dimension	See text.	
Largest Lyapunov Exponent	$LLE = \mathbb{S}(\delta(i), i)$	$\mathbb{S}$ is a slope of a line fitted to the first argument with respect to the second argument, $\delta(i)$ is the average distance between every pair of points after the i-th time sample.
Correlation Dimension	$CD = \mathbb{S}(\log(C(\epsilon)), \log(\epsilon))$	$C(\epsilon)$ is a correlation sum, and ϵ is a distance.

matrix, with ones representing a recurrence (distance smaller than r). The Recurrence Plot (RP) is a visualization of the binary matrix resulting from the applied threshold. Therefore, each point in this plot represents a pair of time samples, at which the phase space trajectory passes near the same location. In another words, it visualizes time samples at which the system, which generated the RPS repeats its state.

Figure 5.3 includes both the distance matrices (left column) and the corresponding RPs (right column) for a RPS of the same signal as in Figure 5.2. Both methods of delay selection were plotted (τ_{zc} and τ_e, upper and lower rows, respectively), and for both methods, embedding dimension of $m = 4$ was chosen using the False Nearest Neighbors algorithm.

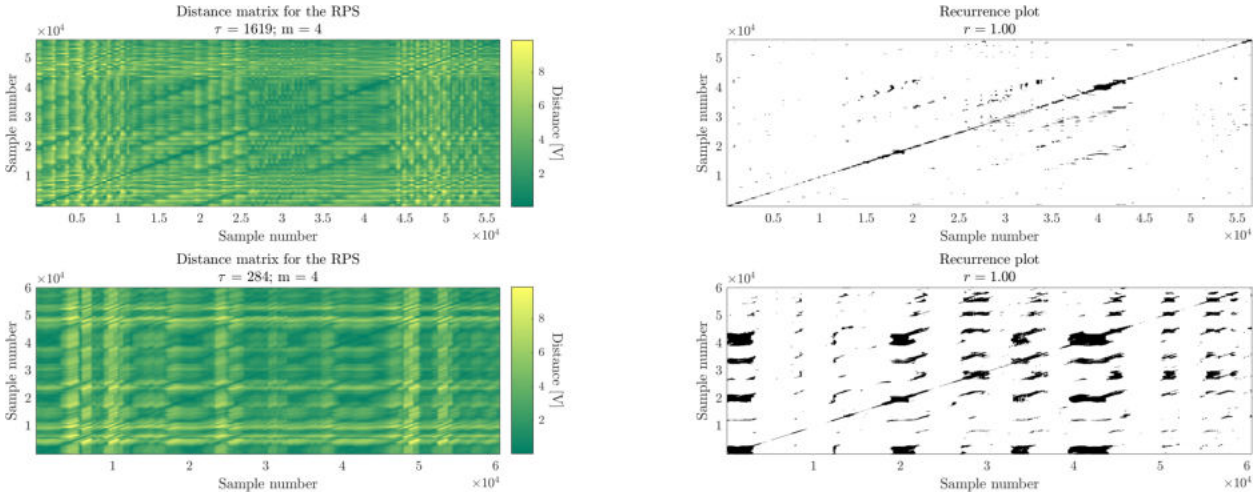

Figure 5.3: Exemplary distance matrices (left) and corresponding Recurrence Plots (right).

Choice of the parameter r is not easy, and several rule-of-thumb methods for this choice exist. In the current work, a 10% of the maximum distance will be used [157], as one of the simplest methods. The maximum distance will also constitute a feature extracted from the RPS.

In the constructed RP a number of characteristic shapes can be spotted. Clusters of points indicate specific regions in the phase space, in which system trajectory tends to remain. Single points indicate states, which are rarely visited and white areas indicate spaces that were not close enough to be considered recurrence. Diagonal lines indicate periods, where the path of trajectory is similar to path previously followed. There is a number of specific measures, that can be extracted from the RP, which are collectively called the Recurrence Quantification Analysis (RQA). Some of those feature were previously used in a number of VAG-related studies [5, 37, 43, 96, 158].

The simplest RQA measure is the Recurrence Rate (RR), defined as a ratio of recurrent points to all points. It measures the density of recurrent points in the RPS. A higher RR indicates a higher likelihood of the system returning to previous states. Similar measure is the Determinism (DET), defined as the ratio of recurrent points forming a diagonal lines to all recurrent points. It measures predictability of the RPS. Another measure based on the the diagonal line quantification is the Average Diagonal Line (ADL), which measures the average duration of similar behaviors in a system. Higher values of the ADL indicate longer periods of similar dynamics in a system, on average. Finally, the Laminarity (LAM) is a ratio of recurrent points forming vertical lines to all recurrent points. It quantifies the duration the system spends in a single state. The key difference between DET, ADL and LAM is that the DET measures the predictability of the system, ADL measures the duration of similarity in the system's behavior, while LAM measures the duration the system remains in the same state.

Since the RP is constructed for the each pair of points in the RPS, the calculation of the RQA measures is extremely computationally expensive for long signals. Therefore, instead of exact measures, their approximations will be used, utilizing the approach of Webber et al. [156]. This method can be used assuming that the distance used for the RQA is the Chebychev (maxiumum) distance. Summary of the RQA-based features, with their corresponding definitions, was provided in Table 5.10.

Table 5.10: Summary of the RQA-based features.

name	definition	remarks
Maximal Distance	$md = \max(\Delta)$	Δ is a matrix of distances between each point pair in the RPS.
Recurrence Rate	$RR = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{n} R_{i,j}$	R is the Recurrence Plot.
Determinism	$DET = \frac{\sum_{l=2}^{N} l \cdot D(l)}{\sum_{i=1}^{N} \sum_{j=1}^{N} R_{i,j}}$	D is the histogram of the diagonal lines' lengths.
Average Diagonal Line	$ADL = \frac{\sum_{l=2}^{N} l \cdot D(l)}{\sum_{l=2}^{N} D(l)}$	
Laminarity	$LAM = \frac{\sum_{l=2}^{N} l \cdot V(l)}{\sum_{i=1}^{N} \sum_{j=1}^{N} R_{i,j}}$	V is the histogram of the vertical lines' lengths.

5.5.3 Entropy measures

Another approach to analyze signal in the delay embedding context is quantification of similar patterns. This can be measured by various entropy measures, such as the Approximate

Entropy (ApEn) [159] and the Sample Entropy (SampEn) [160]. Those entropy measures are qualitatively different than previously described Shannon Entropy estimators, since they are inherently calculated for the temporally ordered signal, not just distribution of its values. The main idea of both ApEn and SampEn is similar. It first involves counting the number of similar sub-signals of a certain length in the signal. Similarity is determined based on a predefined tolerance level r. Next, this process is repeated, but this time for sub-signals that are one sample longer. The final step is to compare the frequency of the shorter sub-signals to the frequency of the longer sub-signals. If short patterns often repeat in a longer context, it suggests a higher regularity and lower complexity of the signal [161]. Main differences between the ApEn and SampEn, and their implications, are clearly visible comparing definitions of those measures.

The short sub-signals repetitions for the i-th sub-signal are usually called B_i and are calculated as follows [161]:

$$B_i^m(r) = \frac{1}{N-m-1} \sum_{j=1, j\neq i}^{N-m} \mathbb{I}\left(d(x_m(j), x_m(i)) < r\right), \tag{5.5}$$

where N is the length of the entire signal, m is the length of the sub-signal, $\mathbb{I}$ is the indicator function that equals to one, when the condition is met and zero otherwise, d is the Chebyshev distance function, x_m is the m-samples long sub-signal and r is the similarity threshold.

Similarly, longer sub-signals repetitions for the i-th sub-signal are called A_i and are obtained using the following equation [161]:

$$A_i^m(r) = \frac{1}{N-m-1} \sum_{j=1, j\neq i}^{N-m} \mathbb{I}\left(d(x_{m+1}(j), x_{m+1}(i)) < r\right). \tag{5.6}$$

Both ApEn[6] and SampEn can be defined using those values:

$$ApEn(m, r, N) \approx -\frac{1}{N-m} \sum_{i=1}^{N-m} \log\left(\frac{A_i^m + 1}{B_i^m + 1}\right) \tag{5.7}$$

$$SampEn(m, r, N) = -\log\left(\frac{\sum_{i=1}^{N-m} A_i^m(r)}{\sum_{i=1}^{N-m} B_i^m(r)}\right) \tag{5.8}$$

The main difference between the ApEn and SampEn is that the fraction in the logarithms calculated for the ApEn has *plus one* in both numerator and denominator. It means that in both Equations 5.5 and 5.6, the sum is calculated for each j, *including* i-th value. In other words, each sub-signal, in addition to being compared to each other sub-signal, is compared also to itself. Such definition makes the Approximate Entropy biased toward zero, i.e., it underestimates complexity of a signal [161]. This bias can be reduced with increasing length of a signal, and dependence of the entropy value on the signal's length is another difference between ApEn and SampEn [161]. According to the work by Bajić

[6]Strict definition of the Approximate Entropy used in the original publication [159] is slightly different, but the provided approximate definition allows easier comparison of the ApEn and SampEn [161].

and Japundžić-Žigon [162], the ApEn produces more biased values, but with lower variance, compared to the SampEn. For long signals, however, this bias is reduced and therefore the ApEn could be a preferable choice. Therefore, both ApEn and SampEn will be used in this work.

Both entropy measures are heavily dependent on their parameters: r and m. The r parameter can be understood similarly to the r parameter in the RQA. However, within the entropy estimation literature, it is usually additionally multiplied by the standard deviation of the signal, which ensures that the r parameter is invariant to the signal scale. That is $r = 0.1$ means similarity threshold equal to $0.1 \cdot \sigma$ with σ being the standard deviation. This convention will also be followed in this work. In the foundational publication by Pincus [159], he proposes r values to range from 0.1 to 0.2. Lu et al. [163], however, argued that utilization of this range can lead to inaccurate results and one should rather use r value that maximizes the Approximate Entropy measure. In the current study, the approach of maximum entropy will be utilized, as well as the range of r values from 0 to 1 with the step size of 0.1. Also, the r from this range that maximizes the entropy, will be used as a feature.

The optimal value of the m parameter is also not straightforward to obtain. Usual values are $m = 2$ or $m = 3$ [161]. There is also a rule-of-thumb that the systems can be differentiated using signals of length $N = 10^m \dots 30^m$ [161]. While this is not a direct recommendation for the optimal m value, it could potentially be used as a guideline. Since signals in this work are 61440 samples long, according to this rule, the m should be between 3 and 5. One could also obtain the m parameter using previously mentioned False Nearest Neighbors algorithm [163] with τ parameter equal to one. All those options will be used and compared in the remaining part of this work.

Both the ApEn and SampEn were previously used in VAG-related studies. Wu et al. [114] used ApEn with m parameter equal to 4 and r equal to 0.2. It was obtained as a parameter set with the lowest p-value of Wilcoxon rank-sum test. Gong et al. [41] used SampEn with parameters $m = 2$ and $r = 0.2$. The same parameters were used in study by Nalband et al. [158] to obtain both ApEn and SampEn. Moreover, Kręcisz and Bączkowicz [5] used Mulsiscale Sample Entropy, i.e., the SampEn obtained for multiple time scale signals. Time scale signal was defined as a n-sample non-overlapping moving average, with n varying from 7 to 30. They used $m = 2$ and $r = 0.15$.

Table 5.11: Summary of the entropy-based features.

name	definition	remarks
Approximate Entropy	See Equation 5.7.	
Sample Entropy	See Equation 5.8.	
Distance Maximizing the Entropy	$r_{max} = \{r \mid \max(Ent(m, r))\}$	Ent being either ApEn or SampEn.
Embedding Dimension	See text.	Determined using False Nearest Neighbors algorithm.

Summary of the entropy features was provided in Table 5.11.

Chapter 6

Signal analysis in frequency domain

Some general information about the signal's periodicity can be extracted directly from the time domain. For example, frequency-related features, such as MoAD, ZCR, TCR or some ACF-based ones, provide some general insights about the rate of change of the signal's amplitude. However, these methods are difficult to interpret for more complex signals, i.e., signals with multiple frequency components. Signal analysis in the frequency domain, that is, on the basis of spectral density, enables more direct analysis of specific frequency components separately, rather than mixed, as the in time domain.

In the VAG domain, the frequency analysis is broadly used, usually consisting of feature extraction from the estimated spectral density. Various spectral estimation methods were used in VAG-related research, including the periodogram [116] or Welch's method with various parameters [37, 75]. In some studies, the frequency features were obtained directly from the Discrete Fourier Transform (DFT) [101] or rolling DFT [5, 76, 77]. More in-depth description of the rolling frequency analysis approach will be included in the following chapter.

This chapter includes brief explanation of the Power Spectral Density estimation, followed by the comparison of some of the selected estimation methods. The chosen method will be used to obtain some frequency-related features. Most of those features can be thought of as being parametric, with the parameters being lower and upper frequency of the band, for which the features are being calculated. Therefore, the Frequency Range Map will be introduced, as a tool to visualize and choose the most informative frequency band for each feature.

Time domain features were obtained separately for three differently preprocessed versions of the vibroarthrogram: the raw, the filtered and the derivative signal. Similar approach will be used in the frequency analysis. Some features, specifically the ones describing the power, will not be influenced by the preprocessing much. Both filtering and differencing operations have direct impact on the power distribution over the frequency of the signal. However, since this impact is almost the same for each signal, the information content should not change dramatically across the raw signal and the pass-band in the filtered signal. However, features describing *shape* of the frequency distribution will be greatly affected by such operations. Therefore, each feature will be obtained for all three methods of the signal preprocessing.

6.1 Power Spectral Density estimation

The Energy Spectral Density of deterministic signal provides information about the distribution of power across the frequencies present in this signal [164]. However, real-world applications, such as vibroarthrography, often handle finite-length signals that are best treated as random sequences. Their random nature implies that the energy of specific frequency band

cannot be directly determined, as in a deterministic signal. Therefore, the exact Energy Spectral Density, possible to obtain for deterministic and infinite-length signals, cannot be directly computed for the random, finite-length signals. As a consequence, the Power Spectral Density (PSD) estimation is employed, which allows to analyze the average distribution of power over frequencies within these finite sequences [164].

To perform this estimation, multiple techniques can be used, some of which are based on the DFT. The DFT provides a way to encapsulate the information embedded in the signal into a series of sinusoids of different frequencies, each with its own magnitude and phase, transforming the signal from the time domain to the frequency domain. It is defined as follows [164]:

$$X(k) = \sum_{t=1}^{N} x(t) e^{-i\frac{2\pi}{N}\cdot(t-1)k}, k = 0, ..., N-1, \tag{6.1}$$

where $X(k)$ is the k-th frequency bin of the complex DFT's output, $x(t)$ is the t-th sample of a signal, and N is number of signal samples. The DFT can therefore be interpreted as a degree to which the original signal $x(t)$ can be represented by the certain sine wave, defined by the complex exponential sequence $e^{-i\frac{2\pi}{N}\cdot(t-1)k}$ for a specified value of the frequency bin k. Squared magnitude of this complex representation yields the PSD estimate called the periodogram [164]. The periodogram of the exemplary vibroarthrogram showed in previous chapters was included in Figure 6.1.

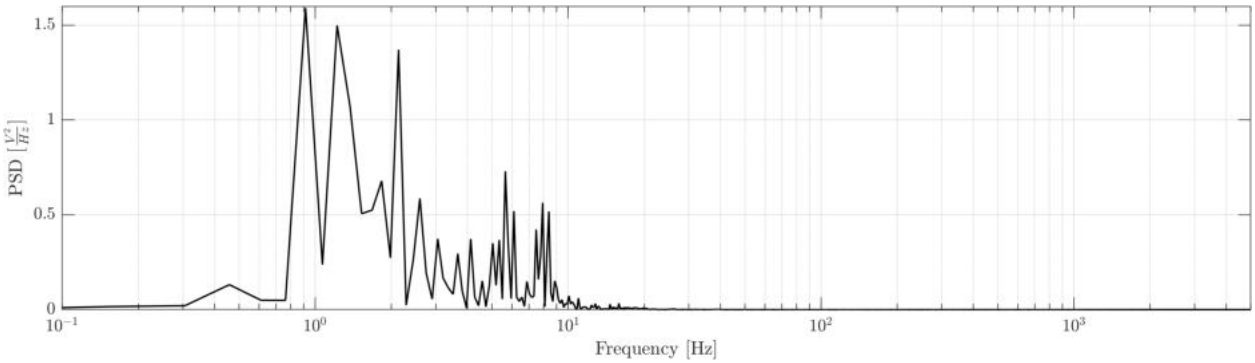

Figure 6.1: Periodogram of the exemplary vibroarthrogram.

The periodogram is an estimator which produces relatively high-variance PSD estimates. See, in Figure 6.1, that the spectrum is relatively "spiky" and in some biomedical signal analysis cases, an estimate with lower variance could be more useful [165]. There is a number of methods used to obtain estimates of lower variance, with the most widely used being the Welch's method and the Multitaper method, both based on the periodogram [165].

The Welch's method consists of first splitting the signal into N-sample, potentially overlapping segments. Then, a windows function is applied to each segment and a periodogram of each segment is calculated. Finally, the obtained periodograms are averaged, yielding the PSD estimate of less variance, compared to the traditional periodogram. Segments' size is of great importance, since it directly affect the frequency resolution of the obtained PSD. Lower sizes provide PSD estimates of lower variance, but also lower resolution. On the other hand, greater sizes provide smaller reduction of variance and greater frequency resolution.

Figure 6.2 shows comparison of PSD estimates obtained using the Welch's method, with window size varying from 2^9 to 2^{15} samples, resulting in the PSD estimate of about 20 Hz to about 0.3 Hz resolution, respectively (the signal is sampled at 10 kHz). All estimates were obtained using Hamming window with 50 percent overlap between segments.

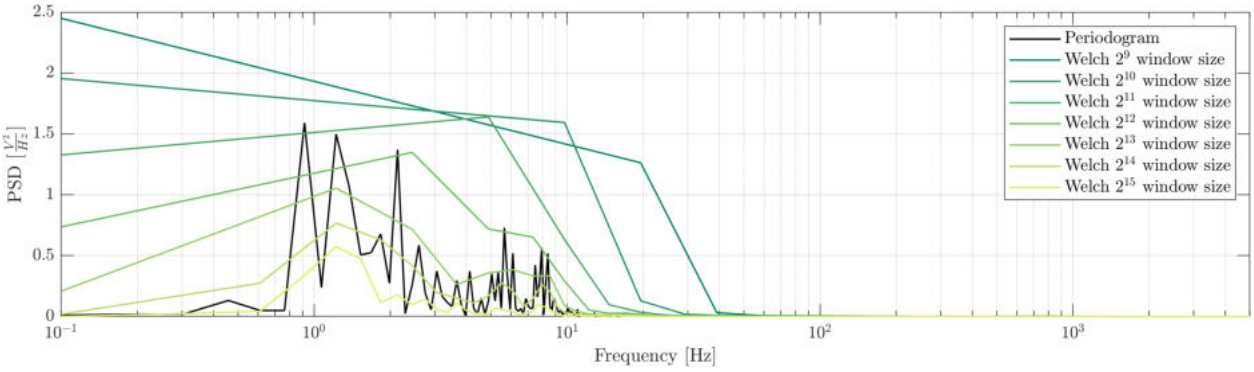

Figure 6.2: Comparison of the Welch's method to the periodogram.

Another method used to lower the variance of the periodogram PSD estimate is the Multitaper method [165]. Instead of splitting signal into segments and averaging their PSD estimates, the Multitaper method consists of obtaining the periodogram for multiple copies of the whole signal. Each copy is first multiplied by a unique Slepian Taper, which is a window function with optimal time-frequency concentration (windows across the signal copies are orthogonal) [165]. Then, for each *tapered* copy, the periodogram is obtained. Finally, periodograms of all the copies are averaged, yielding the Multitaper PSD estimate. Number of such copies is the main characteristic of the estimate, and with increasing number of copies, variance of the PSD decreases. Figure 6.3 shows comparison of PSD estimates obtained using the Multitaper method with varying number of signal copies (n).

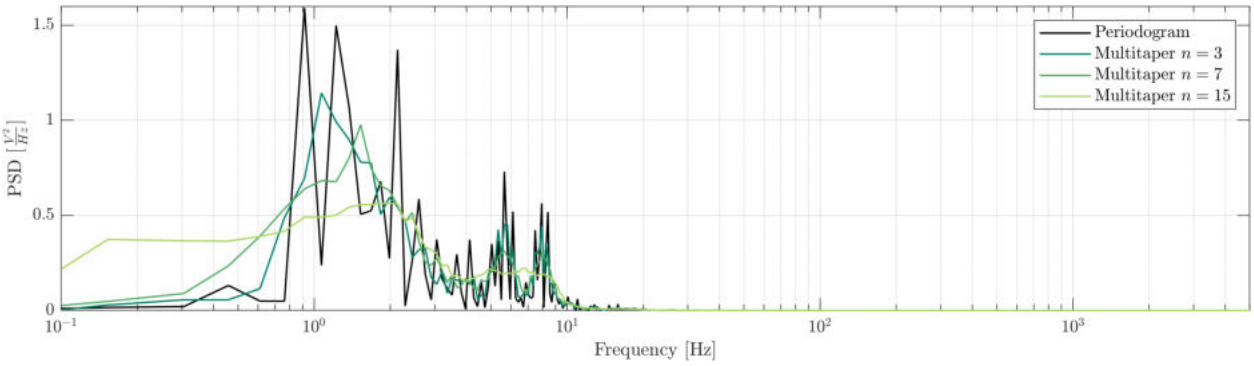

Figure 6.3: Comparison of the Multitaper method to the periodogram.

Choice of the parameters for both Welch's method (window size) and the Multitaper method (number of tapers) is not trivial, and in this work the following methodology was implemented to decide which PSD estimation method should be used for the frequency analysis. A comprehensive set of 22 VAG frequency feature sets was obtained for each of eleven methods:

- the raw periodogram,
- the Welch's method with seven widow sizes: from 2^9 to 2^{15}, like in Figure 6.2,
- the Multitaper method with three numbers of tapers: 3, 7 and 15, like in Figure 6.3.

Each one of 1275 features in each one of 22 feature sets, obtained via each one of the 11 PSD estimation methods, was evaluated using the $coef_{BMC}$ coefficient[1]. Then, the best results for each feature set were ranked across the methods, and the lowest average rank indicated which PSD estimator is potentially the most informative in the analyzed context. The specific feature sets were the Frequency Range Maps with 100 Hz resolution. They will be described in the following section.

6.2 Frequency Range Maps

With the signal being explicitly represented in the frequency domain, features can be extracted directly from the PSD estimate. However, there is one more ambiguity regarding the frequency analysis present in the VAG domain. It is the previously discussed range of the informative frequencies (see Figure 3.3 in the literature review). In other words, given some specific feature describing the spectrum, one could calculate it just for the specific frequency range. For example, instead of calculating the frequency of maximum PSD value across the whole PSD, one could be interested only in the frequency of maximum PSD in the frequency range from 100 Hz to 150 Hz. One approach to resolving this ambiguity is filtering the signal in the specified frequency range, as described in Chapter 4. However, this approach *assumes* specific frequency range to be informative and other potential ranges are not considered. For example, one of the preprocessing methods used in this work is a filtering in 50 Hz to 1000 Hz range. However, maybe more informative frequency content could be found in the 60 Hz to 990 Hz, or any other range.

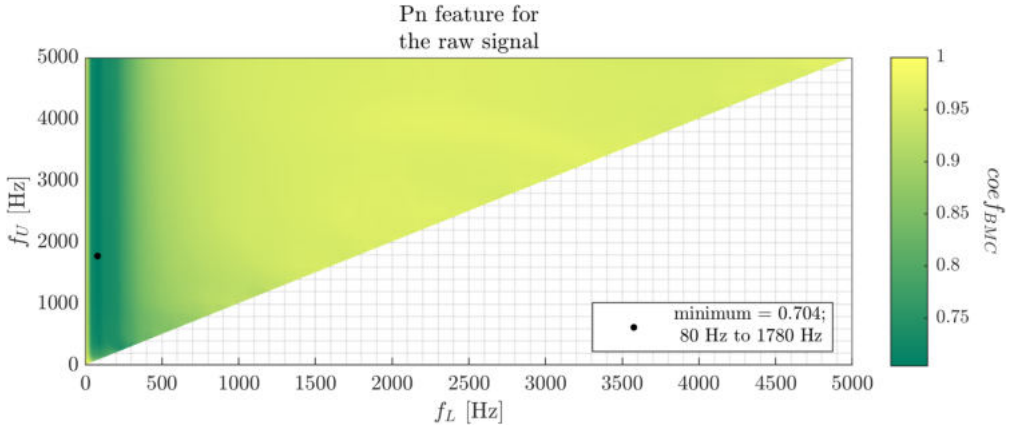

Figure 6.4: Exemplary Frequency Range Map calculated for the average power feature, with 10 Hz resolution.

[1]The $coef_{BMC}$ is the feature informativeness measure used in this work. More detailed description was provided in Section 8.1.

To visualize informativeness of the *whole* spectrum summarized by some specific feature, the FRM can be used, proposed by Łysiak et al. [116]. This map is essentially a three dimensional plot, with x- and y-axis indicating lower (f_L) and upper (f_U) frequency, respectively, and the color indicating informativeness ($coef_{BMC}$) of the feature calculated on the PSD in the f_L–f_U range. Exemplary FRM calculated for the average power feature (described in more detail later), with 10 Hz resolution (frequency range step) was visualized in Figure 6.4.

See, that the FRM from Figure 6.4 seem continuous, that is, there are no sharp changes in color. It is caused by the fact, that proximate points include mostly the same information. For example, average power in range 110 Hz–1000 Hz is *almost* the same as the average power in range 120 Hz˘1000 Hz. For features more sensitive to outliers (such as the maximum value), this continuity will not necessarily hold.

6.3 Scalar features used to calculate FRMs

In this section, all features extracted from the specific frequency range of the estimated PSD will be described. There is one important point, however, regarding values inside this range. They can be calculated directly on the PSD estimate, indicated by $Fn(k)$, or on the PSD estimate normalized to sum up to one. Befrui et al. [75] argued that this normalization is crucial, significantly improving informativeness of the frequency content in the context of power features. Such normalized spectrum will be indicated by $Fr(f)$:

$$Fr(k) = \frac{Fn(k)}{\sum_{k=1}^{N} Fn(k)}, \tag{6.2}$$

where N is number of frequency bins in the estimated PSD. The letter r will also be indicative of features calculated on such normalized PSD. This r stands for relative, indicating that the feature is calculated on the part of the spectrum relative to the whole PSD. The information content of such normalized PSD will be different than not-normalized PSD, since spectra across signals will be normalized by different values. Some features will also be calculated on spectrum normalized to one, but only inside the particular frequency range. Such normalization will be indicated by $Fs(k)$, with s standing for self-normalization:

$$Fs(k) = \frac{Fn(k)}{\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fn(k)}, \tag{6.3}$$

where $\mathbb{B}$ is a mapping function, that maps frequency to the closest frequency bin, and f_L and f_U are the lower and upper frequency ranges, respectively. Similarly to the relative normalization (r), features obtained for the self-normalized spectrum Fs will be indicated by the letter s. The self-normalization provides different information content than Fn and Fr, since again, values inside the range are normalized differently across the signals.

The normalization procedures change the power density inside the selected frequency range, but have no impact on the *shape* of the spectrum inside that range. Therefore, features measuring the shape will be calculated on the non-normalized spectrum only. The first such feature, and perhaps the most noticeable feature of the entire PSD, is the Frequency of the

most prominent power (F_{max}), which is around 0.9 Hz in the periodogram from Figure 6.1. It was used in vibroarthrogram analysis in an early study by Kenohan et al. [128], as well as in more recent ones, for example by Scherpereel et al. [121] or Shidore et al. [71]. In those later studies, however, it was used in the rolling context, which will be described in more detail in the following chapter. One could also extract value of the PSD for this specific frequency, i.e., the maximum value in the considered frequency range. This peak value can be calculated for all three normalization approaches, yielding three distinct features: Peak Value (non-normalized) (PVn), Peak Value (relatively-normalized) (PVr), and Peak Value (self-normalized) (PVs).

Treating the PSD estimate like a probability distribution, a number of distribution-based features can be calculated. One of such features is the Median Frequency (MedF), i.e., the frequency, above and below which, approximately half of the spectrum power is contained. This feature was widely used in VAG-related research in earlier works, such as the study by Zhang et al. [36], as well as more recent ones, such as the study by Kalo et al. [92]. Another central tendency measure is the Spectral Centroid (μ_f), which measures the barycenter of the spectrum [166]. It was also used in vibroarthrography field, by Andersen et al. [37] and Madeleine et al. [84].

The Spectral Spread (σ_f) is defined as the variance [121] or the standard deviation [73] of values around the Spectral Centroid. In the current work, the standard deviation definition will be used. Similarly to the Spectral Spread, Spectral Skewness ($skew_f$) and Spectral Kurtosis ($kurt_f$) can be used, measuring the asymmetry and peakiness of the spectrum around the Spectral Centroid [166]. Specific definitions of those features were included in Table 6.1.

In all feature definitions used in this section, two mapping functions will be used: $\mathbb{F}(k)$ is the function that maps frequency bin k to the frequency value at this bin, and $\mathbb{B}(f)$ is the function that maps specific frequency value f to the closest frequency bin. Moreover, since all FRM features are being calculated on the specific frequency range, f_L and f_U will be used to indicate lower and upper frequencies of this range, respectively. For the sake of clarity, they will be omitted in the left side of each definition. Finally, when the Fr, and Fs will be used, it will be assumed, that they are also normalized for frequency rage from f_L to f_U.

Probably the most widely used frequency features in the vibroarthrography field are the power features, i.e., features defined as the sums (or average values[2]) of the PSD for a given frequency range. There are multiple ranges used across VAG literature, including 50 Hz–250 Hz and 250 Hz–450 Hz, called *P1* and *P2*, respectively. Those ranges were introduced by Bączkowicz and Kręcisz [123] and later used in multiple other studies [5, 35, 40, 42, 90, 91, 93, 95, 97, 123]. In the original study analyzing signals used in this dissertation, Kręcisz and Bączkowicz [5] used also power at the frequency of 470 Hz and 780 Hz, called, respectively, *F470* and *F780*. Reddy et al. [101] used average power in 100 Hz–500 Hz range. Much lower frequencies, from 2 Hz to 10 Hz, were used by Faria et al. [102]. Multiple frequency ranges were used by Tanaka and Hoshiyama [76, 77]. In both studies, they used 50 Hz wide frequency bands, from 50 Hz–99 Hz to 300 Hz–349 Hz. Similar approach of

[2]Since the average value is just a linear transformation of the sum, informativeness of the features defined as the sum and the average is identical.

Table 6.1: Summary of the PSD distribution features.

name	definition
Frequency of the most prominent power	$F_{max} = \mathbb{F}\left(\max_k \{Fn(k)\}\right)$
Peak Value (non-normalized)	$PVn = Fn\left(\mathbb{B}\left(F_{max}\right)\right)$
Peak Value (relatively-normalized)	$PVr = Fr\left(\mathbb{B}\left(F_{max}\right)\right)$
Peak Value (self-normalized)	$PVs = Fs\left(\mathbb{B}\left(F_{max}\right)\right)$
Median Frequency	$f_{median} = \mathbb{F}\left(min_k \left\{\left\lvert\left(\sum_{i=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fs(i)\right) - 0.5\right\rvert\right\}\right)$
Spectral Centroid	$\mu_f = \frac{\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} \mathbb{F}(k) \cdot Fn(k)}{\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fn(k)}$
Spectral Spread	$\sigma_f = \sqrt{\frac{\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} (\mathbb{F}(k) - \mu_f)^2 \cdot Fn(k)}{\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fn(k)}}$
Spectral Skewness	$skew_f = \frac{\sum_{i=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} (\mathbb{F}(k) - \mu_f)^3 \cdot Fn(k)}{\sigma_f^3 \cdot \sum_{i=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fn(k)}$
Spectral Kurtosis	$kurt_f = \frac{\sum_{i=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} (\mathbb{F}(k) - \mu_f)^4 \cdot Fn(k)}{\sigma_f^4 \cdot \sum_{i=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fn(k)}$

multiple frequency ranges was used by Gharehbaghi et al. [100] and Richardson et al. [103]. In both studies, power spectrum of 29 logarithmically spaced ranges between 250 Hz and 10 kHz was used. Semiz et al. [98] used analogous strategy, for ranges between 250 Hz and 20 kHz. In all referenced studies, the frequency ranges were exclusive, i.e., they did not overlap.

Befrui et al. took similar approach of multiple frequency ranges, but in their study, the ranges were not exclusive. They extracted power from ranges defined on frequencies of 25 Hz, 50 Hz, 75 Hz, 100 Hz, 250 Hz, 500 Hz, 750 Hz and 1000 Hz. Similarly to the Frequency Range Maps, they visually represented the whole spectrum. However, instead of the specific informativeness coefficient describing specific feature (like the $coef_{BMC}$, for example), informativeness in their study was defined as the Area Under the ROC Curve (AUC) of the Support Vector Machine classifier trained on two features: power in frequency range f_L to f_U and in f_U to 8 kHz, with f_L and f_U being frequencies from the mentioned set.

Given all those different frequency ranges analyzed in the literature, the FRM approach seem to be able to provide valuable insights that current state-of-the-art VAG-related studies seem to lack. The power in this work will be calculated as the average value of the given frequency range on the non-normalized and the relative-normalized PSD, since self-normalized power would result in the same value for every signal (that is, one over the number of frequency bins). Additionally, the geometric mean power will be calculated, indicating the multiplicative central tendency of the PSD. It will potentially provide a more robust and representative measure of the central power within a selected frequency band, especially in ranges with substantial power variations. The geometric mean will be calculated for each normalization. Finally, also for every normalization, the RMS value of the specific frequency range will be calculated, indicating the "effective" power, taking into account the impact of peaks in the PSD. All the frequency power features, with their respective definitions, were

summarized in Table 6.2.

Table 6.2: Summary of the PSD power features.

name	definition
Average Power (non-normalized)	$Pn = \frac{1}{N} \cdot \sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fn(k)$
Average Power (relatively-normalized)	$Pr = \frac{1}{N} \cdot \sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fr(k)$
Geometric Mean (non-normalized)	$Gn = \left(\prod_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fn(k)\right)^{\frac{1}{n}}$
Geometric Mean (relatively-normalized)	$Gr = \left(\prod_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fr(k)\right)^{\frac{1}{n}}$
Geometric Mean (self-normalized)	$Gs = \left(\prod_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fs(k)\right)^{\frac{1}{n}}$
Root Mean Square (non-normalized)	$RMSn = \sqrt{\frac{1}{N} \sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fn(k)^2}$
Root Mean Square (relatively-normalized)	$RMSr = \sqrt{\frac{1}{N} \sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fr(k)^2}$
Root Mean Square (self-normalized)	$RMSs = \sqrt{\frac{1}{N} \sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fs(k)^2}$

Using the power measures, more complex shape features can be defined, such as the Spectral Crest, defined as the maximum value of a given range, divided by its average value [166] or the Spectral Flatness, defined as the geometric mean divided by the arithmetic mean [166]. The Spectral Flatness, also known as the Wiener Entropy, was applied in the VAG literature, in the preprocessing step. It served as a threshold to determine whether further filtering was necessary for a sub-signal of the decomposition. Besides the Spectral Crest and the Spectral Flatness, and similarly to the time-domain statistical features, the Spectral Form Factor could be defined, as a ratio of the RMS value to the arithmetic mean. Note, that all those features describe the shape of the distribution and as such, normalization does not affect their information content. Therefore, they could be calculated on the self-normalized spectrum. And, since the summation and the average value contain the same information, and the summarized self-normalized power is equal to one, all three features could be defined without denominator in their respective definitions. Therefore, in terms of the information content, the Spectral Crest is equivalent to the Geometric Mean of the self-normalized PSD (Gs_f), the Spectral Flatness is equivalent to the Peak Value of the self-normalized PSD (PVs_f), and the Form Factor is equivalent to the RMS of the self-normalized PSD ($RMSs_f$). Therefore, separate FRMs of the Spectral Crest, the Spectral Flatness and the Spectral Form Factor will not be calculated.

Other features describing shape of the spectrum are the Spectral Slope ($slope_f$) and the Spectral Decrease ($decrease_f$) [166]. They both measure how fast the spectrum decreases for the higher frequencies. The $slope_f$ is defined as the slope of the line fitted to the PSD. In VAG-related research, it was used in the context of the Fractal Dimension estimation by Rangayyan et al. [61] and Cai et al. [60]. On the other hand, $decrease_f$ measures how fast the power spectrum of a signal decreases from the lowest frequency. The main difference in interpretation of the $slope_f$ and $decrease_f$ is that the $decrease_f$ is a measure of how the power spectrum drops off from the first frequency bin, while the $slope_f$ measures the average decrease across the whole spectrum. In other words, the $decrease_f$ is more sensitive

to changes in the lower frequencies, while the $slope_f$ gives a more global view of how power changes with frequency.

The relative-normalized and the self-normalized PSDs can be treated as the Probability Density Function. As such, it allows for the calculation of Shannon Entropy, measuring how concentrated, or spread across the frequencies, the power is. The Spectral Entropy (relatively-normalized) (Hr) indicates the distribution of the power relative to the entire spectrum, while the Spectral Entropy (self-normalized) (Hs) measures distribution of the power within that range only. The difference between them is that the Hr value for a given frequency range is influenced by the power content of the entire frequency spectrum, providing a comparative measure, while the Hs value represents a standalone measure of how the energy within the frequency range is distributed, not affected by power levels outside of the range. Finally, there is another entropy measure defined on the PSD, defined in the electroencephalography domain, called the Irregularity Index (II) [167]. Similarly to the Hr it measures entropy of the specific range relative to the whole spectrum. However, it differs in that in the definition of II, the entropy of the specific range is explicitly divided by the entropy of the whole relative-normalized PSD. Specific definitions of the entropy measures, as well as the $slope_f$ and the $decrease_f$ were provided in Table 6.3.

Table 6.3: Summary of the PSD shape features.

name	definition	remarks
Spectral Slope	$slope_f = \frac{\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)}(Fn(k)-Pn_f)(\mathbb{F}(k)-\mu_f)}{\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)}(\mathbb{F}(k)-\mu_f)^2}$	
Spectral Decrease	$decrease_f = \frac{\sum_{k=\mathbb{B}(f_L)+1}^{\mathbb{B}(f_U)} \frac{Fn(k)-Fn(1)}{k-1}}{\sum_{k=\mathbb{B}(f_L)+1}^{\mathbb{B}(f_U)} Fn(k)}$	
Spectral Entropy relative-normalized	$Hr = -\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fr(k) \cdot \log(Fr(k))$	
Spectral Entropy self-normalized	$Hs = -\sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fs(k) \cdot \log(Fs(k))$	
Irregularity Index	$II = \frac{Hr_f+\alpha \cdot Hs_f}{H_f}$	$H_f = -\sum_{k=1}^{N} Fr(k) \cdot \log(Fr(k))$, and $\alpha = \sum_{k=\mathbb{B}(f_L)}^{\mathbb{B}(f_U)} Fr(k)$

6.4 Other frequency features

In a couple of VAG-related publications, the frequency power ratio was used to describe the signal. For example, Lee et al. [81] used a feature defined as the power in 20 Hz to 100 Hz divided by the power of 5 Hz to 500 Hz. Also, VAG signals registered in different context, i.e., for the temporomandibuar joint, are often described by the ratio of power below 300 Hz to to power above 300 Hz [168]. It would be interesting to somehow visualize various ratios using Frequency Range Maps. However, such visualization would require additional dimension(s). Therefore, in this work, various ratios will be extracted, but for whole PSD only. That is, for each frequency in the estimated PSD, a feature will be extracted defined as the power equal or below this frequency divided by the power above this frequency. For the sake of completeness, additional feature will be defined, as the reciprocal of the described ratio, i.e.,

the power above the frequency divided by the power equal or below it.

Similar feature, also extracted for one-dimensional parameter, is the Spectral Generalized Roll-off Point, similar in its definition to the percentile feature described in previous chapter. The Spectral Roll-off Point feature is defined as a frequency below which 95 percent of power is contained [166]. It was used in a number of VAG-related studies, but in the rolling context, described in more detail in the following chapter. Instead of taking this specific 95 percent value, in this work, a number of such thresholds will be tested, from 0 percent to 100 percent, with 1 percent step size.

Table 6.4: Summary of the frequency features not following the FRM methodology.

name	definition	remarks
Spectral Low-High Ratio	$LHR(f) = \frac{\sum_{k=1}^{\mathbb{B}(f)} Fn(k)}{\sum_{k=\mathbb{B}(f)+1}^{N} Fn(k)}$	
Spectral High-Low Ratio	$HLR(f) = \frac{\sum_{k=\mathbb{B}(f)+1}^{N} Fn(k)}{\sum_{k=1}^{\mathbb{B}(f)} Fn(k)}$	
Spectral Generalized Roll-off Point	$SGRP(r) = \mathbb{F}\left(\min_k \left\{\left\lvert\left(\sum_{i=1}^{k} Fs(i)\right) - r\right\rvert\right\}\right)$	

Both the frequency ratio features and the Generalized Spectral Roll-off Point, with their respective definitions, were summarized in Table 6.4.

Chapter 7

Time-frequency analysis

Signal analysis in time domain allows to draw conclusions from temporal changes of some specific value, like the acceleration. If this value changes periodically, analysis of the period, or periods could be of great value. Extracting information about the periodicities directly from the time domain, however, is difficult. In such cases, transforming the signal from time to frequency domain allows much more straightforward analysis. Frequency analysis, however, provides information about the specific frequencies that can be found *somewhere* in the signal. That is, the PSD shows *average* frequency content over the *entire* duration of the signal, i.e., without temporal localization. For example, a 10-second sinusoidal signal that has frequency of 100 Hz for the first 5 seconds and a frequency of 200 Hz for the remaining 5 seconds, would yield the same PSD estimate as the signal that fluctuates between 100 Hz and 200 Hz uniformly throughout its entire duration. Therefore, while the PSD provides valuable insights into the frequency composition of a signal, it does not give any information on how these frequencies change over time.

Time domain and frequency domain analyzes provide complementary information about the signal. Combining those approaches, that is, not analyzing signal in those two domains separately, but together, seems to be able to provide more comprehensive information. However, measuring changes of the frequency components in the time domain seems impossible, or contradictory. Since frequency is the inverse of time, multiple time samples are required to transform the signal into the frequency domain. That is, to draw *any* conclusions about the frequency content, *multiple* time samples need to be analyzed *simultaneously*, thereby losing temporal information. Thus, the frequency analysis cannot be conducted on one-sample time resolution. However, measuring changes of the frequency components in time domain is not necessarily contradictory. Rather, there is a tradeoff between resolution in one domain and the second domain, called the Uncertainty Principle [169]. That is, the more time samples are used to transform the signal, the more frequency bins will be possible to distinguish in the frequency domain up to the Nyquist frequency, i.e., half of the sampling frequency.

7.1 The Spectrogram

The general idea behind time-frequency analysis used in this work is the same as with the rolling features described in the Chapter 5. That is, to split a signal into multiple short sub-signals and perform analysis on each sub-signal. Obtaining the periodogram of each sub-signal and stacking the consecutive periodograms, yields the time-frequency distribution called the spectrogram [169]. Extracting a frequency feature from each short-time PSD estimation creates a time-varying Spectral Fluctuation Signal (SFS). Each such SFS is further analyzed using some time domain methods described in Chapter 5, finally providing scalar,

time-frequency features. As in the separate time and frequency analyzes, time-frequency analysis was conducted on all three preprocessed versions of the VAG signal.

There are three crucial parameters that need to be chosen for spectrogram calculation, with the most important being length of the sub-signal, or the size of the window. As mentioned, wider windows provide better resolution in frequency, while short windows allow for more precise time localization. Another parameter is the overlap between windows. Greater overlap provides more time samples, making the spectrogram more continuous in time domain. Note, that greater overlap does not increase the time resolution related to the uncertainty principle. Time samples are shared between windows, and therefore high overlap increases continuity of the spectrogram, but information contained in the overlapped parts of the signal is redundant. Finally, curve by which each sub-signal is multiplied, or shape of the window, is of great value. It determines the balance between frequency resolution and the spectral leakage. That is, it influences how finely distinct frequency components can be resolved[1], and how much spectral leakage appears. Spectral leakage is an artifact where power in the estimated PSD spreads into frequencies that are not in the actual signal, potentially masking or distorting the true spectral content [170].

In the VAG-related literature, various window sizes were used, varying from 2 ms [127] to 500 ms [98], with overlap between windows varying from 20 percent [71] to over 97 percent [115]. The most popular shape of the window used in vibroarthrography is the rectangular window, however, other windows, such as Hamming [115], Tukey [69] and Hanning [5] were also used. Table 7.1 includes a short summary of windowing approaches used in some VAG-related studies. Alongside window shape, size and overlap between windows, the table includes also the frequency resolution inside each window, resulting from window size and sampling frequency[2]. It is worth noting that in all studies included in the table, rolling approach to frequency analysis was utilized, but not necessarily the spectrogram approach.

Note, that not only the window size varies significantly between studies, but also the frequency resolution, rendering both parameters difficult to choose. Therefore, multiple combinations will be compared in the current work, with the window size varying from $2^4 = 16$ to $2^{14} = 16384$ samples, with consecutive powers of two. With the sampling frequency of 10 kHz, time of each window will vary between 1.6 ms and 1638.4 ms, and the frequency resolution will vary from 625 Hz to about 0.6 Hz, respectively. The overlap will vary from 0 percent (no overlap) to 90 percent of the window size, with 10 percent step. For percentages resulting with non-integer sample sizes, the overlap will be rounded down to the nearest integer.

The window size used in the current work will be the Hamming window. It was chosen over the Hanning, Rectangular, or Tukey windows also present in the VAG-related studies, since it provides higher practical frequency resolution[3] than the Hanning window, and significantly decreases spectral leakage compared to the Rectangular and Tukey windows [170].

Three exemplary spectrograms were included in Figure 7.1, with window sizes of 16, 512, and 16384 samples, using 90-percent overlap. All spectrograms were obtained for the

[1]Not in the sense of the uncertainty principle, but rather in a practical sense of distinguishing closely-spaced frequencies.

[2]Frequency resolution is calculated as the sampling frequency divided by the window size.

[3]The ability to distinguish close frequencies resulting from the effects of the windowing process, not the uncertainty principle.

Table 7.1: A review of time-frequency analysis window sizes used in the VAG domain.

references	window shape	window width		overlap		sampling	resolution
		samples	time [ms]	time [ms]	percentage	[kHz]	[Hz]
[127]	ns	100	2.0	1.8	90.0	50.0	500.0
[5]	Hanning	150	15.0	10.0	66.7	10.0	66.7
[41]	Rectangular	30	15.0	3.0	20.0	2.0	66.7
[71]	Rectangular	60	30.0	20.0	66.7	2.0	33.3
[73]	Rectangular	4000	40.0	36.0	90.0	100.0	25.0
[121]	Rectangular	1250	50.0	25.0	50.0	25.0	20.0
[115]	Hamming	512	51.2	50.0	97.7	10.0	19.5
[69]	Tukey	128	64.0	ns	ns	2.0	15.6
[100]	Rectangular	20000	200.0	100.0	50.0	100.0	5.0
[72]	Rectangular	8820	200.0	100.0	50.0	44.1	5.0
[103]	Rectangular	20000	200.0	180.0	90.0	100.0	5.0
[99]	Rectangular	40000	400.0	ns	ns	100.0	2.5
[98]	Rectangular	54000	500.0	250.0	50.0	108.0	2.0

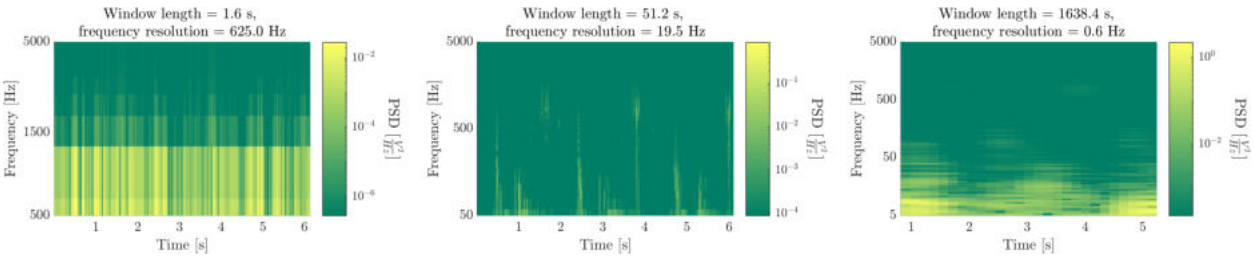

Figure 7.1: Exemplary spectrograms with varying window sizes.

exemplary raw signal used in earlier examples. It is clear how window size influences both time and frequency resolution. See, that the spectrogram with the shortest window has the lowest frequency resolution, with hardly any changes along the y-axis. On the other hand, the one with the greatest window size has much better frequency resolution, but time resolution is much worse. It is difficult to predict which resolution should be sacrificed to which, in order to obtain the most informative features. It is also possible that different features could benefit, in terms of the informativeness, from different time-frequency tradeoffs.

7.2 Spectral Fluctuation Signals

In this work, the specific, scalar features are obtained from the spectrograms in two steps. First, the spectrogram is reduced to *some* SFS using one of the specific frequency features described in the previous chapter. That is, for each time sample, the corresponding PSD estimate is reduced to some frequency feature, that creates time-varying SFS. Then, the SFS is further reduced to specific scalar values using *some* time domain features described in the

Chapter 5. This process was visualized in the Figure 7.2, using 512-sample spectrogram from the Figure 7.1.

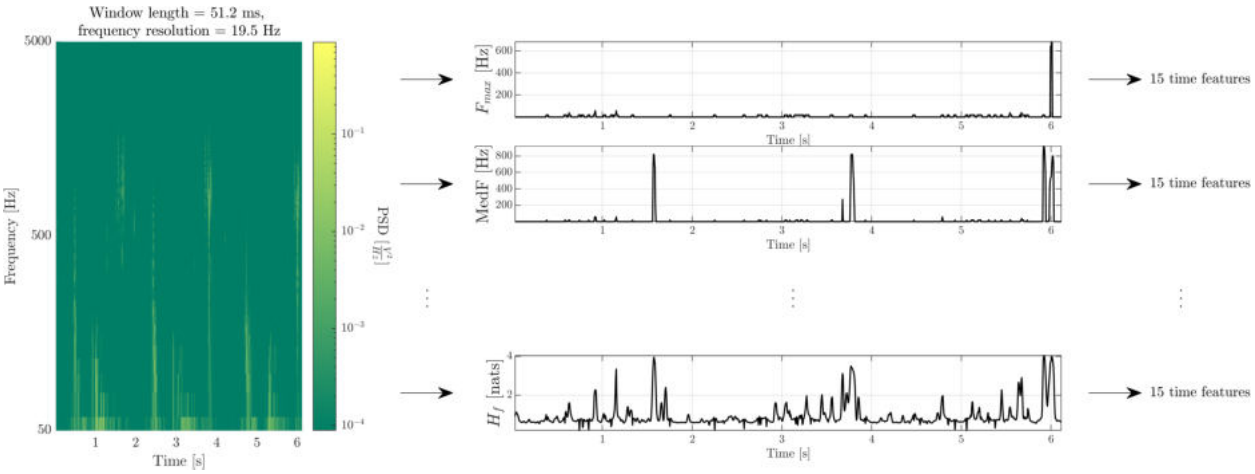

Figure 7.2: Exemplary time-frequency feature extraction.

The frequency features described in the previous chapter were calculated for specific frequency range. To reduce complexity of the analysis, however, time-frequency features will be obtained for the full frequency range, i.e., from 0 Hz up to 5 kHz. As a result, self-normalization and relative-normalization features will be the same. Therefore, features will be calculated either without normalization, on the spectrogram $SP(k,t)$ where k is the frequency bin and t is the time sample, or on normalized spectrogram: $SPr(k,t) = \frac{SP(k,t)}{\sum_{k=1}^{n} SP(k,t)}$. Summary of the features used to obtain SFSs, with their respective definitions, was included in Table 7.2.

Note, that there are two new features in Table 7.2, called the Spectral Flux. They are defined as the euclidean distance between consecutive samples of the spectrogram [171]. Both Absolute and Relative Spectral Flux are used to quantify the changes in the PSD of a signal from one time sample to the next. However, the normalization in the Relative Spectral Flux makes it sensitive to the proportional changes in different frequency bins. The Absolute Spectral Flux measures total change in power across the PSD, while the Relative Spectral Flux measures changes in the relative distribution of power across the PSD, discarding changes in total power.

7.3 Time analysis of the Spectral Fluctuation Signals

After obtaining all 21 time-varying SFSs, each one of them was summarized using 15 selected time features described in Chapter 5. Those features included non-temporal statistics measuring the central tendency, spread and shape of the distribution, as well as temporal measures based on the analysis of the derivative of the SFS. Table 7.3 includes summary of those features.

Only a subset of the features described in the Chapter 5 was extracted from each SFS, and there are two reasons for such approach. First, the described time-frequency analysis is

Table 7.2: Summary of the frequency features used for SFS calculation.

name	definition
Frequency of the most prominent power	$F_{max}(t) = \mathbb{F}\left(\max_k \{SP(k,t)\}\right)$
Peak Value (non-normalized)	$PVn(t) = SP\left(\mathbb{F}\left(f_{max}(t)\right), t\right)$
Peak Value (relatively-normalized)	$PVr(t) = SPr\left(\mathbb{F}\left(f_{max}(t)\right), t\right)$
Median Frequency	$f_{median}(t) = \mathbb{F}\left(min_k \left\{\left\lvert\left(\sum_{i=1}^{k} SP(i,t)\right) - \sum_{k=1}^{n} SP(k,t)\right\rvert\right\}\right)$
Spectral Centroid	$\mu_f(t) = \frac{\sum_{k=1}^{n} \mathbb{F}(k) \cdot SP(k,t)}{\sum_{k=1}^{n} SP(k,t)}$
Spectral Spread	$\sigma_f(t) = \sqrt{\frac{\sum_{k=1}^{n} (\mathbb{F}(k) - \mu_f(t))^2 \cdot SP(k,t)}{\sum_{k=1}^{n} SP(k,t)}}$
Spectral Skewness	$skew_f(t) = \frac{\sum_{i=1}^{n} (\mathbb{F}(k) - \mu_f(t))^3 \cdot SP(k,t)}{\sigma_f(t)^3 \cdot \sum_{i=1}^{n} SP(k,t)}$
Spectral Kurtosis	$kurt_f(t) = \frac{\sum_{i=1}^{n} (\mathbb{F}(k) - \mu_f(t))^4 \cdot Fn(k)}{\sigma_f(t)^4 \cdot \sum_{i=1}^{n} SP(k,t)}$
Average Power (non-normalized)	$P_f(t) = \frac{1}{n} \cdot \sum_{k=1}^{n} SP(k,t)$
Geometric Mean (non-normalized)	$Gn_f(t) = \left(\prod_{k=1}^{n} SP(k,t)\right)^{\frac{1}{n}}$
Geometric Mean (relatively-normalized)	$Gr_f(t) = \left(\prod_{k=1}^{n} SPr(k,t)\right)^{\frac{1}{n}}$
Root Mean Square (non-normalized)	$RMSn_f(t) = \sqrt{\frac{1}{n} \sum_{k=1}^{n} SP(k,t)^2}$
Root Mean Square (relatively-normalized)	$RMSr_f(t) = \sqrt{\frac{1}{n} \sum_{k=1}^{n} SPr(k,t)^2}$
Spectral Crest	$crest_f(t) = \frac{PV_f(t)}{P_f(t)}$
Spectral Flatness	$flatness_f(t) = \frac{G_f(t)}{P_f(t)}$
Spectral Form Factor	$ff_f(t) = \frac{RMS_f(t)}{P_f(t)}$
Spectral Slope	$slope_f(t) = \frac{\sum_{k=1}^{n} (SP(k,t) - P_f(t))(\mathbb{F}(k) - \mu_f(t))}{\sum_{k=1}^{n} (\mathbb{F}(k) - \mu_f(t))^2}$
Spectral Decrease	$decrease_f(t) = \frac{\sum_{k=2}^{n} \frac{SP(k,t) - SP(1,t)}{k-1}}{\sum_{k=2}^{n} SP(k,t)}$
Spectral Entropy	$H_f(t) = -\sum_{k=1}^{n} SPr(k,t) \cdot \log(SPr(k,t))$
Spectral Flux (non-normalized)	$SFn(t) = \sqrt{\sum_{k=1}^{n} \left(SP(k,t) - SP(k,t-1)\right)^2}$
Spectral Flux (relatively-normalized)	$SFr(t) = \sqrt{\sum_{k=1}^{n} \left(SPr(k,t) - SPr(k,t-1)\right)^2}$

already rather complex, including variation of window size and overlap in the spectrogram calculation, as well as multiple PSD-based measures (or SFSs). Second, all SFSs, except for the Spectral Slope, can take only positive values, rendering some time-domain features useless. For example, the average value and the rectified average value, or the CV and the MCV, would be equal and therefore redundant.

Table 7.3: Summary of the time features extracted from each SFS.

name	definition	remarks
mean	$\mu(SFS) = \frac{1}{n}\sum_{i=1}^{n} x(i)$	$x(i)$ is the i-th sample of the n-sample SFS.
median	$M(SFS) = \frac{(SFS)(\frac{n}{2})+(SFS)(\frac{n}{2}+1)}{2}$	(SFS) is the sorted set of SFS's values.
std	$\sigma(SFS) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(SFS(i) - \mu(SFS))^2}$	
spread	$spread(SFS) = \max(SFS) - \min(SFS)$	max and min denote maximum and minimum value of the SFS.
skewness	$skew(SFS) = \frac{\frac{1}{n}\sum_{i=1}^{n}(SFS(i)-\mu(SFS))^3}{s^3}$	
kurtosis	$kurt(SFS) = \frac{\frac{1}{n}\sum_{i=1}^{n}(SFS(i)-\mu(SFS))^4}{s^4} - 3$	
CV	$CV(SFS) = \frac{\sigma(SFS)}{\mu(SFS)}$	
form factor	$FF(SFS) = \frac{RMS(SFS)}{RAV(SFS)}$	
crest factor	$CF(SFS) = \frac{\max(SFS)}{RMS(SFS)}$	
impulse factor	$IF(SFS) = \frac{\max(SFS)}{RAV(SFS)}$	
margin factor	$MF(SFS) = \frac{\max(\lvert SFS\rvert)}{(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert SFS(i)\rvert})^2}$	
MoAD	$MoAD(SFS) = \frac{1}{n-1}\sum_{i=1}^{n-1}\lvert SFS(i+1) - SFS(i)\rvert$	
mobility	$mobility(SFS) = \frac{\sigma(SFS')}{\sigma(SFS)}$	$\sigma(SFS)$ denotes standard deviation of the SFS, and $\sigma(SFS')$ denotes standard deviation of the derivative of the SFS.
complexity	$complexity(SFS) = \frac{\sigma(SFS'')}{\sigma(SFS')} / \frac{\sigma(SFS')}{\sigma(SFS)}$	$\sigma(SFS'')$ denotes standard deviation of the double-derivative of the SFS.
TCR	$TCR(SFS) = \frac{1}{n-2}\sum_{i=1}^{n-2}\mathbb{I}\left((SFS(i+2) - SFS(i+1)) \cdot (SFS(i+1) - SFS(i)) < 0\right)$	

Chapter 8

Dimensionality reduction and classification

In previous chapters, specific features extracted from each VAG signal were introduced. Some of those features, however, provide very similar, or even identical information. See, for example, that the Frequency Range Maps are constructed by multiple features, that are highly correlated, rendering the maps visibly "smooth", or continuous. In the classification context, such redundancy leads not only to increased computational complexity, but also overfitting and worsened classification performance in general [172]. Measuring the *informativeness* of the feature, however, is not a trivial task. Therefore, in the first section of this chapter, the specific feature evaluation method will be described. The next section will be devoted to the method of reducing dimensionality of the extracted features. In the second section, specific classifiers will be described, that were used to classify reduced feature set.

8.1 Feature evaluation

Selecting *proper* features is a crucial step of any classification problem. These features should describe the object being classified as clearly as possible. However, evaluation of the feature quality is a difficult task. In modern research that heavily relies on Big Data, the emphasis on quantity often results in disregarding the importance of data quality [173]. Numerous techniques are available for evaluation purposes [172], and are usually divided into three categories: filter, wrapper, and embedded methods [172]. Wrapper and embedded methods use some additional models (such as classifiers or regressors) to evaluate feature in specific context. Filter methods, on the other hand, evaluate feature quality independently of the classification algorithm, which is desirable for two major reasons. First, the classification algorithms are usually best suited for specific distribution of data. For example, the Support Vector Machine is best suited for linearly separable data [174], while the k-Nearest Neighbors classifier can be used for nonlinearly separable data [142]. Usage of some filter method to evaluate quality of the feature ensures that specific characteristics of the classification algorithm do not influence evaluation of the feature's quality. Second, since they do not use additional models, filter methods are usually much less computationally expensive [172].

In an earlier study by the author of this work and his supervisor [133], some of the most popular feature evaluation methods were compared in the context of classification. The comparison was conducted as follows. Every method evaluated features from some feature sets. The best feature subset from each feature set was chosen and then used to train a set of classifiers. Results of this classification were used as a quality measures of each evaluation method. This comparison was conducted, besides others, on the sets of features

extracted from VAG signals analyzed in this work. Results indicated, that in the context of classification, the most useful measure proved to be the Battacharyya coefficient. It led to the most accurate classification of the VAG signals, and as such, will be used in the rest of this work.

The Bhattacharyya coefficient, or the $coef_B$, is a measure based on the probability density function with wide usage across the literature [175,176]. Intuitively, the $coef_B$ measures the overlap between distributions of values between two classes, and is defined as follows:

$$coef_B = \sum_{x \in X} \sqrt{P_1(x) \cdot P_2(x)}, \tag{8.1}$$

where P_1 and P_2 and distributions of features within two classes in the feature values domain, X[1]. Visualization of this concept was provided in Figure 8.1.

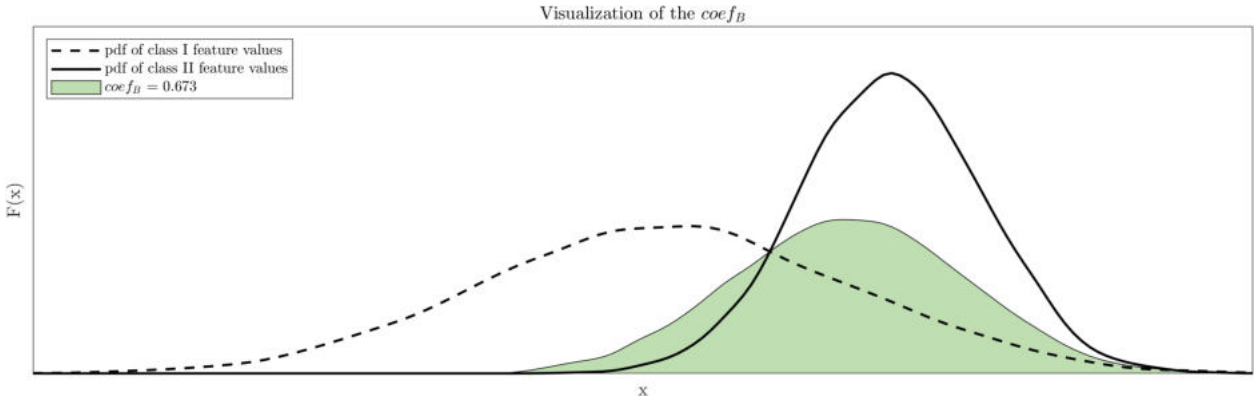

Figure 8.1: Visualization of the $coef_B$.

The $coef_B$ varies between 0, for the "perfect" feature, i.e., feature which values do not overlap between classes, to 1, for the feature which values overlap between classes completely.

Since in this work multiple classes are considered, a multi-class version of the $coef_B$ will be used, called $coef_{BMC}$. It is simply defined as the average of $coef_B$ across the pairs of classes, weighted by their prevalence:

$$coef_{BMC} = \frac{1}{\binom{k}{2}} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} w(i) \cdot w(j) \cdot coef_B(P_i, P_j), \tag{8.2}$$

where k is the number of all classes, $\binom{k}{2}$ is the number of all class pair combinations, P is the Probability Distribution Function of feature values within a class, and w is a class weight proportional to its prevalence. This coefficient also ranges from 0 to 1 with smaller values indicating less overlap between the classes and larger values indicating higher overlap. Note, that $coef_{BMC}$ only provides information about separation of *some* classes, without information about which classes are separated and to what extent.

[1]Since explicit distributions of features are not know, they are estimated using the Kernel Density Estimation, as described in more detail in [133].

8.2 Dimensionality reduction

The current section will be divided into three subsections. In the first one, a summary of all extracted features will be provided, as well as the general description of the two-stage dimensionality reduction process. Two following subsections will be devoted to the description of the specific approaches took in order to minimize the dimensionality, namely the Maximum Relevance Minimum Redundancy algorithm and Principal Component Analysis.

8.2.1 Summary of all extracted features and general dimensionality reduction approach

Features extracted in this work were broadly divided into three categories: time, frequency and time-frequency. All features were extracted for three preprocessing methods, i.e., the raw, the filtered, and the derivative signal. Time domain analysis yielded 3142 features per preprocessing. Frequency analysis resulted in 22 FRMs, 125 250 features each, and 16 485 features not following FRM approach. Time-frequency analysis allowed for extraction of 34 650 features. Therefore, 2 809 777 features were extracted for each preprocessing, summing up to the total of 8 429 331 features. Overwhelming majority of those features, however, is highly redundant. Remember, for example, mentioned continuity of most FRMs. It is a result of highly redundant information contained in proximate points on the map.

Reducing dimensionality of such number of features at once would be quite computationally expensive. In order to reduce computational cost, dimensionality reduction was conducted in two stages. In the first stage, features after each preprocessing were divided into 25 groups: time-domain features, 22 FRMs, other frequency-domain features and time-frequency features. In summary, counting all preprocessing methods, 75 groups were created. Best feature subset in each group was then reduced using Maximum Relevance Minimum Redundancy (MRMR) algorithm. That ensured easy interpretation of the obtained feature subset. In the second stage, all the subsets from the previous stage were combined and their dimensionality was further reduced using two approaches: again, the MRMR algorithm or the Principal Component Analysis (PCA).

This approach was visualized in Figure 8.2.

8.2.2 The Maximum Relevance Minimum Redundancy algorithm

The Maximum Relevance Minimum Redundancy [177] is a feature selection algorithm, that iteratively sorts features based on their informativeness (relevancy) in the analyzed context, while also ensuring that they are consecutively the least possibly redundant. It works as follows [178]:

1. initialize empty feature subset,
2. calculate relevance of every feature with respect to the classes,
3. add the most relevant feature to the subset,
4. for each remaining feature, calculate redundancy with the features in the chosen subset,
5. calculate the MRMR score of each remaining feature by subtracting its redundancy score from the relevance score (see Equation 8.3),

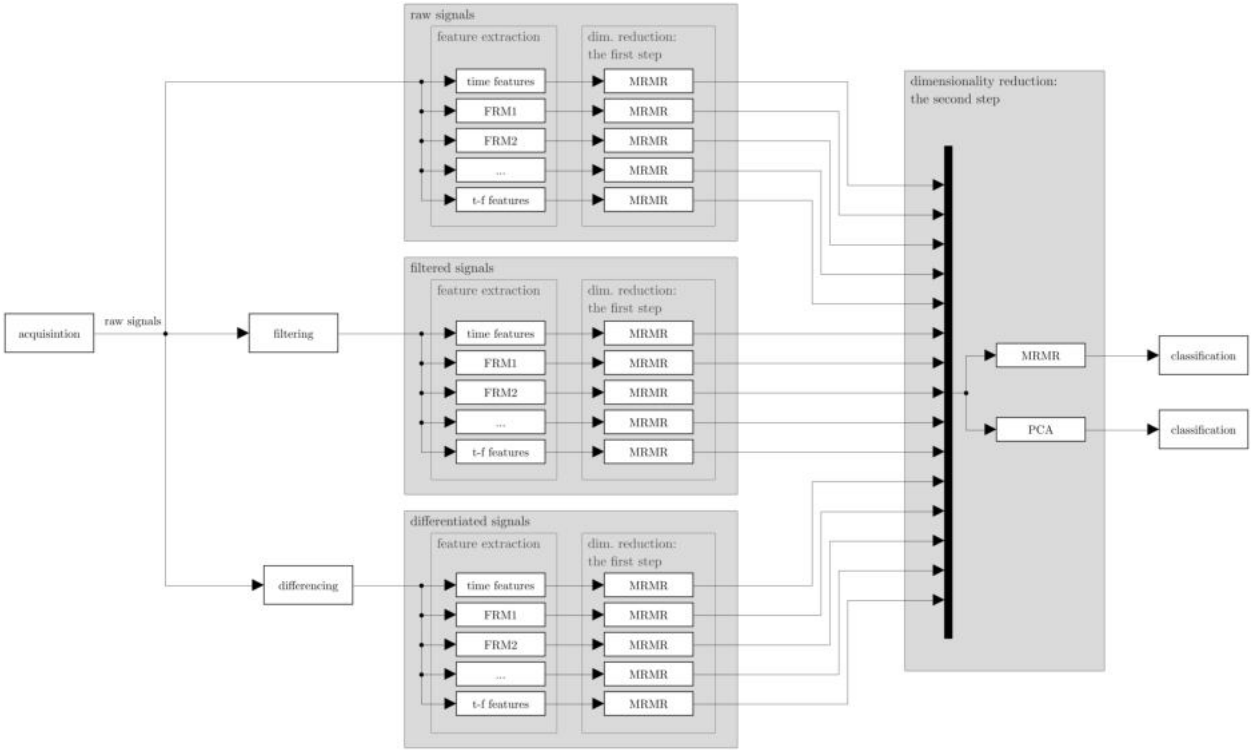

Figure 8.2: Visualization of the two-stage dimensionality reduction approach employed in this work.

6. add feature with the biggest score to the feature subset,
7. repeat steps 4-6 until the desired number of features is selected.

Because the relevance should increase for better (more relevant) features, in this study, the relevance was defined as the $coef_{BMC}$ subtracted from one. As for the redundancy, as in [178], it was calculated as the average value of the absolute Pearson's correlation between the features inside the subset and one of the remaining features. Therefore, the MRMR score was calculated as follows:

$$score(f) = 1 - coef_{BMC}(f) - \frac{1}{k}\sum_{i=1}^{k} |\rho(f, f_i)|, \tag{8.3}$$

where f is a feature remaining in the original chosen subset, k is a number of features in a chosen subset and ρ is Pearson's correlation coefficient.

One remaining factor is the number of features in the chosen subset (point 7 in the list above). There are couple of approaches to chose this number, and perhaps the most straightforward one is the elbow method [179]. When the MRMR scores are plotted against the number of selected features, such as in the Figure 8.3, a characteristic shape emerges, resembling an elbow. Features on the left-side of this elbow are considered relatively informative, while features after it represent diminishing returns.

Note, that while this plot generally decreases with consecutive features, it locally may increase. It can happen when a feature is highly informative, yet more redundant compared

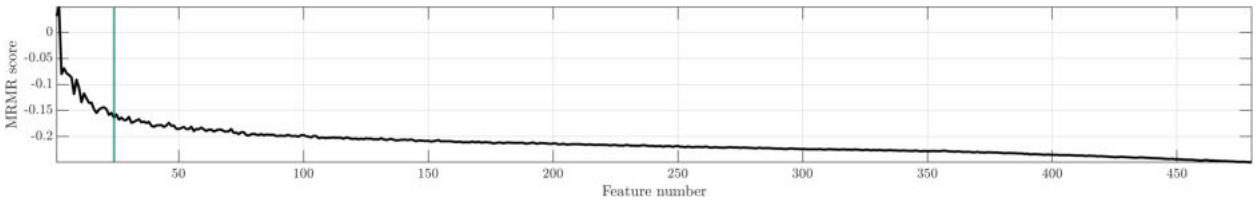

Figure 8.3: Visualization of the exemplary MRMR score for consecutively selected features. Green line indicates the elbow point.

to the previously selected features. Therefore, the elbow point might change after some iterations. In this study, the number of features to keep was based on the elbow point, that did not change for twenty consecutive iterations. The elbow point itself was found by fitting separate linear trends before and after each point of the curve and identifying the location where the difference between the two trends was greatest [179].

8.2.3 The Principal Component Analysis method

The Principal Component Analysis algorithm is another widely used tool for the dimensionality reduction task [180]. In contrast to the MRMR, instead of *selecting* the features, it *transforms* them to create a set of new, uncorrelated features that are linear transformations of the original features. It works as follows.

1. The original features are standardized, to ensure comparable influence of them on the new features.
2. The covariance matrix of the standardized features is calculated.
3. The eigenvalues and eigenvectors of the covariance matrix are calculated, representing the magnitude of variance and the directions of the principal components, respectively.
4. The eigenvectors are sorted in descending order based on their corresponding eigenvalues. This order represents the amount of variance each principal component captures, with the first being the most significant.
5. A desired number of top eigenvectors (principal components) are chosen to reduce dimensionality, based on the cumulative variance they capture.
6. The standardized feature matrix is then projected onto the selected principal components to transform it into the lower-dimensional space.
7. The result is a feature matrix with reduced dimensionality, where each new feature is a linear combination of the original standardized features, weighted by the eigenvector's coefficients.

A disadvantage of using the PCA dimensionality reduction is poor interpretability of the reduced feature matrix. Since each new feature is a linear combination of *all* original features, it is difficult to interpret specific values of the new features.

One additional concern while using the PCA for the dimensionality reduction is that it does not utilize any information about the relevance of the feature in a specific classification context. It creates new features, that capture the most *variability* of the original features matrix. However, not all variability in this matrix is necessarily informative in the context of classification. If the original feature matrix contains variability not related to the classified groups, the PCA might prioritize that, creating new features that capture the most

variability, while actually reducing informativeness understood in the classification context.

8.3 Classification algorithms

In the first dimensionality reduction step, an elbow method was used to determine number of features selected from each feature set. After the second step, i.e., after combining all the features from all selected subsets, and sorting them in order of importance, they could be already used for classification. Therefore, instead of selecting specific number of features and then using this specific number of selected features to train a classifier, multiple classifiers were trained, with increasing number of features. This number varied from one feature to 183, chosen to be one smaller than the number of samples.

In order to reduce the potential impact of a specific classifier on the distribution of the selected features, a total of eleven distinct classification algorithms were trained: the Linear Discriminant Analysis (LDA), the naive Bayes, two k-Nearest Neighbors (KNN), two decision trees, two decision forests, two Support Vector Machine (SVM)s, and a neural network classifier. They will be briefly described below.

The Linear Discriminant Analysis [181] works by finding a linear combination of features that maximizes the difference between means of classes, while minimizing the variance within each class, based on assumptions of equal class covariances and normally distributed features. Since it is based on maximizing distance between *means* of classes, it can be sensitive to outliers.

The Naive Bayes Classifier [182] assumes independence of features. It computes the probability of specific class based on the product of individual features' probabilities of belonging to that class.

The k-Nearest Neighbors classifier [142] works by comparing samples that are being classified to already classified ones. First, the distances of the classified sample to the training samples are being calculated. In this work, the euclidean distance is used. Then, k closest samples are chosen. The sample is classified to the class, that majority of k-nearest neighbors belong to. For small k, the classifier can model more flexible decision boundaries and local patterns in the features' values. However, it also becomes more sensitive to outliers and noise. On the other hand, greater k leads to more robust and generalized classification, but also lower sensitivity to local patterns. In this work, k values equal to 5 and 20 were used. Those values were chosen based on the principle that k should be smaller than the size of the least populated class to avoid potential bias towards more dominant classes, and to ensure that even the minority class can have a fair representation in the decision-making process. While the KNN with k equal to 5 aims to capture local patterns in the features, k equal to 20 provides more generalized classification.

Decision tree algorithms [183] create hierarchical classifiers by consecutively splitting the feature values, with each split being determined by the most informative feature. Greater number of splits allows more complex decision boundaries, but also increases the risk of overfitting. In this work, 5 and 10 splits were used. With a small number of highly informative features, decision trees perform well. However, with increasing number of features, especially of low informativeness, risk of overfitting increases. Decision trees can be considered robust against outliers, since they partition the features' values based on hierarchical splits,

rather than on their distribution, ensuring that small number of extreme values have limited influence on the general structure of the tree.

Random forest classifiers are ensemble methods. That is, they are based on multiple decision trees that are trained separately, and then their separate classifications are somehow aggregated. In this work, two approaches were used. First, bagging approach, trains multiple classifiers separately (64 in this work), on bootstrapped samples. That is, training samples as randomly drawn (with replacement) from the training data. The classification is done separately on each tree, and the sample is classified to the class, that was chosen by the majority of trees. The second random forest approach used in this work is RUSboost [184]. Boosting, in general, also consists of training multiple trees, however, not in parallel. Instead, subsequent trees are trained to correct the errors of their predecessors. RUSboost algorithm puts additional focus on class imbalance. In datasets where the numbers of samples in each class significantly differ, the algorithm adjusts the weights of misclassified instances and performs random under-sampling of the classes with more instances, ensuring better representation of class with fewer instances. In this work, the difference between number of samples in the most and the least numerous classes is not great (see Table 2.1). However, using RUSBoost could potentially ensure more balanced consideration of all classes, improving the generalization capabilities of the classifier.

The Support Vector Machine [174] is a linear classifier that finds an optimal hyperplane separating classes, while maximizing the distance between the hyperplane and the nearest point from either class. The Gaussian SVM first transforms the feature values into a higher-dimensional space, enabling a linear separation in the transformed space. The SVM is inherently a binary classifier. In this work, classification of multiple classes is done with the one-vs-one approach. That is, separate SVM is trained for each class pair. In the actual classification, voting approach is used: the final classification result is the class which was chosen by the majority of classifiers.

The Neural Network (NN) classifier's design is slightly more complex, but enables modeling of nonlinear dependencies between the features' values and classes [182]. It consists of multiple layers of neurons, all of which have multiple inputs and one output. Output of each neuron serves as an input to all neurons in the next layer, and is calculated as a weighted sums of its inputs transformed by the activation function. The first layer consists of the features' values, and the last layer contains one neuron for each class. Number of layers between those two, the number of neurons within them, as well as the activation functions of those neurons are the most important NN parameters. More complex networks, i.e., with more layers and neurons are able to model more intricate patterns, but require more training samples and increase the risk of overfitting. In this work, the middle layer was designed to have 10 neurons (one neuron for possible class pair) with Rectified Linear Unit activation function.

In order to prevent any classification model from overfitting, i.e., loosing its ability to generalize, the dataset is usually partitioned into training and testing subsets [185]. The testing subset, which the classifier did not use during training, is then used to determine the final classification accuracy. That ensures that new data will be classified with similar accuracy. The accuracy of the model is dependent on the division of the original dataset, with a greater impact when a smaller number of data points is available. To address any potential biases resulting from this fact, the dataset is typically split multiple times with a

separate training of the classifier for each split. The classification accuracy is then reported as the average accuracy across the splits. This approach is called cross-validation [185]. This study utilized the leave-one-out cross-validation, which involved training the model on all data points but one, on which the model was tested. The accuracy was an average from n models, where n was the number of signals in a signal base (i.e., 184).

Each classification model has its strengths and weaknesses in the context of different feature distributions. Multiple various features were extracted in this work. Analysis of specific characteristics of each one, and their various combinations would be unfeasible. Therefore, employing diverse set of classification algorithms enabled to utilize varied feature characteristics, increasing the likelihood of achieving high classification accuracy.

In summary, eleven classification algorithms were trained on increasing number of features sorted by the MRMR algorithm, as well as the PCA. Accuracy of each classifier was calculated as the average classification of 184 classifiers trained in leave-one-out cross-validation approach.

Part III

Results, Discussion and Conclusions

Chapter 9

Preprocessing

In the current chapter, results of the preprocessing methods comparison will be provided and discussed, in the context of normality and stationarity. Then, a short summary will be provided, including limitations of the chosen methods, as well as the conclusions.

9.1 Comparison of the preprocessing methods

For the visualization of the distributions, all signals in each knee joint condition classes were concatenated. Figure 9.1 shows QQ plots and histograms of the concatenated signals for the raw signal, Butterworth filtered signal and the derivative signal. In the QQ plots, the dots correspond to the signal quantiles, while the solid black line represents corresponding quantiles of the normal distribution. Note, that histograms were normalized to sum up to one, creating estimation of the probability density function. The probability density function of the normal distribution with the mean and standard deviation equal to those of corresponding signals, was plotted with solid black line.

Values of kurtosis and skewness for each signal and each preprocessing method were visualized using boxplots in the Figure 9.2. Note logarithmic y-scale in both plots[1].

The QQ plots of any preprocessing method does not seem to resemble normal distribution. The raw signal seem to be closest, however, its values seem to deviate from the line at about negative and positive one. Also, a saturation on positive and negative five is very pronounced. This is also confirmed by the histogram. Moreover, one can notice the right skewness and heavy tails of the raw signal values compared to the normal distribution. See also negative skewness and positive kurtosis in Figure 9.2. This right skewness can also be seen in the exemplary signal from Figure 4.3, where negative parts of the peaks seem to be of smaller magnitude than the positive ones. The average value close to zero, however, does not seem to reflect this phenomenon, because of some counterbalancing effect from negative values. Notice two peaks at the extremes of the histogram. Values of negative and positive five correspond to the maximal value of the measurement setup and probably indicate analog to digital converter saturation. For the future studies, this might suggest registering the vibroarthrogram using greater voltage ranges or using smaller gain. What is interesting, however, is the fact that the negative extremes seem to be more pronounced, pulling the mean value close to zero.

The QQ plots of both filtering methods seem similar in shape, both having much more extreme values than the normal distribution. Histograms, however, indicate more extreme

[1]Logarithmic scale of the skewness plot was obtained using transformation proposed by Webber [186], using C constant of zero.

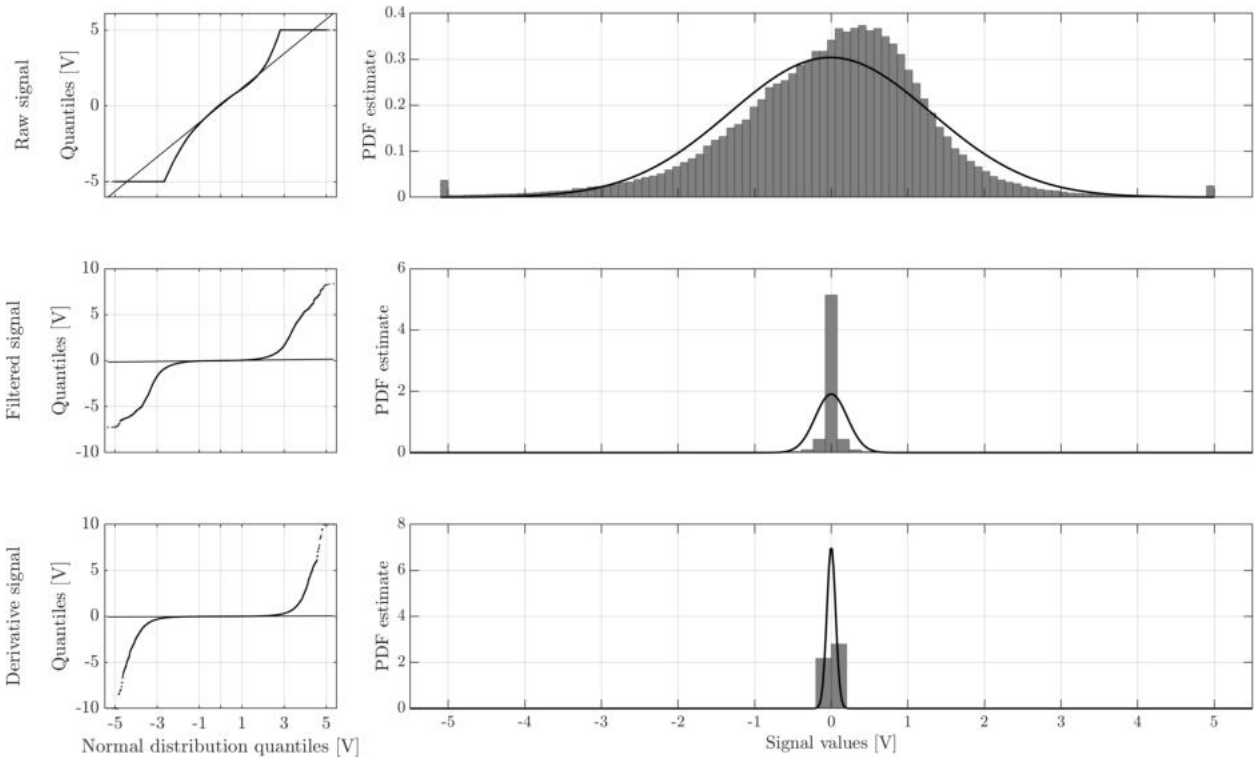

Figure 9.1: QQ plots (left column) and histograms (right column) for the raw (upper row), Butterworth filtered (middle row) and derivative (lower row) signals.

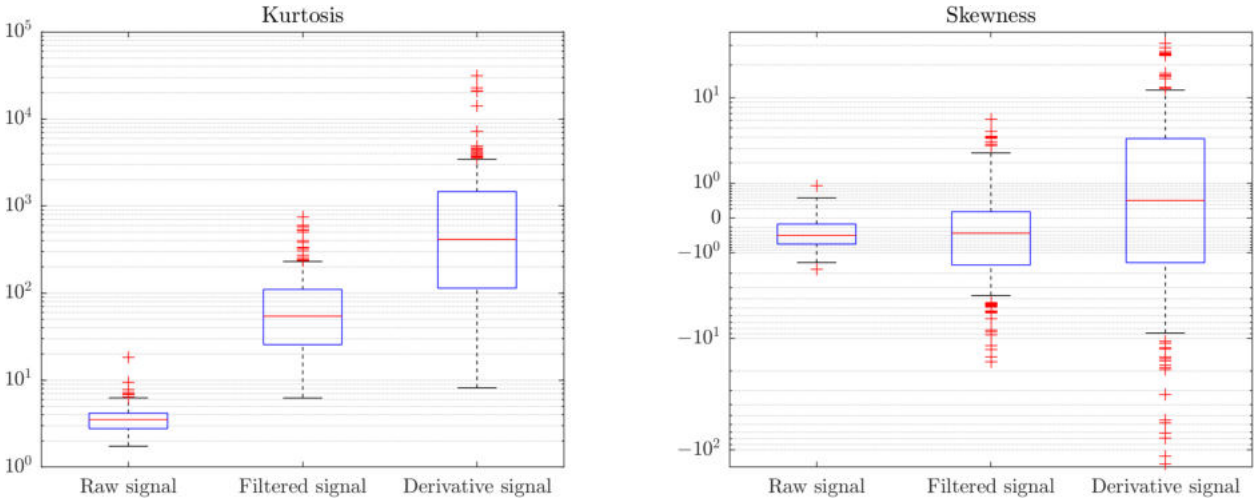

Figure 9.2: Boxplots of kurtosis (left) and skewness (right) of signal values after each preprocessing.

values in the derivative signals, which is also confirmed by the boxplots from Figure 9.2—notice median kurtosis in the derivative signals being almost ten times higher. Notice also much bigger spread of skewness values across derivative signals compared to both filtered and raw signals. That could indicate less "stable" preprocessing, leading to less consistent signals. Those boxplots were composed of signals from all the knee joint condition classes, though,

and high spread could be caused by differences of skewness between those conditions. This issue will be further investigated in the following chapter.

Stationarity of the signals was quantified using CVNMA and CVNMV values. Exemplary normalized signals of moving statistics for each preprocessing method were plotted in Figure 9.3. The moving average signals seem quite different across the compared methods,

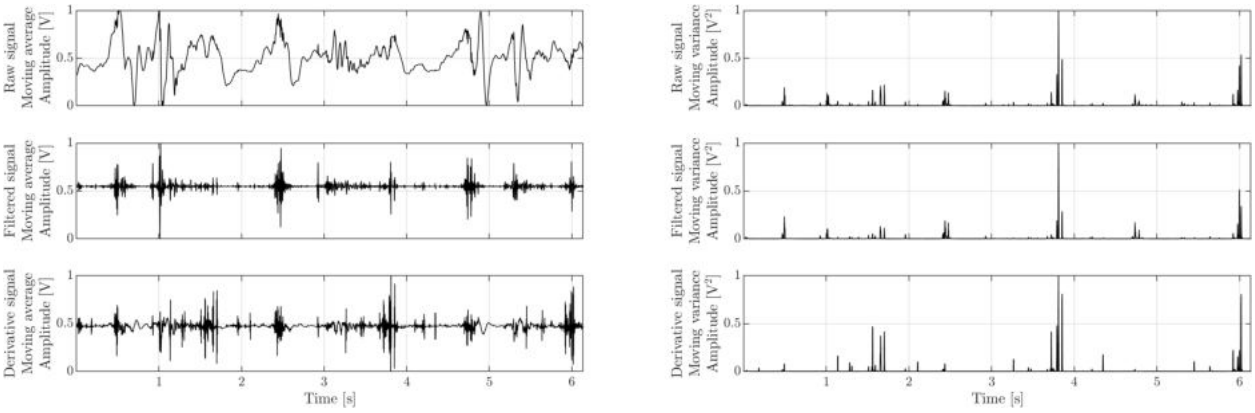

Figure 9.3: Exemplary normalized moving average value (left column) and variance (right column) for the raw (upper row), Butterworth filtered (middle row) and derivative (lower row) signals. This example uses the same signal as the example from Figure 4.3.

with highly lowered variability in both preprocessed signals, compared to the raw signal. Note, that the moving average is effectively a low-pass filter. Interestingly, the derivative signal seem to contradict the intuition and contain more low-frequency oscillations compared to the Butterworth filtered signal. This can be attributed to the normalization. According to the exemplary signals in Figure 4.3, majority of peaks in the derivative signal are really high-frequency, or short-duration. Moreover, they seem to occur rather singularly than in clusters, compared to the Butterworth filtered signal. Therefore, in the derivative signal, in which the peaks are more surrounded by the low-amplitude components, each peak has more relative influence on the samples in the window, making the low-frequency oscillations more visible. In the Butterworth filtered signal, on the other hand, the peaks are more clustered and therefore their influence on the window's samples is much less pronounced.

Variability of both moving averages seem similar, though driven by different factors. In the Butterworth filtered signal, there seem to be more clusters of high variance, separated by relatively quiet fragments, while in the derivative signal, the variance seem to by more or less pronounced, but everlasting.

The moving variance, on the other hand, seem to be very similar in shape across the signals. It seems to have the highest amplitude in the derivative signal, which could be explained by previously mentioned singular occurrences of peaks. Even though the plot in Figure 9.3 is just an example, it seem to be representative. Boxplots for both CVNMA and CVNMV coefficients, based on all the signals, were plotted in Figure 9.4. The CVNMA coefficient is significantly smaller after both processing, while CVNMV seem to be highest in the derivative signal and lowest in the raw signal.

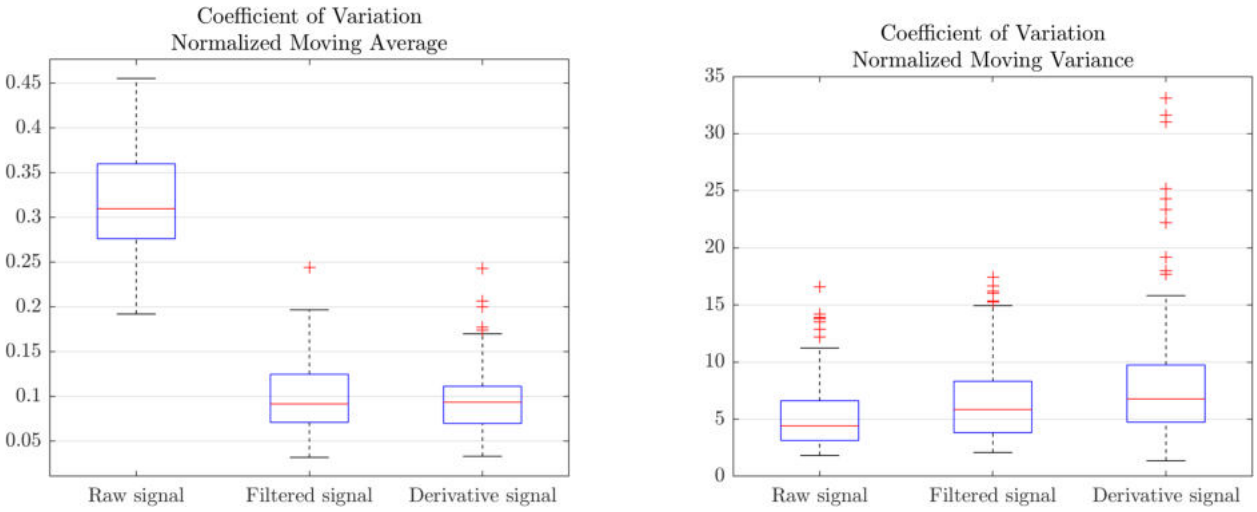

Figure 9.4: Boxplots of the CVNMA (left) and CVNMV (right) coefficients for each preprocessing method.

9.2 Summary

9.2.1 Preprocessing methods limitations

Even though differentiation is not frequency-selective, both Butterworth filtering and differentiation can be interpreted as being focused on the frequency components of the vibroarthrogram and treating all signal samples the same. Another approaches, like decompositions, could potentially also be of great value, as indicated in the literature.

Since this dissertation focuses on the feature extraction, the preprocessing step in the signal classification pipeline was chosen to be extremely intuitive and interpretable, even at the cost of potentially lower performance. As listed in the Chapter 3, there is a plethora o preprocessing methods, all having their own advantages and disadvantages. In the future studies, an additional, more comprehensive comparison between various methods could be conducted, perhaps indicating another preprocessing method as more suitable for the task.

9.2.2 Summary and conclusions

As mentioned earlier, vibroarthrograms are by nature non-stationary. However, in terms of the moving average variability, or the CVNMA coefficient, both Butterworth filtering and differencing seem to improve their stationarity. Relative variance, on the other hand, quantified by the CVNMV coefficient, seem to increase after the preprocessing. Influence of those characteristics on specific features will be discussed in corresponding chapters.

Various frequency ranges seem to be dominated by the signal after each preprocessing. The lowest frequency values are the most pronounced in the raw signal, while the middle part of the spectrum seem to be occupied by the Butterworth filtered signal, as per its design. The highest frequencies were amplified by the differencing operation, and therefore clearly stand out in its spectrum.

Chapter 10

Time domain features

In this chapter, results of time-domain features will be presented. Informativeness of all features will be provided and discussed. For features of relatively high informativeness, a deeper analysis will be conducted, including visualization in form of boxplots, as well as additional informativeness in terms of $coef_B$, i.e., in the class-pair-wise context. First, specific features in each type (i.e., statistical, rolling, etc.) will be analyzed. Then, their correlations will be discussed. The chapter will end with a summary and conclusions, which can be drawn from the conducted research.

10.1 Statistical features

Figure 10.1 includes results of the central tendency measures' informativeness. As expected, since VAG signals are, in general, symmetrical, mean and median values are not informative in any preprocessing approach. However, the Rectified Average of the filtered signal proved to be quite informative with $coef_{BMC}$ of around 0.58.

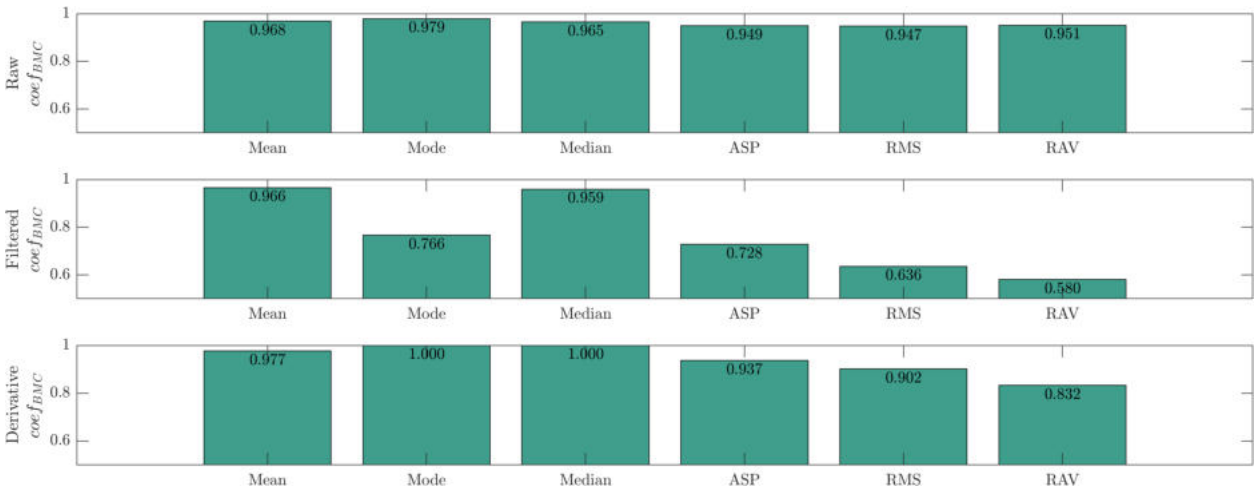

Figure 10.1: Central measures informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

Figure 10.2 includes boxplot of this feature with corresponding class-pair $coef_B$coefficients. It can clearly be seen, that the RAV increases with progressing stages of the chondromalacia. However, there is practically no difference between the cmp3 and the oa groups. In general, neighboring classes are not very well differentiated using the RAV, with the smallest $coef_B$of 0.69 being determined for cmp1 and cmp2 neighboring pair. Less similar conditions of knee

are differentiated more easily with the smallest $coef_B$ of 0.24 for the ctrl and cmp3 pair, and 0.33 for ctrl and oa.

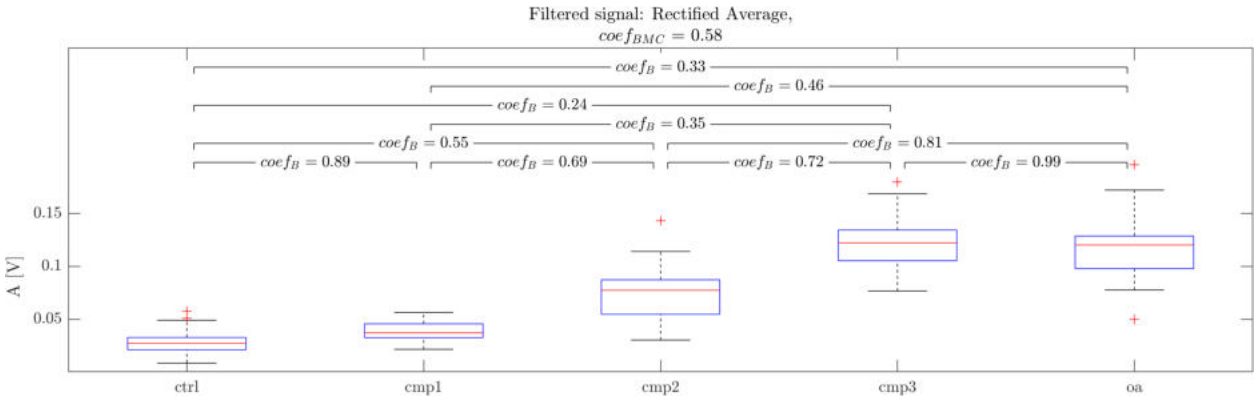

Figure 10.2: Boxplot of the Rectified Average feature calculated for the filtered signal.

Results of the spread measures' informativeness were included in the Figure 10.3. For the features that were extracted with multiple parameters (like the IPerR or the percentiles), maximal and minimal informativeness scores were showed, indicated by the darker and lighter color, respectively.

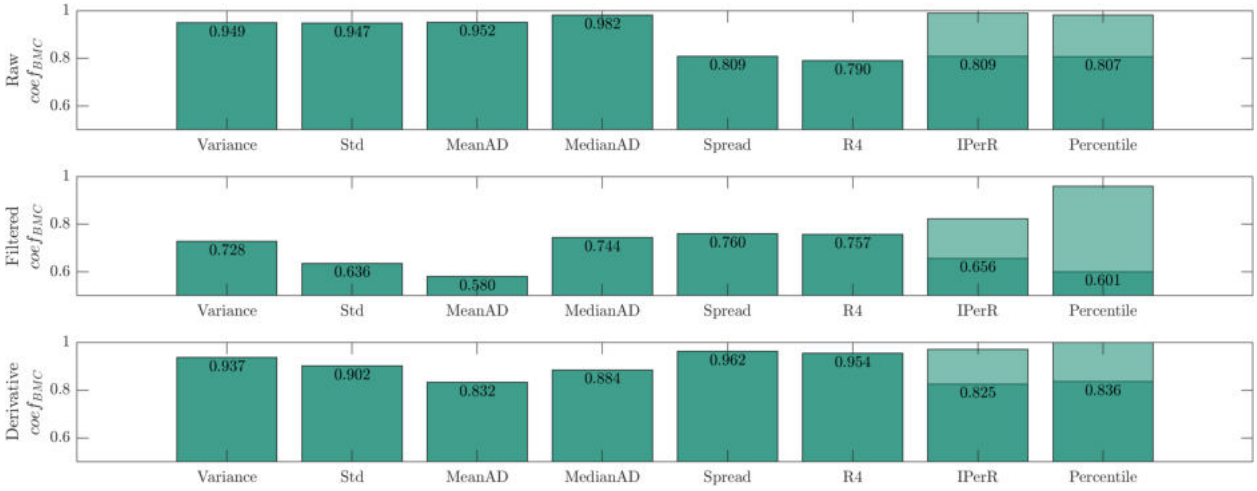

Figure 10.3: Spread measures informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

The analysis of the raw signal indicated low informativeness ($coef_{BMC} > 0.9$) for the features of variance, standard deviation, MeanAD and MedianAD. This observation implies limited insight into the inherent structure or characteristics of the raw signal via these measures. Notably, the spread and the R4 showed slightly improved informativeness, reflected $coef_{BMC}$ around 0.8.

The filtered signal shows a clear improvement in the informativeness of all spread features. Decrease of $coef_{BMC}$ to approximately 0.73, 0.64, and 0.58 in the informativeness of variance, standard deviation, and MeanAD shows that the frequency range of the filter, i.e., 50 to

1000 Hz, effectively isolates key information of the signal in the studied classification context. The MedianAD, spread, and R4 features also decreased in $coef_B$ informativeness, though not as drastically, all achieving values around 0.75. Boxplot of the MeanAD feature was

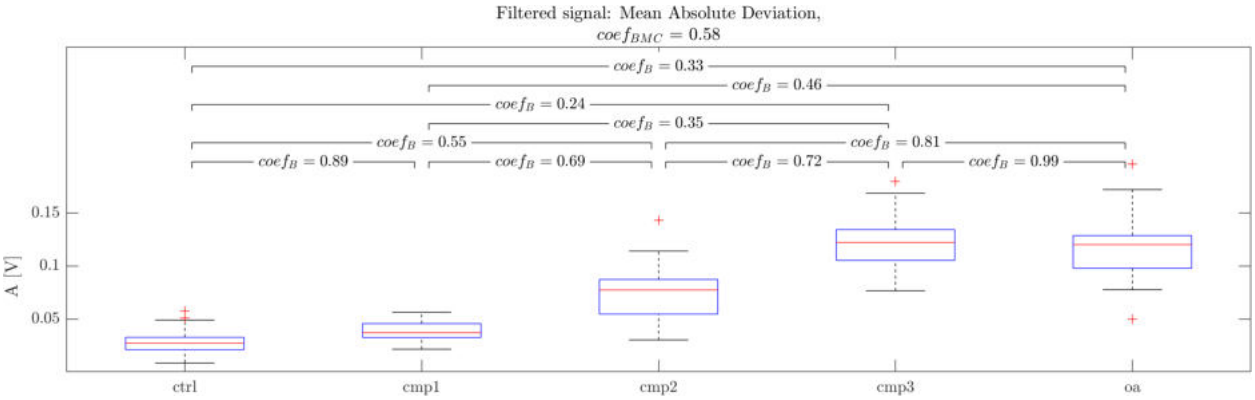

Figure 10.4: Boxplot of the Mean Absolute Deviation feature calculated for the filtered signal.

included in Figure 10.4. Note, that it is virtually the same as the RAV feature, plotted in Figure 10.2. It is the result of the VAG signals' symmetry. Since the mean value is close to zero, the RAV feature and the MeanAD feature yield practically the same value for each signal.

The derivative signal presented a more diverse results. While the informativeness scores were generally superior to the raw signal, they were not as high as the filtered signal's scores. Variance and standard deviation showed rather low informativeness, achieving $coef_{BMC}$above 0.9. Both MeanAD and MedianAD proved do be slightly more informative, with the $coef_{BMC}$ between 0.9 and 0.8. Interestingly, a loss of informativeness can be observed for spread and R4 features compared to the raw signal, suggesting that these measures might be more sensitive to absolute values in the signal rather than changes between successive points, which differentiation emphasizes.

The percentile values, both singular and IPerR, show similar picture, with the values for the filtered signal being the most informative, followed by the raw and the derivative one. Detailed informativeness of percentiles was included in Figure 10.5. Plots in the upper part of the Figure visualize the IPerR features informativeness, while lower part corresponds to the singular percentile values.

The lowest $coef_{BMC}$, i.e., the highest informativeness of the percentile values in the raw signal is indeed greater than in the derivative signal. However, both the IPerR and the singular percentile features seem to be more informative *in general* compared to the raw signal. Moreover, the informativeness of singular percentiles of the derivative signal seem to be lowest in the middle, i.e., around the median value. Lower and higher percentiles seem more informative. This trend is much more visible in the filtered signal, with values of the $coef_{BMC}$ around 0.6 for the 2-nd and the 98-th percentile. The boxplots of both features were included in Figures 10.6 and 10.7.

It can be seen that with progressively worse condition of the joint, the range of filtered signal's values grows, i.e., the 2-nd percentile decreases, while the 98-th percentile increases.

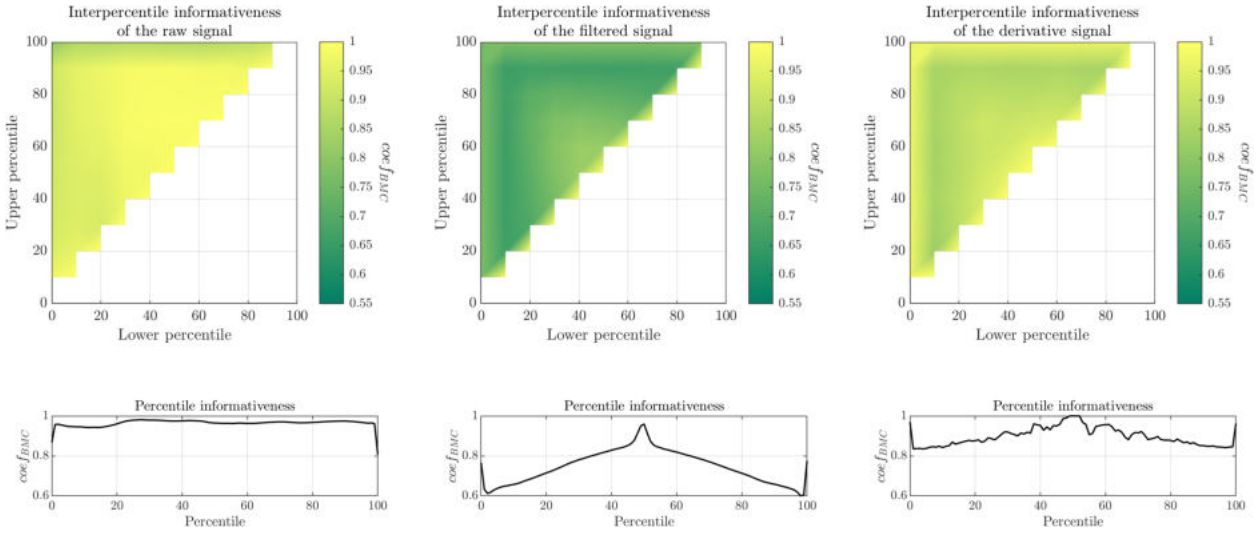

Figure 10.5: Detailed informativeness results of the Interpercentile Range (upper part) and the singular percentiles (lower part), for the raw signal (left column), filtered signal (middle column) and the derivative signal (right column).

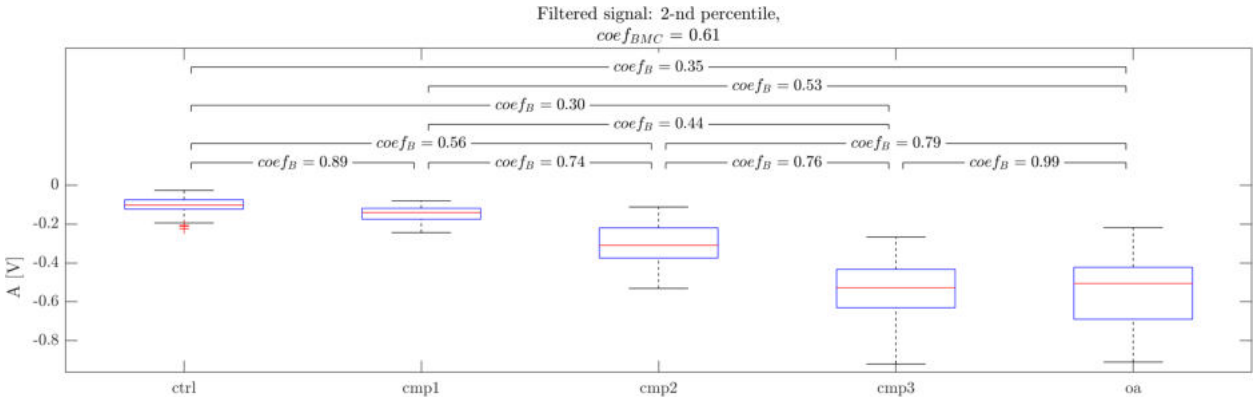

Figure 10.6: Boxplot of the 2-nd percentile feature calculated for the filtered signal.

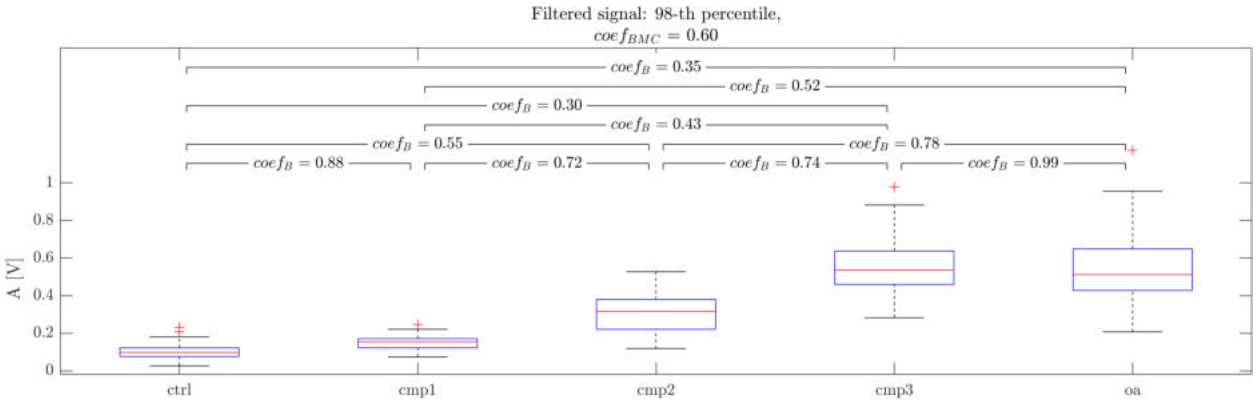

Figure 10.7: Boxplot of the 98-th percentile feature calculated for the filtered signal.

That suggests that maybe redefining R4 value to incorporate different number of extreme samples would yield better results. On the IPerR plot, however, it can be seen that feature defined as the difference between high and low percentiles is not necessarily more informative than the feature defined as the singular percentile. The most informative percentage range, i.e., the 10-th to 40-th percentile, achieved $coef_{BMC}$ of 0.66. Boxplot of this feature was showed in Figure 10.8.

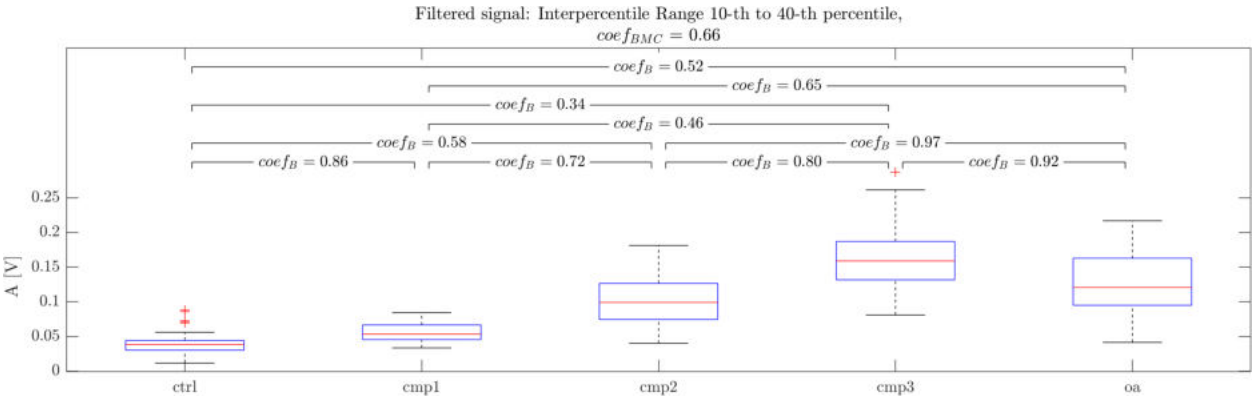

Figure 10.8: Boxplot of the 10-th to 40-th Interpercentile Range feature calculated for the filtered signal.

The boxplot for the most informative IPerR feature seems to confirm, that range of filtered signal's values increases with the for progressing knee joint condition. Interestingly, for the discussed feature, the oa group seem more similar to the cmp2 group than to the cmp3 ($coef_B$ of 0.97 vs 0.92 with the $coef_B$ between the cmp2 and cmp3 being 0.8), which was the case in all previously discussed features.

Simple shape features' informativeness was included in Figure 10.9. Both skewness and kurtosis proved to be uninformative, despite of the preprocessing approach. It seems surprising given their extreme popularity in VAG-related studies.

Spiky Index, and its Rectified version also turn out to be of rather poor informativeness, with the most informative version being the derivative signal. Since the difference operation is virtually the high-pass filter, it amplifies the "spikyness" of the signal, which, apparently, contains *some* information about the classified knee joint condition. Specific informativeness values of of both SI and RSI, as functions of the threshold value, were plotted in Figure 10.10.

The SI and RSI yielded generally quite high values of $coef_{BMC}$ for the raw signal, with the lowest value, i.e., the most informative, for the threshold of 0.99. It means that the raw signal provides limited differentiation capacity, unless it is counting the instances where the signal is close to its maximum value. This might be because the raw signal carries a lot of noise or irrelevant information that obscures the informative patterns. On the other hand, the filtered signal, provides somewhat higher informativeness, at a low threshold ($\alpha = 0.01$). The increase in informativeness could be due to the filtering process reducing the noise in the signal, thereby enhancing the differentiation power among the condition groups. Interestingly, with a threshold of 0, SI and RSI become completely uninformative. This behavior seem completely opposite to the derivative version of the signal, which is the most informative precisely at the $\alpha = 0$, i.e., with no threshold. However, raising

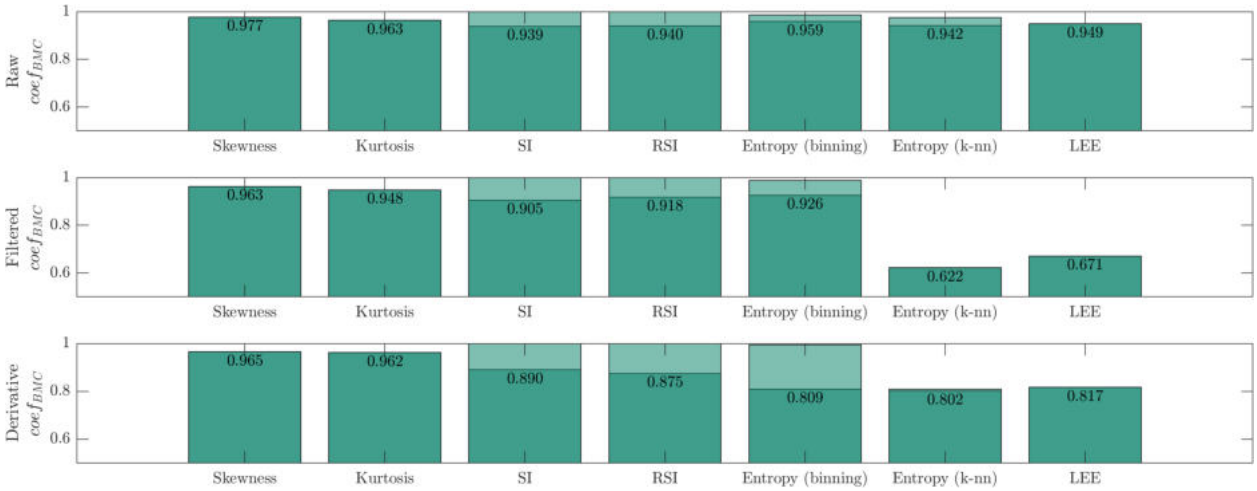

Figure 10.9: Simple statistical shape measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

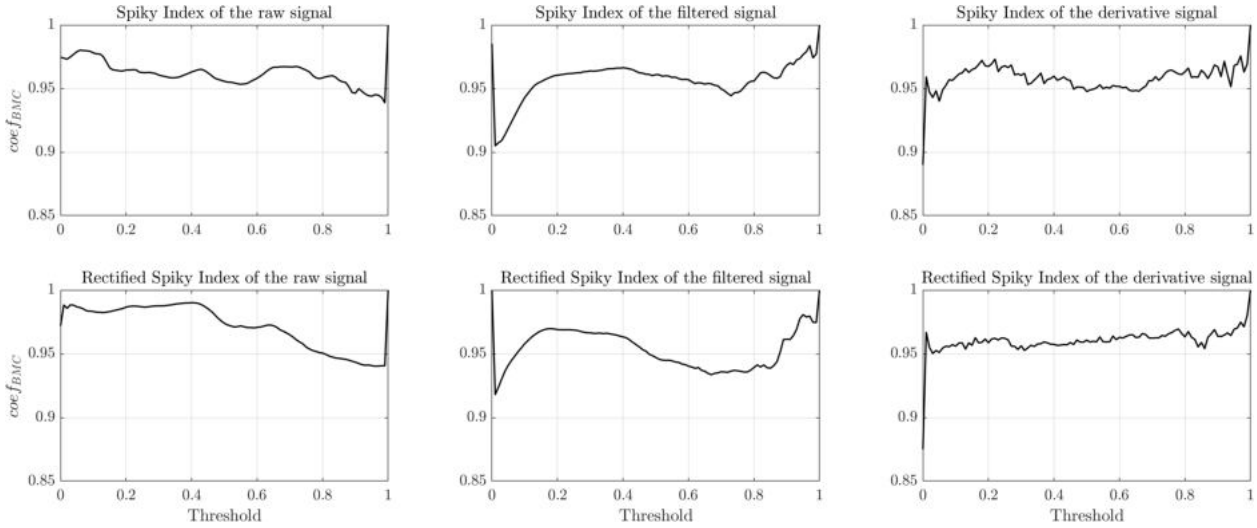

Figure 10.10: Detailed informativeness results of the Spiky Index (upper part) and the Rectified Spiky Index (lower part), for the raw signal (left column), filtered signal (middle column) and the derivative signal (right column).

the threshold even slightly makes the features significantly less informative. This indicates that the differencing process, in which change in signal value from one sample to the next is calculated, provides the most distinguishable information among the condition groups. Moreover, calculating the proportion of the values greater than zero in the derivative signal is equivalent to counting proportion of the consecutive values in the raw signal, that are not equal. Therefore, the most informative features, i.e., the SI and RSI with no threshold, calculated on the derivative signal, measure how much the signal changes over time, rather than the absolute values of the signal's spikes.

Within each preprocessing approach, the SI and the RSI appear to behave similarly as functions of the threshold level. This indicates, that the rectification, despite altering the

signal by converting all negative values into positive ones, does not drastically change the informativeness of the features extracted from the signal. This observation further confirms symmetry of the vibroarthrogram around the zero value.

Another simple shape measure, i.e., the entropy, seem to be more informative. The LEE feature seem quite informative for the filtered signal, while derivative signal and raw signal show progressively higher $coef_{BMC}$ values. The binning entropy estimator achieved highest values for the derivative signal, while the nearest neighbors one proved to be the most informative for the filtered signal. More in-depth visualization of the entropy informativeness, as a function of both the number of bins and neighbors, was provided in the Figure 10.11.

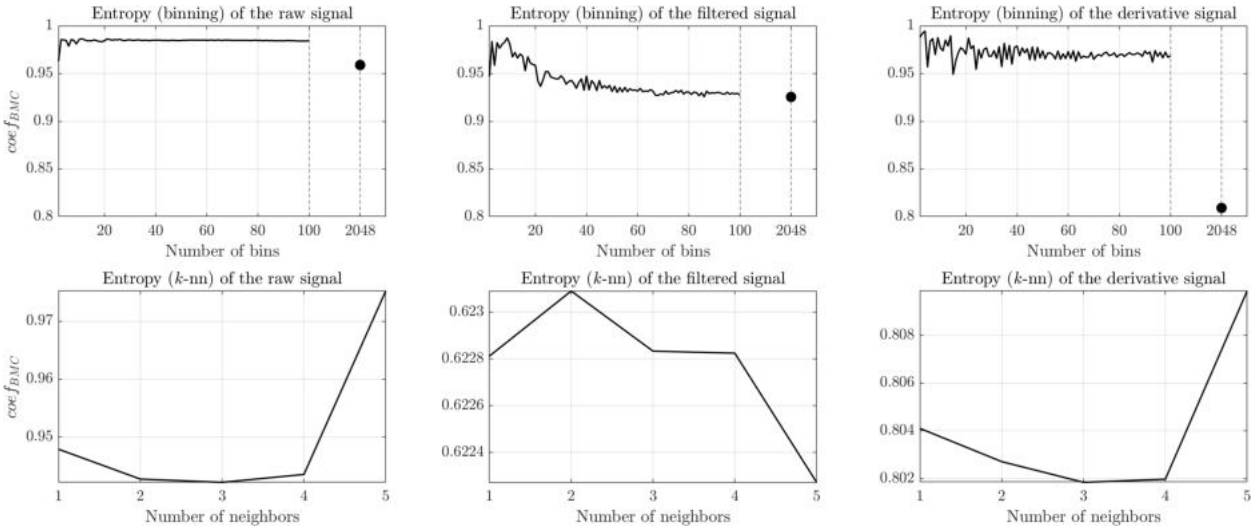

Figure 10.11: Detailed informativeness results of the entropy estimation using the binning (upper part) and the nearest neighbors (lower part) approach, for the raw signal (left column), filtered signal (middle column) and the derivative signal (right column).

The binning entropy estimation for raw signal proved to be of very low informativeness, with $coef_{BMC}$ consistently achieving around 0.98 for up to a hundred bins. Slightly higher score, of around 0.96, was achieved for 2048 bins. Informativeness of the same feature improved after filtering. Nevertheless, $coef_{BMC}$ proved to be quite high, of over 0.9, for all bin numbers. The derivative signal, up to a hundred bins, also provided rather uninformative binning entropy estimation, yielding results somewhere between the raw and the filtered signal. However, for 2048 bins, significant improvement was observed, with $coef_{BMC}$ decreasing to around 0.81. For all preprocessing approaches, despite being rather uninformative in general, the informativeness increased with increased number of bins. Notably, however, all curves seem to plateau well before a hundred bins, but for the raw signal and the derivative signal, informativeness of the 2048 bins seem to break this plateau. This can be explained by the fact that when the number of bins is relatively small, each bin is broad enough to encompass a wide range of signal values, potentially oversimplifying the signal and reducing informativeness. As the number of bins increases, the informativeness also increases as the binning approach starts to capture more intricate patterns within the data. However, beyond a certain number of bins, each bin becomes so narrow that additional granularity does not contribute to new meaningful information, resulting in a plateau. For the raw and

derivative signal, the significant increase in informativeness at 2048 bins suggests that these signals may have contained high-resolution features or patterns that only became discernible when the binning resolution was significantly increased, breaking the previous plateau.

The nearest neighbors entropy estimator yielded much better results for all preprocessing approaches except for the raw signal. The filtered signal provided much more informative entropy estimation, with $coef_{BMC}$ as low as around 0.62. Derivative signal, again, allowed for a little more informative feature extraction compared to the raw signal, but worse than the filtered one. Interestingly, informativeness of the entropy estimation did not change significantly with the number of neighbors for neither preprocessing approach. More informative results the nearest neighbors approach indicate that the characteristics of the VAG signals are better captured using a method that considers proximity of points, rather than simply frequency distribution as in binning. Those results confirm that vibroarthrograms contain complex or nonlinear components [10].

Note, however, that the binning and the nearest neighbors entropy estimators differ in the types of entropy they estimate, which drives their varying degrees of informativeness. Binning, a histogram-based approach, estimates Shannon entropy based on the frequency distribution within predefined bins. However, this discretization can lead to an oversimplification of complex signals, potentially resulting in lost information if the number of bins does not adequately capture the signal's structure. On the other hand, the nearest neighbors method estimates differential entropy, suited for continuous distributions, by considering distances to neighboring points rather than discretizing the signal. This approach inherently maintains the signal's continuity, potentially capturing more nuanced and intricate patterns and proving more informative for complex signal data. The boxplot of the most informative entropy estimation was included in Figure 10.12.

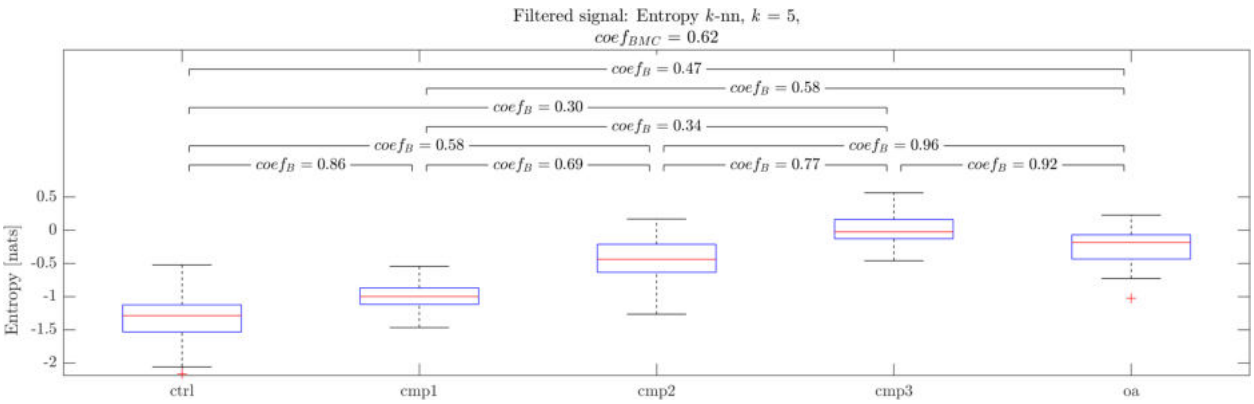

Figure 10.12: Boxplot of the k-nn Entropy feature calculated for the filtered signal.

For the healthy joint, i.e., the control group, entropy values are lowest, scoring from around -2 to around -0.5. The progressing chondromalacia stages are characterized by increased entropy values, with the third stage values being in approximate range of -0.5 to 0.5. Finally, the osteoarthritis group includes entropy values of around -0.7 to around 0.2. The nearest neighbors entropy estimation provides a measure of "spread" of a distribution. When differential entropy is negative, it usually means that the probability density function of the variable is more concentrated, i.e., the random variable is more likely to be around

certain values than others, making the signal more predictable. That is, negative entropy value indicates that the signal samples are relatively predictable or more regular. This can seen in the healthy joint group with entropy values between -2 to -0.5. As entropy values increase, predictability of the signal samples decreases, indicating more random or complex vibroarthrogram. This is observed in chondromalacia progression and can be explained by increased structural and functional irregularities in the knee joint. Interestingly, in this entropy feature, second stage chondromalacia is more similar to the osteoarthritis group than the third stage chondromalacia. Similar pattern was observed in the IPerR feature.

Informativeness of the complex shape features was provided, in form of the barplot, in Figure 10.13.

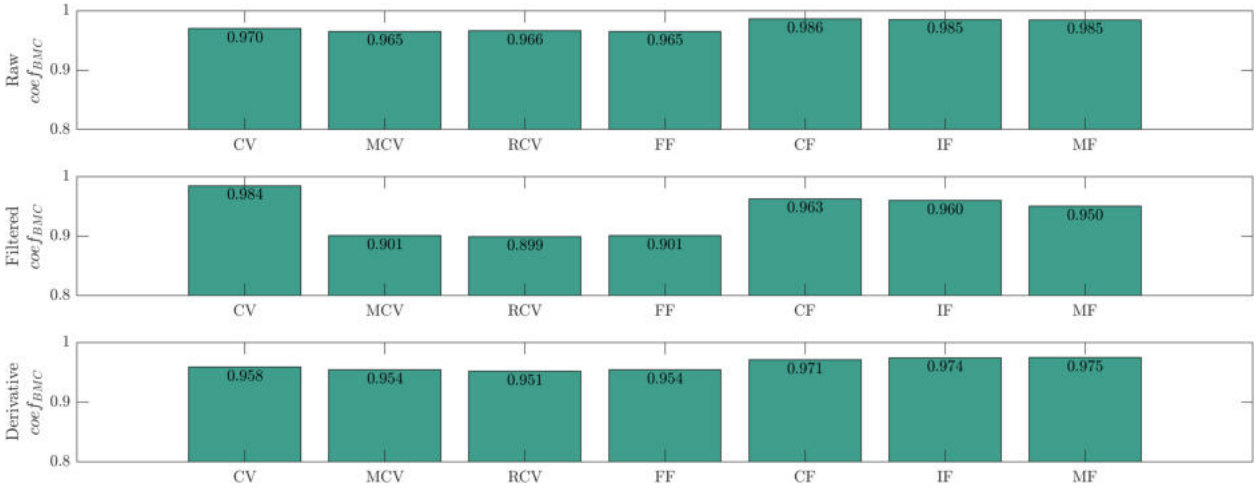

Figure 10.13: Complex statistical shape measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

For the raw signal, high $coef_{BMC}$ scores indicating low informativeness were observed across all complex shape features. The CV, MCV, and RCV indicate similar variation across the knee joint classes. Low informativeness of the CV features is to be expected to some extent across all preprocessing approaches, because of the earlier mentioned symmetry of the VAG signals. Since their average value is close to zero, the CV can become very large or undefined, rendering its information content potentially misleading or less relevant for assessing the true variability within the signal. The FF, defined as the ratio of the RMS to the RAV, also proved to be of low informativeness, suggesting similarity of the ratio between classes. The rest of the factors, i.e., the CF, IF, and the MF also achieved high $coef_{BMC}$ scores, indicating similarity of classes in the context of extreme value ratios.

After filtering the signal, the informativeness of some features improved, as in the previous results. Notably, $coef_{BMC}$ decreased the most for the MCV, the RCV and the FF. Both the MCV and RCV achieved better informativeness scores. This can be explained by their definitions: the Modified Coefficient of Variation is the standard deviation divided by the mean of the rectified signal, and Rectified Coefficient of Variation is the standard deviation of the rectified signal divided by the mean of the rectified signal. Filtering the vibroarthrogram likely emphasized specific frequency components that enhanced variations crucial for class distinction. The similarity in their informativeness stems from their shared

sensitivity to these variations, with both metrics responding to the absolute values of the signal's magnitude. Note, that the Form Factor feature achieved the same informativness as the MCV feature for all preprocessing approaches. This can be attributed to the symmetry of the vibroarthrogram. Remember, that the standard deviation is defined as the square root of the mean squared deviations from the mean, while the RMS is defined as the square root of the average of the squared values of the signal. Since the average value is close to zero, the standard deviation is *almost* equal to the RMS value, which results in almost identical definitions of the MCV and the FF.

Also similarly to previous results, features extracted from the derivative signal achieved informativeness scores better than the raw signal, but worse than the filtered signal. This indicates that the difference operation has revealed *some* important characteristics of the signal, but not as effectively as filtering. Nevertheless, complex shape features proved to be of rather low information content, regardless of the employed preprocessing approach. In other words, described complex shape features proved to be unable to emphasize or capture the distinctions between the knee joint classes within the VAG signals under analysis.

10.2 Rolling features

Results of the rolling features, i.e., the VMS, the CVNMA and the CVNMV, were provided in the Figure 10.14.

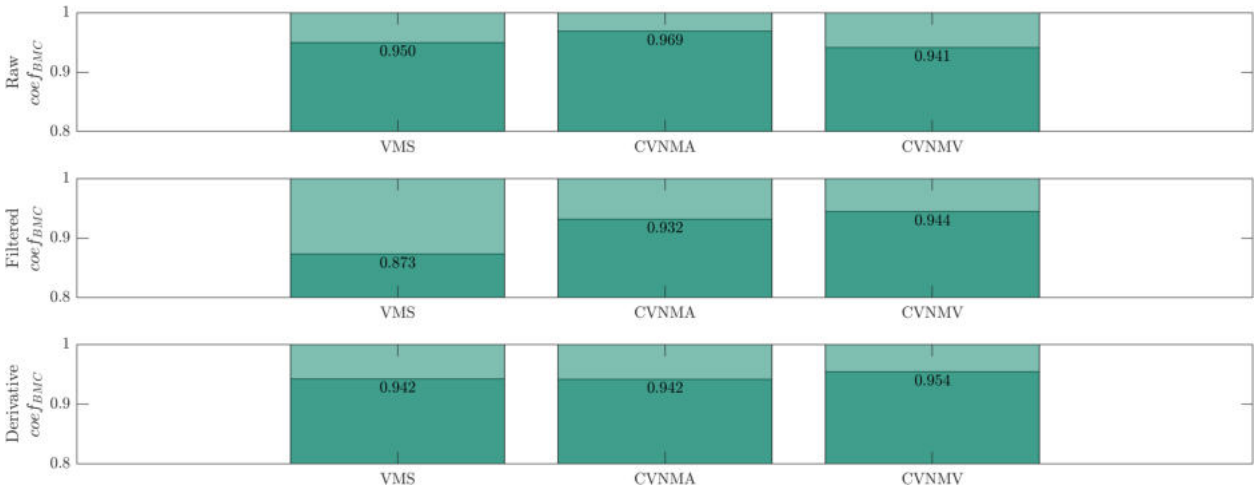

Figure 10.14: Rolling signal measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

Detailed results of the rolling features' informativeness as a function of window size were provided in Figure 10.15. The VMS feature proved to be the most informative for the filtered signal. For all preprocessing approaches, smaller window sizes seem to provide more informative VMS measures, however, informativeness does not seem to strongly depend on the window size. In general, the VMS informativeness seem to be low, with the lowest $coef_{BMC}$ score equal to 0.87, achieved for the filtered signal and window size of 0.2 ms. For 5-ms window, commonly used in literature, the VMS feature yielded $coef_{BMC}$ score of 0.89.

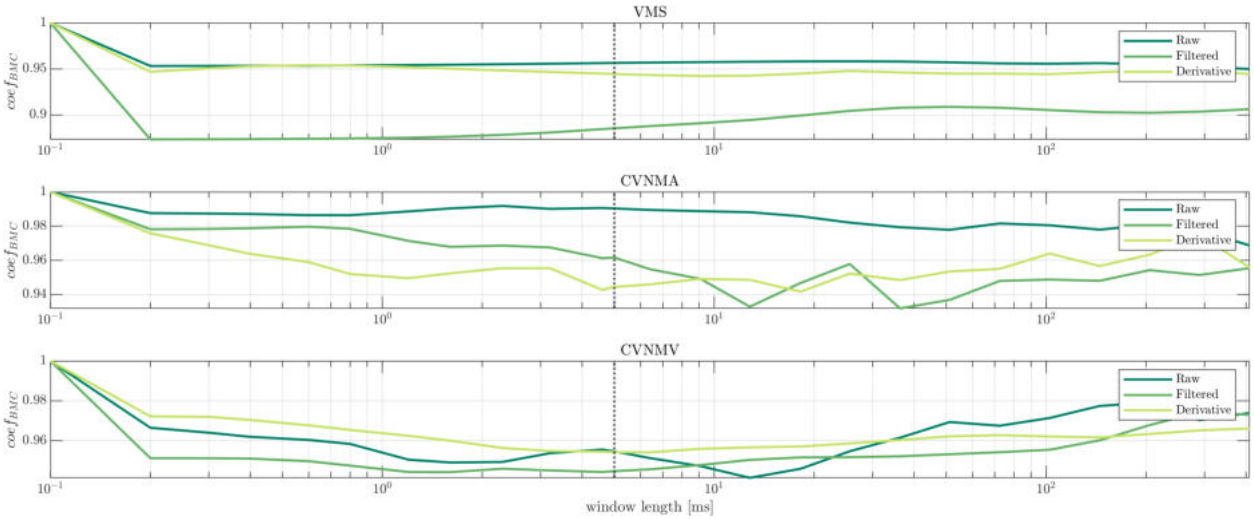

Figure 10.15: Detailed informativeness results of the rolling features: the Variance of the Mean Squared (upper plot), the Coefficient of Variation of the Normalized Moving Average (middle plot), and the Coefficient of Variation of the Normalized Moving Variance (lower plot). In all plots, window size of 5 ms was indicated by the dotted line.

The stationarity rolling features, i.e., the CVNMA and the CVNMV seem to achieve better informativeness scores for windows sizes of around 1 ms. However, their informativeness is poor in general, indicating that stationarity of the VAG signals, at least measured by those features, is not different across knee joint conditions.

10.3 Difference and frequency features

Figure 10.16 includes results of the information content in difference and frequency features calculated from time domain. Features extracted from the raw signal yielded moderate $coef_{BMC}$ scores, with the most informative being Hjorth's complexiy, achieving 0.798. It is not surprising, given its popularity across VAG-related literature. Mobility proved to be not very informative, while the Mean of Absolute Differences scored 0.832. From the Zero Crossing Rate and the Turns Count Rate with both thresholds, only TCR_β, i.e., with threshold dependent on the spread of values, proved to be decently informative, scoring $coef_{BMC}$ of 0.862. On the other hand, TCR_α and ZCR_α proved to be only two features in this category, which informativeness increase after differentiation. Detailed results of the threshold-dependent features will be provided later.

Filtering of the signal improved informativeness of two features: the TCR_β and the Mean of Absolute Differences. With the $coef_{BMC}$ score of around 0.6, the MoAD proved to be one of the most informative features extracted from time domain. Boxplot of the this feature, with corresponding $coef_B$ scores for each class pair, was included in Figure 10.17. The variability of the signal, measured by the MoAD feature, seem to increase along with level of the knee joint degradation. Similar to other features, neighboring classes seem to be difficult to differentiate, with the $coef_B$ varying from 0.7 to 0.97 for the cmp3 and oa pair. Differentiation between cmp2 and oa, and cmp2 and cmp3, yields $coef_B$ values of 0.85 and

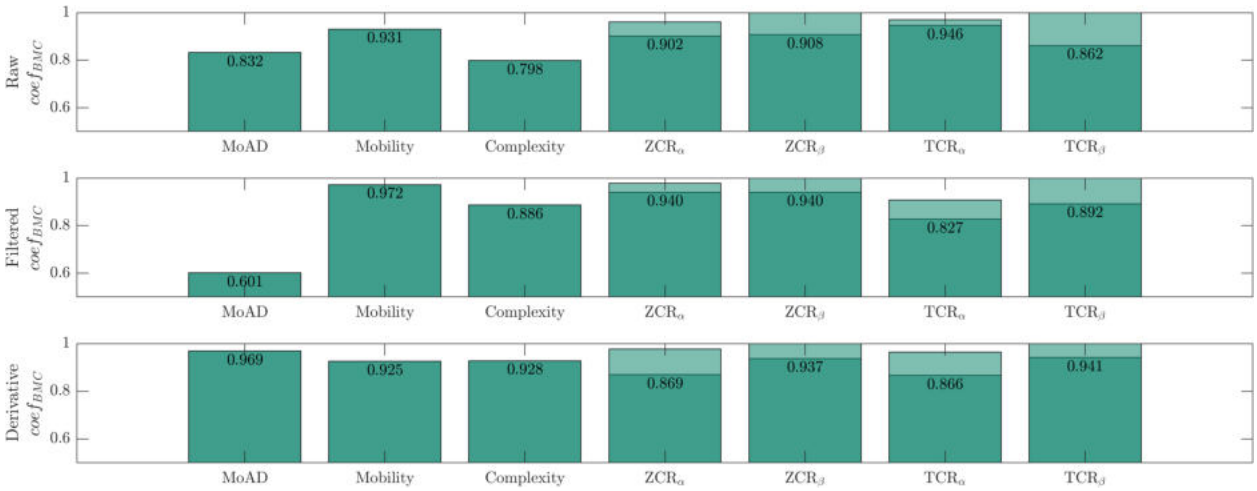

Figure 10.16: Difference and time-domain frequency measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

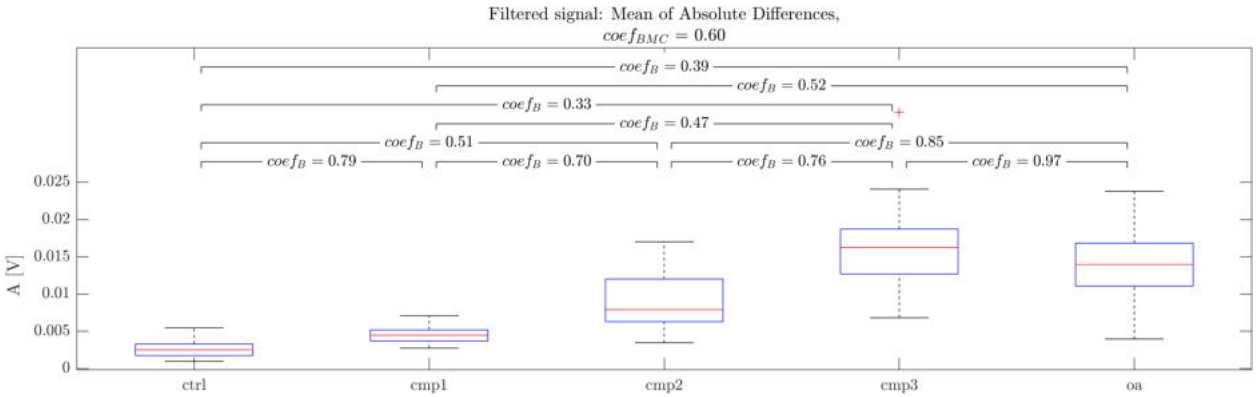

Figure 10.17: Boxplot of the Mean of Absolute Differences feature calculated for the filtered signal.

0.76, respectively. This suggests that the Mean of Absolute Differences complements the previously discussed entropy feature, which was lacking in differentiation ability between cmp2 and oa classes.

Figures 10.18 and 10.19 include informativeness scores of the ZCR and the TCR features, respectively, as functions of a threshold. The Zero Crossing Rate with α threshold, i.e., dependent on the standard deviation of signal, seems to be uninformative for the raw signal and the filtered signal. Slight informativeness fluctuations can be observed, but regardless of threshold, $coef_{BMC}$ is above 0.9, indicating low information content. The derivative signal seems to achieve somewhat better $coef_{BMC}$ scores for threshold values of around 5 to 15 percent of the standard deviation. For lower threshold values, zero crossings are counted, that do not differ between knee joint conditions signals. On the other hand, higher thresholds render ZCR_α feature *less* sensitive to zero crossings that can differentiate classes. The ZCR with threshold β, that is, dependent on the range of values, seem to be uninformative in general. In other words, it cannot differentiate between analyzed VAG signals. However, an

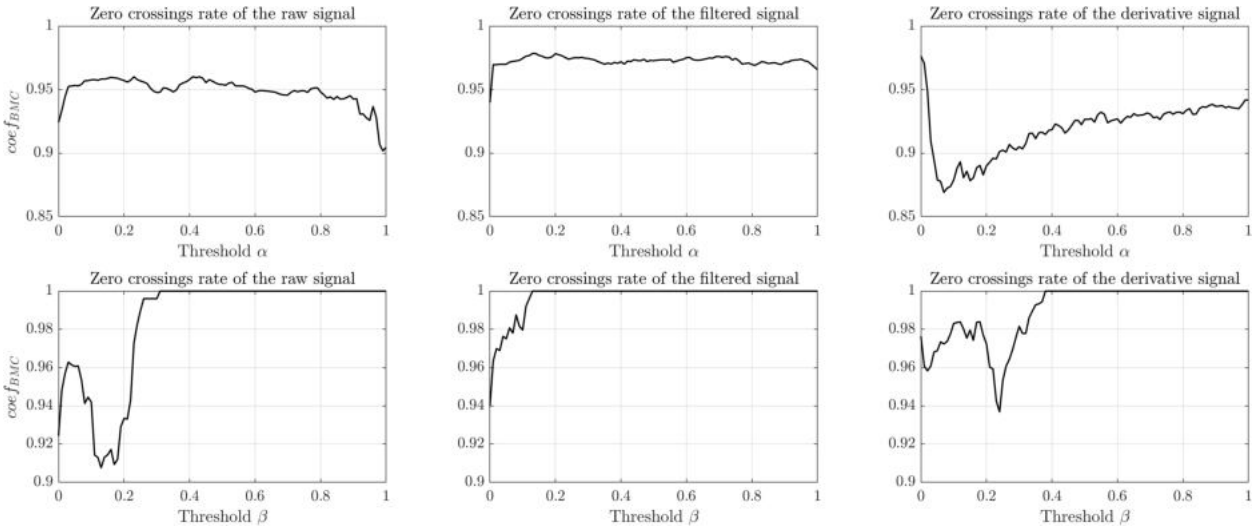

Figure 10.18: Informativeness of the Zero Crossing Rate as a function of the threshold α (upper row) and β (lower row) for the raw signal (left column), filtered signal (middle column) and the derivative signal (right column).

interesting insight can be drawn from inspection of the Figure's 10.18 lower row: it seems that VAG signals of *all* classes seem to have similar Zero Crossing Rate when counting only zero crossings greater than around 30, 10 or 40 percent of the spread, when calculated for the raw, filtered or derivative signal, respectively. It reflects that these thresholds likely capture fluctuations or noise common across all classes, rather than features that distinguish them, thus leading to the observed values of $coef_{BMC}$ equal to 1.

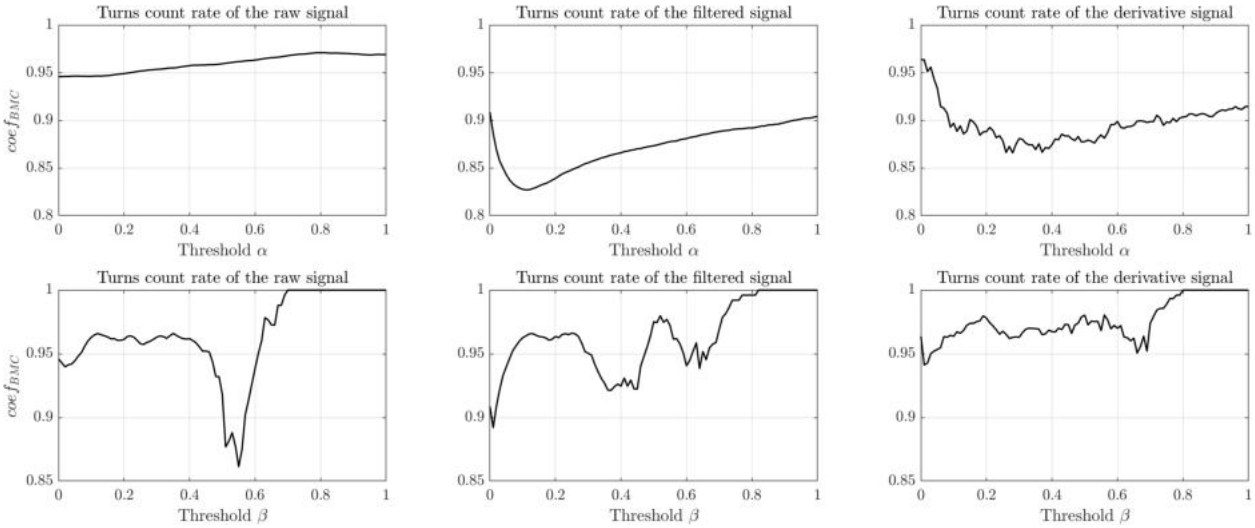

Figure 10.19: Informativeness of the Turns Count Rate as a function of the threshold α (upper row) and β (lower row) for the raw signal (left column), filtered signal (middle column) and the derivative signal (right column).

The Turns Count Rate with threshold α seem to be uninformative for the raw signal,

regardless of the threshold. The filtered signal, however, seem to follow a pattern similar to the ZCR_α of the derivative signal, yet being a little more informative (lowest $coef_{BMC}$ of around 0.83). See also, that the TCR_α seem much "smoother" than the ZCR_α of the derivative signal. It can be explained by the high-frequency fluctuations (above 1 kHz), surpassed by the filter. The TCR_α of the derivative signal, however, also seems decently informative compared to the ZCR_α, with seemingly smaller dependence on the threshold.

The TCR_β of the raw signal seem decently informative for threshold equal to around 50 percent of the spread. It means that direction changes greater than around 50 percent of the spread of values differ between classes. Nevertheless, with $coef_{BMC}$ of around 0.86, those between-the-classes differences are not the most pronounced. For the filtered signal, the TCR with spread-based threshold achieved $coef_{BMC}$ above 0.9 for most thresholds. The derivative signal provided the least informative TCR_β, with hardly any $coef_{BMC}$ scores below 0.95.

10.4 Self-similarity features

Informativeness results of the Autocorrelation Function features were showed in Figure 10.20. All three central tendency measures, i.e., the Average Autocorrelation Function, the Rectified Average Autocorrelation Function, and the Mean Squared Autocorrelation Function proved to be similarly uninformative for all preprocessing approaches, achieving $coef_{BMC}$ value above 0.9. Values slightly lower than 0.9 were scored by the First Zero Crossing Time and the Autocorrelation Function Zero Crossing Rate features of the filtered signal, but for both the raw and the derivative signal, $coef_{BMC}$ achieved values well over 0.9. Finally, the Autocorrelation Function Turns Count Rate feature proved to be somewhat informative, with the $coef_{BMC}$ of around 0.81, calculated for the derivative signal. It means that there are some minor but consistent differences between knee joint conditions, in the number of times that the signal's Autocorrelation Function changes its direction. Generally low informativeness of

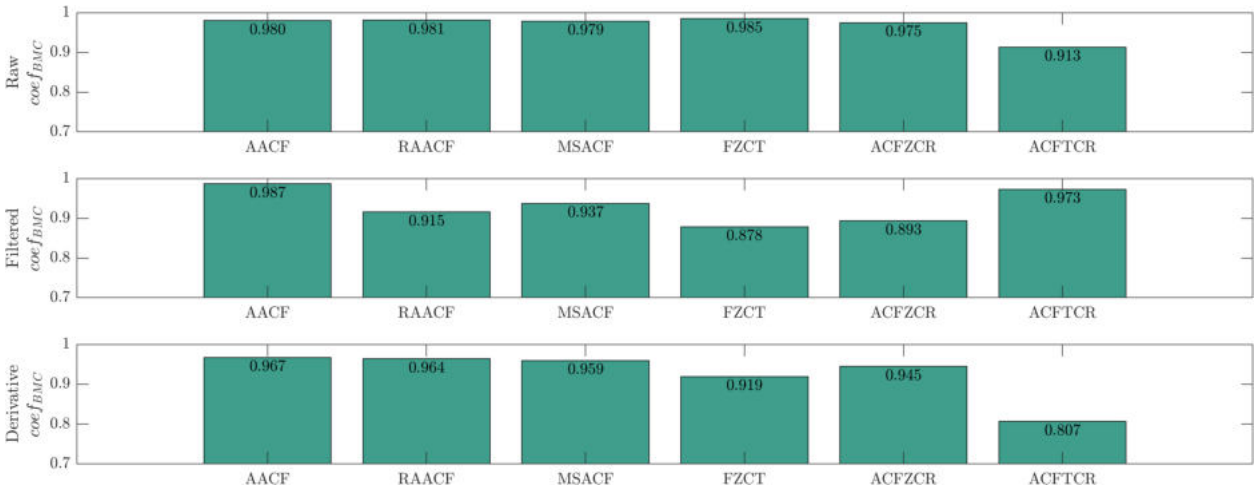

Figure 10.20: Autocorrelation Function measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

the features based on the Autocorrelation Function can be attributed to a couple of reasons. First, the features were calculated for the full ACF, i.e., for the ACF of length one sample smaller than the signal itself. This can result in a considerable amount of noise, because for increasing number of samples in the ACF, sub-signals length, for which the correlation is calculated, decreases. Another reason could be that the sensitivity of the ACF to the specific characteristics of the signal might be inherently limited. This limitation might stem from the ACF properties, which could constrain its ability to capture nuanced variations in the signal's structure. That is, the ACF can be of limited use for VAG signals in general, at least with preprocessing approaches followed in this work.

In the Figure 10.21, summarized informativeness results of the features based on the Multifractal Detrended Fluctuation Analysis were provided. Most features were calculated for specific value of the q order. Detailed results of those features, with the informativeness as a function of the q order were provided in Figure 10.22. One feature calculated for *whole* multifractal spectrum is the width of the spectrum. It indicates the multifractality of the signal: a wider spectrum signifies higher multifractality, while a narrow one suggests the contrary. In other words, a wide spectrum indicates that the signal displays diverse patterns or behaviors at different scales, while a narrow spectrum means that the signal is more uniform across the scales. Low informativeness of the spectrum's width suggests, that VAG signals of all condition classes exhibit similar degree of multifractality.

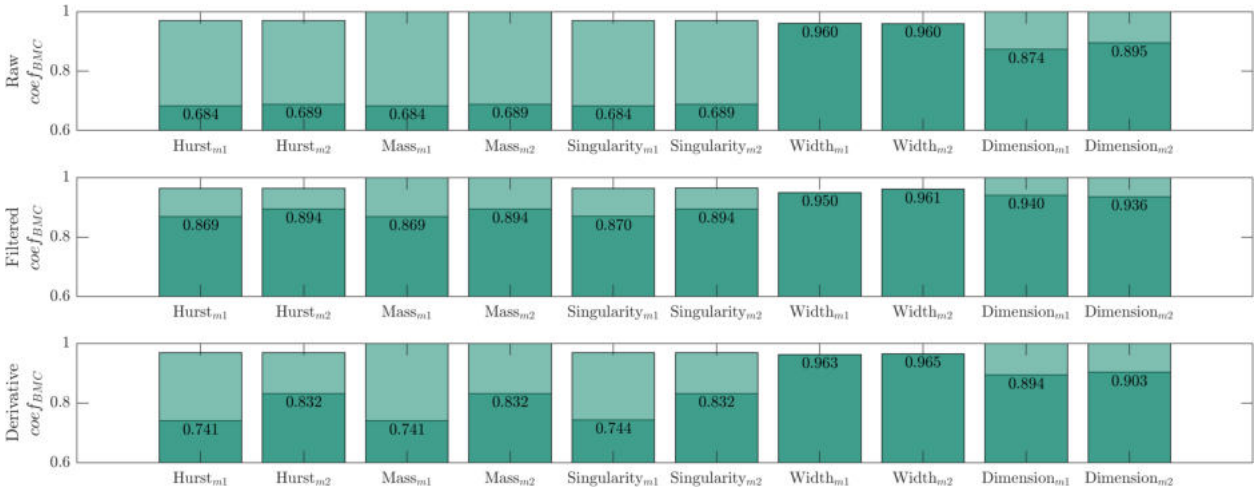

Figure 10.21: Multifractal Detrended Fluctuation Analysis measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

Note, that in Figure 10.22, both polynomial order values (m) were included, with the linear trend ($m = 1$) plotted in black, and quadratic trend ($m = 2$) plotted in green. The polynomial order determines the nature of trends removed from the signal. For $m = 1$, linear trends within each segment are subtracted, eliminating any simple drifts or steadily increasing or decreasing tendencies in the signal. On the other hand, $m = 2$ corresponds to removing quadratic trends, offering a more flexible detrending process. Note, that for most features, the informativeness results are similar, or consistent, for both orders. It indicates the vibroarthrogram's intrinsic fractal properties are robust against the detrending order. In other words, the multifractality and scaling behaviors are inherent characteristics of the

VAG signal. In cases, when informativeness between the trend orders differ, the lower trend usually provides better results. It suggests that the primary trends in the vibroarthrograms are predominantly linear, and that a simple linear detrending is adequate to reveal the core fractal behaviors. Increasing the order can result in over-detrending, i.e., removing fluctuations that are actually informative, or distinctive between the knee joint conditions.

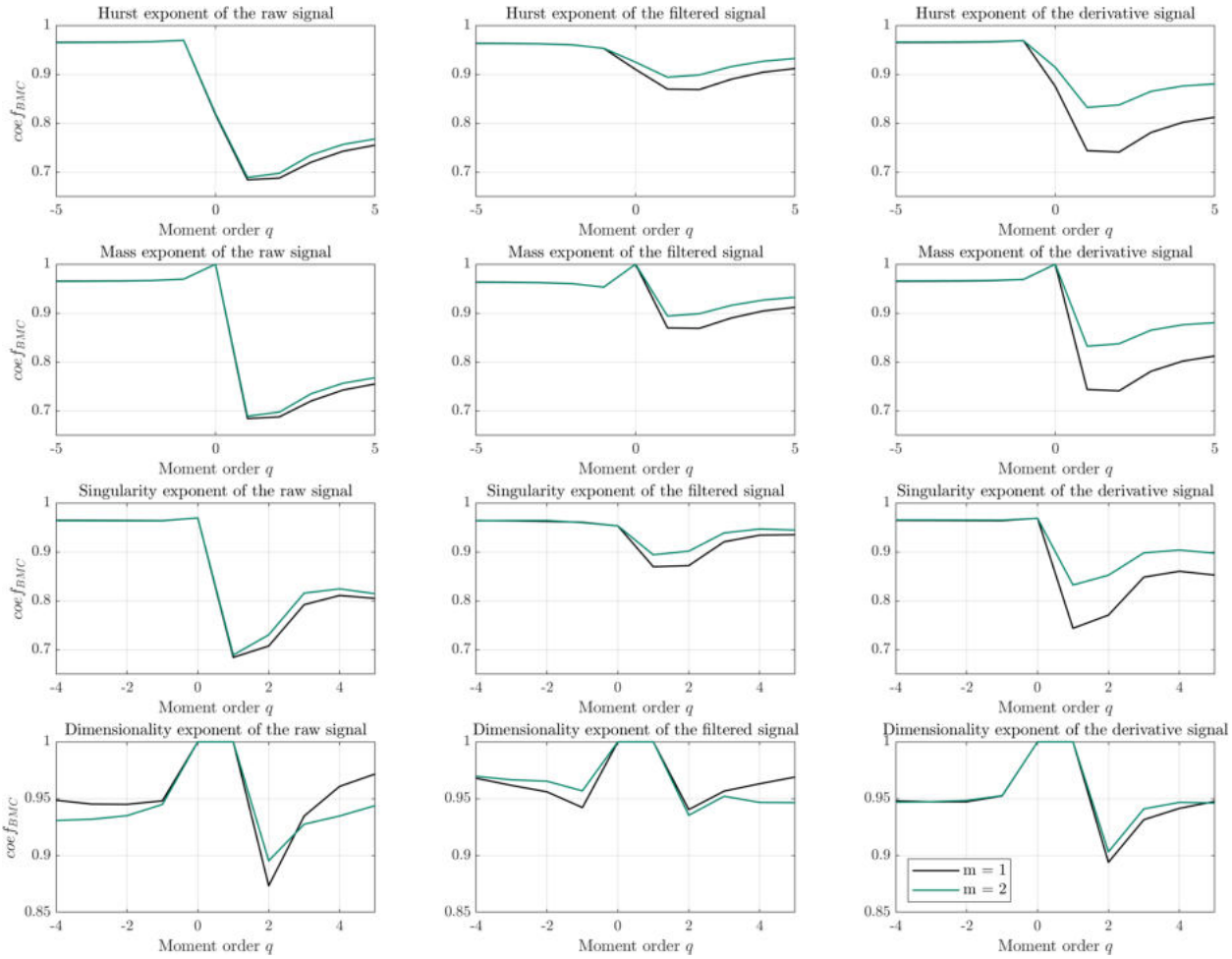

Figure 10.22: Informativeness scores of the Multifractal Detrended Fluctuation Analysis features in a function of the moment order q: the Hurst, mass, singularity, and dimensionality exponents from top to bottom row. Left, middle and right column correspond to the raw, filtered and derivative signal, respectively.

The Hurst exponent of the raw signal was relatively uninformative for negative q values, with the $coef_{BMC}$ of around 0.96. This suggests that small fluctuations, emphasized by the negative q values, do not hold much differentiating power between condition classes. For $q = 0$, which gives an equal weight to all fluctuations, the informativeness increases, with the $coef_{BMC}$ slightly above 0.8. It means that while considering average behavior across all scales, the Hurst exponent provides more distinction between knee joint conditions. For $q = 1$ the fluctuations are weighted linearly based on their size. In this case informativeness reaches maximum, yielding $coef_{BMC}$ of around 0.68 and making it the most distinguishing

feature from the Multifractal DFA. It captures more common or average-sized fluctuations, which seem to hold the most distinguishing information in the analyzed context. As q increases beyond 1, informativeness begins to drop, reaching $coef_{BMC}$ of around 0.75 for $q = 5$. This shows that emphasizing larger fluctuations does not seem to provide as much differentiation between classes as more commonly observed fluctuations. However, comparing informativeness scores between negative and positive values of q in general, shows that the information content of larger fluctuations seem to be greater than the information content of smaller fluctuations.

For the filtered signal, the Hurst exponent showed a similar informativeness pattern. However, the values of informativeness were significantly lower, with the lowest $coef_{BMC}$ of around 0.88 for $q = 1$. This reduction in informativeness suggests that the filtering process likely attenuated or removed from signals specific information content which could have been useful for class differentiation. The derivative signal also followed pattern similar to the raw and filtered signals, with informativeness greater than the filtered signal, but smaller than the raw one.

The mass exponent, for all preprocessing approaches, provided informativeness scores identical to the Hurst exponent, except for the $q = 0$ case. It is caused by the definition of the mass exponent, which is equal to the Hurst exponent multiplied by the q order and decreased by one. Therefore, for all q values, the mass exponent is a quasi-linear transformation of the Hurst exponent, preserving its information content, with an exception for $q = 0$ case, which makes the mass exponent for each signal equal to 0, completely removing its information content.

The singularity exponent, defined as the difference of consecutive mass exponents, also provided the most informative feature for $q = 1$. It is the case, in which from the mass exponent of q order 1, the mass exponent of q order 0 is subtracted. And, since the latter is always equal to 0, the singularity exponent for q order 1 is equal to the mass exponent of q order 1. For other cases, the informativeness was decreased or equal compared to the mass exponent for corresponding order q. It suggests that the nuanced variations between segments, captured by the singularity exponent, are not as discriminative between the classes as the direct fractal properties represented by the mass exponent.

The dimensionality exponent shows how the distribution of singularities, as quantified by the singularity exponent, varies across different scales of the signal. In the current context, for negative q values, its information content proved to be low across all preprocessing approaches. This suggests that small fluctuations in the signal don't significantly vary across the different knee joint condition classes and hence are not effective for differentiation. Interestingly, for q orders of 0 and 1, informativeness dropped to zero. It is, again, a result of the definition of the dimensionality exponent, which is equal to the mass exponent subtracted from the order q multiplied by the singularity exponent (see Table 5.8). This definition, for q order 0, makes the value of singularity exponent zero for all signals. On the other hand, for q order equal to 1, the definition of dimensionality exponent reduces to the $1 \cdot \alpha_1 - \tau_1 = \tau_1 - \tau_0 - \tau_1 = -\tau_0$, where the α_q is q-order singularity exponent, and τ_q is the q-order mass exponent. As discussed previously, τ_0 is equal to zero for all signals, rendering it uninformative.

Boxplot of the most informative DFA feature, i.e., the Hurst exponent for q order 1 and m order 1, was provided in Figure 10.23.

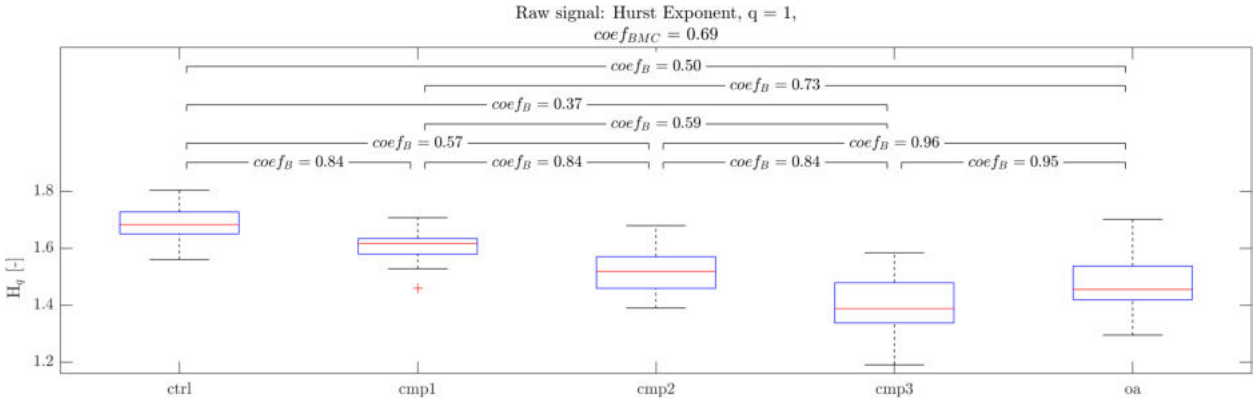

Figure 10.23: Boxplot of the Hurst Exponent feature calculated for the raw signal.

Values of the Hurst exponent greater than 1 indicate signals similar to random walks in their structure [150]. Moreover, greater values of the Hurst exponent indicate more robust trend-following behavior of the signal. In the studied context, the Hurst exponent seem to decrease with the joint damage, achieving greatest values for the control group and smallest for the cmp3. It means that in less damaged joints, the trend is more pronounced compared to the noise or random fluctuations, than in more damaged ones. Interestingly, the Hurst exponent is another VAG signal feature that is more similar within the cmp2 and oa class pair ($coef_B = 0.96$) than the cmp3 and oa ($coef_B = 0.95$). Though the difference is not great, the cmp2 and cmp3 class pair is differentiated better with $coef_B$ of 0.84.

10.5 Phase Space Reconstruction features

Informativeness of the Reconstructed Phase Space features, for both methods of choosing the embedding dimension, were plotted in Figure 10.24.

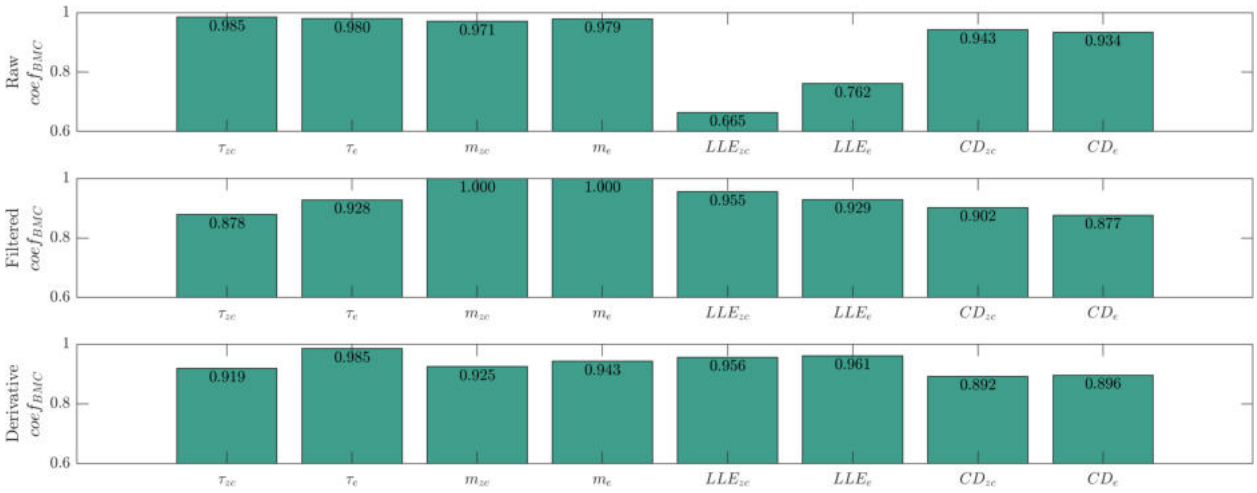

Figure 10.24: Reconstructed Phase Space measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

Features defined as the optimal time delay for the RPS, i.e., the τ_{zc} and the τ_e, did not turned out to have much differencing power between VAG classes, with the smallest $coef_{BMC}$ of around 0.88 for the τ_{zc} of the filtered signal. Since the τ_{zc} is the linear transformation of the FZCT of the ACF, their informativeness scores are the same (compare to Figure 10.20). The second method, i.e., the first time the ACF crosses $1-\frac{1}{e}$ value also proved to be rather uninformative. Similar values were achieved by the embedding dimension calculated by the False Nearest Neighbors method. For both the embedding dimension, and the time delay, the first zero crossing of the ACF proved to be a better method of optimal time delay estimation, in terms of information content. The Correlation Dimension feature, contrary to the delay itself, proved to be more informative when calculated for the τ_e delay. However, in general, its informativeness scores proved to be low, with $coef_{BMC}$ of around 0.88 for the most informative setup of filtered signal and τ_e delay.

On the other hand, the Largest Lyapunov Exponent proved to be quite informative, especially for the raw signal, scoring $coef_{BMC}$ of around 0.67 and 0.76 for the τ_{zc} and the τ_e delays, respectively. Both filtered and derivative signals proved to be rather uninformative, no matter which delay was used in the phase space reconstruction process. It can be explained by the low-frequency component of the raw signal, that is generated by the movement of the knee joint (see Figure 2.2). In the calculation of the LLE, proximate points are used to estimate the rate of divergence or convergence of trajectories over time, providing insights into the underlying dynamical properties of the system that generated the signal. Keeping the movement-generated low-frequency component enables points to be considered proximate only, when their high-amplitude low-frequency component is in similar phase. In other words, points are considered proximate, when they are both similar in their "spikydness" around the neighborhood, and in their movement phase. That way, the raw signal incorporates the information about the actual knee joint movement phase, which is lost in both filtering and differencing operations. Boxplot of the most informative LLE was provided in Figure 10.25.

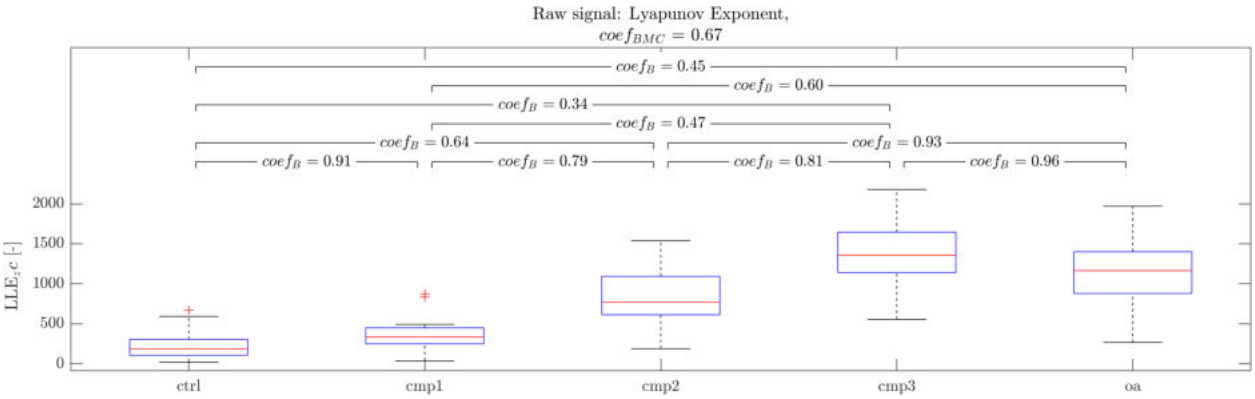

Figure 10.25: Boxplot of the Largest Lyapunov Exponent feature calculated for the raw signal.

The values of the Largest Lyapunov Exponent vary from 0 to around 2000. Those values can be interpreted as measures of the VAG-generating system's sensitivity to initial conditions, with 0 indicating stable and predictable dynamics with trajectories within the reconstructed phase space that remain close, and higher values signifying the average ex-

ponential rate at which system trajectories diverge, reflecting increasing levels of chaos and unpredictability in the system. Clearly, complexity of the VAG-generating system increases with progressing conditions of the knee joint, with the highest values of LLE scored, again, by the third stage CMP group.

Informativeness of the features based on the Recurrence Quantification Analysis were plotted in Figure 10.26.

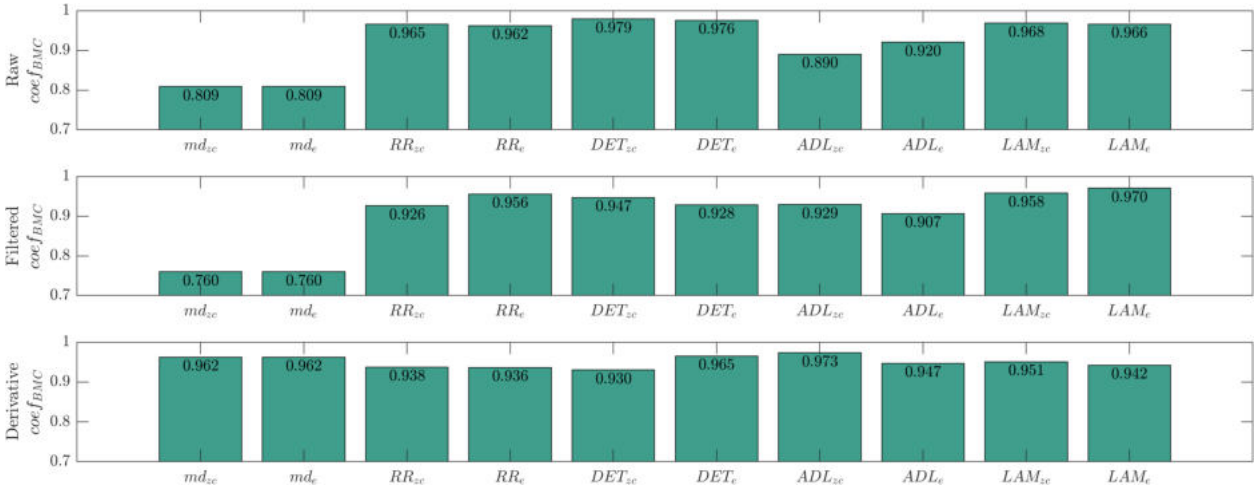

Figure 10.26: Recurrence Quantification Analysis measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

The Maximum Difference feature proved to be the most informative for the filtered signal, with the $coef_{BMC}$ of 0.76. For all preprocessing approaches, this feature did not differ in its information content between two methods of optimal time delay estimation, since it is basically equivalent to the statistical spread feature (see Figure 10.3). Boxplot of this feature was provided in Figure 10.27. Other methods, however, proved to be uninformative, with the lowest $coef_{BMC}$ of 0.89 scored by the Average Diagonal Line feature of the raw signal with the time delay of the ACF's first zero crossing. Their low informativeness can be explained by the time delay estimation method, as for time delay of 1, used by Kręcisz and Bączkowicz [5], they proved to be informative enough to be chosen by the feature selection algorithms. Another reason could be that in the current work, their values were approximated, which potentially could diminish their differencing capabilities.

In the boxplot of the Maximum Distance feature, it can clearly be seen that the spread of values increases with progressing knee joint degeneration, achieving highest values for the oa group. Values close to zero, extracted from some of the VAG signals from the control and cmp1 classes, mean that there are hardly any fluctuations observed in the signal. On the other hand, the median value of the Maximal Distance feature of the oa group is above 10, with the greatest value as high as 14. Note, that the VAG signals were recorder in a range from negative to positive 5 volts, meaning that the filtering process amplified some of the extreme values.

In Figure 10.28 summary of embedded entropy measures' informativeness were included. Bars ApEn (r_{max}) and ApEn r_{max} correspond to values of the Approximate Entropy that

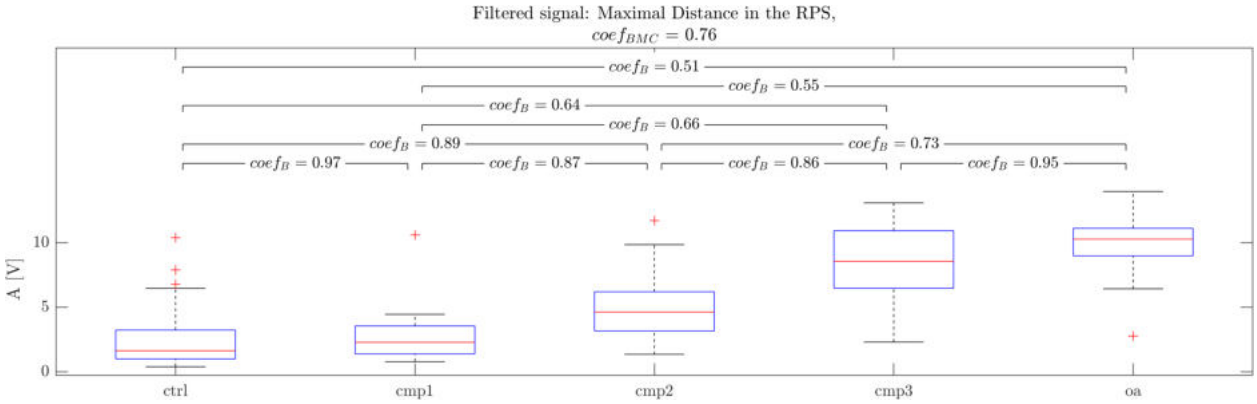

Figure 10.27: Boxplot of the Maximum Distance in the RPS (statistical spread) feature calculated for the filtered signal.

maximize the entropy value, and the radius r that maximizes this value, respectively. The same stands for the Sample Entropy measure.

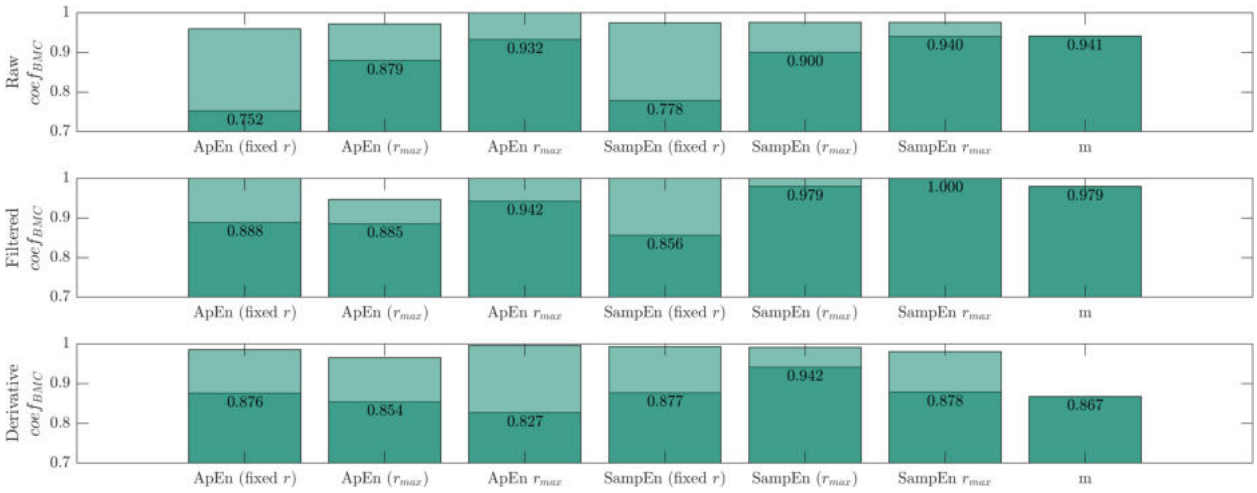

Figure 10.28: Entropy measures' informativeness of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

Detailed results of both the Approximate Entropy and the Sample Entropy, calculated for fixed r value, were plotted in Figure 10.29. Clearly, informativeness proved to be highest for the raw signal. It can be explained similarly to the Largest Lyapunov Exponent measures. The embedded-signal analyzes seem to be able to utilize information contained in the low-frequency component, related to the movement phase of the knee joint. In the Approximate Entropy and the Sample Entropy, the similarity of short sub-signals is used to quantify the unpredictability of the signal. It is done by calculating the logarithmic likelihood that sub-signals that are similar for a certain length will remain close when the length is increased. High-amplitude low-frequency component ensures that those sub-signals are considered similar only when they are in similar movement phases of the knee joint, increasing information extracted by both ApEn and SampEn.

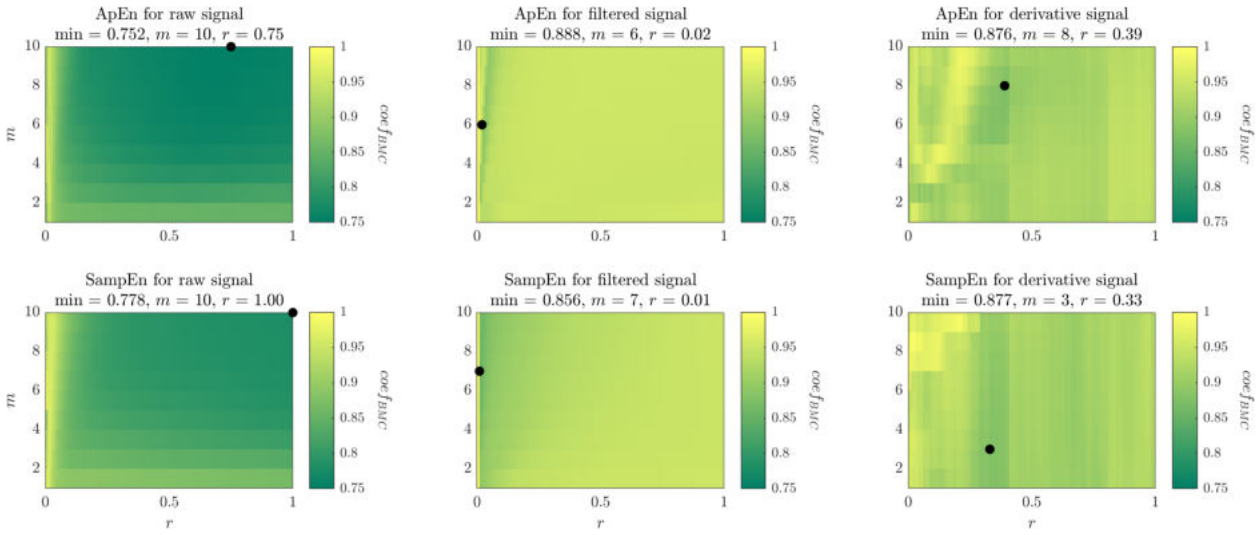

Figure 10.29: Detailed informativeness results of the Approximate Entropy (upper row) and the Sample Entropy (lower row) of the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot). Results for the fixed r parameter across the signals.

Comparing plots of the raw and filtered signals, it can be seen that for the raw signal, the informativeness of both entropy measures increases with greater r, and the opposite can be observed for the filtered signal. It can be explained, again, by the low-frequency component. The radius r, serving as a threshold to consider sub-signals similar, needs to be greater when considering movement-generated component present in the raw signal because of its high-amplitude. The derivative signal seem to be similarly informative to the filtered signal. Patterns observed in right-column plots of Figure 10.29, however, seem to be much less smooth than those found in left and middle columns. It can be explained by the much more noisy nature of the derivative signal. One more thing to consider is the spread and the standard deviation of the signal's values after each preprocessing. Their boxplots were included in Figure 10.30.

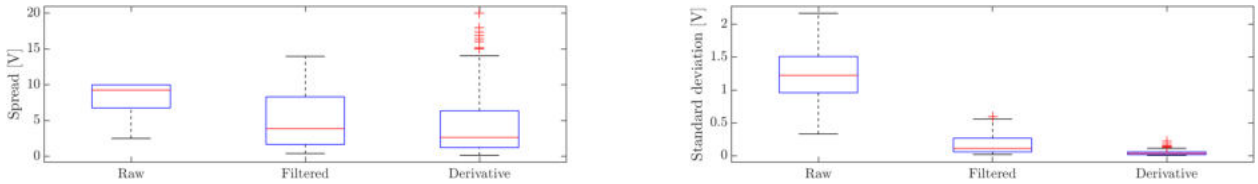

Figure 10.30: Spread (left) and standard deviation (right) of signals after each preprocessing.

Clearly, standard deviation of both filtered and derivative signals is much less compared to the raw signal. Their spread, however, is less affected, and can potentially increase. In the ApEn and SampEn features, the radius r is used as a multiplier of the std, and the std of the derivative signal is much lower than the other preprocessing approaches. Therefore, right-column plots of Figure 10.29 could be though of as corresponding to "stretched out" slices of low-r parts of plots for raw and filtered signals.

Interestingly, in most plots, length of sub-signals (m) used in the most informative ver-

sions of the features can be seen to be much greater than values of 2 or 3 suggested in literature. This could indicate, that VAG signals used in the current work could potentially be over-sampled and smaller sampling frequencies could be enough to be descriptive, at least in terms of Approximate Entropy and Sample Entropy. In previous work using the same signal base, Kręcisz and Bączkowicz [5] used Multiscale Sample Entropy, i.e., SampEn calculated for non-overlapping moving-average signal, with window sizes varying from 7 to 30. In the feature selection step, they selected window sizes of 14 and 30, which could confirm over-sampled VAG signal conclusion. Note, that this conclusion is to be considered in embedding entropy measures context only.

In all entropy results analyzed to this point, entropy measures were obtained for the same r across all signals. However, another approach is to choose r for each signal separately, in a way to maximize the entropy measure [163]. In Figure 10.31, informativeness of this approach was presented, for varying m parameter. However, in this approach, the m is constant across all signals. One more option could be to first choose m using the False Nearest Neighbors algorithm, and only then choose the r parameter maximizing the entropy. This approach was also included in each plot of Figure 10.31, indicated by the dot marker. Finally, informativeness of the r parameter treated as a feature were also included.

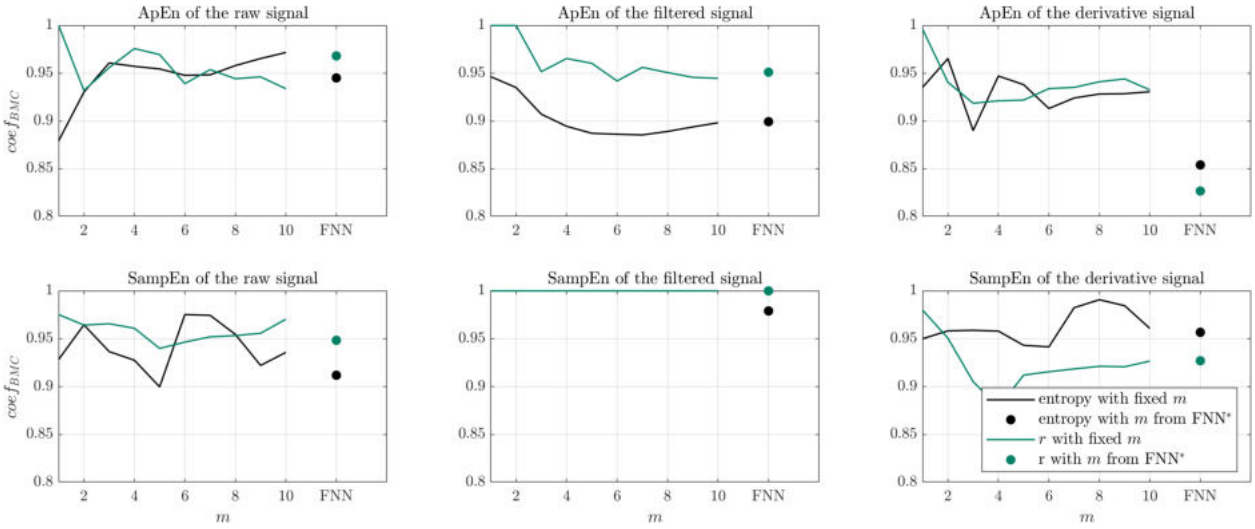

Figure 10.31: Detailed informativeness results of the Approximate Entropy (upper row) and the Sample Entropy (lower row) of the raw signal (left column), filtered signal (middle column), and the derivative signal (right column). Results for the r parameter maximizing the entropy measure. *False Nearest Neighbors algorithm.

Approach of maximizing r does not seem to provide more informative entropy measures compared to fixed r. Moreover, first choosing m using the False Nearest Neighbors also does not seem to improve informativeness, with a exception for the ApEn calculated for the derivative signal. Interestingly, for the filtered signal, maximizing SampEn seem to provide perfectly *uninformative* feature, i.e., scoring the $coef_{BMC}$ of 1. It means that the same r maximizes the SampEn of the filtered signals across all signals. Moreover, the maximum SampEn also seem to be the same across all vibroarthrograms, for fixed m. For the m first chosen by the False Nearest Neighbors algorithm, informativeness is extremely small, but

non-zero ($coef_{BMC} \neq 1$), meaning that there are *some* differences across vibroarthrograms.

In general, informativeness of ApEn and SampEn is very similar within all preprocessing approaches. Slightly lower values of $coef_{BMC}$ in the ApEn can be explained by lower variance of ApEn values compared to the SampEn [162]. Since VAG signals used in this work are rather long, bias of the ApEn is greatly reduced, apparently making it more informative feature compared to the SampEn. Boxplot of the most informative embedding entropy, i.e., the Approximate Entropy with $m = 10$ and $r = 0.75$ calculated for the raw signal, was included in Figure 10.32.

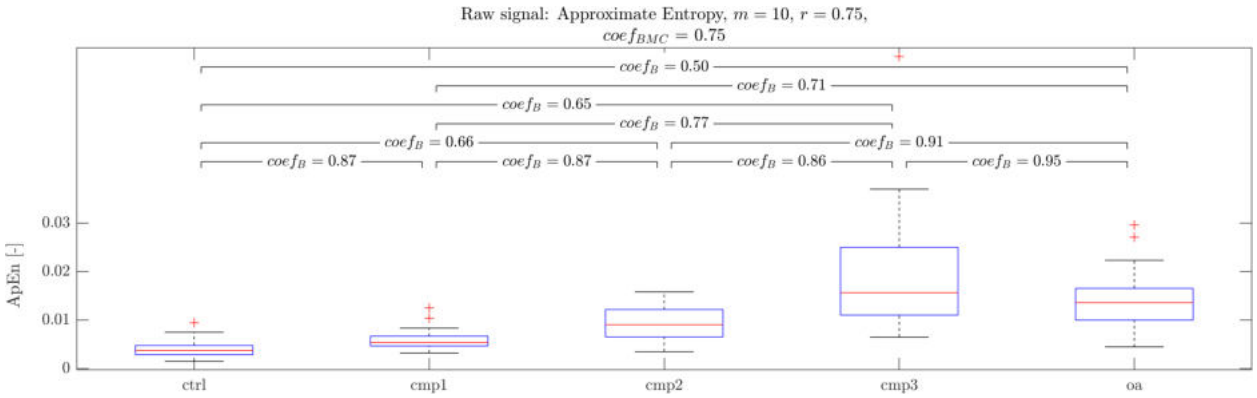

Figure 10.32: Boxplot of the Approximate Entropy feature calculated for the raw signal.

The Approximate Entropy generally increases with progressing knee joint conditions, achieving greatest values for the third stage Chondromalacia Pattellae class. For the Osteoarthritis class, however, ApEn values are very similar to the cmp3 ($coef_B = 0.95$). Interestingly, those results seem to be exactly opposite to the Multiscale Sample Entropy used by Kręcisz and Bączkowicz [5]. In their study, the entropy values seem to decrease with progressing knee joint degeneration, achieving smallest values for the oa group. It can be explained by the filtering process used in [5], which removed low-frequency components from the signals. Therefore, in their work, the SampEn measured similarity of sub-signals within the signal in general. In the current work, on the other hand, similarity of sub-signals was additionally "filtered" by the knee joint movement phase.

10.6 Correlation of the features

In this section, correlation of the time-domain features within each preprocessing, and between preprocessings will be discussed. For the parametric features, i.e., ones that were calculated for multiple parameters, only the most informative feature was used. Informativeness scores of those features were earlier presented as the lowest bars in barplots. Plots 10.33, 10.34, and 10.35 include correlations of features within the raw, filtered and derivative signals, respectively.

In each preprocessing approach, central tendency measures, i.e., the mean, mode and median values proved to be relatively uncorrelated. It can seem surprising, and is an effect of the skewed distribution of values discussed earlier. However, for the all preprocessing

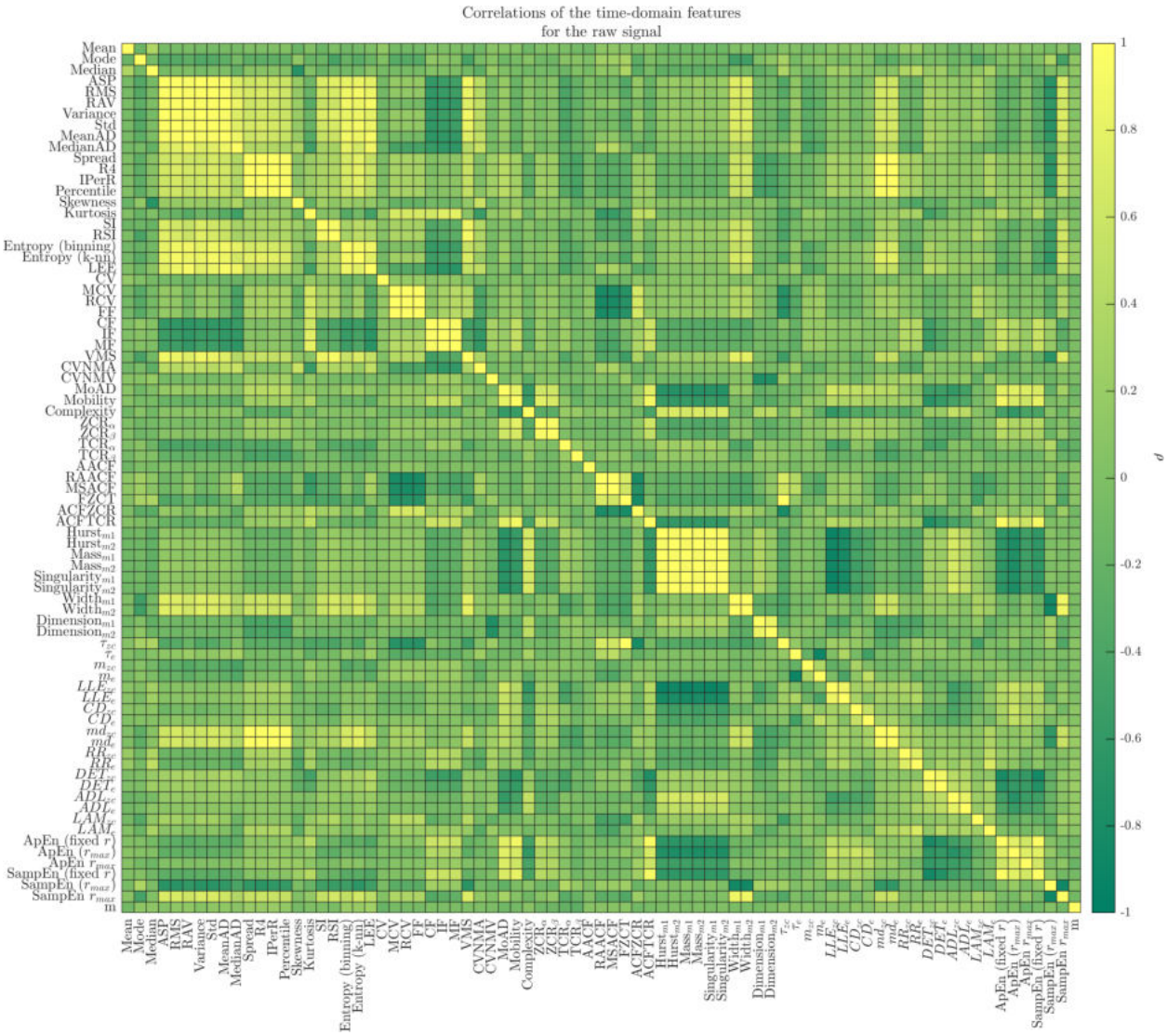

Figure 10.33: Correlation of time domain features obtained for the raw signal.

approaches, the power features, i.e., the ASP, the RMS and RAV proved to be highly correlated. It is not very surprising given their similar definitions. And since VAG signals are symmetrical, the power features are also highly correlated to the spread features. One exception is the percentile feature of the derivative signal, which is negatively correlated to power and spread features. It can be explained by the low percentile value that proven to be most informative in the derivative signal. With increased variability of the signal, the lower percentiles decreased.

For all preprocessing approaches, skewness proved to be correlated to other features rather poorly, proving uniqueness of its information content. Kurtosis proved to be correlated to some complex shape features, especially Crest Factor, Impulse Factor and Margin Factor. All three factors are sensitive to the extreme or peak values of a signal due to the inclusion of the maximum value in their numerators. This shared sensitivity to extremes aligns with kurtosis's measurement of the "tailedness" or extremeness of a distribution, leading to their

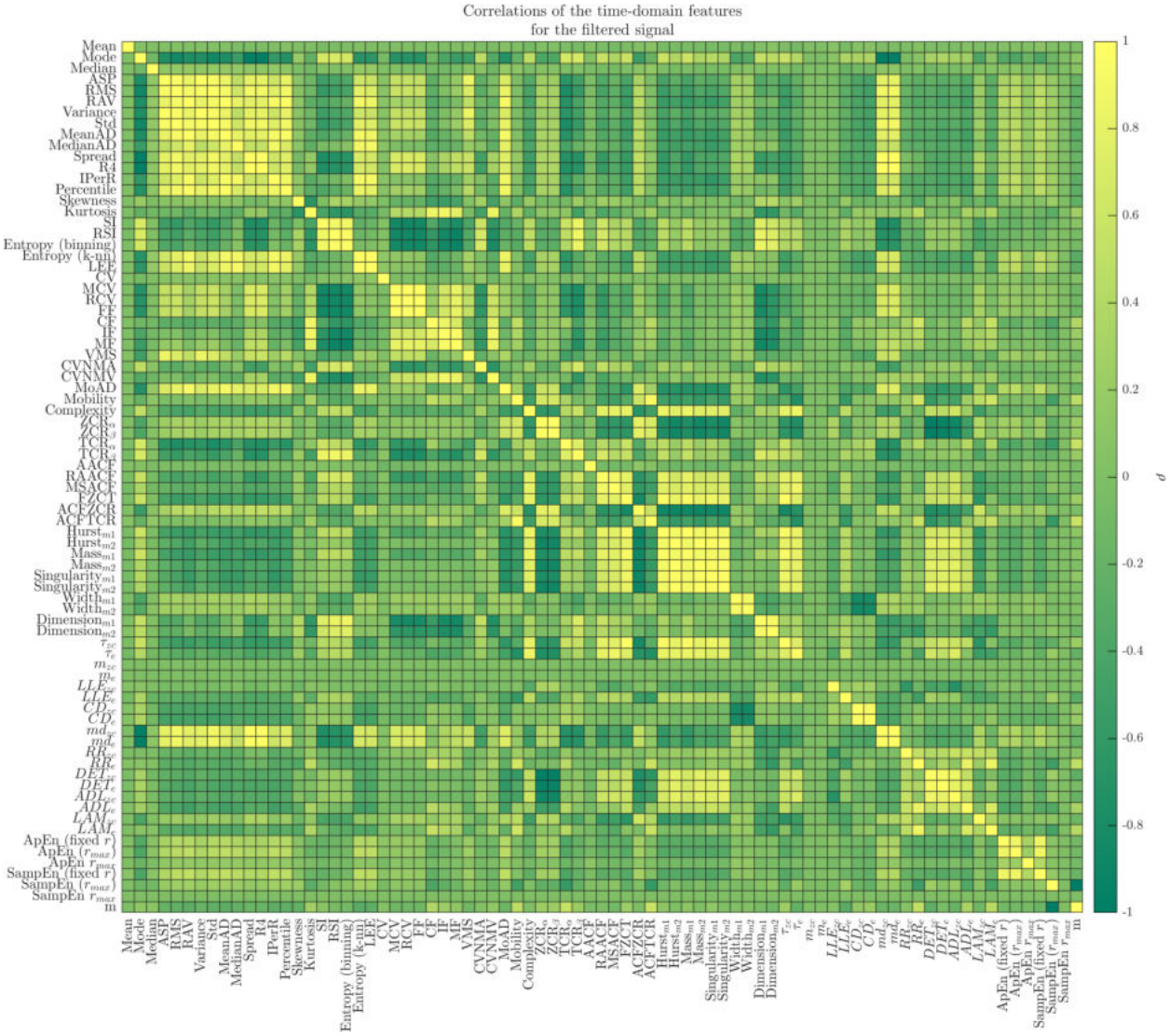

Figure 10.34: Correlation of time domain features obtained for the filtered signal.

high correlation. Surprisingly, for the filtered and derivative signals, kurtosis was also highly correlated to the Coefficient of Variation of the Normalized Moving Variance feature, which is one of features that are qualitatively different, since are defined considering temporal order of signal samples. When a signal has high kurtosis, indicating frequent extreme deviations from the mean, it naturally leads to pronounced changes in its moving variance. These fluctuations, when normalized and expressed relative to their mean, yield a high CVNMV. In the raw signal, however, a low-frequency high-amplitude component is present. While the kurtosis, focused on extreme values or "tailedness", might not significantly change due to this component, the CVNMV, which captures local fluctuations in variance, will be influenced. The low-frequency component smooths out the local variance, decreasing the CVNMV. As a result, even if kurtosis remains relatively high due to other extreme values in the signal, the CVNMV decreases, leading to a diminished correlation between the two metrics in the presence of low-frequency high-amplitude components.

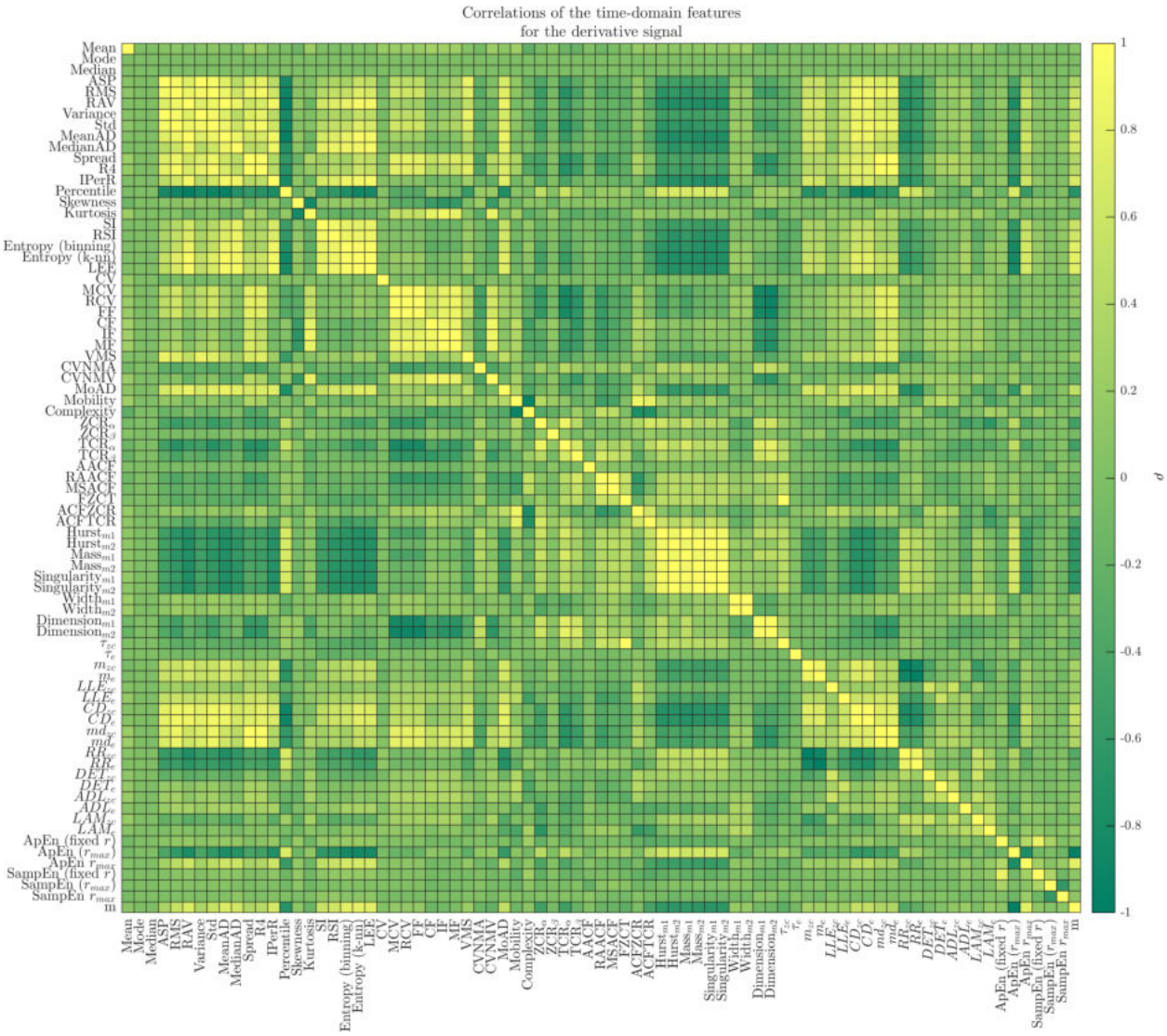

Figure 10.35: Correlation of time domain features obtained for the derivative signal.

Other simple statistical shape features, including the spiky indices and entropies proved to be correlated to each other within the raw and the derivative signals. In the filtered signal, however, both spiky indices and the binning entropy estimator create a highly correlated cluster, that is less correlated to the k-nn and LEE, which are more correlated to each other. Interestingly, the Log Energy Entropy feature proved to be correlated to the most informative k-nn entropy estimator, despite differences in their definitions.

Complex shape features mostly correlate with each other within each preprocessing approach. In the raw signal, two distinct clusters can be observed: features defined using std or RMS in the numerator (MCV, RCV, FF) and features defined using maximal value (CF, IF, MF). It can be explained by great influence of both filtering and differentiating on the std, but not maximal value (see Figure 10.30). Note, that the CV does not really correlate with any other feature, which is a result of its definition using average value in the denominator. Since VAG signals are relatively symmetrical across the y-axis, their average is close to zero,

rendering the CV inflated or undefined for most signals.

The first feature calculated using the rolling approach, so the VMS, proved to be relatively correlated with power and variability features, for all preprocessing methods. Two other features, so the CVNMA and the CVNMV turned out to be relatively uncorrelated with other features, with previously mentioned exception of CVNMV's correlation with kurtosis calculated for the filtered and derivative signals.

Difference and frequency features proved to be rather uncorrelated in most part, with an exception of the Mean of Absolute Differences being correlated with power and shape features withing the filtered and derivative signals. Also, Hjorth's complexity showed some correlation with the Multifractal DFA features, but only for the raw and filtered signals. Complexity, being defined using two differences, can be considered rather sensitive to noise. In the context of already derivative signal, its low informativeness, and low correlation with other features can be explained by extremely high noise content.

Features obtained using Autocorrelation Function show no considerable correlations, with two exceptions. First, the RAACF and MSACF show correlation to each other within each preprocessing. Their definition is very similar, so this correlation is not surprising. Both measures are relatively correlated to the first zero crossing time of the ACF. Both RAACF and MSACF reflect a signal's overall repetitiveness, while the first zero crossing time indicates the duration this repetitiveness remains significant. Their high correlation suggests that more repetitive signals maintain their self-similarity longer before the autocorrelation diminishes.

The Multifractal Detrended Fluctuation Analysis features, within each preprocessing approach, show extremely high correlation between the linear and quadratic detrending process ($m = 1$ and $m = 2$). Also, the Hurst, the mass and the singularity exponents are extremely correlated. This is a result of their definition and was discussed earlier. Interestingly, for the filtered and derivative signals, those exponents showed high negative correlation to power features. In the absence of low-frequency, high-amplitude components, the negative correlation between the the DFA features and power features indicates that pronounced high-frequency fluctuations are predominantly anti-persistent. That is, when the signal exhibits more high-frequency variations, they tend to reverse direction more frequently rather than continuing in the same trend. However, in raw signals, the low-frequency components can skew power values, masking this relationship.

Width of the fractal spectrum, on the other hand, proved to be positively correlated to power features, but only for the raw signal. It further confirms the link between signal's power and complexity, greatly influenced by high-amplitude low-frequency components. After removing these components in filtered and derivative signals, the high-frequency fluctuations, though anti-persistent, do not significantly impact the signal's multifractality, leading to a reduction of correlation with power.

For the raw and the derivative signals, optimal time delays, τ_{zc} and τ_e showed no major correlations to other features. However, for the filtered signal, those times were highly correlated with the DFA features. It makes sense, since high Hurst exponent indicates long-term dependencies in a signal, causing the ACF to decay more slowly, and rendering values of both τ_{zc} and τ_e greater. On the other hand, lower Hurst exponent indicates short-term memory, resulting in a quicker decay of the autocorrelation function and smaller values of both τ_{zc} and τ_e. The filtering process proved to emphasize the memory, or persistence, of the signal, making the relationship between the DFA features the ACF more evident.

The embedded dimension obtained using the False Nearest Neighbors algorithm proved to be uncorrelated to other features for the raw signal. For the filtered signal, correlation could not be calculated, since all m values were the same across all signals. Derivative signal, however, showed high correlation of both m_{zc} and m_e to power features. It means that with greater power of a signal, number of dimensions needed to reconstruct its phase space increases.

The Largest Lyapunov Exponent proved to be negatively correlated to the DFA measures for the raw signal, and did not exhibited significant correlations in either filtered or derivative signals. The high-amplitude low-frequency component in the raw signal can be understood as a *guiding* structure, ensuring that points close in knee joint movement phase, also remain proximate in phase space. That results in a reduced sensitivity to initial conditions, leading to lower LLE values. Furthermore, this guiding structure also makes the signal exhibit long-range correlations, which is reflected in a higher Hurst exponent. However, when this low-frequency component is removed through filtering or differencing, both the long-term correlations, measured by the Hurst, mass and singularity exponents, and the chaotic behavior, measured by the LLE, are significantly altered.

The Correlation Dimension feature showed positive correlations to power features of the derivative signal. It can be explained by the fact, that the power of the derivative signal, primarily emphasizing high-frequency fluctuations, is directly tied to its complexity. As this power increases, so does the complexity, leading to a positive correlation with the CD feature.

The Recurrence Quantification Analysis measures did not show major correlations with other features within the raw and the derivative signals. However, for the filtered signal, they proved to be positively correlated to Detrended Fluctuation Analysis measures and negatively correlated to the ZCR. The positive correlation between RQA measures and the Hurst, mass and singularity exponents suggests that the removal of specific frequency components might expose some underlying recurrent nature of the signal, that aligns with its long-range dependence. The negative correlation with the ZCR is not surprising. It indicates that that high-frequency fluctuations are inversely related to the recurrence and determinism of the signal. Those relations, however, seem to be masked in both raw and derivative signals.

Embedding entropy measures, i.e., the Approximate Entropy and the Sample Entropy measure unpredictability of the signal. In the raw vibroarthrogram, the ApEn showed negative correlation to the Hurst, mass and singularity exponents, indicating that stronger long-term correlation within the signal is inversely proportional to its unpredictability. On the other hand, variability measured by the MoAD feature, proved to be positively correlated with ApEn.

Correlations between features *across* the preprocessing approaches were included in Figures 10.36 (for the raw and the filtered signals), 10.37 (for the raw and the derivative signals), and 10.38 (for the filtered and the derivative signals).

Power measures of the raw signal proved to be negatively correlated to the Correlation Dimension of the filtered signal. It indicates that as the raw signal becomes more energetic, the filtered signal exhibits a reduction in complexity. The filtering process may be simplifying the inherent dynamics of the signal, reducing its dimensionality. In other words, higher energy components in the raw signal may correspond to features that, once filtered, reduce the diversity or complexity in the filtered signal's structure.

The Mean of Absolute Differences and Hjorth's mobility of the raw signal showed negative

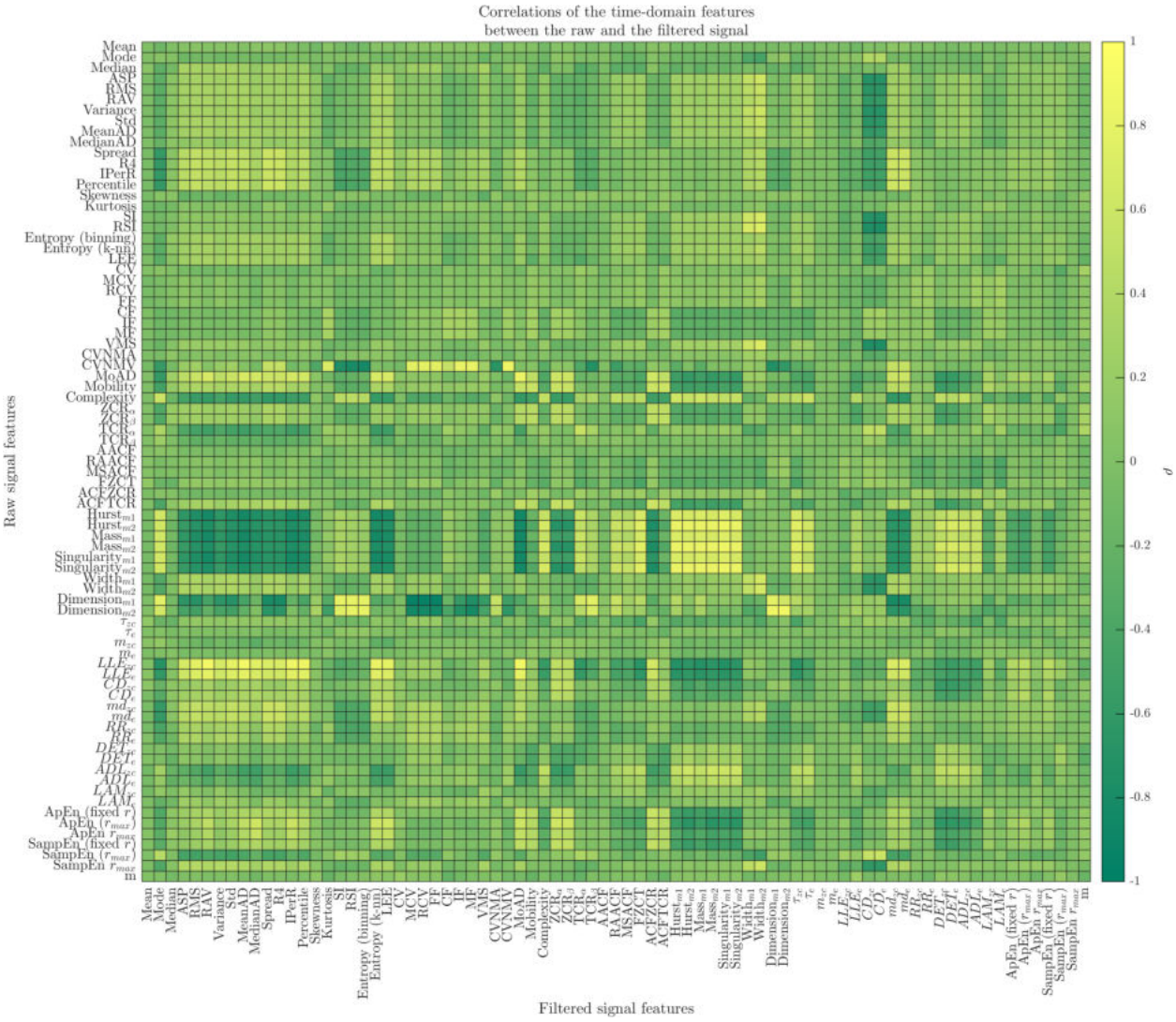

Figure 10.36: Correlation of time-domain features between the raw (y-axis) and the filtered (x-axis) signals.

correlation to the Hurst, mass and singularity exponents calculated for the filtered signal. When the raw signal has greater variability, the filtered signal tends to exhibit lower long-term memory. The filtering process, therefore, seem to remove components in the raw signal, that contribute to long-term memory when the raw signal has larger absolute differences. It means, that low-frequency high-amplitude component of the raw signal may not be the only factor determining the persistence characteristics of the filtered signal. Instead, the high-frequency components of the raw signal also play *some* role in shaping its post-filtering long-term memory characteristics. On the other hand, Hjorth's complexity proved to be positively correlated to the DFA measures of the filtered signal. In can seem unintuitive, given the previously mentioned relationship. However, it indicates that while immediate fluctuations in the raw signal can influence long-term persistence post-filtering, the overall complexity of the raw signal, as quantified by Hjorth's complexity, also has a significant

impact on how the filtered signal's memory properties are *preserved*.

The Hurst, mass, and singularity exponents of the raw signal seem to be positively correlated to the same measures of the filtered signal. In other words, the memory characteristics inherent in the raw signal are preserved to a certain extent in the filtered signal. On the other hand, the same measures of the raw signal are negatively correlated to its post-filtering power. In other words, a raw signal with a higher degree of long-term dependencies or persistence results in a filtered signal with reduced energy. This can be explained by the low-frequency component of the raw signal, that contributes to its long-term memory measured by the DFA features. Since it is also quite high-amplitude, when it gets removed in the filtering process, the resulting power features are reduced. Intuitively, the bigger the low-frequency component, the more power is lost in the filtering process.

The Largest Lyapunov Exponent of the raw signal positively correlates with the power of the filtered signal, while negatively correlating with its Hurst, mass and singularity exponents. Those results reveal that as the raw signal becomes less deterministic (higher LLE values), the corresponding filtered signal's energy increases, suggesting that chaotic components in the raw signal directly contribute to the energy in the filtered version. However, this increase in chaotic behavior reduces the long-term memory features of the filtered signal. This indicates that the filtering process modifies the chaotic characteristics from the raw signal in such a way that it reduces long-term dependencies in the filtered domain while enhancing its overall power.

The general patterns in the correlation between the raw and the derivative signal are similar to the raw-filtered correlations. There are some interesting differences, however. The Mean of Absolute Differences, as well as Hjorth's mobility, and the ZCR feature showed high correlation with the derivative signal power. It is not really surprising, since the MoAD of the raw signal has exactly the same definition as the Rectified Average of the derivative signal. Less obvious positive correlations can be observed between the embedded entropy measures of the raw signal and the derivative signal power. It suggests that the differentiation magnifies the power traits tied to a signal's inherent irregularities. On the other hand, negative correlation between the Hjorth's complexity of the raw signal and the power of the derivative signal can be observed. Hjorth's complexity measures the signal's frequency composition in relation to its waveform's shape. A higher complexity value implies a signal that has diverse frequency components. When such a signal is differentiated, low-frequency components get reduced, yielding derivative signal of lowered power.

Correlation between the filtered and the derivative signals' power features seem to be positive, though not extremely high. Both filtering and differentiating modify a signal's frequency content, reducing its low-frequency component. The fact that a moderate correlation exists implies that this low-frequency component has considerable power, as expected from previous discussions.

Positive correlation was also observed between the derivative signal's simple statistical shape features and the power of the filtered signal. This result indicates that with increased variability of the derivative signal, power of the filtered signal also increases. It is quite intuitive result, given that the variability of the derivative signal is proportional to high-frequency content of the raw signal. Since this content is preserved by the filter, it contributes to the resulting signal's power. On the other hand, negative correlation was observed between the Hurst, mass and singularity exponents of the derivative signal and the power of the

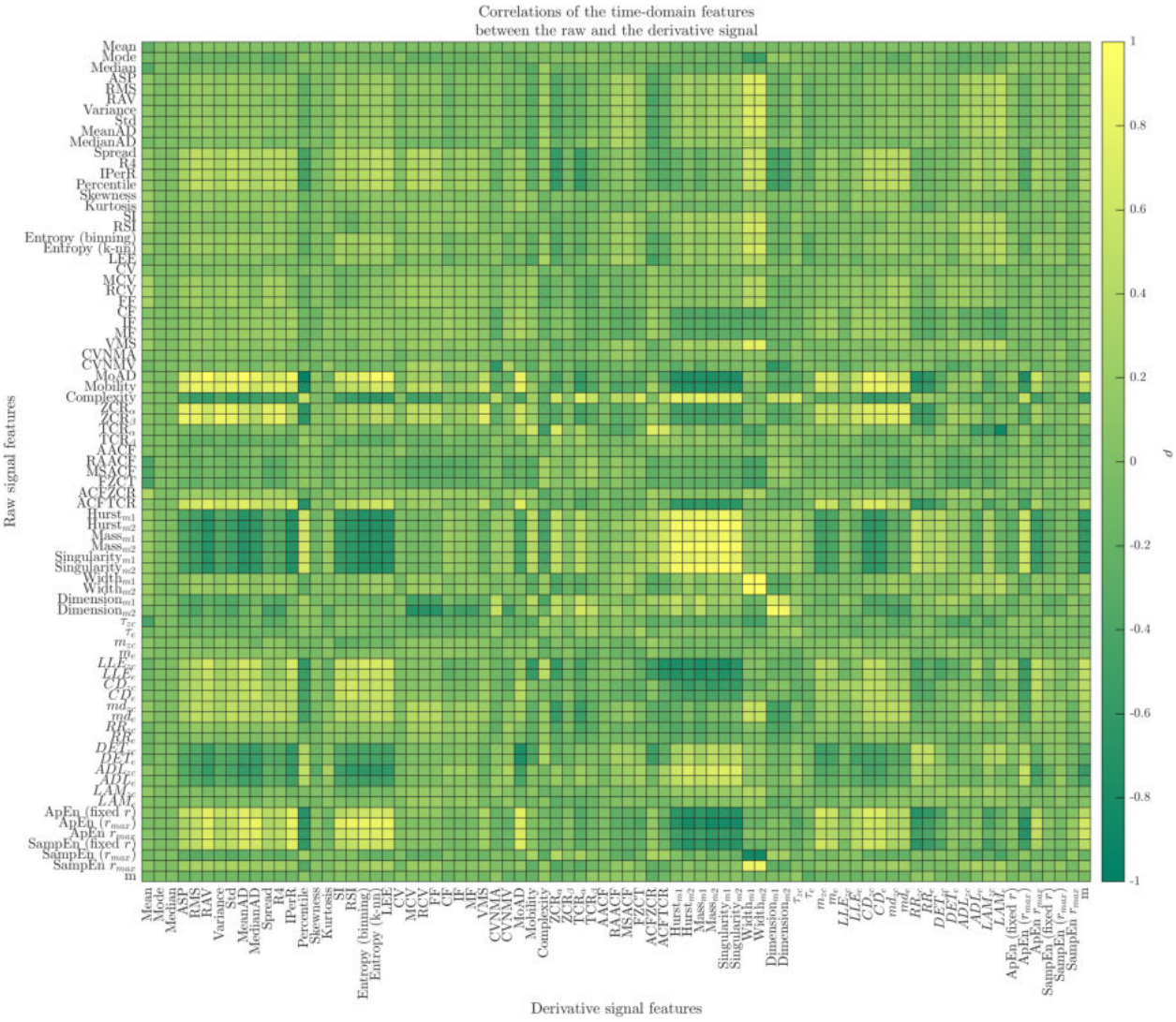

Figure 10.37: Correlation of time-domain features between the raw (y-axis) and the derivative (x-axis) signals.

filtered signal. The DFA measures of the derivative signal represent its long-term memory. Higher values in this context suggest more pronounced long-term dependencies in the raw signal that may predominantly lie above 1000 Hz range, making them amplified by the differentiation process and reduced by the filtering. When these dependencies are strong, the signal post-filtering has lesser power. And inversely, when the DFA measures calculated from the derivative signal are lower, indicating less pronounced long-term characteristics, the power of the signal post-filtering tends to be greater. This suggests that when rapid changes and short-term dynamics are more prominent in the raw signal (reflected in a lower Hurst value of the derivative signal), power within the 50-1000 Hz range is more significant after filtering.

Finally, the DFA measures proved to be positively correlated between the filtered and the derivative signals. This further confirms that certain aspects of long-term memory or

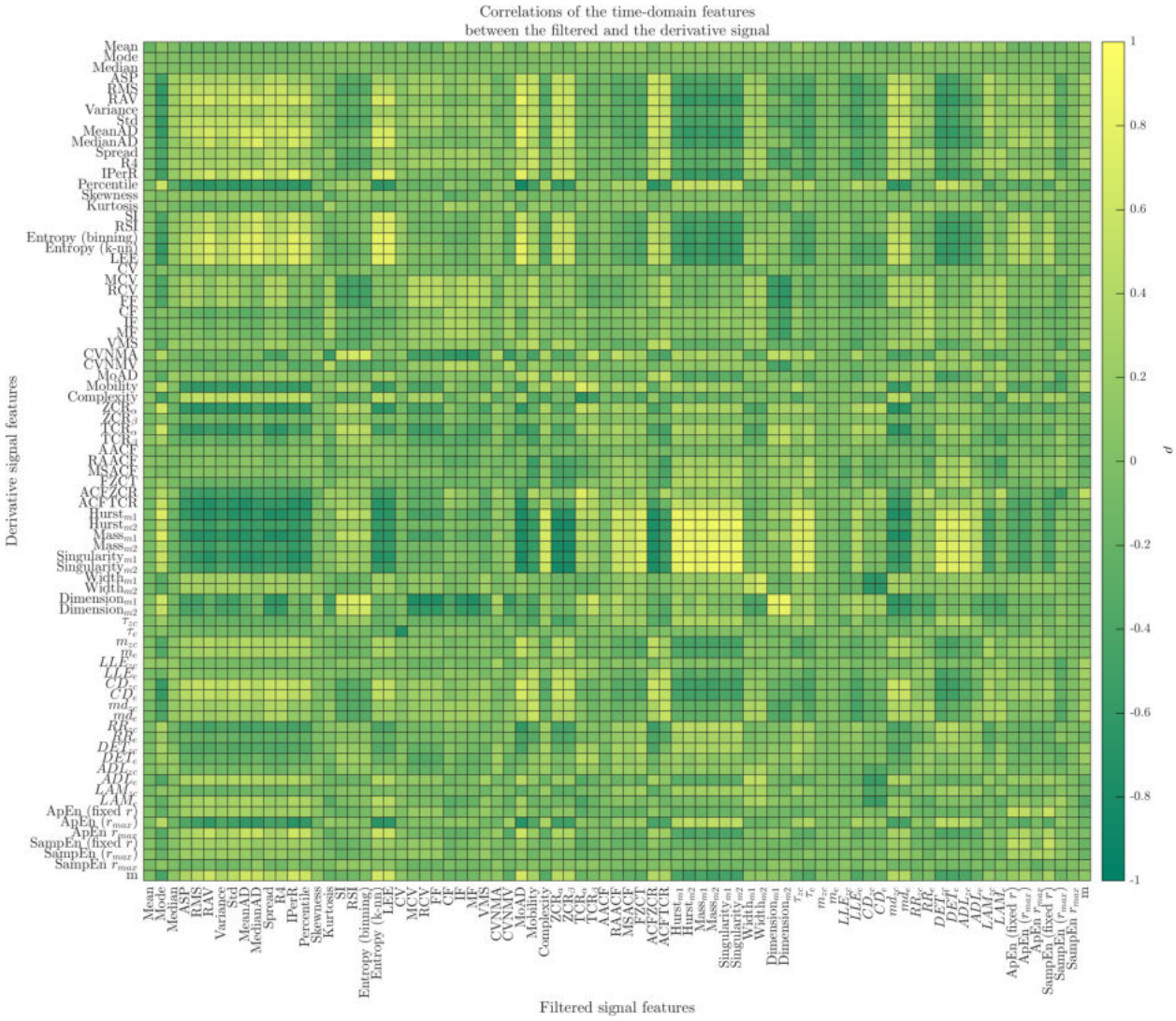

Figure 10.38: Correlation of time-domain features between the derivative (y-axis) and the filtered (x-axis) signals.

persistence present in the signal's structure are retained across both differentiation and filtering. That is, if a signal displays pronounced long-term dependencies post-differentiation, these attributes are similarly evident post-filtering. It highlights the robustness of the DFA measures, as their characteristics seem to be consistent across all preprocessing approaches.

10.7 Summary

In summary, for each preprocessing approach, 79 main features were extracted. While some of these were singular scalar values, others were calculated for multiple parameters, leading to a cumulative count of 3142 distinct features for each preprocessing approach.

Statistical features, i.e., features that are calculated for a signal's values without considering their temporal order, turned out to be relatively informative. Power features of the

vibroarthrogram increase with progressing knee joint conditions, achieving highest values for the cmp3 and oa classes. Since VAG signals are symmetrical along the y-axis, the same pattern can be observed in spread features.

From the simple statistical shape measures, the k-nn entropy estimator calculated for the filtered signal proved to be highly informative, without great influence of k on the informativeness, at least in the 1 to 5 range. Similarly to power measures, this entropy increases with progressing knee joint conditions. The Log Energy Entropy, despite being calculated qualitatively different, also achieved high informativeness and proved to be highly correlated with k-nn entropy estimator. Complex shape measures did not show high informativeness.

Rolling features also scored relatively low informativeness scores. Influence of the window size on the informativeness turned out to be limited for most features. For the VMS calculated for filtered signal, window size of the highest informativeness proved to be around 2 ms, which is significantly shorter than commonly used in literature.

Time-domain frequency features achieved relatively high informativeness, with the highest being the Mean of Absolute Differences feature calculated for the filtered signal. Similarly to previously mentioned features, MoAD increased with progressing knee joint condition. The Zero Crossing Rate feature proved to be most informative in the derivative signal context, calculated for threshold of around 10 percent of the signal's standard deviation. Similar threshold was most informative for the Turns Count Rate of the filtered signal. Both features proved to be positively correlated to each other and negatively correlated to the power of the filtered signal. Therefore, they can be expected to decrease with progressing knee joint conditions.

Features defined using the Autocorrelation Function proved to be mostly uninformative, with moderately informative exception of the Autocorrelation Function Turns Count Rate feature calculated for the derivative signal. Because of its positive correlation with both derivative signal's ZCR and filtered signal's TCR, as well as negative correlation with filtered signal power and spread features, it can be expected to decrease with progressing knee joint conditions.

The Multifractal DFA features proved to be most informative for both q and m orders equal to one, calculated for the raw signal. With those parameters, with progressing knee joint conditions, the Hurst exponent proved to decrease.

From the RPS features, the Largest Lyapunov Exponent of the raw signal showed highest informativeness, calculated for the phase space reconstructed using first-zero-crossing method of obtaining the time delay. The feature proved to increase with progressing knee joint conditions. Measures derived from the Recurrence Quantification Analysis did not show high informativeness in the analyzed context.

Embedded entropy measures showed relatively high informativeness in the raw signal context, for high r values ($r = 0.75$ for the ApEn and $r = 1.0$ for the SampEn) and m equal to 10. Both measures calculated for the r parameter maximizing the entropy value proved to be of rather low informativeness.

10.8 Conclusions

From the conducted time-domain analysis of VAG signals, following conclusions can be derived.

1. From the implemented approaches, filtering seem to be the most informative preprocessing for time-domain feature extraction.
2. Features, which definition is based on proximity of points or sub-signals, seem to gain considerable class differentiating power when the low-frequency component of the signal is kept. It could be attributed to knee joint movement phase localization capabilities of the low-frequency component.
3. In general, with increasing degeneration of the knee joint, variability and power of the signal increases. At the same time, decreasing of features measuring long-range correlations can be expected.
4. Further research could benefit from first filtering the VAG signal, and the augmenting it with movement-phase information, derived from the VAG signal itself (its low-frequency component) or additional external sensor (such as goniometer).
5. In the current study, filtering approach was chosen based on the literature. Further research could benefit from analyzing informativeness of the features, based on different filtering frequency bands.

Chapter 11

Frequency domain features

In this chapter, results of the frequency analysis will be provided and discussed. Similarly to previous chapter, for the most informative features, their boxplots will be provided. The chapter will begin with the PSD estimation methods comparison. Then, features obtained using the FRM approach will be discussed, followed by features not following that methodology. Next, correlations between specific features will be discussed. The chapter will end with a summary and conclusions which can be derived from the conducted frequency analysis.

11.1 Estimation of the Power Spectral Density

To compare Power Spectral Density estimation methods, 22 Frequency Range Maps were obtained using each method. Then, informativeness of each feature in each FRM was determined. Next, the best feature from each FRM and for each PSD estimator was chosen. Finally, for each PSD estimator, the best features were ranked according to their informativeness. For example, for the F_{max} feature, Welch method with 2^{15} window size provided most informative feature, achieving rank 1. Second most informative feature was obtained using Multitaper method with $n = 15$, achieving rank 2, etc. For the PV_n feature, the Welch method with 2^{14} window size proved to be most informative, achieving rank 1. Such ranking was created for each feature. Average rankings per PSD estimation methods, across all features, were provided in Table 11.1. Columns named period, w512, w1k, w2k, w4k, w8k, w16k, w32k, mt3, mt7 and mt15 correspond to the following PSD estimation methods, respectively: the periodogram, the Welch method with window sizes of 2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{14}, and 2^{15}, and the Multitaper method with $n = 3$, $n = 7$ and $n = 15$. Full tables, including the informativeness of the best feature per FRM, per PSD estimation method, were included in Appendix A.

Table 11.1: Average rankings of the Power Spectral Density estimation methods.

preprocessing	period	w512	w1k	w2k	w4k	w8k	w16k	w32k	mt3	mt7	mt15
raw	6.09	4.95	4.77	4.77	3.73	9.09	5.32	5.82	5.82	8.09	7.55
filter	6.95	5.91	6.36	6.82	5.82	6.86	3.59	7.86	4.64	5.50	5.68
difference	5.73	6.45	5.59	6.32	6.55	7.41	2.50	7.82	5.23	5.91	6.50
average rank	6.26	5.77	5.58	5.97	5.36	7.79	3.80	7.17	5.23	6.50	6.58

In the Table 11.1, average rankings for each preprocessing method were included. In the last row, average rank across all preprocessings was added for easier analysis. In the context of this work, the best Power Spectral Density estimation method proved to be the

Welch method, with window size of 2^{14} (16384) samples. Therefore, it was used for the frequency feature extraction, including Frequency Range Maps with greater resolution, as well, as features not following the FRM methodology.

11.2 Frequency Range Maps

In Figure 11.1, a summary of the FRM distribution features' informativeness was provided. Similarly to barplots of time-domain features analyzed in the previous chapter, two bars are provided for each feature: the most and the least informative feature in a given set.

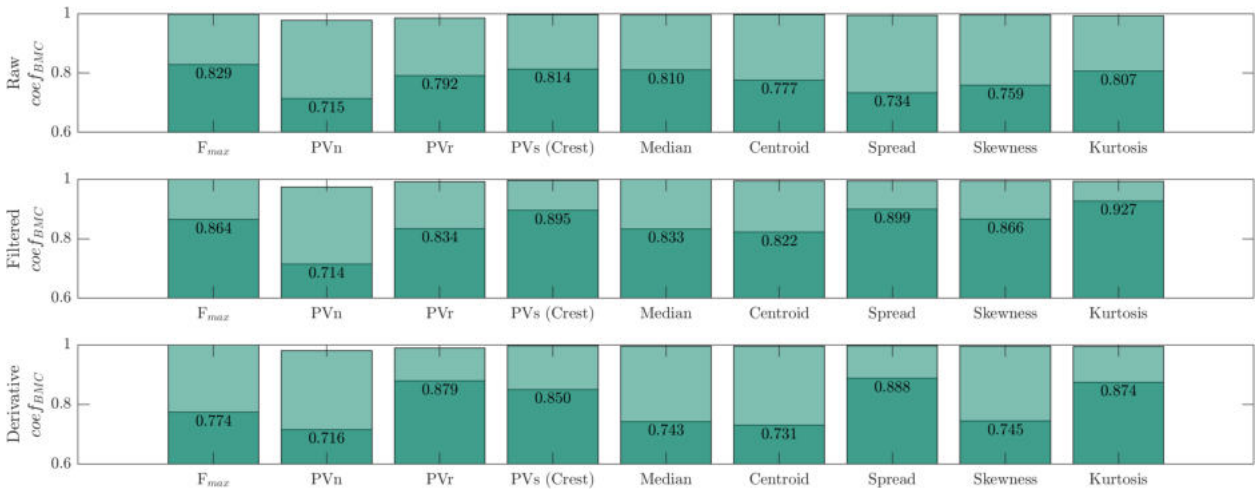

Figure 11.1: Informativeness of the Frequency Range Map distribution features.

In Figure 11.2, the Frequency Range Maps of the F_{max} were provided. For this feature, the most informative preprocessing approach seems to be the differentiation. Despite its high-pass filtering characteristics, the lowest $coef_{BMC}$ values can be found in rather low frequencies, i.e., the left part of the FRM, with the most informative range being from 0 to 130 Hz. Similarly low frequencies were found to be most informative for the raw and filtered signals, with 30 to 240 Hz and 0 to 100 Hz, respectively. For the raw and the filtered signals, characteristic vertical lines can be spotted, indicating that the lower frequency has a greater influence on informativeness of the feature. Those lines are not present in the derivative signal FRM. It can be explained by the definition of the F_{max} feature. Energy of VAG signals is mostly contained in the lower frequencies of the spectrum. Therefore, for most ranges, the peak value can be found in lower frequencies in the given range. Considering wider spectrum, therefore, does not change the frequency value, at which the peak is found, rendering its information content identical to narrower frequency range. The differentiation, however, increases power of the higher frequencies, minimizing this phenomenon.

Figure 11.3 includes FRMs of the actual non-normalized power of the peak value in a given range. Maps for all preprocessings look very similar, and in all three, almost the same ranges seem to be the most informative: from 220 to 270 Hz (with an exception for the derivative signal, for which the lower frequency is 10 Hz smaller). Different pattern seems to emerge from the relative-normalized part of the PSD. Maps for this feature were

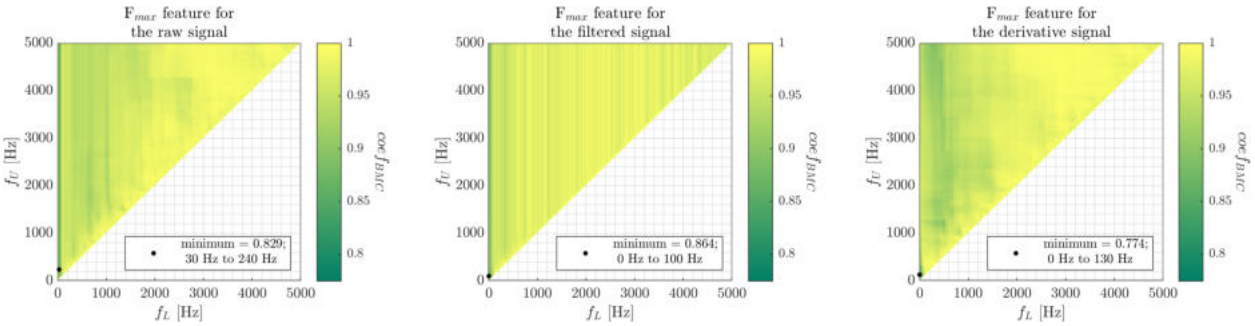

Figure 11.2: Frequency Range Map of the F_{max} feature.

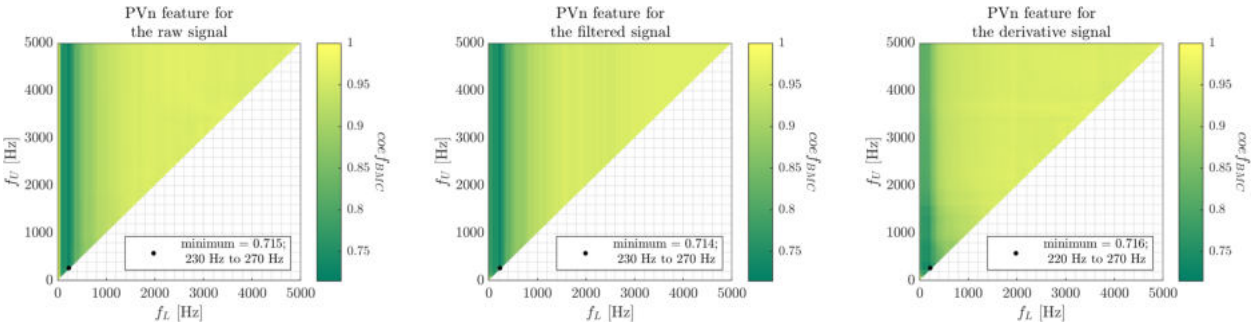

Figure 11.3: Frequency Range Map of the PVn feature.

included in Figure 11.4. The raw signal proved to be most informative, with the lowest $coef_{BMC}$ of around 0.79. It was scored for lower frequency range than the PV_n, i.e., from 100 to 140 Hz. In general, for all maps generated for the PV_r, the best features were obtained for much lower frequency ranges compared to the PV_n. And the informativeness of the relative-normalized spectrum was, in general, much lower compared to non-normalized one. Even lower informativeness was obtained for the self-normalized part of the PSD, FRMs of which were included in Figure 11.5.

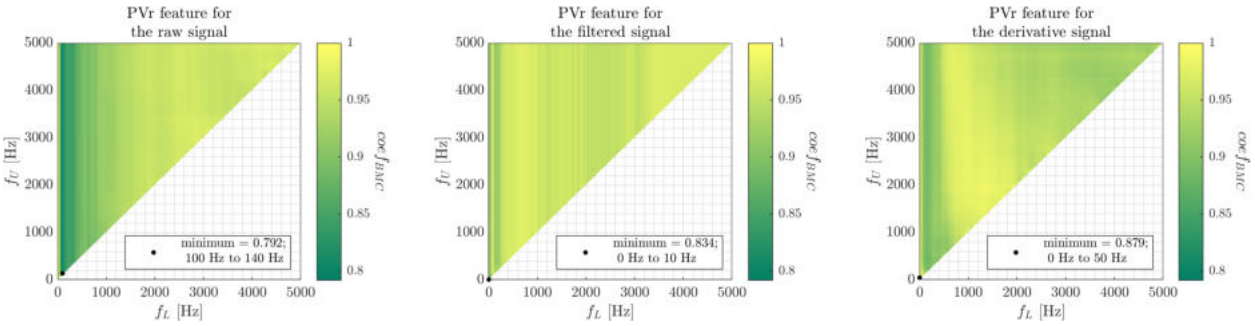

Figure 11.4: Frequency Range Map of the PVr feature.

Peak value of the self-normalized PSD, or the Spectral Crest Factor, scored the lowest $coef_{BMC}$ values for the raw signal, calculated for almost the whole frequency spectrum. Achieving $coef_{BMC}$ of around 0.81, it proved to be not very informative, compared to non- or relative-normalized parts of the spectrum. In all thee peak value FRMs, in the raw and

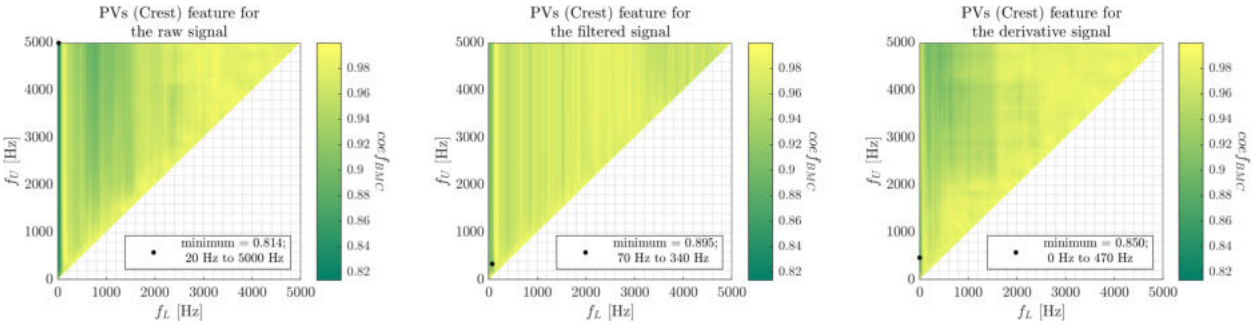

Figure 11.5: Frequency Range Map of the PVs (Spectral Crest Factor) feature.

filtered preprocessing approaches, vertical lines similar to the F_{max} feature can be found. They can be attributed to the same causes, i.e., the general decrease of power in higher frequencies of the VAG signal. Also, in all four features, i.e., the F_{max} and differently defined peak values at this frequency, lower frequency ranges proved to be much more informative than the higher ones. This effect is the most pronounced in the non-normalized PSD.

Boxplot of the most informative feature, i.e., the Peak Value of the non-normalized PSD in the 230-270 Hz range calculated for the filtered signal, was provided in Figure 11.6. Clearly, power in this range increases with progressing knee joint condition. Similarly to time-domain features, the most difficult classes to differentiate are healthy and cmp1, and the cmp3 and oa.

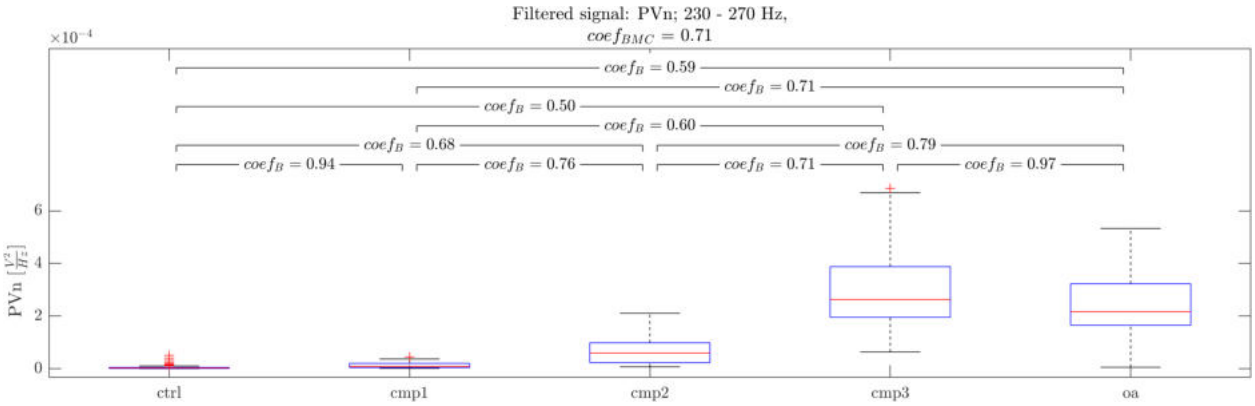

Figure 11.6: Boxplot of the Peak Value of the non-normalized Power Spectral Density feature calculated for the filtered signal.

Figure 11.7 includes Frequency Range Maps for the Spectral Median feature. Again, similar ranges proved to be the most informative in all preprocessing approaches, with the lower frequency ranges being more informative compared to the higher ones. Derivative signal proved to allow extraction of the most informative feature, with the $coef_{BMC}$ of around 0.74. Similar maps, included in Figure 11.8, were created using the Spectral Centroid feature. It proved to be slightly more informative compared to the Spectral Median, but nevertheless, it provided very similar results. It is not really surprising, given their definitions.

Specific values of both Spectral Median and Spectral Centroid in the most informative frequency ranges, in the form of boxplots, were included in Figures 11.9 and 11.10, respec-

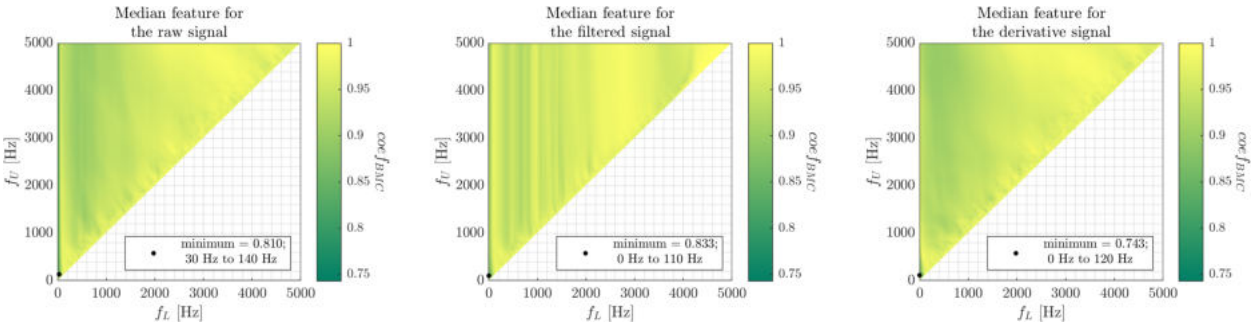

Figure 11.7: Frequency Range Map of the Spectral Median feature.

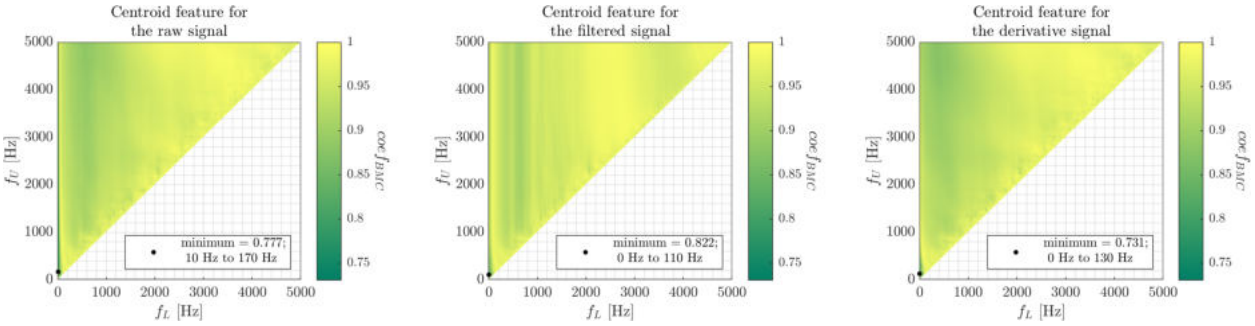

Figure 11.8: Frequency Range Map of the Spectral Centroid feature.

tively. With progressing knee joint conditions, the Spectral Median increases, indicating that in the 0-120 Hz range, higher frequencies are more pronounced in the more degenerated knee joints. This phenomenon was emphasized by the differencing operation, allowing power in lower and higher frequencies to be more comparable to each other. Very similar trend can be observed in the boxplot of the Spectral Centroid feature. Note, however, that the specific frequencies are slightly different. For example, the median value of the Spectral Median for the control class is about 20 Hz, while the Spectral Centroid is above 30 Hz. This can be caused by the fact, that the boxplot of the Spectral Centroid was obtained for slightly wider range. However, it could also indicate non-zero Spectral Skewness of the given frequency range.

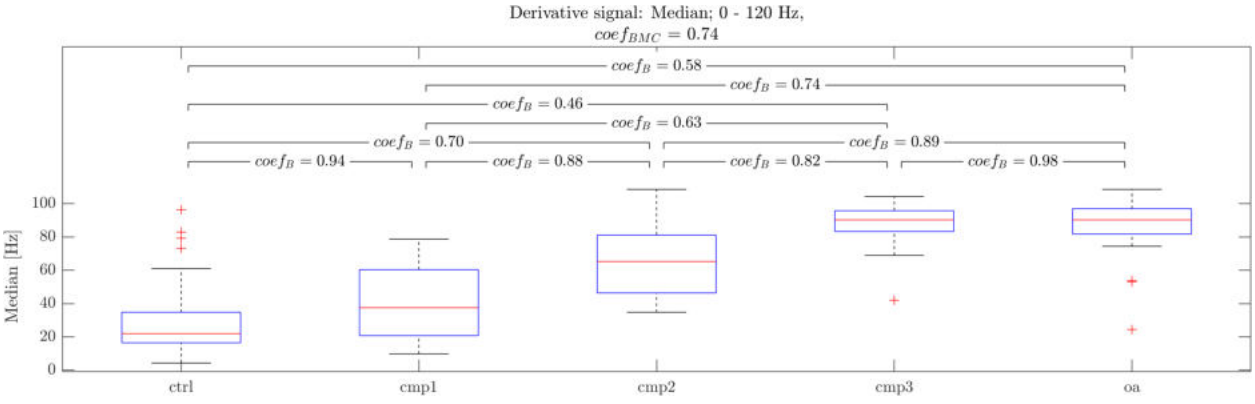

Figure 11.9: Boxplot of the Spectral Median feature calculated for the derivative signal.

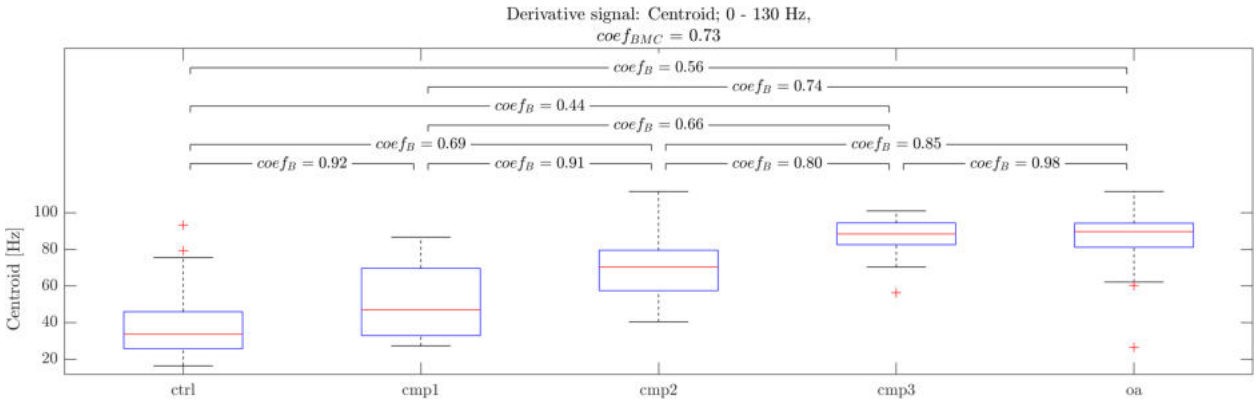

Figure 11.10: Boxplot of the Spectral Centroid feature calculated for the derivative signal.

The Spectral Spread around the Centroid was used to generate FRMs included in Figure 11.11. Both differentiation and filtration seem to significantly decrease informativeness of the signal, with the most informative feature being calculated for the 0 to 360 Hz range of the raw signal. Boxplot of this feature was provided in Figure 11.12. Clearly, spread increases with progressing knee joint conditions, indicating that the distribution of the spectral components widens, suggesting greater variability and dispersion in the signal's frequency in the 0 to 360 Hz range.

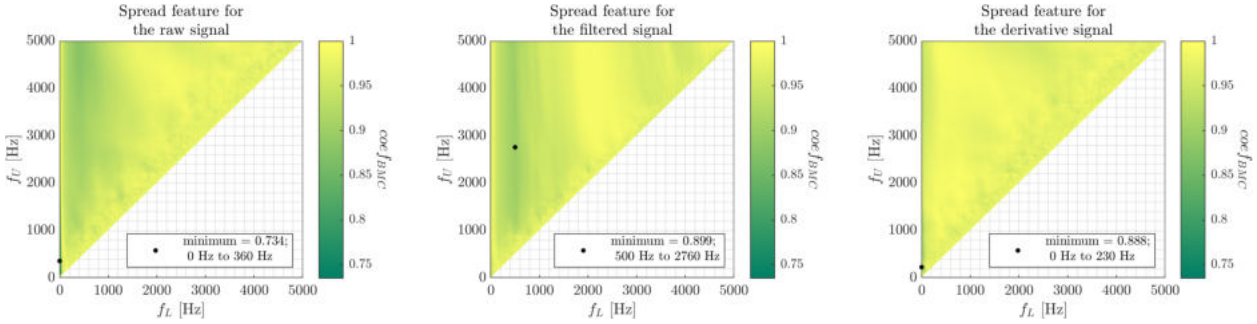

Figure 11.11: Frequency Range Map of the Spectral Spread feature.

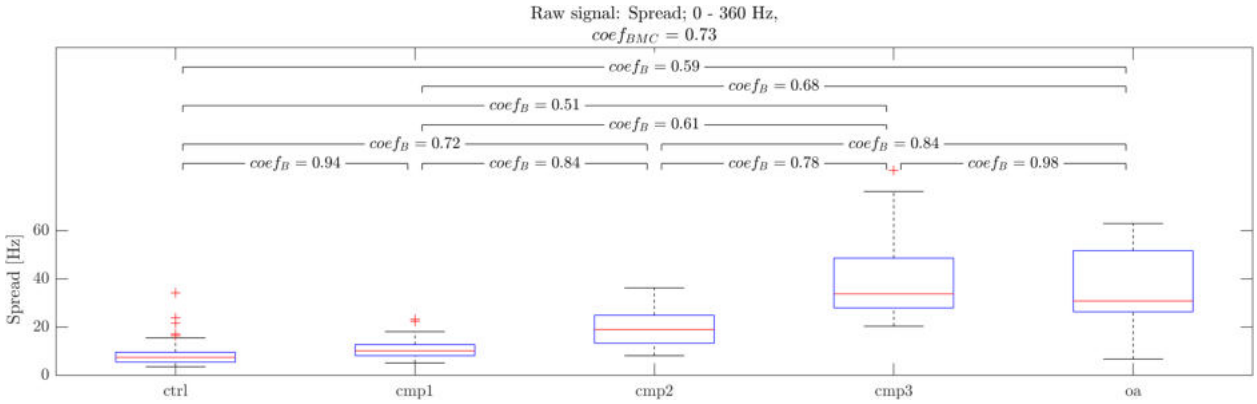

Figure 11.12: Boxplot of the Spectral Spread feature calculated for the raw signal.

Frequency Range Maps generated using the Spectral Skewness feature were included in Figure 11.13. Interestingly, patterns of informativeness seem similar to those of the Spectral Median and Centroid. The most informative preprocessing approach proved to be differentiation, with the same range as the Median, i.e., from 0 to 120 Hz. Boxplot of this feature was included in Figure 11.14.

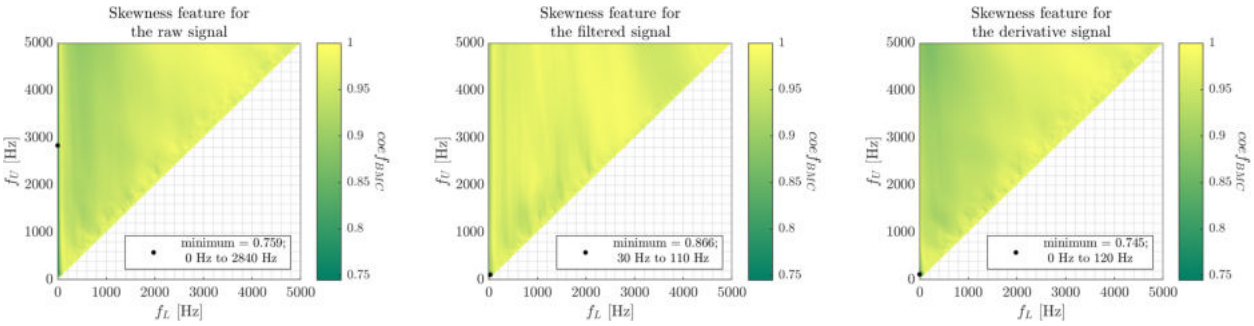

Figure 11.13: Frequency Range Map of the Spectral Skewness feature.

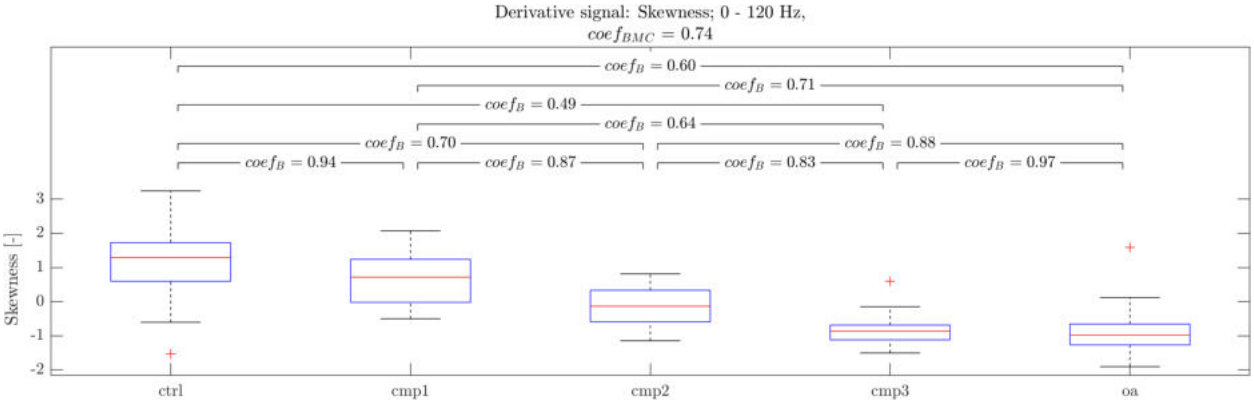

Figure 11.14: Boxplot of the Spectral Skewness feature calculated for the derivative signal.

Spectral Skewness seem to decrease with progressing knee joint conditions. For the control group, it proved to be mostly positive (median above positive one), as expected from the discrepancy between the Spectral Median and Centroid mentioned earlier. For the oa class, however, it proved to be negative (median around negative one). Note, that both Spectral Median and Centroid increased with progressing conditions. Apparently, however, the rates of those increases differed: Spectral Centroid increased slower than the Spectral Median, rendering the Spectral Skewness negative. It indicates that with progressing knee joint conditions, frequency range from 0 to 120 Hz in the derivative signal, seem to have differently distributed power, shifting more towards the higher frequencies.

Frequency Range Maps of the Spectral Kurtosis feature were provided in Figure 11.15. Compared to other features, it proved to be moderately informative. It indicates that the "flatness" of the PSD does not vary greatly across the knee joint conditions. Greatest differentiation was obtained in the 0 to 330 Hz range, for the raw signal, scoring $coef_{BMC}$ of around 0.81. Interestingly, in the derivative signal's FRM, increase in informativeness can be seen for wide frequency range, i.e., in the upper-left part of the map.

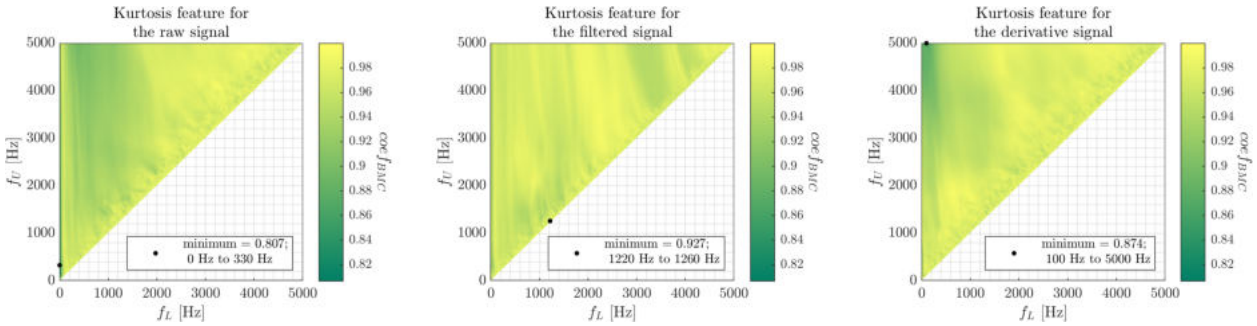

Figure 11.15: Frequency Range Map of the Spectral Kurtosis feature.

Barplot summary of the power features' informativeness was included in Figure 11.16. In general, they seem more informative compared to the distribution features, with the lowest $coef_{BMC}$ of around 0.69 being scored by the geometric mean.

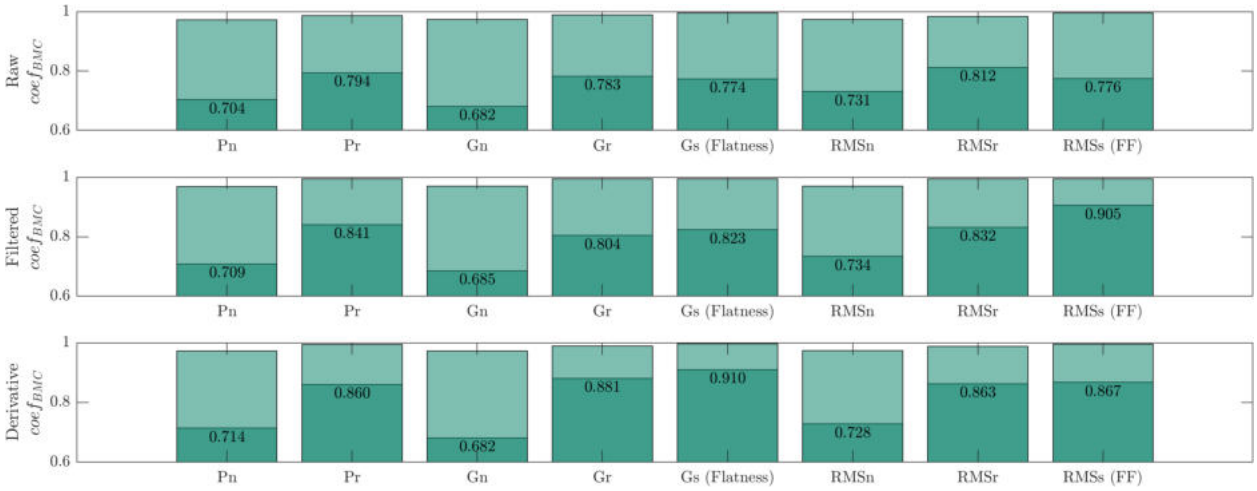

Figure 11.16: Informativeness of the Frequency Range Map power features.

Figure 11.17 includes Frequency Range Maps of the Average Power feature calculated for the non-normalized PSD. The most informative feature was extracted from the raw signal, scoring $coef_{BMC}$ of around 0.7 for the relatively wide range of 80 to 1780 Hz. However, in general, lower frequency ranges seem more informative than higher ones, with the lower part of the map being visibly darker compared to the right part. This pattern is clear in both raw and filtered signals. Derivative signal, on the other hand, seem to produce highly informative features only in lower-left part of the map. Since power of the VAG signals is, in general, concentrated in lower frequencies of the spectrum, features calculated for the raw and filtered signals are similarly informative when they include lower frequencies, with inclusion of higher frequencies being of little difference to the information content, at least in the given context. Since difference works like a high-pass filter, it can be observed that those higher frequencies do not carry much information, even when amplified.

Boxplot of the most informative non-normalized Average Power feature, i.e., calculated for the raw signal in 80 to 1780 Hz frequency range, was included in Figure 11.18. Clearly, power increases with progressing knee joint conditions, being close to 0 for most signals from

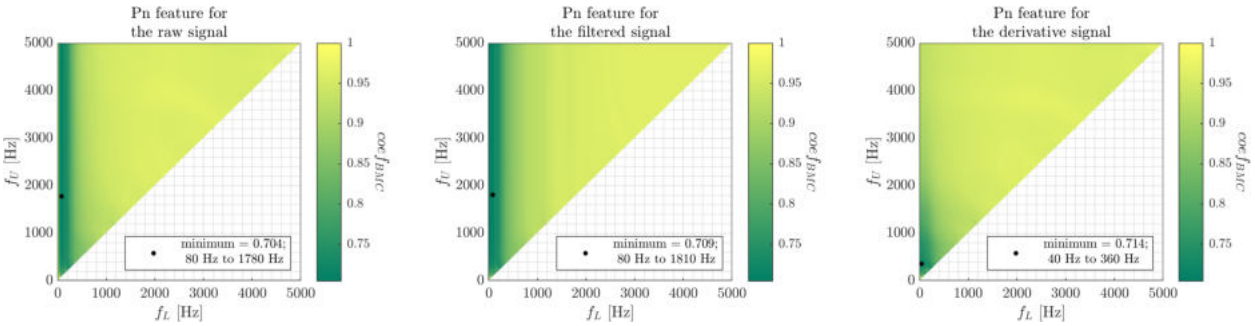

Figure 11.17: Frequency Range Map of the Average Power (non-normalized) feature.

control group, and achieving the highest values for both cmp3 and oa classes. Note, however, that difference between those classes is very low, with $coef_B$ of 0.95.

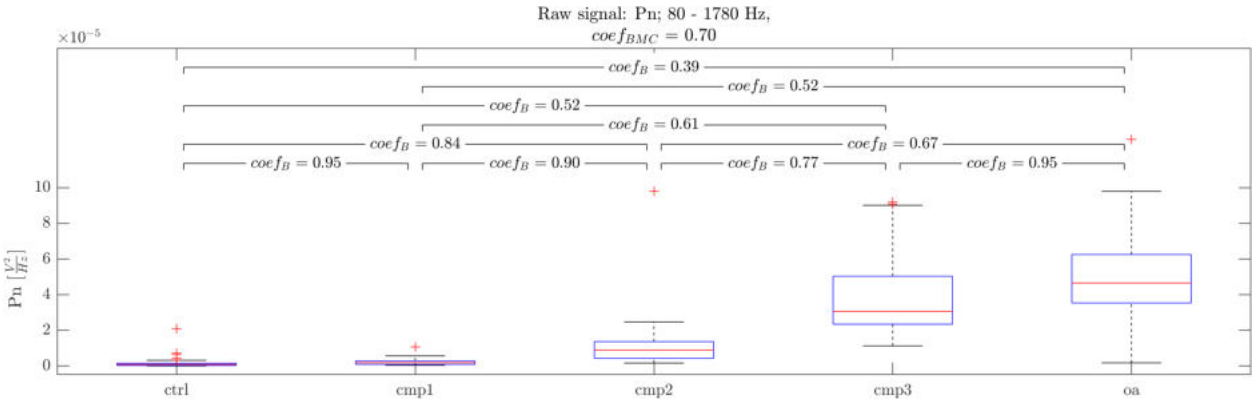

Figure 11.18: Boxplot of the Average Power of the non-normalized Power Spectral Density feature calculated for the raw signal.

Frequency Range Maps calculated for the Average Power of the relative-normalized PSD were included in Figure 11.19. Informativeness of the relative-normalized power features is generally lower than their corresponding non-normalized versions. Again, the most informative feature was extracted for the raw signal, with the $coef_{BMC}$ of around 0.79. Interestingly, for the filtered signal, range of 0 to 10 Hz proved to be the most effective discriminator between classes. Note, that this range is below pass-band of the filtering (from 50 to 1000 Hz). It may seem counterintuitive that, after the filtration, any frequencies outside the pass-band contain any information at all. Note, however, that the filtering does not *remove* frequency components outside the pass-band completely, just reduces their amplitude (see upper part of Figure 4.1).

Map generated for the derivative signal seem to be the most different to its non-normalized equivalent. Specifically, whole left part of the map seem informative, not only the lower-left. Observed patterns, however, can be explained similarly. In the non-normalized PSD, higher, non-informative frequencies were largely influencing the average power feature, rendering it uninformative. In the relative-normalized PSD, however, their influence was reduced by the normalization. In other words, frequency ranges containing highly-informative low frequencies keep their information content even after being "contaminated" with now high-amplitude

higher frequencies, when being considered relatively to the whole spectrum. Nevertheless, the most informative feature calculated for the derivative signal can be considered rather uninformative, scoring $coef_{BMC}$ of 0.86.

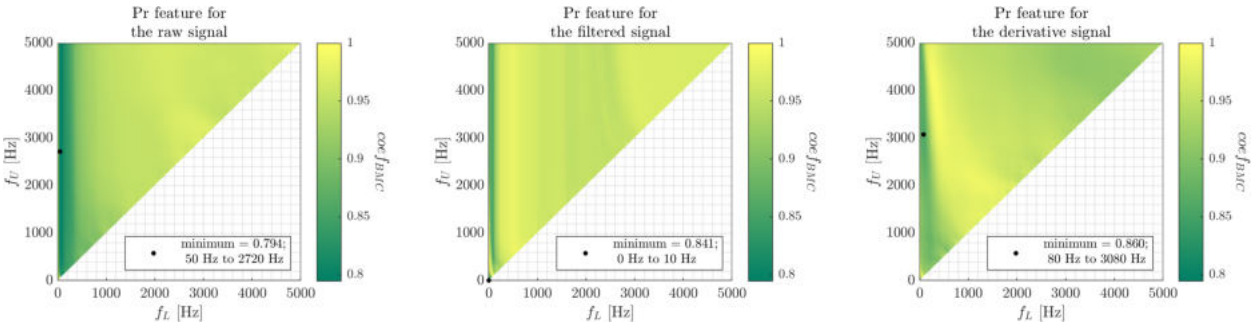

Figure 11.19: Frequency Range Map of the Pr feature.

In Figure 11.20, Frequency Range Maps for the geometric mean feature of the non-normalized PSD were included. All three maps look very similar, with the lower-left part of the map being the most informative. The geometric mean, defined as the n-th root of the product of n power values, has a great sensitivity to low values, especially those approaching zero. It means that even slight variations in the low-amplitude power values can significantly influence its outcome. On the other hand, while values of the higher-amplitude power contribute to its increase, they do not greatly skew it, rendering the geometric mean relatively robust against high amplitude outliers.

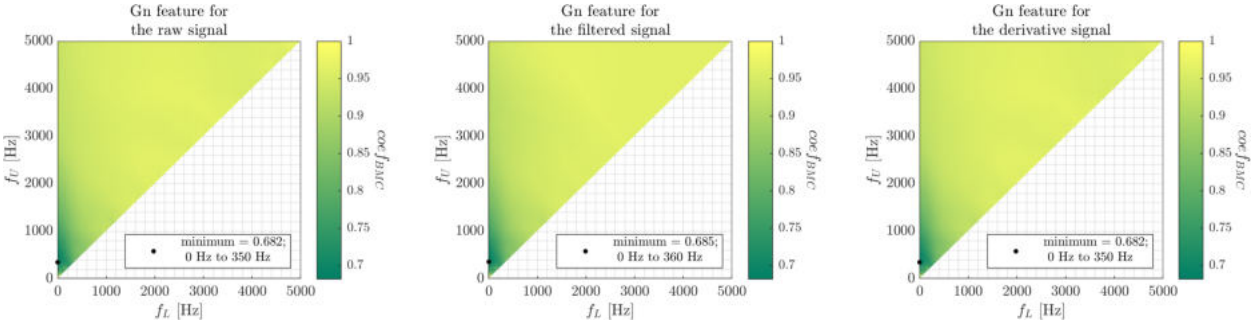

Figure 11.20: Frequency Range Map of the Gn feature.

The consistent informativeness of the feature across the preprocessing approaches can indicate that it captures robust characteristics of the VAG signal in the frequency domain. Moreover, the captured features are consistently contained in the lower frequency ranges, generally below 1 kHz. High-frequency low-amplitude parts of the PSD proved to be uninformative in the given context. It might seem that it can be attributed to the definition of geometric mean and its sensitivity to low-amplitude values. When just one frequency in the calculated range has a power of 0, the geometric mean itself also becomes zero, rendering it extremely uninformative for the ranges containing zero or near-zero values. However, even narrow frequency ranges (closer to the diagonal than to the y-axis) proved to contain low information content. Therefore, it can be concluded, that the higher frequencies generally

do not contain much information specific to classes analyzed in this work, at least when their absolute, i.e., non-normalized values are considered.

After the normalization, some higher frequencies became slightly more informative. Frequency Range Maps calculated for the geometric mean using relatively-normalized and self-normalized PSD were provided in Figures 11.21 and 11.22, respectively. Similarly to the arithmetic mean, the relatively-normalized maps of the geometric mean seem generally much less informative compared to non-normalized ones, especially in lower, more informative frequency ranges. Similarly to their non-normalized equivalents, all three maps seem to be much less informative in higher frequency ranges. While the difference in informativeness between lower and higher frequency ranges seem lower in the geometric mean calculated for the relative-normalized PSD, higher frequencies can still be considered rather uninformative.

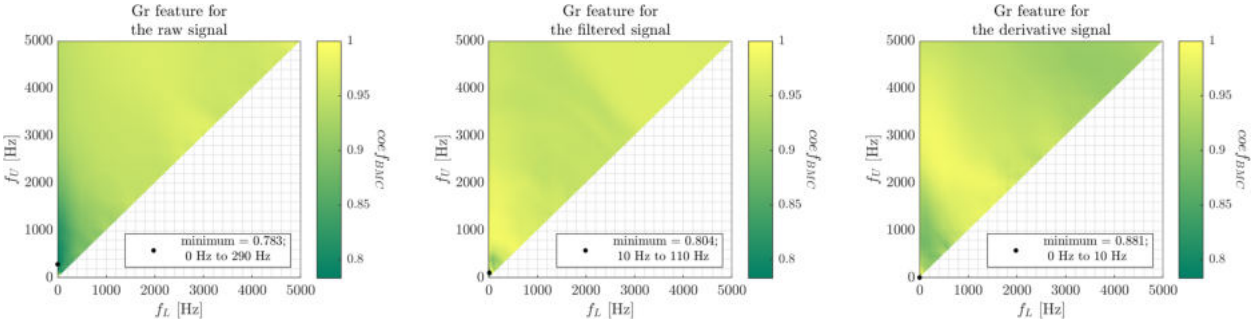

Figure 11.21: Frequency Range Map of the Gr feature.

Interestingly, in case of the self-normalized spectrum, FRMs for which were included in Figure 11.22, wider frequency ranges, i.e., upper-left parts of the FRMs, proved to be of relatively high information content. Note, however, that values of $coef_{BMC}$ in those regions are still rather high. This pattern is mostly visible in the raw signal, in ranges with lower frequency of around 200 Hz to 1 kHz and upper frequency above around 2 kHz. When the frequency range is self-normalized, values in ranges that do not include high-amplitude low-frequency components are not so much diminished, rendering the geometric mean less contaminated by low-amplitude noise and more informative in the context of this work.

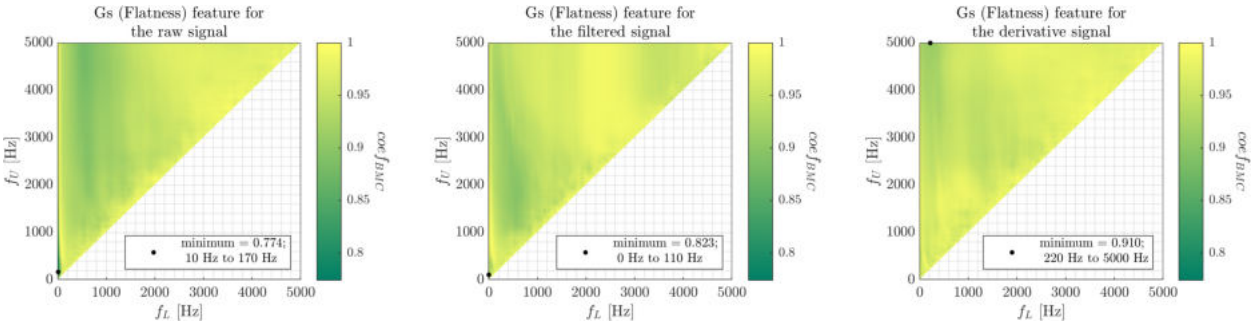

Figure 11.22: Frequency Range Map of the Gs (Spectral Flatness) feature.

Boxplot of the most informative geometric mean feature, i.e., the one calculated using non-normalized PSD for the raw signal in the 0 to 350 Hz frequency range, was provided in Figure 11.23.

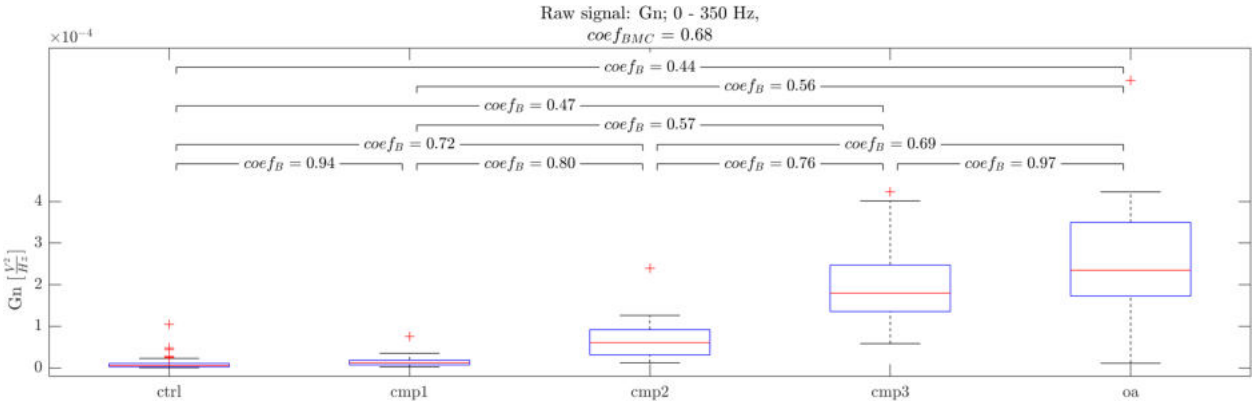

Figure 11.23: Boxplot of the Geometric Mean Power of the non-normalized Power Spectral Density feature calculated for the raw signal.

Clearly, power of the VAG signal in 0 to 350 Hz frequency range increases with progressing knee joint conditions, being close to zero in the control group, and above 2 $\frac{V^2}{Hz} \cdot 10^{-4}$ in the oa group. It indicates, that with more degenerated knee joints, PSD in this range *consistently* increases, i.e., not in just some sub-range of this frequency, but rather in most frequency components included in this range.

The FRMs for Root Mean Square values of the non-normalized, relatively-normalized, and self-normalized PSD were included in Figures 11.24, 11.25, and 11.26, respectively. Since calculation of the RMS includes squaring operation, it is sensitive to large-amplitude power values: large values squared become larger, while small values become even smaller. In this regard, the RMS can be considered opposite of the Geometric Mean. In the FRMs generated for the non-normalized PSD, patters seem similar to those found in the Average Power (see Figure 11.17). In both raw and filtered signals, small-amplitude high-frequency components do not contribute much to the value of RMS and therefore informativeness is mostly dependent on the high-amplitude low-frequency components, causing visible vertical lines of high informativeness. After the differentiation, high-frequency components are amplified, and including them in the calculation of the RMS can significantly decrease its information content, rendering wider frequency ranges less informative.

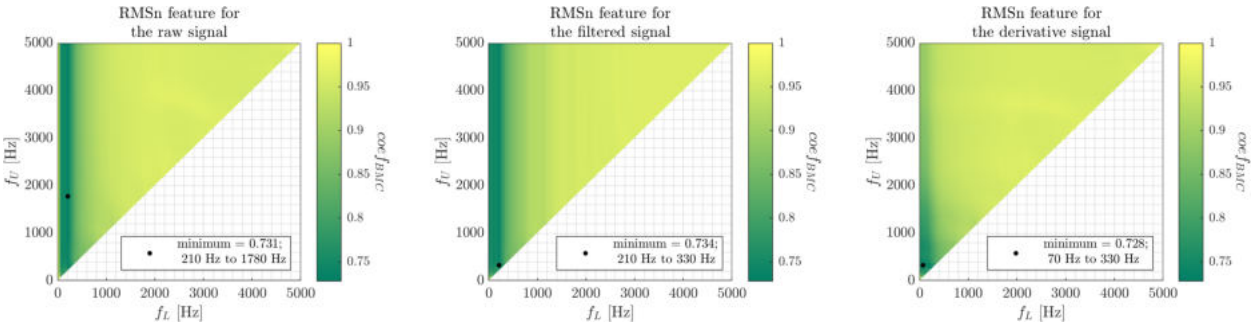

Figure 11.24: Frequency Range Map of the RMSn feature.

In the FRMs calculated using RMS of the relatively-normalized PSD (Figure 11.25), the vertical informativeness pattern emerges additionally in the derivative signal. After

the normalization, frequency ranges containing higher frequencies (upper part of the map) become more informative compared to the non-normalized PSD. It indicates that for the derivative signal, power of the wider frequency ranges is more informative in the context of the entire spectrum. Nevertheless, informativeness of the RMS feature in the relative-normalization setup proved to be rather low for all preprocessing approaches.

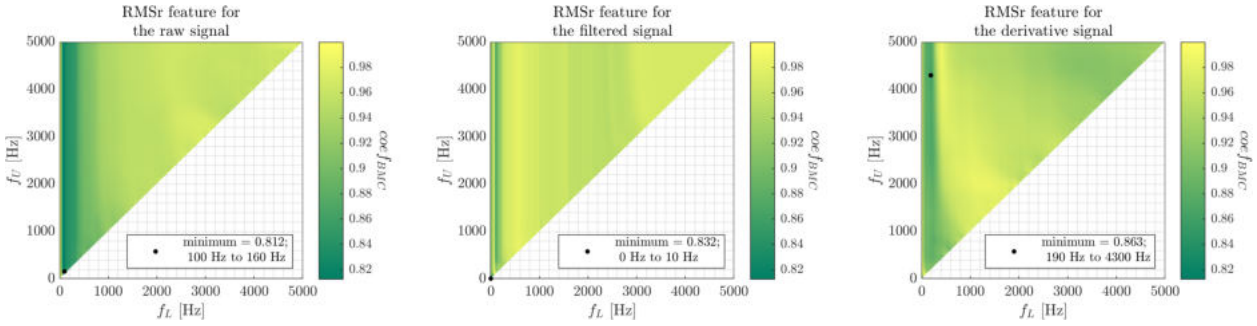

Figure 11.25: Frequency Range Map of the RMSr feature.

Even smaller informativeness was observed in the self-normalized FRMs (Figure 11.26), with an exception for the raw signal, whose most informative feature proved to achieve a little lesser $coef_{BMC}$ value compared to its relative-normalized version.

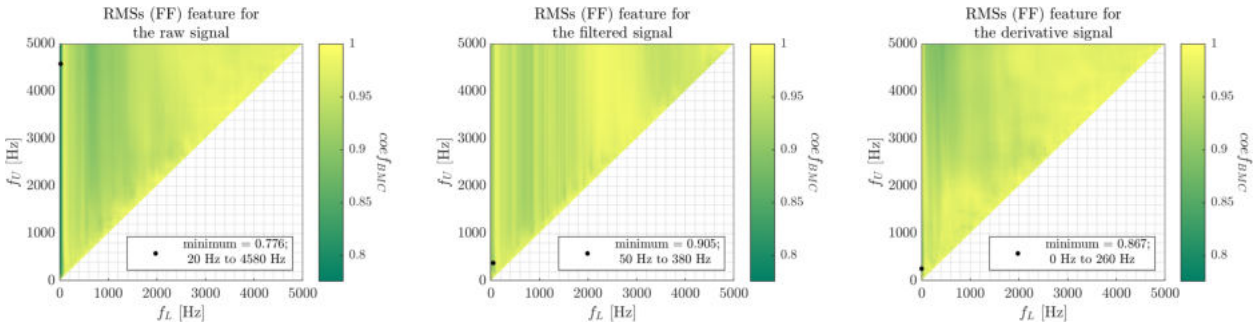

Figure 11.26: Frequency Range Map of the RMSs (Spectral Form Factor) feature.

Informativeness analysis of the power features in this work indicated that normalization of the Power Spectral Density estimate generally seem to reduce informativeness of the features, for all preprocessing approaches. It is rather unexpected result, given that in earlier frequency analysis of the VAG signals conducted by Befrui et al. [75], normalization of the spectrum led to significantly increased informativeness. It can be explained by different frequency analysis approach utilized in [75], in which multiple spectra were obtained for the signal divided into segments, and then averaged.

Barplot summary of the PSD shape features' informativeness was provided in Figure 11.27. The most informative feature proved to be Spectral Slope, which FRMs were included in Figure 11.28.

Across the preprocessing approaches, the Spectral Slope proved to be the most informative when lower frequencies were included in its calculation. See significantly lower $coef_{BMC}$ values in the left part of all maps, the most clearly visible in raw and filtered

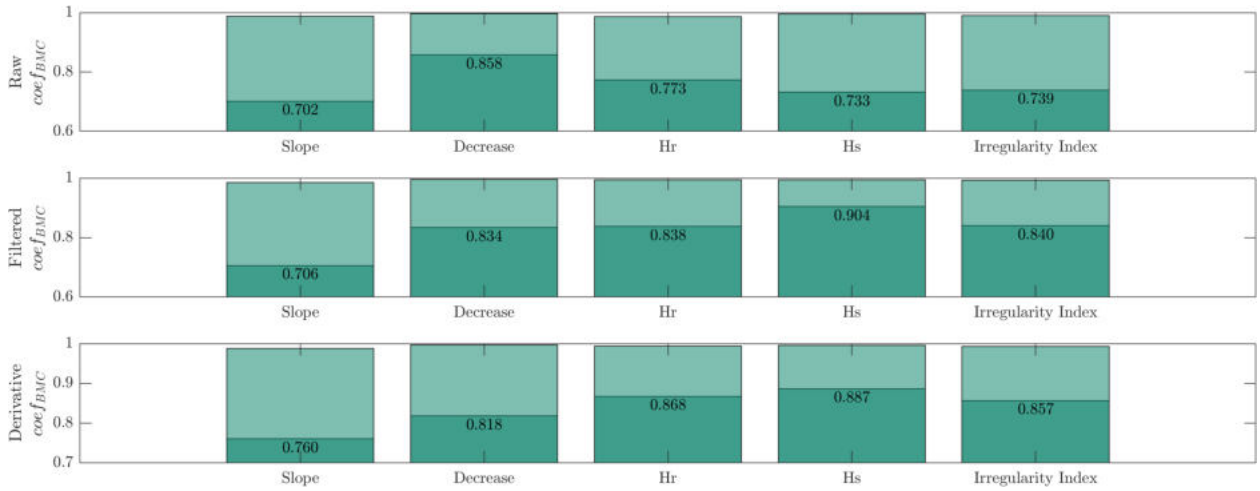

Figure 11.27: Informativeness of the Frequency Range Map shape features.

signals. Above around 500 Hz of the higher frequency, informativeness does not change significantly, as long as the lower frequency is kept below around 200 Hz. Preprocessing did not significantly altered this pattern. However, derivative signal provided the least informative features. It can be explained by its amplification of higher frequencies, which, when measuring the Spectral Scope, lowered informativeness. Boxplot of the most informative Spectral Scope was provided in Figure 11.29.

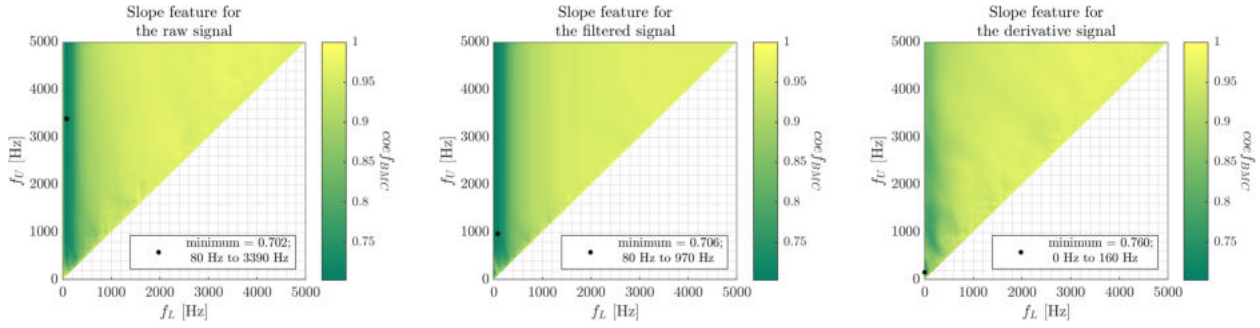

Figure 11.28: Frequency Range Map of the Spectral Slope feature.

Clearly, Spectral Scope in the 80 to 3390 Hz decreases significantly with progressing knee joint conditions. For the healthy and cmp1 classes, slope is slightly negative, but close to zero. It achieves the lowest values for both cmp3 and oa classes, which can be explained by increased power of lower frequencies with progressing knee joint degeneration. Note, in the boxplot of the Average Power feature calculated for the raw signal in 80 to 1780 Hz range (Figure 11.18), that power measured in almost exactly half of the lower range of the Spectral Slope's range, proved to increase with progressing conditions.

Feature similar to Spectral Slope, i.e., the Spectral Decrease, proved to be much less informative, with lowest $coef_{BMC}$ of around 0.82 achieved by the derivative signal. Its Frequency Range Maps were provided in Figure 11.30. See, that similar pattern of high informativeness emerged for the features calculated including lower frequencies. However,

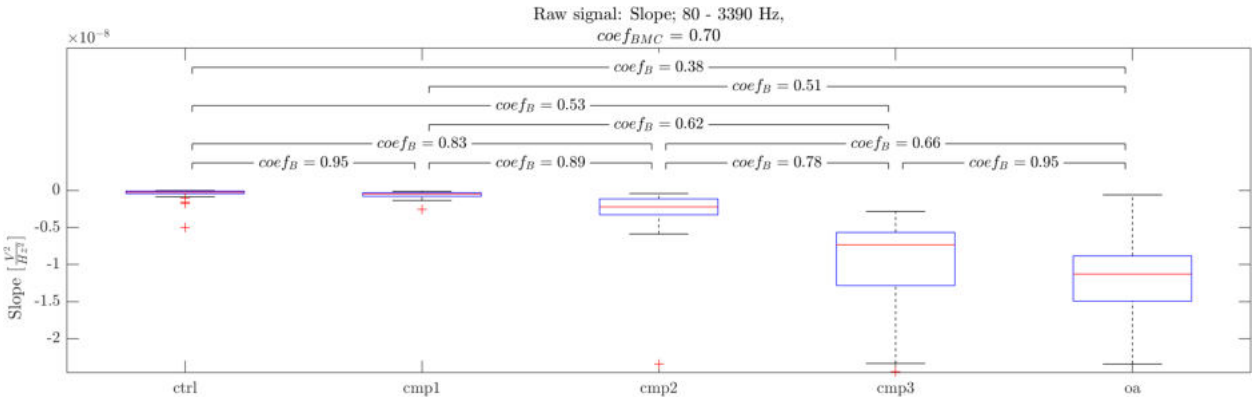

Figure 11.29: Boxplot of the Spectral Slope feature calculated for the raw signal.

compared to the Spectral Slope, relatively high informativeness was achieved, when the lower frequency of the range was not greater than around 100 Hz.

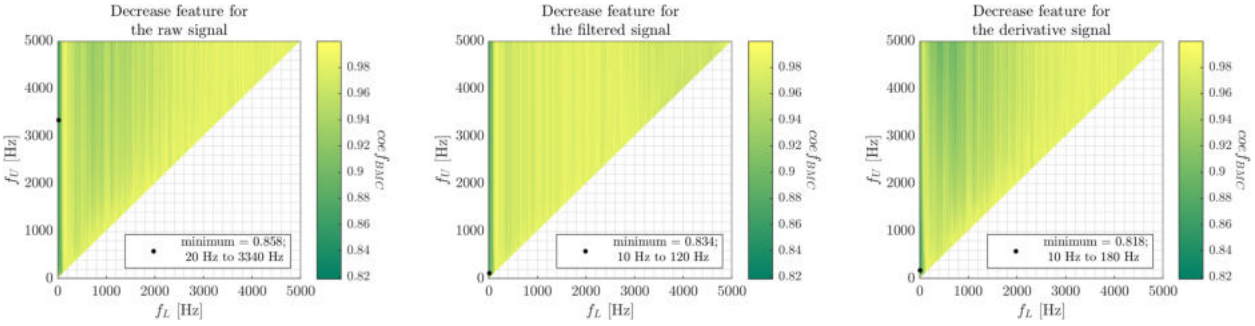

Figure 11.30: Frequency Range Map of the Spectral Decrease feature.

Frequency Range Maps of the Spectral Entropy feature calculated for the relatively-normalized PSD were provided in Figure 11.31. Raw signal's map show similar patterns to previous features, that is, relatively high informativeness of ranges that include low frequencies, i.e., with lower frequency up to about 300 Hz. Both filtered and derivative signals proved to be much less informative, with the lowest $coef_{BMC}$ well above 0.8. Boxplot of the feature with the lowest $coef_{BMC}$ was provided in Figure 11.32.

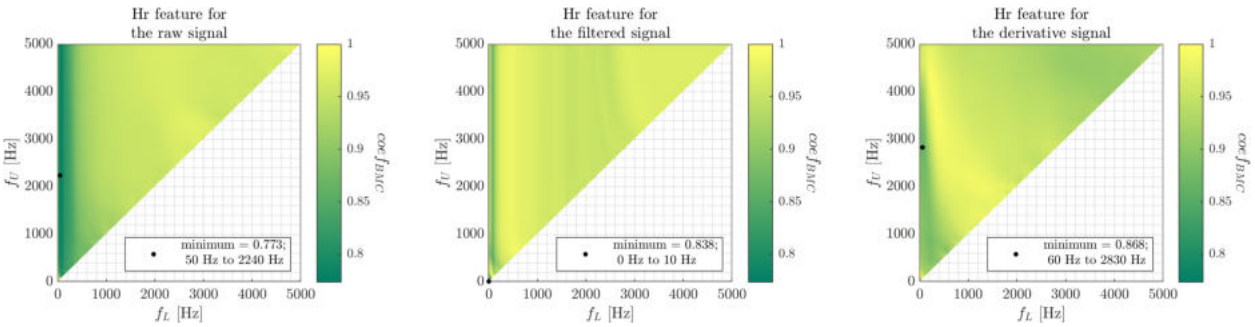

Figure 11.31: Frequency Range Map of the Hr feature.

Entropy of the relatively-normalized spectrum, measured in the 50 to 2240 Hz frequency

range for the raw signal, seem to increase with progressing knee joint conditions. It means that frequencies in this range become more evenly spread out, when the power values are normalized relatively to the whole PSD. It might seem contradictory to previous analyzes, which clearly showed that power increases in lower frequencies, which intuitively should decrease the entropy: since power of the lower frequencies is greater, the signal should become more predictable, or less uncertain. However, power in the higher frequencies could influence the entropy measure by becoming more spread out along the frequencies, while still allowing decrease in Spectral Slope and increase in low-frequency power.

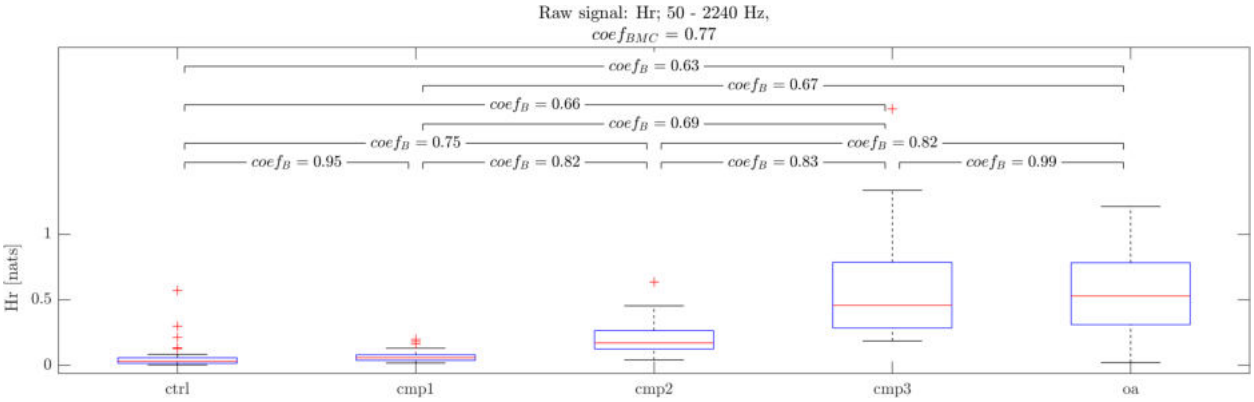

Figure 11.32: Boxplot of the Spectral Entropy of the relative-normalized Power Spectral Density feature calculated for the raw signal.

Informativeness of the self-normalized PSD frequency ranges' entropy was plotted in Figure 11.33. Clearly, in most part of all preprocessing approaches informativeness significantly decreased in comparison to the relative-normalization. One exception is the raw signal. Again, inclusion of low frequencies, i.e., ranges with lower frequency not higher than around 20 Hz, seem to allow extraction of the most informative features. The lowest $coef_{BMC}$ was achieved for the range of 10 to 2000 Hz, and its boxplot was provided in Figure 11.34. General tendency of the entropy calculated for the self-normalized range seem to be similar to the relative-normalized one, i.e., increasing with the progressing knee joint condition.

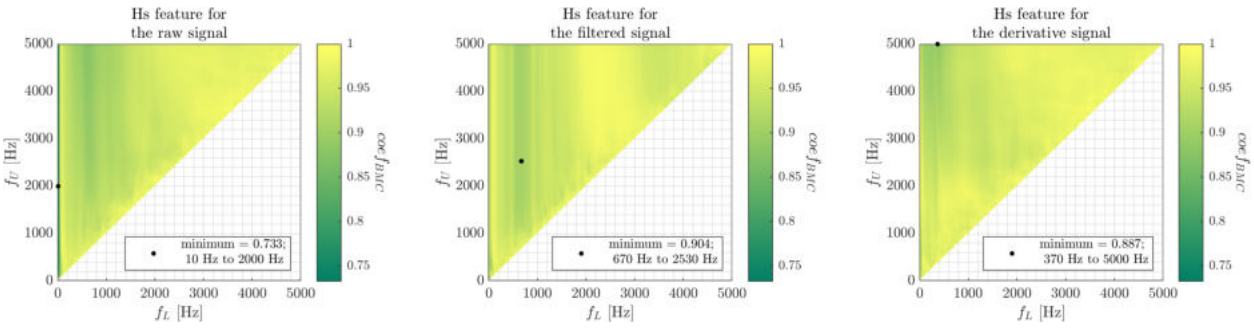

Figure 11.33: Frequency Range Map of the Hs feature.

Frequency Range Maps of the last entropy feature, i.e., the Irregularity Index, were provided in Figure 11.35. The Irregularity Index is defined as the summation of the entropy of

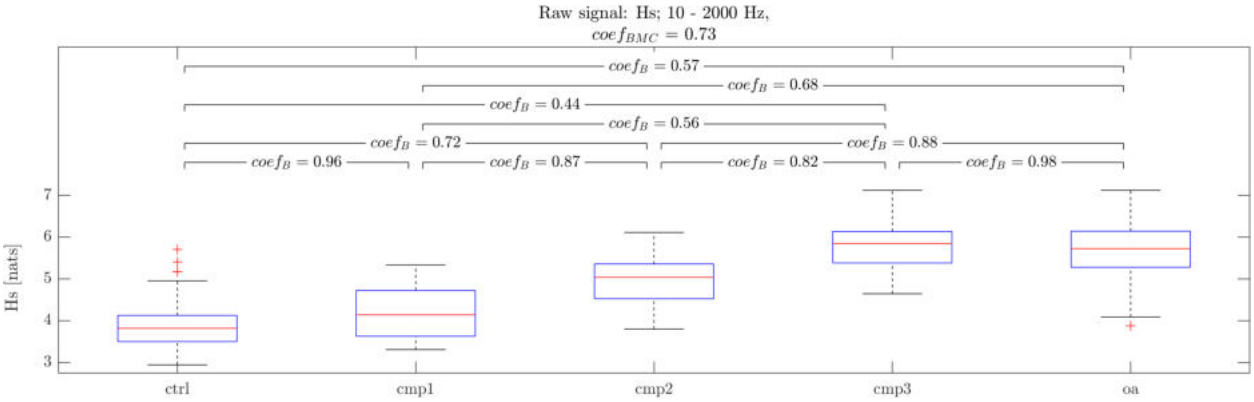

Figure 11.34: Boxplot of the Spectral Entropy of the self-normalized Power Spectral Density feature calculated for the raw signal.

the relative-normalized range, and scaled entropy of the self-normalized range, divided by the entropy of the entire PSD (see Table 6.3). General patterns observed in maps seem to closely follow patterns seen in the relative-normalized Spectral Entropy feature (Figure 11.31). It can be explained by the greatest influence of this part of the definition. Informativeness of the entire spectrum's entropy (see the most upper-left point in the FRMs from Figure 11.31) is quite low ($coef_{BMC}$ of around 0.94), which means that it does not consistently vary between analyzed classes, and has low influence on the informativeness of the Irregularity Index. Interestingly, in the raw signal, much narrower frequency range proved to be the most informative: from 0 to 40 Hz. Boxplot of the Irregularity Index calculated for this range was provided in Figure 11.36.

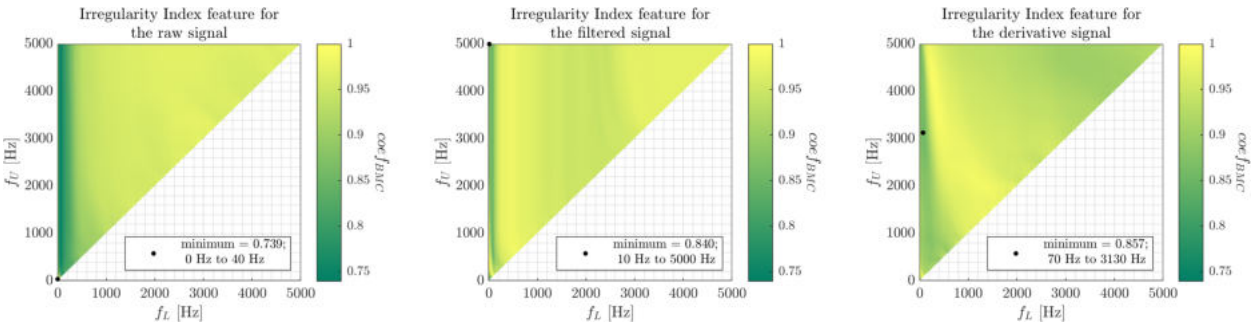

Figure 11.35: Frequency Range Map of the II feature.

Irregularity Index of the very low frequency range, i.e., from 0 to 40 Hz, seem to decrease with the progressing knee joint degeneration. It can be explained by the disproportionate increase in some specific frequency bin in the 0 to 40 Hz range, overshadowing other bins, and decreasing the overall uncertainty, measured by the Irregularity Index.

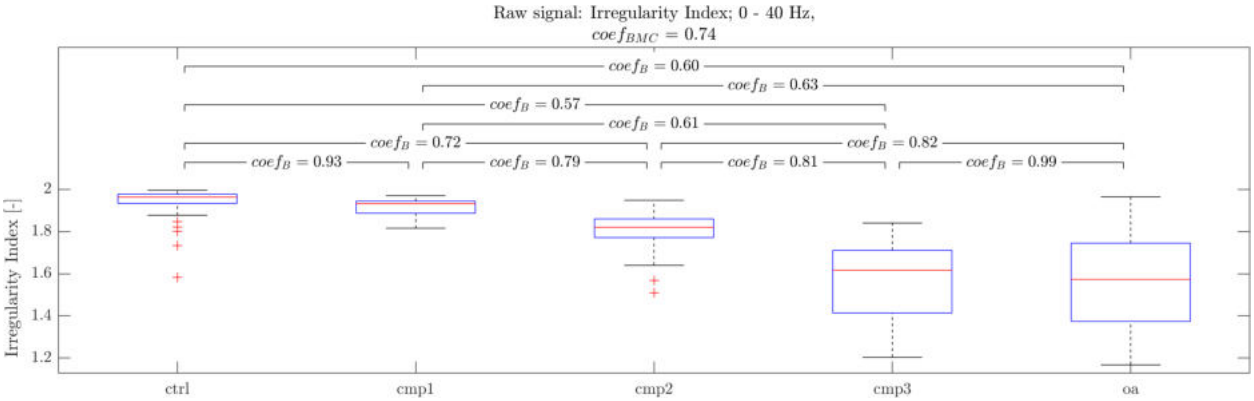

Figure 11.36: Boxplot of the Irregularity Index feature calculated for the raw signal.

11.3 Other frequency features

Figure 11.37 includes a barplot summary of the informativeness of frequency features, that were obtained not following the FRM methodology. In general, those features do not seem very informative. Detailed results of the ratio features were provided in Figure 11.38.

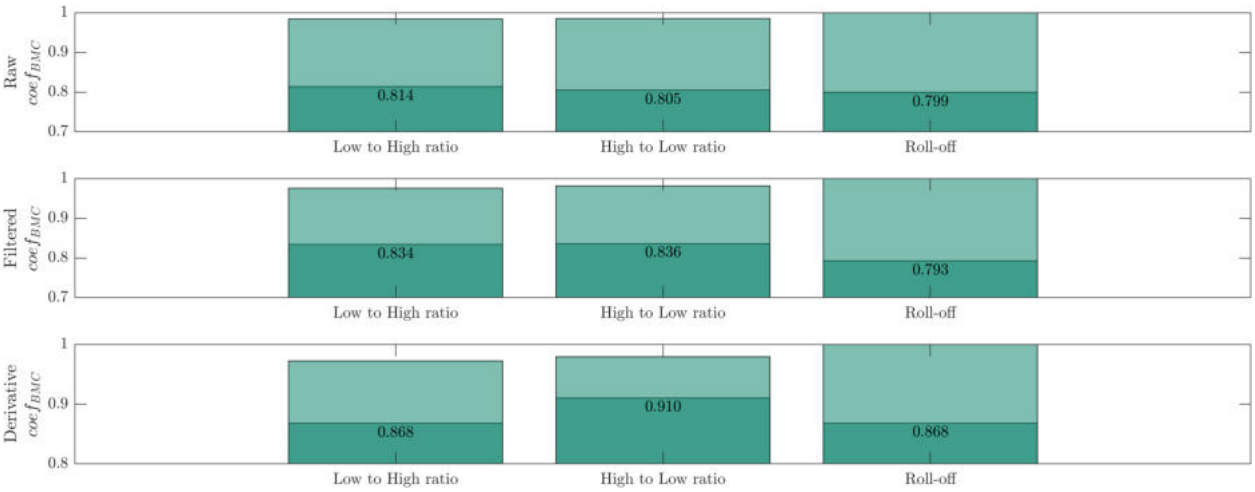

Figure 11.37: Informativeness of the frequency features not following the Frequency Range Map methodology.

Clearly, informativeness of both ratio features seem very correlated within all preprocessing approaches. The lowest $coef_{BMC}$ values were obtained for the high to low ratio of the raw signal, with the frequency threshold of around 40 to 80 Hz. Much less information proved to be extracted for the filtered and derivative signals. Boxplot of the most informative ratio was included in Figure 11.39.

As expected, with progressing knee joint conditions, ratio of frequencies above to frequencies below around 43 Hz decreases, indicating that power in the lower frequencies increases much faster compared to the higher frequencies. Note, that this ratio is greater than 1 for most signals in all classes, indicating that more power is contained in the upper part of the

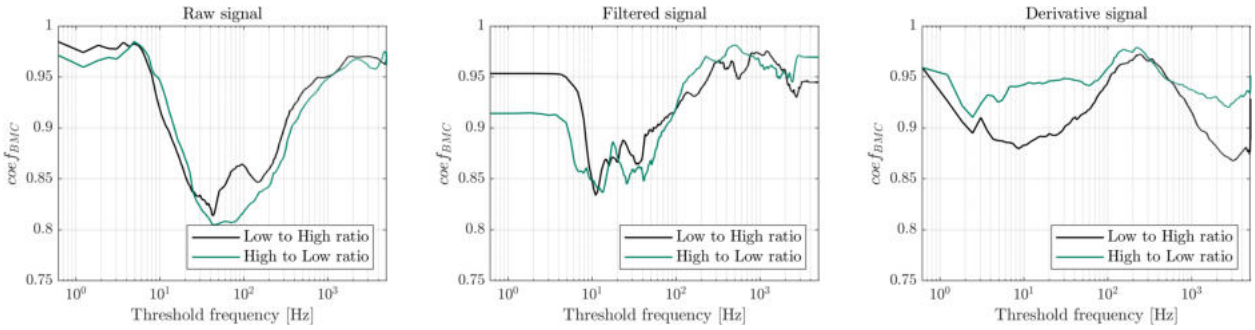

Figure 11.38: Detailed results of the frequency ratios' informativeness.

PSD, no matter the knee joint group. It can be explained by the summation used in the definition of the ratio. Since the threshold is relatively low compared to the Nyquist frequency, power of the PSD is greater above the threshold. Nevertheless, as stated previously, power in lower parts of the PSD seem to increase much faster compared to higher frequencies.

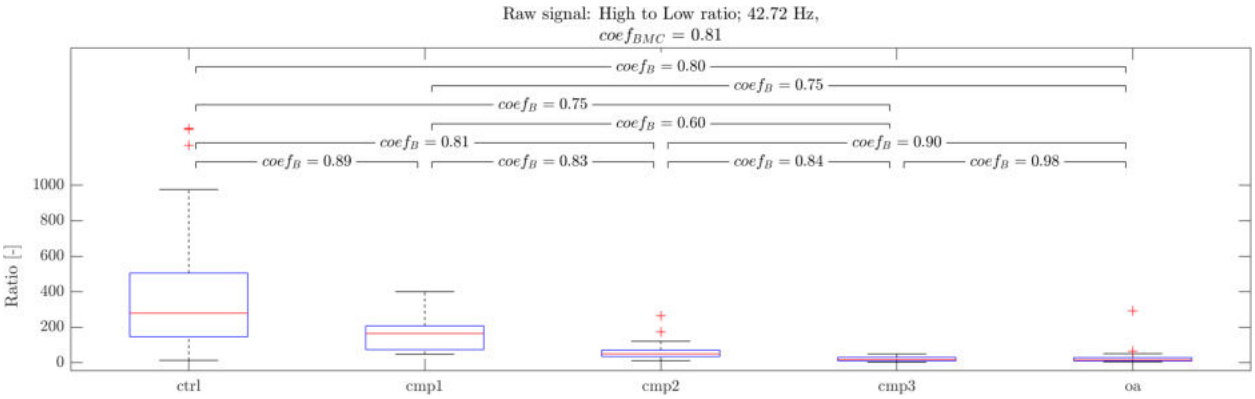

Figure 11.39: Boxplot of the High-to-Low ratio feature calculated for the raw signal.

Detailed results of the Spectral Roll-off features' informativeness were included in Figure 11.40.

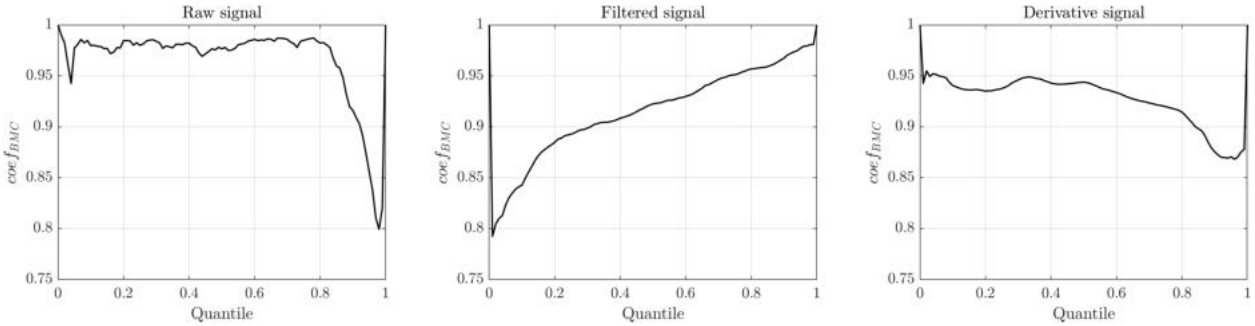

Figure 11.40: Detailed results of the frequency Roll-off Points informativeness.

In general, this feature proved to be of relatively low informativeness. For the raw signal, the lowest $coef_{BMC}$ was achieved for the 98-th percentile, which is very close to

95-th percentile used in the literature. It means, that the frequency below which most of the signal power is contained is a somewhat decent class discriminator. However, for the filtered signal, the dependence of the informativeness seem to follow the opposite trend. The lowest $coef_{BMC}$ was achieved for the second percentile, and increased for higher percentiles, achieving around 0.98 for the percentile that proved to be the most informative in the raw signal. Differentiation turned out to lower informativeness of the Spectral Roll-off feature. Boxplots of the 98-th and 2-nd percentiles of the raw and filtered signals were included in Figures 11.41 and 11.42, respectively.

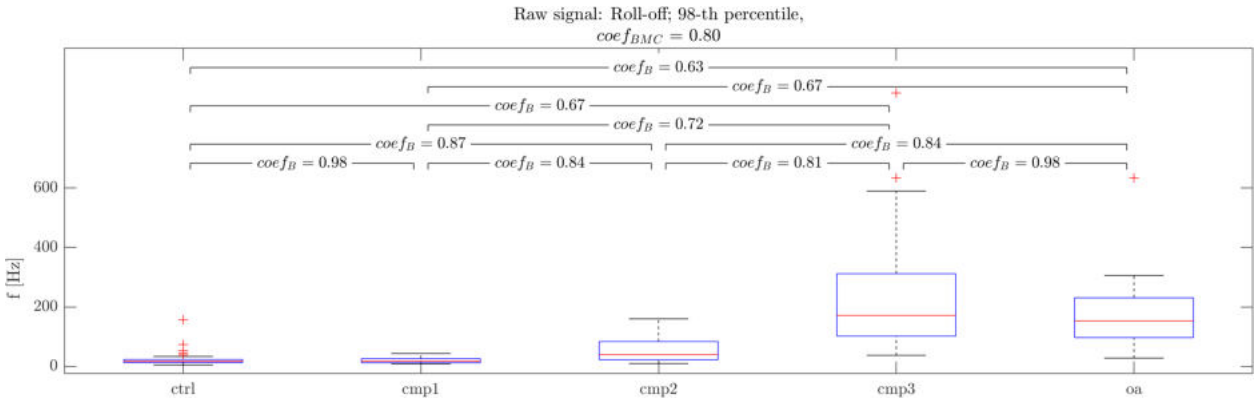

Figure 11.41: Boxplot of the Roll-off Point feature calculated for the raw signal.

For both raw and filtering preprocessing approaches, the Roll-off frequency seem to increase with progressing knee joint conditions. For the 98-th percentile of the raw signal (Figure 11.41), frequency increases from median value of around 20 Hz in control group to around 150 Hz in the oa. This result complements previous results, indicating that the power increases not *only* in lower frequencies, but along wider frequency range spectrum. Increase in the Roll-off point, along with the previous results suggests that the signal increases in power along wide frequency range, but lower ranges tend to increase more compared to higher ones.

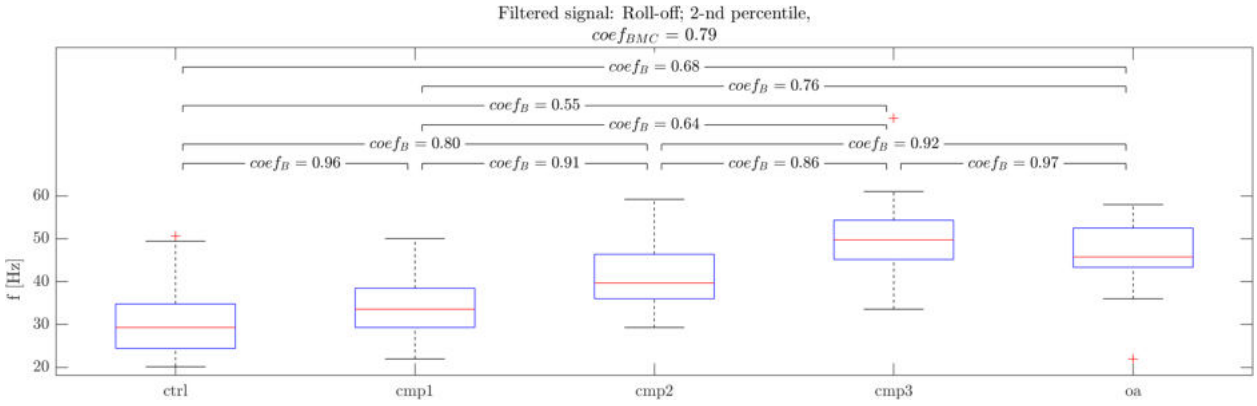

Figure 11.42: Boxplot of the Roll-off Point feature calculated for the filtered signal.

The Roll-off frequency calculated for the 2-nd percentile of the filtered signal seem to vary across studied groups much less. It indicates frequency, below which 2 percent of the

power spectrum is contained. It is expected, that this frequency is below 50 Hz for most classes, since the filter used in this work has relatively slow transitions between the pass-band and stop-band (see Figure 4.1). Its increase with the progressing knee joint degeneration confirms that power generally increases in lower frequencies faster than higher ones.

11.4 Correlations of the features

Correlations of the most informative feature in each FRM, as well as the most informative frequency ratios and Roll-off point, were provided in Figure 11.43. Only features obtained for the raw signal were analyzed in such way, since filtering and differentiating operations influence the Power Spectral Density in relatively predictable ways. Note, that most FRM features were extracted from different frequency ranges.

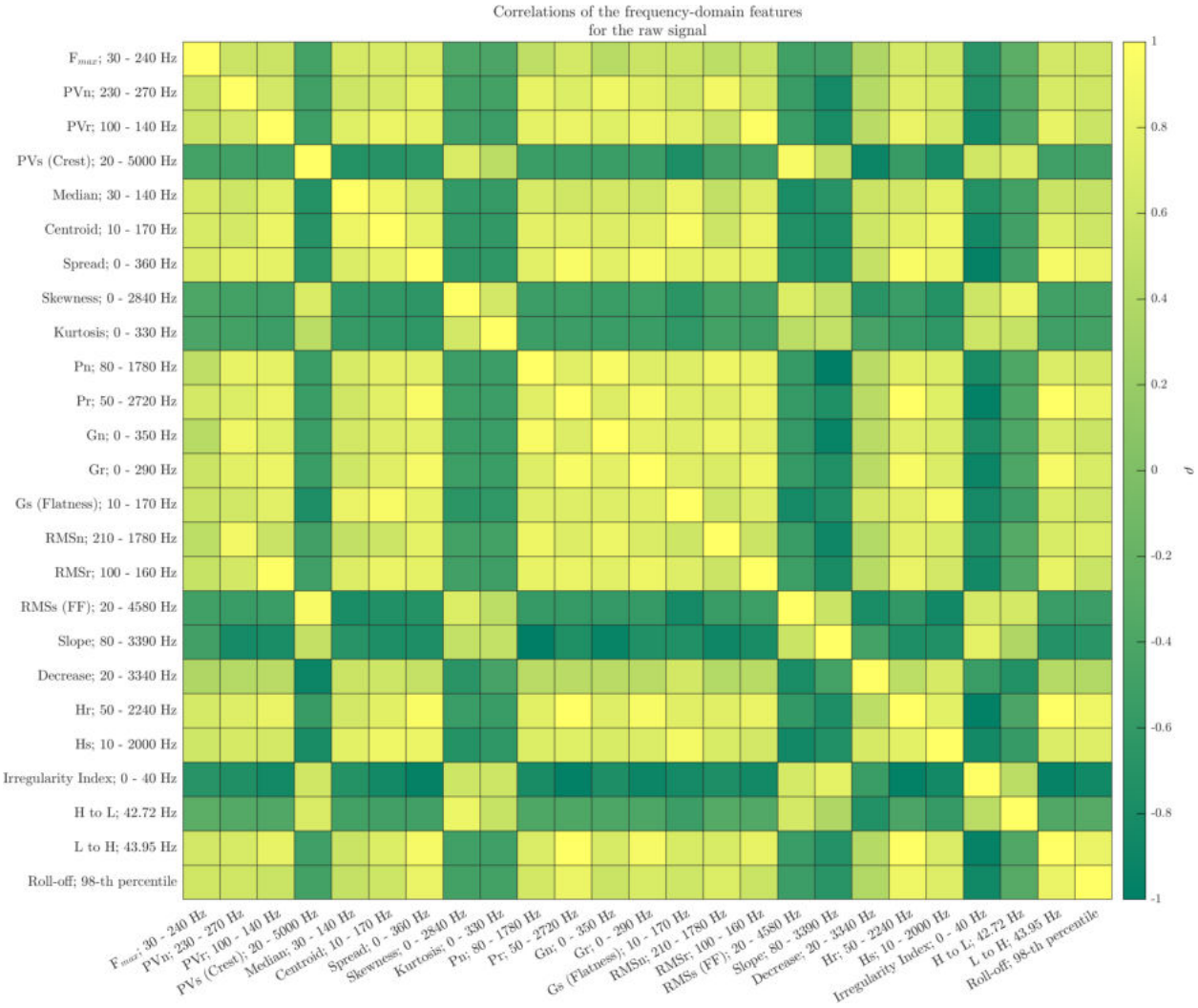

Figure 11.43: Correlation of frequency domain features obtained for the raw signal.

In the raw signal, the F_{max} feature proved to be positively correlated with the non-normalized and relatively-normalized Peak Values. Interestingly, correlation of those features

with the self-normalized Peak Value proved to be negative. It indicates, that when the Peak Value in non-normalized or relatively-normalized range increases, it decreases in the self-normalized range. This decrease could be explained by the increase in all frequency bins in a given range, rendering the prominent bin *less* prominent relative to other bins within this range, especially when considering quite wide range, like 20 to 5000 Hz.

Slight positive correlations can be observed also for the non-normalized and relative-normalized Peak Values with central tendency measures, i.e., the Spectral Median and Centroid. Notably, positive correlation can be observed also for the Spectral Spread value, confirming, that more frequency bins increase along with the most prominent one.

Spectral Skewness and Kurtosis proved to be negatively correlated with the central measures, indicating that with the progressing knee joint conditions, distribution of power inside the given frequency range becomes flatter and shifts towards higher frequencies.

All power features, except for the self-normalized PSD's Root Mean Square, proved to be more or less positively correlated. It further confirms that progressing knee joint conditions generally increase power in specific, narrow ranges, as well as wider ones. The negative correlation of the Spectral Form Factor, or the self-normalized Root Mean Square, can be explained by the faster increase of smaller amplitude values (measured by the arithmetic average) compared to large amplitude ones (measured by the RMS), causing the Form Factor to decrease when other power values increase.

Negative correlation of the Spectral Slope to the Peak Values, central tendency features, as well as power features indicate, that with progressing knee joint conditions, power in lower frequencies increases more compared to higher frequencies, lower and higher being understood relatively to the frequency range of 80 to 3390 Hz. As expected, the Spectral Decrease feature proved to be negatively correlated to the Spectral Slope. It can be explained by their definitions: while the Spectral Slope measures slope of line fitted to the power in the analyzed range, the Spectral Decrease measures how the power decreases compared to the first frequency bin in the analyzed range (see Table 6.3).

Both Entropy features, i.e., calculated for the relatively- and self-normalized ranges, proved to be positively correlated to the power, again, indicating that power increases in multiple frequency bins. They also proved to be negatively correlated to the Spectral Kurtosis, which is expected: when the power distribution becomes flatter, the Spectral Kurtosis decreases. That makes the distribution more uncertain, which causes the Entropy measures to increase. Irregularity Index, however, also proved to decrease with progressing knee joint conditions, which can be considered counterintuitive. Note, however, that it was calculated in quite low frequency range: from 0 to 40 Hz. As mentioned earlier, it could indicate that specific frequency bin in this range increased disproportionally to other bins, making distribution in this specific narrow range *more* predictable, decreasing value of the Irregularity Index.

Ratio of frequencies above to frequencies below around 43 Hz showed negative correlation with power features, indicating that power in lower frequencies increases faster with degeneration of the knee joint. Since the other ratio was calculated for almost the same frequency threshold, its positive correlation to the power features is expected. Spectral Roll-off feature, calculated for the 98-th percentile, also showed positive correlation to power features, indicating that power increases in rather wide frequency range.

11.5 Summary

In summary, a couple popular PSD estimation methods were compared, including the periodogram, the Welch method with varying window size, and the Multitaper method with varying number of tapers. The Welch method, with window size of 2^{14} samples, proved to allow extraction of the most informative features. However, features' informativeness did not vary greatly across methods (see Appendix A). Estimated PSD was used for further feature extraction.

For each preprocessing approach, i.e., the raw, the filtered and the derivative signal, frequency analysis provided 22 sets of range-dependent features, called the Frequency Range Maps, two frequency ratio sets and one set based on the Spectral Roll-off point. Each FRM provided $\binom{501}{2} = 125\,250$ features, both frequency ratios provided 8192 features, and the Spectral Roll-off point provided 101 features. In summary, for each preprocessing, 2 771 985 frequency features were extracted, summing up to 8 315 955 features total.

Features defined using extreme values, such as the frequency of the maximal value (F_{max}) and its actual values, absolute or normalized, proved to be relatively informative. For most features, the lowest $coef_{BMC}$ was achieved in low frequency ranges, up to 500 Hz. The non-normalized Peak Value proved to increase in 230 to 270 Hz range with progressing knee joint conditions. The F_{max} feature in 30 to 240 Hz range showed positive correlation to the Peak Value, indicating that, in this range, progressing knee joint conditions make PSD to be more concentrated around higher frequencies. Similar conclusions were drawn from the Spectral Median and Spectral Centroid features, as well as the Spectral Skewness.

The Geometric Mean proved to be the most informative frequency power feature. When measured in 0 to 350 Hz frequency range, it increased with progressing knee joint conditions. However, in general, all power features showed similar increase in power. Between the normalization methods, i.e., without normalization, with normalization relative to the whole spectrum, and relative to the analyzed range of the PSD, features obtained for non-normalized power proved to be the most informative.

The Spectral Slope calculated for the 80 to 3390 Hz frequency range decreased with progressing knee joint degeneration, indicating that power in lower frequencies of this range increases faster compared to the higher frequencies. Spectral Decrease, being negatively correlated to the Spectral Slope allowed to draw similar conclusions. Entropy features indicated that in more degenerated knee joints, power is more spread out across the spectrum when measured in wide frequency ranges. In narrow range of 0 to 40 Hz, however, the Irregularity Index proved to decrease, indicating that one specific frequency bin increases disproportionally faster to other bins.

Frequency ratio features proved to be relatively informative when defined with frequency threshold of around 43 Hz. They indicated that for progressing knee joint conditions, power below this frequency grows much faster compared to the power above it. Spectral Roll-off calculated for the raw signal proved to be the most informative when calculated for 98-th percentile. Values of this feature in specific classes showed that with progressing knee joint conditions, frequency of the Roll-off point shifts from around 20 Hz in the ctrl up to around 150 Hz in the oa.

After relatively detailed analysis of frequency ranges used in the FRMs, it would be

beneficial to somehow quantify how informative are specific frequency bins, i.e., from 0 Hz to 10 Hz, from 10 Hz to 20 Hz, up to 4990 Hz to 5000 Hz, but not understood separately. Rather, how much information can each frequency bin bring to a wider range, that contains it. Such analysis would enable better understanding of the information content across the frequency domain[1]. This, in turn, would allow potentially better design of the filter in the future studies. In order to somehow reduce dimensionality of each FRM down to 2 dimensions, i.e., the frequency bin and the informativeness, the following process has been implemented. First, for each frequency bin of the FRM, mean $coef_{BMC}$ of all frequency ranges containing this bin was calculated. Such a mean value could be understood as taking the average of the rectangular area in the FRM, defined using two points: $(0, f)$ and $(f, 5000)$, where f is the frequency bin. That created a reduced curve of informativeness for each frequency bin of the FRM. Then, the reduced curves for all FRMs were averaged. Figure 11.44 includes this average, calculated for each preprocessing method separately, and additionally, their average.

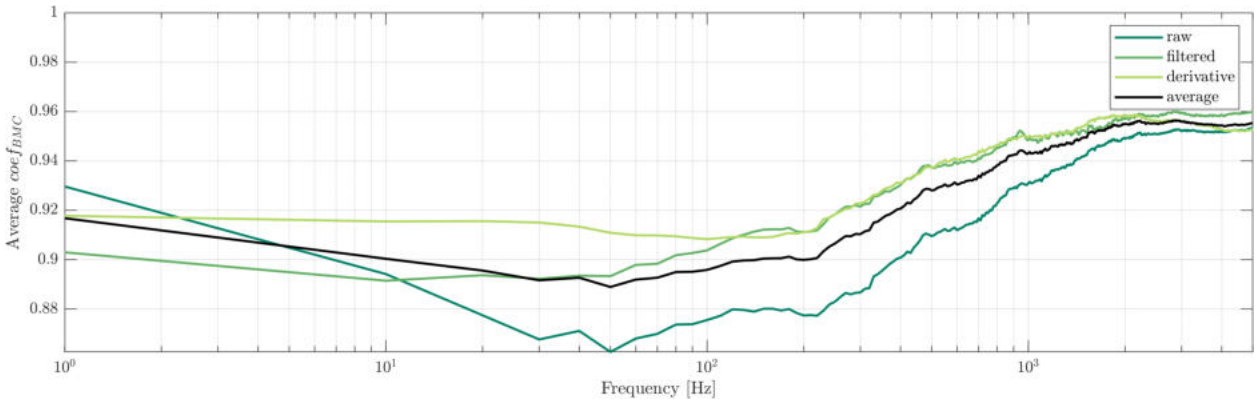

Figure 11.44: Average informativeness of the features calculated on ranges containing each frequency bin.

Clearly, the average $coef_{BMC}$ scores proved to be the lowest for the raw signal. In terms of frequency, all preprocessing approaches showed similar results: average informativeness is the lowest for ranges from around 10 Hz up to around 350 Hz.

Another approach of showing information content of each frequency bin could be to simply count the number of FRMs, in which given frequency bin was included in the most informative frequency range. Such curve, with the frequency bin on the x-axis, and number of FRMs on the y-axis, was included in Figure 11.45. The plot includes separate curves for each preprocessing, as well as their sum.

Curves included in the Figure 11.45 allow to draw similar conclusions compared to Figure 11.44. Frequency bins included most frequently in the most informative frequency ranges proved to span from around 10 Hz to around 350 Hz.

Those results indicate, that the VAG signals could be acquired with sampling frequency much lower than 10 kHz, without great impact on the PSD's informativeness. That could be

[1]Note, that the *information content* is understood in a context of this work only, and means ability to differentiate between analyzed classes.

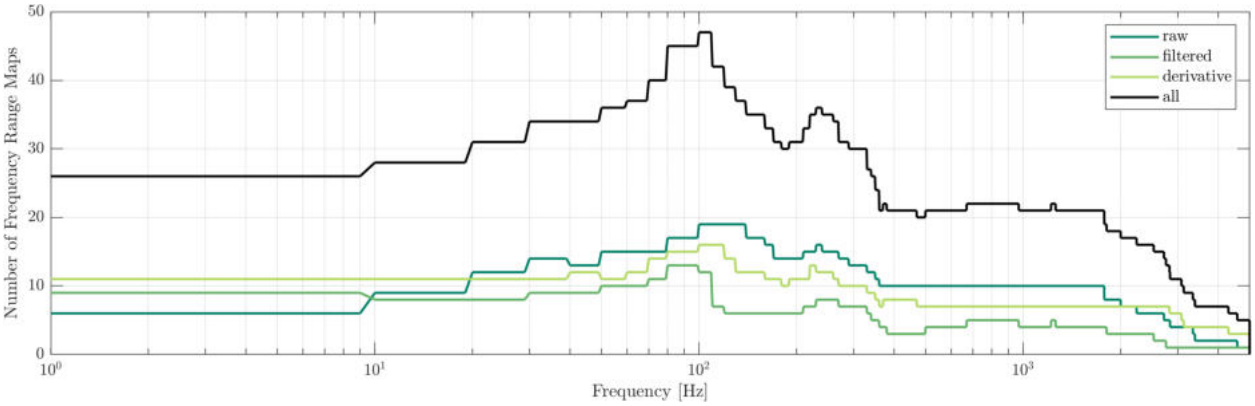

Figure 11.45: Number of the FRMs, in which given frequency bin was contained in the most informative frequency range.

further confirmed with the fact that some other VAG-related studies used sampling frequencies of 1 kHz [37, 96, 114, 125] (see Table 3.3). Moreover, in VAG-related studies utilizing filtering approaches, the most popular frequency bands are from 10 Hz to around 1000 Hz (see Figure 3.3). In future studies, VAG signal could be filtered in range spanning from 10 to 350 Hz, potentially allowing more informative analysis also in time-domain.

11.6 Conclusions

The frequency analysis allowed for drawing the following conclusions.

1. From the compared Power Spectral Density estimation methods, Welch method with 2^{14} window size, provided the most informative features.
2. In the context of this work, frequencies with the highest information content proved to be in range of around 10 Hz to 350 Hz, indicating that the vibroarthrogram could be acquired with sampling frequency much lower than 10 kHz.
3. With progressing knee joint deterioration:
   - power of the VAG signal increased,
   - that increase was observed in most frequency bins, yielding higher values of features such as Spectral Spread or Entropy,
   - lower frequencies generally increased faster than higher ones, yielding lower values of Spectral Slope and Spectral Skewness.
4. With the frequency analysis based on the PSD estimation, analysis of the raw signal provides the most informative features.
5. Both frequency range normalization methods, i.e., relative to the whole PSD or the range itself, decreased informativeness of features extracted from it.
6. In future studies, time-domain analysis could be conducted on the signal filtered in range from 10 Hz to 350 Hz.

Chapter 12

Time-frequency domain features

In this chapter, results of the time-frequency analysis will be provided. Since from each spectrogram (calculated for multiple parameters), 21 Spectral Fluctuation Signals were obtained and summarized using 15 time features, complexity of the research can be considered complex. Therefore, two first sections of this chapter will be devoted to analysis from frequency and time perspective separately. For the most informative features, specific influence of the spectrogram's parameters on the given feature will be discussed. In the following section, average informativeness of all features will be explored, in the context of spectrogram parameters, i.e., the window size and overlap. The chapter will end with a summary of the results and conclusions that can be derived from them.

12.1 Frequency perspective—Spectral Fluctuation Signals

In this section, barplots of features extracted from each Spectral Fluctuation Signal will be provided. Similarly to results provided in previous chapters, each bar shows two $coef_{BMC}$values: of the most and the least informative feature in a set. Here, each set is constructed by all time features extracted for the given SFS, for all spectrogram parameters, summing up to 1650 singular values: 15 time features, times 11 window sizes of the spectrogram, times 10 overlap values.

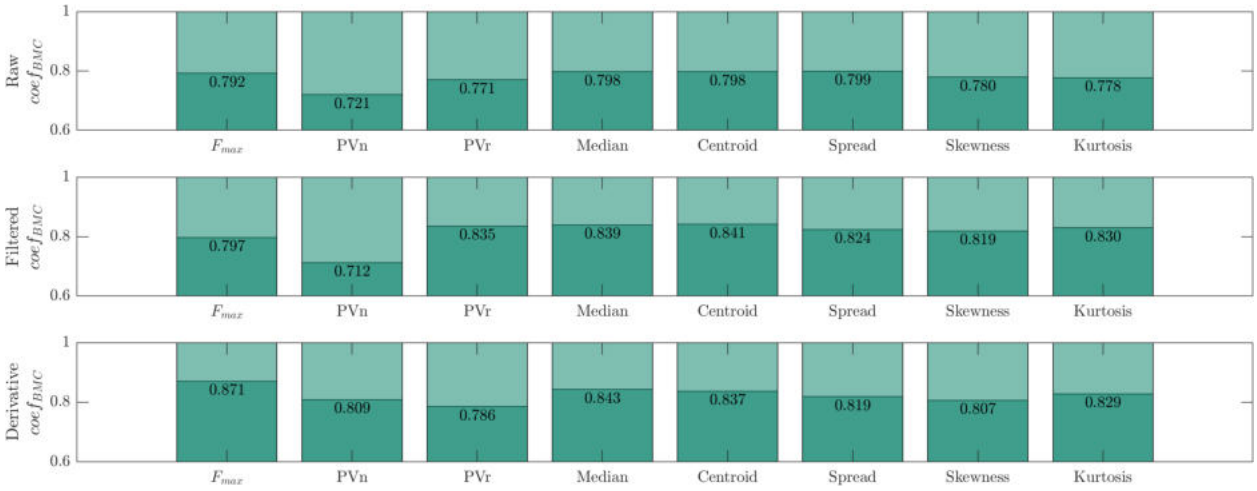

Figure 12.1: Informativeness of all time features extracted from the Power Spectral Density distribution SFSs, from the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

Informativeness of the features based on the PSD distribution SFSs was provided in Figure 12.1. Compared to the frequency-only analysis using the same features, informativeness is lower in most part. One exception is the Peak Value of the non-normalized spectrum. Barplots of the specific time features extracted from this SFS were provided in Figure 12.2.

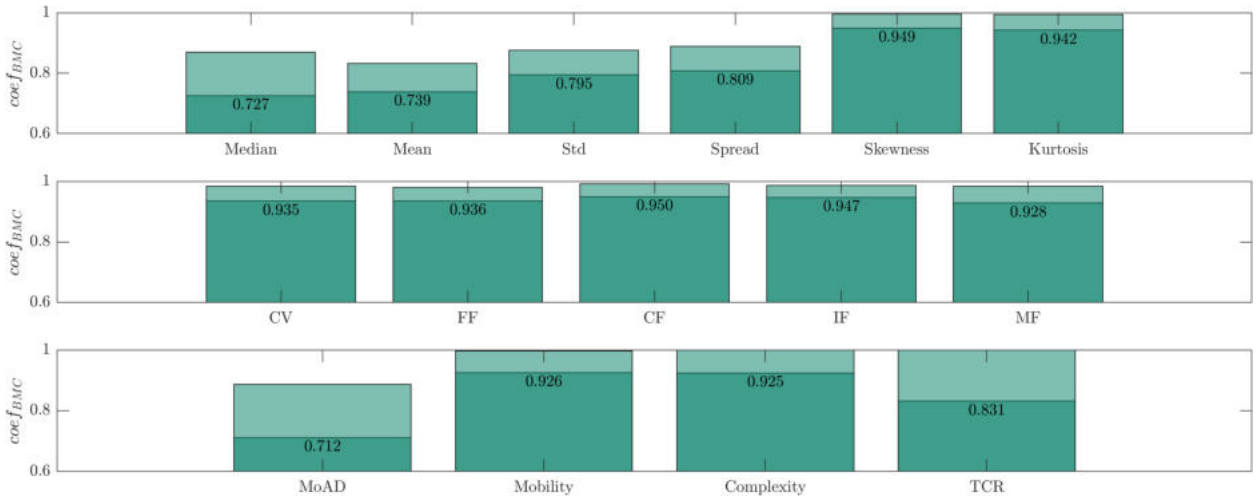

Figure 12.2: Informativeness of specific time features extracted from the non-normalized Peak Value SFS, from the filtered signal.

Median and the Mean of Absolute Differences features achieved the lowest $coef_{BMC}$ values for the PVn SFS. Visualization of those features' dependance on the spectrogram parameters was provided in Figure 12.3. A couple of interesting patterns can be observed. First, for the median feature, higher values of the window size provide much lower $coef_{BMC}$ values. It indicates, that the frequency resolution is more important compared to the time resolution, when measuring median value of the PVn fluctuations. Note, however, that the median calculated for small window sizes, i.e., below 64 samples, provided not completely uninformative features, achieving $coef_{BMC}$ of around 0.8. Medium-sized windows, between around 64 and 2048 samples, achieved the highest $coef_{BMC}$, proving to be the least informative.

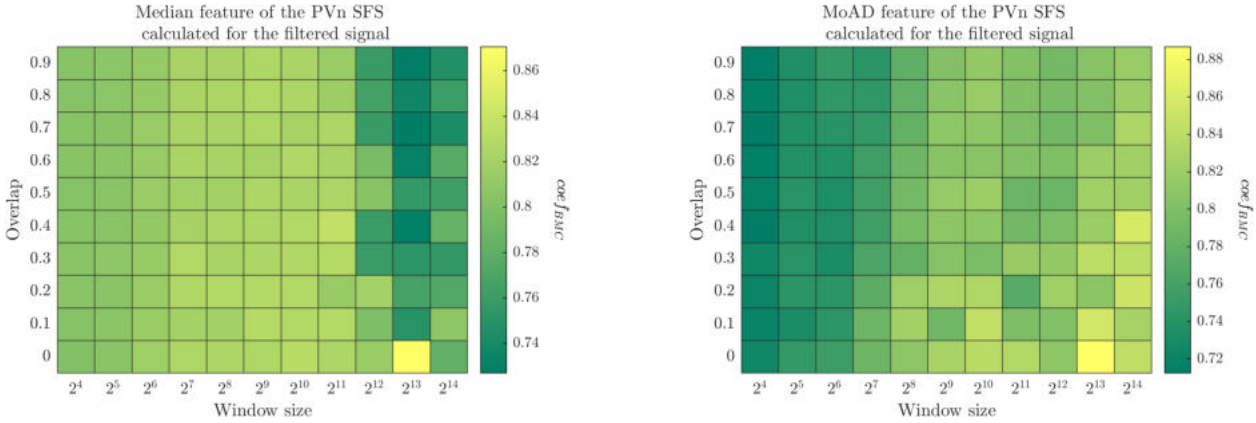

Figure 12.3: Informativeness of the non-normalized Peak Value SFS summarized by the median (left) and MoAD (right) features, for varying spectrogram parameters.

Another interesting pattern emerges from the vertical axis of the plot. Clearly, higher overlap values provide better informativeness scores. This pattern, although less obvious, is also visible in the Mean of Absolute Differences feature. Dependance of the informativeness on the window size, however, proved to be completely opposite compared to the median value. It could be explained by the definition of the MoAD feature, which inherently measures *temporal* variability of the signal. Note, that with 16-sample wide window, frequency resolution is only 625 Hz. Compared to the MoAD feature calculated on the filtered signal directly, i.e., in time domain, informativeness of the PVn SFS's MoAD is worse. Here, the lowest $coef_{BMC}$ proved to be around 0.71, while in the time domain, MoAD achieved $coef_{BMC}$ of around 0.6.

Boxplots of the median and MoAD features calculated on the PVn Spectral Fluctuation Signal were provided in Figures 12.4 and 12.5, respectively. The median clearly increases with progressing knee joint conditions, indicating that the Peak Value, *in general*, grows. It is not very surprising result, given analyzes described in the previous chapter. Perhaps more interesting result is increasing value of the MoAD feature, indicating that with increased power of the most prominent frequency bin, its variability in time also increases.

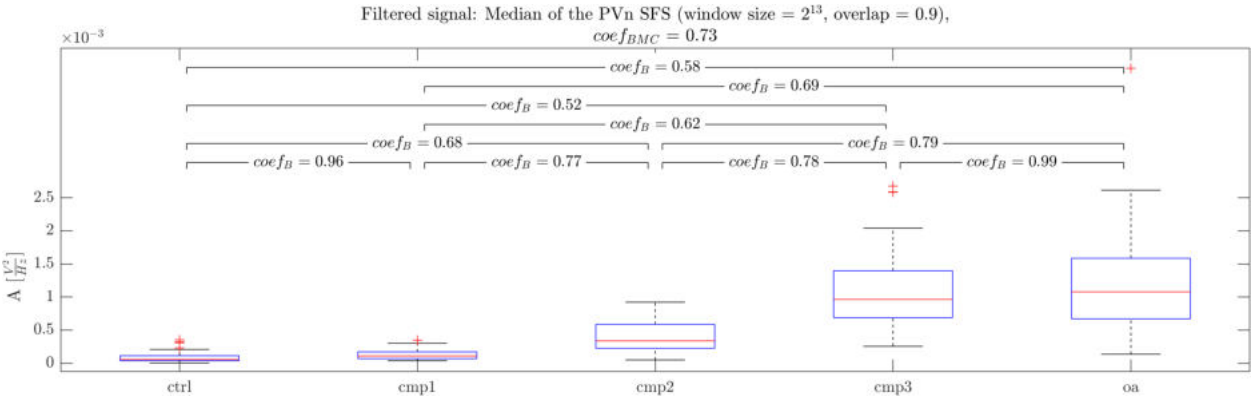

Figure 12.4: Boxplot of the Median feature extracted from the non-normalized Peak Value SFS, calculated for the filtered signal.

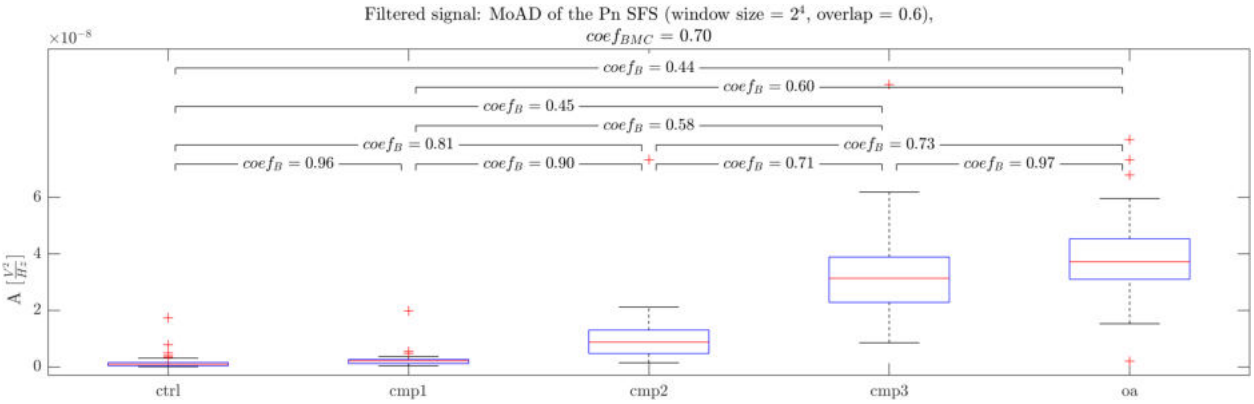

Figure 12.5: Boxplot of the MoAD feature extracted from the non-normalized Peak Value SFS, calculated for the filtered signal.

In Figure 12.6 informativeness summary of the features extracted from the power SFS's was provided. The non-normalized Average Power, calculated from the filtered signal, proved to be the most informative Spectral Fluctuation Signal. Barplot of the time features extracted from it was provided in Figure 12.7.

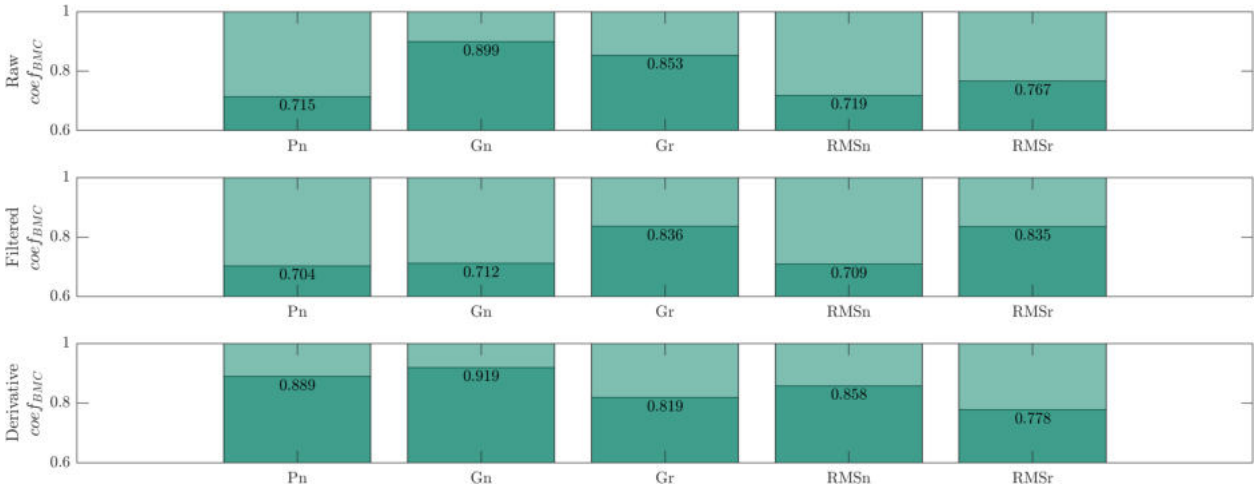

Figure 12.6: Informativeness of all time features extracted from the Power Spectral Density power SFSs, from the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

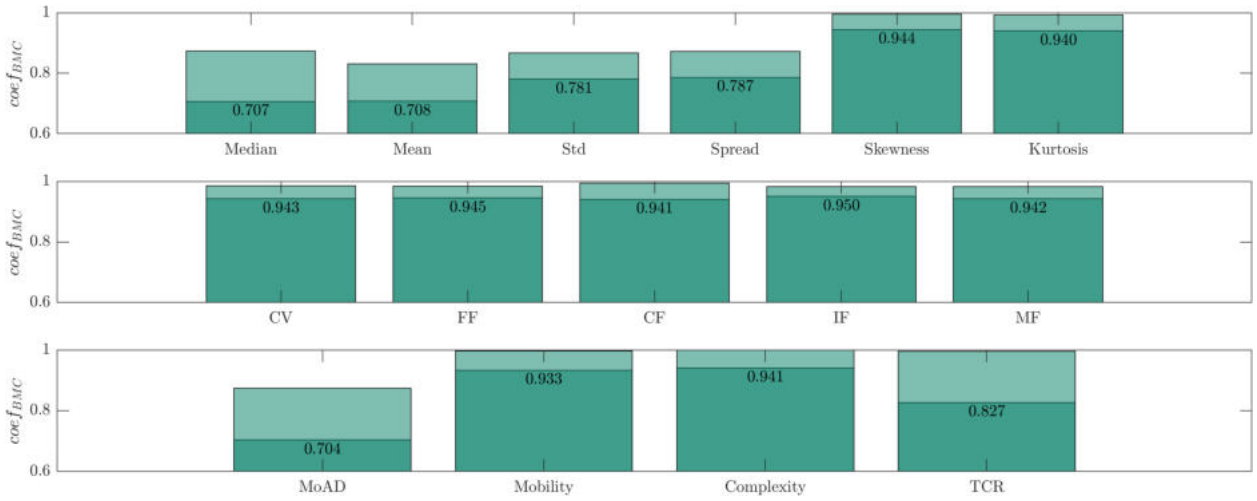

Figure 12.7: Informativeness of specific time features extracted from the non-normalized Average Power SFS, from the filtered signal.

Similarly to the non-normalized Peak Value, the Average Power was best summarized by the median and the MoAD features. Visualization of their dependance on the spectrogram parameters was provided in Figure 12.8. Patterns observed, in general, also seem very similar to the ones obtained for Peak Value. For the median feature, higher window sizes provided better informativeness, while shorter window sizes were better suited for the inherently *temporal* MoAD feature. In both cases, higher values of the overlap provided more

informative features. Boxplots of both median and MoAD were provided in Figures 12.8 and 12.9, respectively.

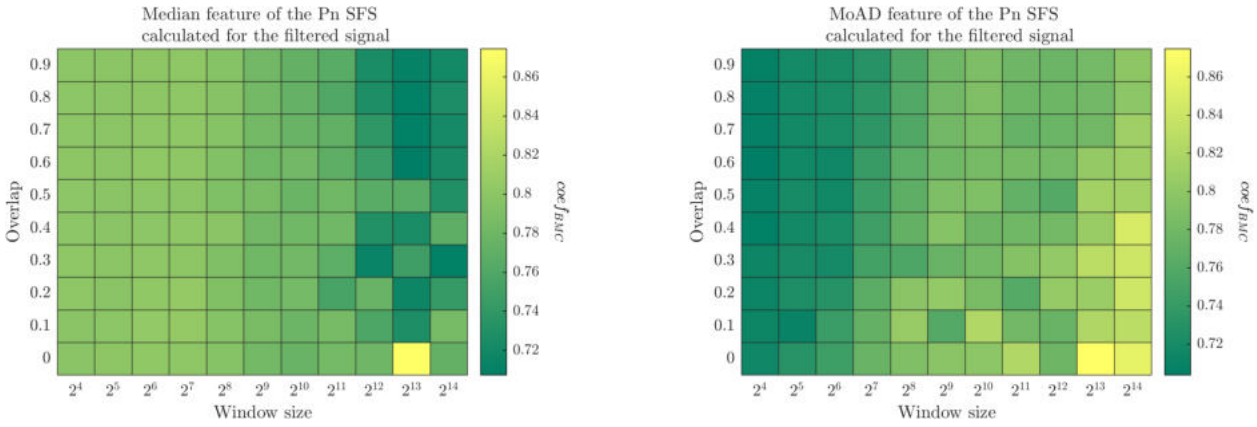

Figure 12.8: Informativeness of the non-normalized Average Power SFS summarized by the median (left) and MoAD (right) features, for varying spectrogram parameters.

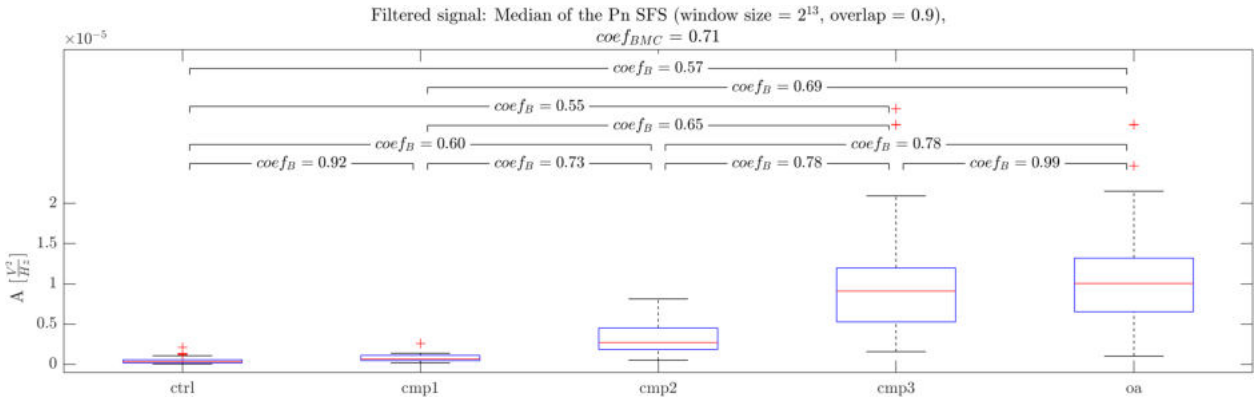

Figure 12.9: Boxplot of the Median feature extracted from the non-normalized Average Power SFS, calculated for the filtered signal.

Again, clear similarity can be observed between the Peak Value and the Average Value of the non-normalized PSD. Values of both median and MoAD features increase with progressing knee joint deterioration, indicating change in both power and its variability in time domain.

Informativeness of the features extracted from the Spectral Fluctuation Signals based on the PSD's shape, was provided in Figure 12.11. Features, in general, proved rather mediocre informativeness. The lowest $coef_{BMC}$ of around 0.7, was achieved by some feature of the Spectral Slope calculated from the filtered signal. Barplots of all time features extracted from this SFS were provided in Figure 12.12.

Yet again, the most informative time features proved to be the median and the MoAD, achieving lowest $coef_{BMC}$ values of around 0.71 and 0.70, respectively. Their dependence on the spectrogram parameters was visualized in Figure 12.13.

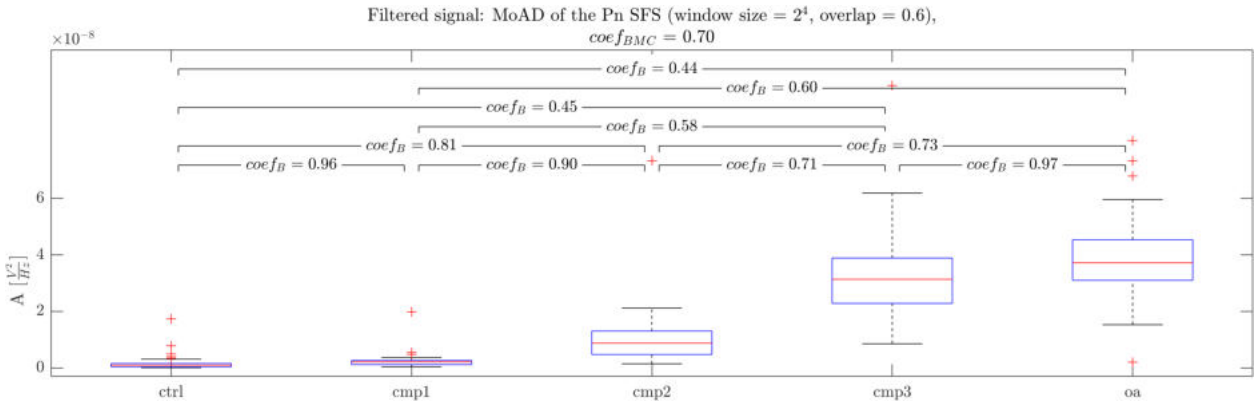

Figure 12.10: Boxplot of the MoAD feature extracted from the non-normalized Average Power SFS, calculated for the filtered signal.

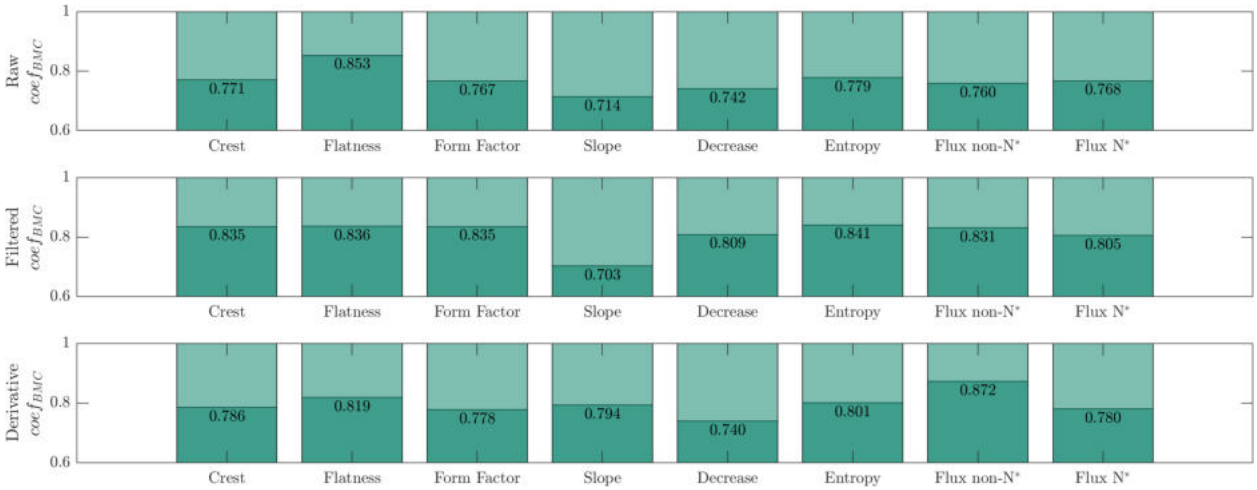

Figure 12.11: Informativeness of all time features extracted from the Power Spectral Density shape SFSs, from the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot). * N stands for Normalized.

Interestingly, both plots seem almost exactly the same as the Average Power plots. Not only in terms of general patterns, but also exact $coef_{BMC}$ values. Boxplots of the features extracted for the most informative spectrogram setup, were provided in Figures 12.14 and 12.15, for the median and the MoAD features, respectively.

12.2 Temporal perspective—Time features

In this section, barplots of time domain features extracted from all Spectral Fluctuation Signals will be provided. Here, each bar is based on one time feature extracted for all SFSs, for all spectrogram parameters, summing up to 2310 singular values: 21 SFSs, times 11 window sizes of the spectrogram, times 10 overlap values. In Figure 12.16 time features based on the values' distribution were provided.

In general, time domain distribution features are not very informative, in most part

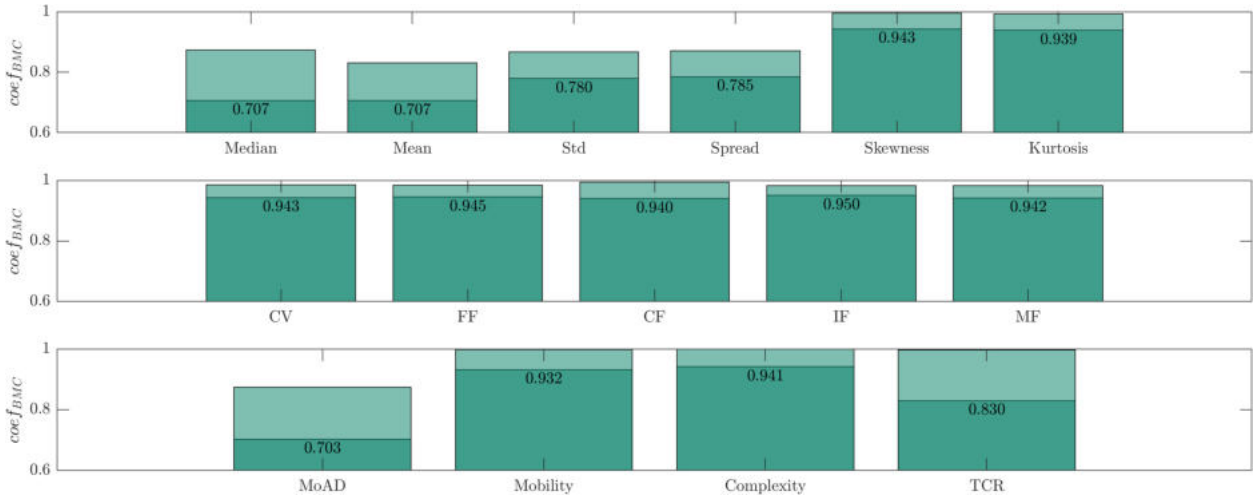

Figure 12.12: Informativeness of specific time features extracted from the Spectral Slope SFS, from the filtered signal.

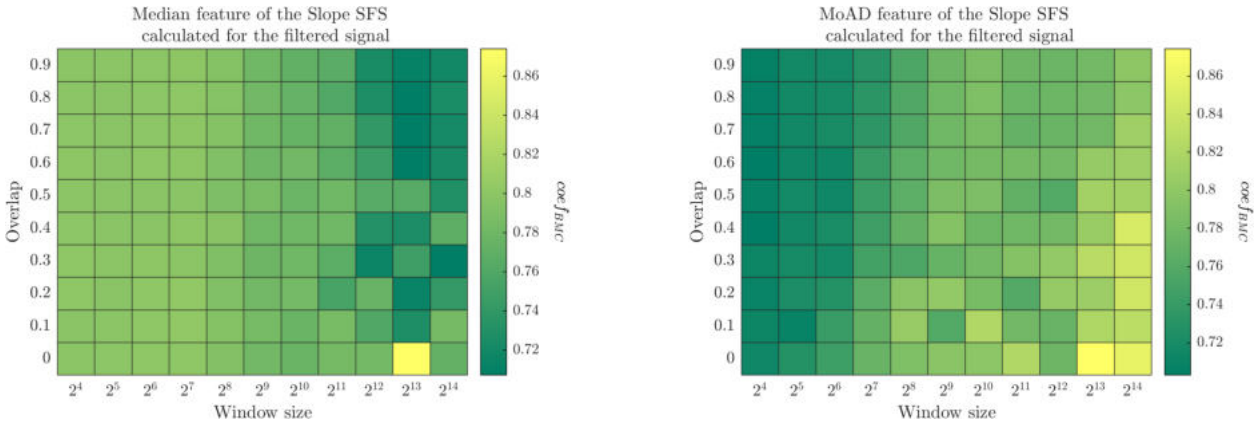

Figure 12.13: Informativeness of Spectral Slope SFS summarized by the median (left) and MoAD (right) features, for varying spectrogram parameters.

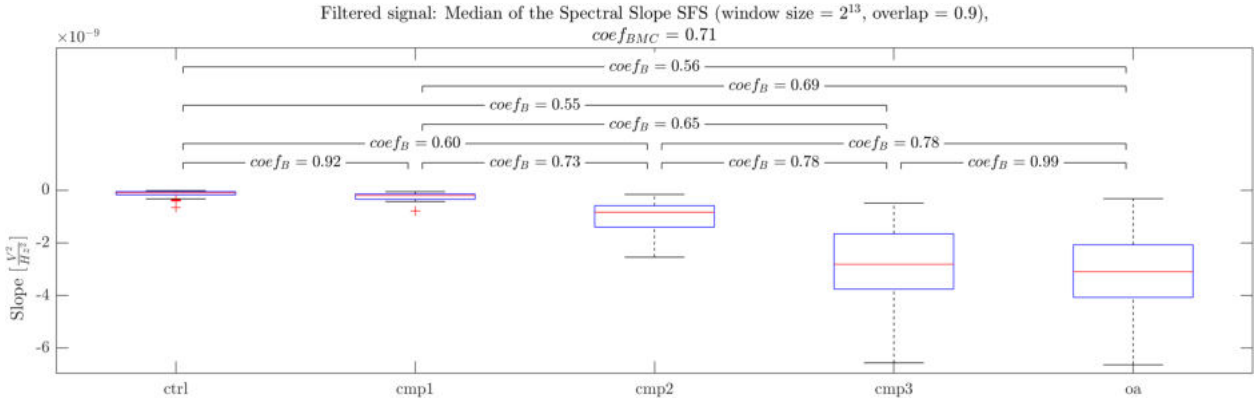

Figure 12.14: Boxplot of the Median feature extracted from the Spectral Slope SFS, calculated for the filtered signal.

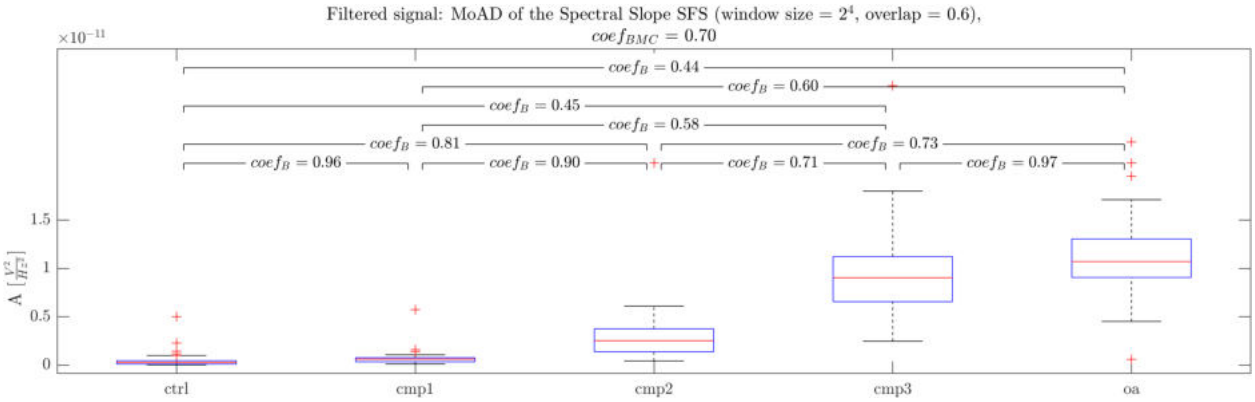

Figure 12.15: Boxplot of the MoAD feature extracted from the Spectral Slope SFS, calculated for the filtered signal.

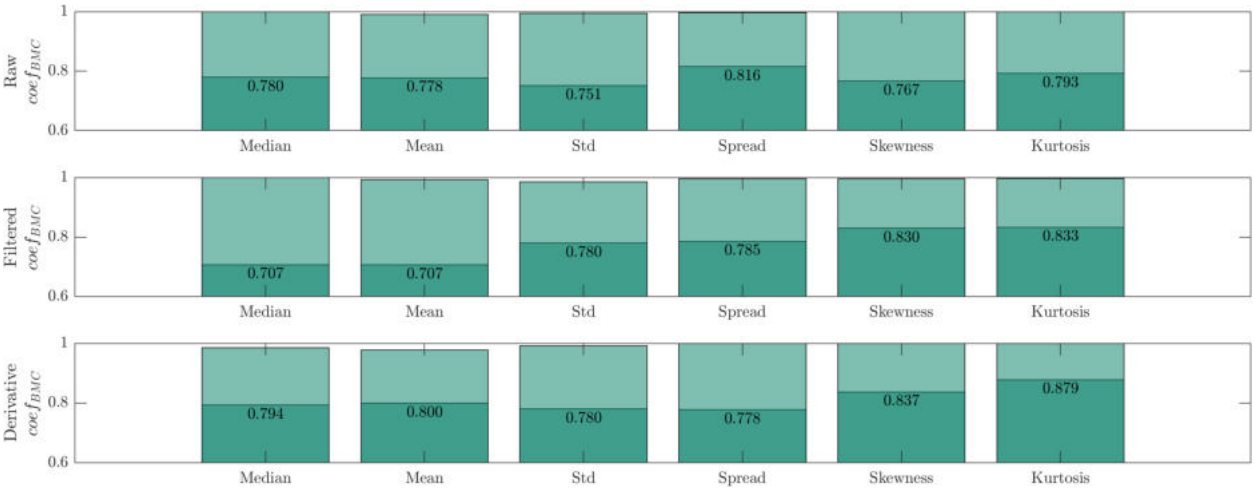

Figure 12.16: Informativeness of specific time domain distribution features extracted from all SFSs, from the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

achieving $coef_{BMC}$ scores well above 0.7. The most informative feature, as expected from the results discussed in previous section, proved to be median of some SFS calculated for the filtered signal. Barplots of specific SFSs summarized by the median were provided in Figure 12.17.

As discussed in previous section, the Peak Value and the Average Power, both calculated on the non-normalized PSD, as well as the Spectral Slope, proved to be the most informative SFSs. However, the RMS value also scored quite low $coef_{BMC}$. Other features turned out to be of relatively low informativeness.

Barplots of the complex time domain statistical shape measures, of each SFS, were provided in Figure 12.18. All features proved to be of rather moderate informativeness, mostly achieving $coef_{BMC}$ more than 0.8.

Informativeness summary of the time domain features actually based on the temporal variation of values, per each SFS, was provided in Figure 12.19. The mobility, complexity and

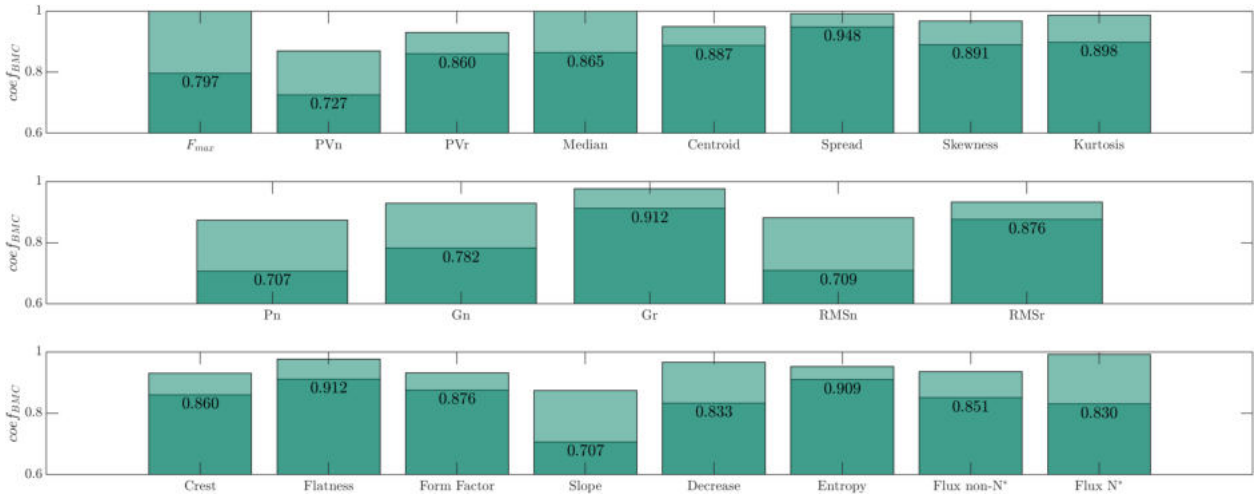

Figure 12.17: Informativeness of the median feature extracted from specific SFSs, from the filtered signal.

Figure 12.18: Informativeness of specific time domain shape features extracted from all SFSs, from the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot). * N stands for Normalized.

Turns Count Rate features proved to be of rather moderate differencing potential, scoring $coef_{BMC}$ between 0.74, for the Turns Count Rate of the derivative signal to around 0.87 for the complexity for the same preprocessing approach.

The MoAD feature proved to be the most informative in the filtered signal, with the $coef_{BMC}$ of around 0.7. As identified it the previous section, it was scored for the Spectral Slope. Informativeness of the MoAD feature, extracted from each SFS separately, was provided in Figure 12.20.

Besides the Peak Value and the Average Power of the non-normalized PSD, as well as the Spectral Slope, the MoAD feature proved to be relatively informative when used in the Geometric Mean and the RMS of the non-normalized spectrum. Dependance of the MoAD's informativeness on the spectrogram parameters, when extracted from those two SFSs, was

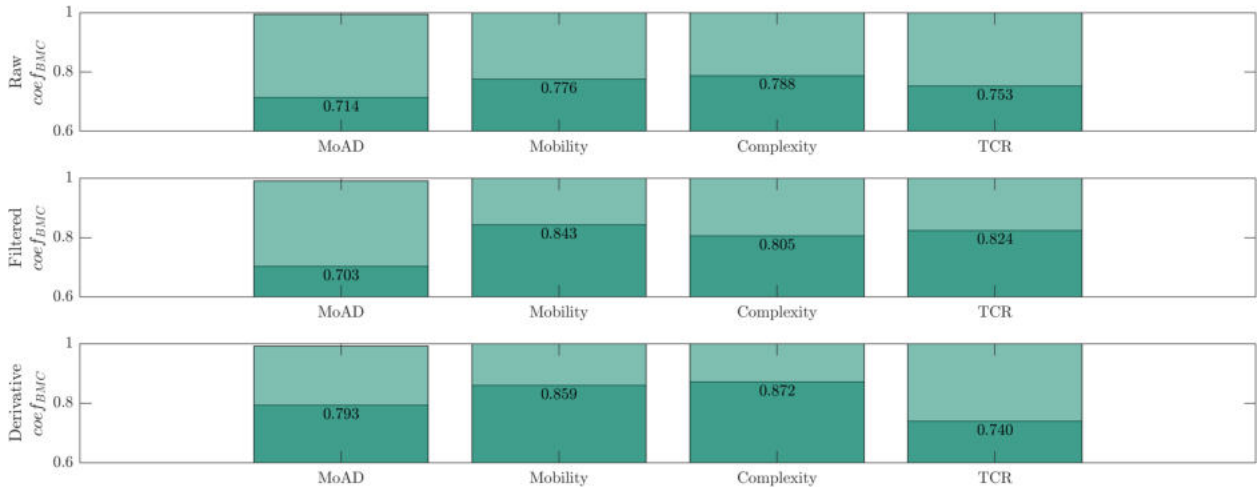

Figure 12.19: Informativeness of specific time domain temporal features extracted from all SFSs, from the raw signal (upper plot), filtered signal (middle plot), and the derivative signal (lower plot).

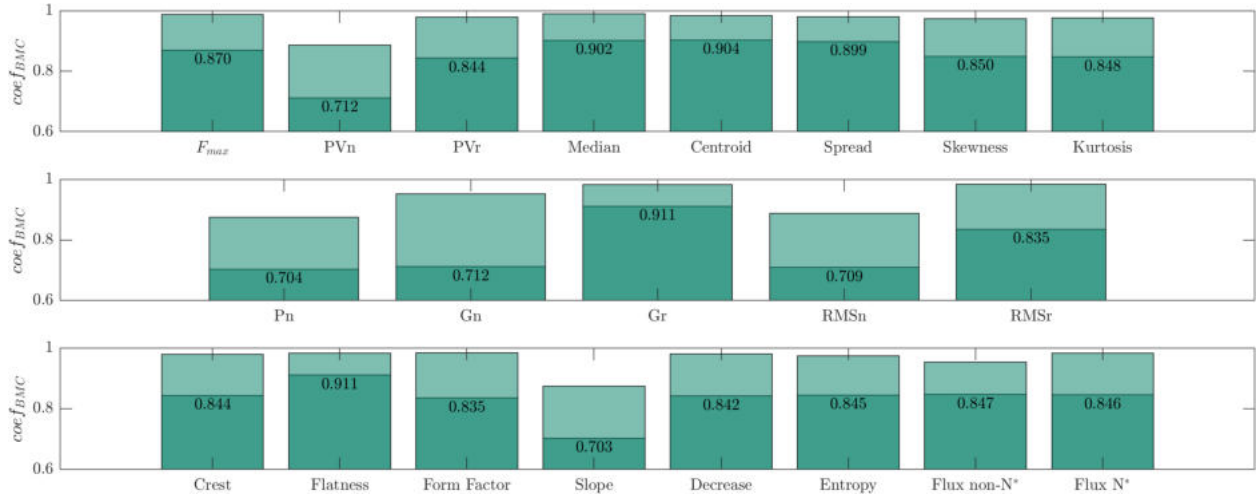

Figure 12.20: Informativeness of the MoAD feature extracted from specific SFSs, from the filtered signal. * N stands for Normalized.

provided in Figure 12.21.

Similarly to previously analyzed features, MoAD showed greater informativeness when calculated for shorter window sizes in both the Geometric Mean and the RMS. Note, that in both SFSs, greater overlap usually provided better results. For the Geometric Mean, window size of 512 samples and the overlap of 70 percent allowed extraction of the most informative feature. For the RMS, those parameters were 16 samples and 60 percent.

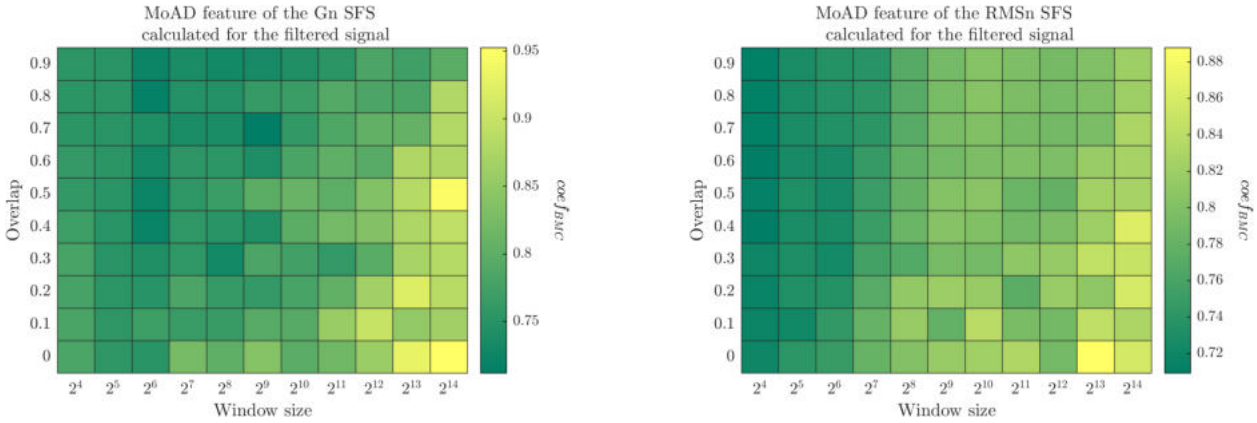

Figure 12.21: Informativeness of the MoAD summary of the Geometric Mean (non-normalized) (left) and the Root Mean Square (non-normalized) SFSs, for varying spectrogram parameters.

12.3 Spectrogram parameters

In this section the influence of the spectrogram parameters on the features' informativeness will be discussed. In the Figure 12.22, the average $coef_{BMC}$ value of all time features was provided, for each Spectral Fluctuation Signal separately. Due to relatively large number of SFSs, axes labels were omitted. In each plot, the axes are identical to previous plots, such as those in Figure 12.21. That is, the window size was indicated on the x-axis, spanning from 2^4 to 2^{14} samples, while the overlap, spanning from 0 to 90 percent, was represented on the y-axis. Since specific informativeness values were discussed earlier, here, the general patterns observed on each plot will be analyzed. Therefore, each plot has color axis (the $coef_{BMC}$) scaled separately. Finally, similarly to previous plots, darker areas indicate lower $coef_{BMC}$ values, i.e., more informative features.

Pattern of higher informativeness achieved for low frequency and high time resolution can be observed in most SFSs, with a couple of exceptions. For example, the F_{max}, seems to follow it up to around 64 samples. With sampling frequency of 10 kHz, it results in frequency resolution of around 150 Hz. Similarly, Spectral Median and Centroid seem to achieve highest informativeness at around 128 samples, with frequency resolution of around 80 Hz. Finally, Spectral Skewness achieves lowest $coef_{BMC}$ values with window size of 256 samples, resulting with frequency resolution of around 40 Hz. In all mentioned features, the frequency resolution of 625 Hz (achieved for the smallest window size of 16 samples used in this study) proved to be too low to extract the most informative features. It can be explained by their definition, measuring inherently frequency-related information, rather than general tendencies of power. For example, as indicated in the previous chapter, the power of VAG signal increases across wide range of frequencies. Therefore, high frequency resolution does not necessarily allow extraction of more informative features, at least with a time-frequency resolution tradeoff.

The SFS calculated for the Geometric Average of the non-normalized PSD is the only exception of this observation, achieving similar informativeness for both low and high frequency resolution. A possible explanation of this could be the sensitivity of geometric mean to small

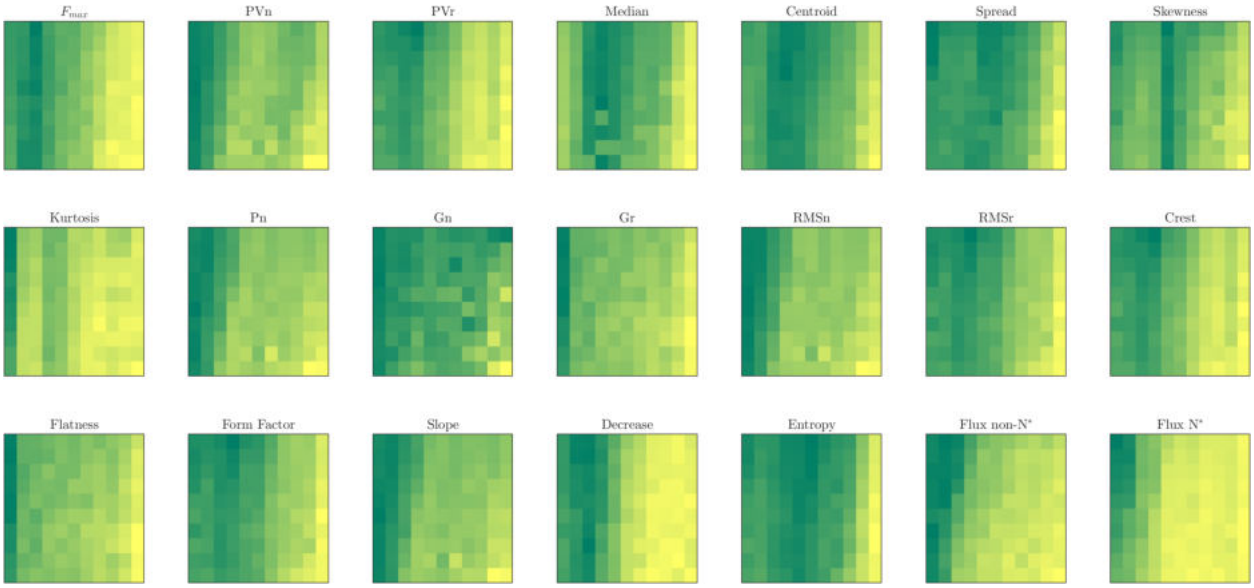

Figure 12.22: Average informativeness of all time features extracted from each SFS, for varying spectrogram parameters.

amplitude values, more of which are present in the PSD, when the frequency resolution is high.

For all SFSs, higher overlap in the spectrogram calculation allows extraction of more informative features. Even though it does not provide any time resolution increase in the Uncertainty Principle sense, it allows to obtain more continuous SFSs. By increasing the number of sub-signals used for the spectrogram calculation, and therefore the number of PSDs, it minimizes the risk of losing information contained in the edges of the window.

In the Figure 12.23, the average $coef_{BMC}$ value obtained for all SFSs, for each time feature was provided.

Again, for most time measures, lower frequency resolution allows extraction of more informative features. It is most visible in inherently temporal ones, i.e., the MoAD, mobility, complexity and TCR. Central tendencies, that is, the median and the mean, seem to achieve lowest $coef_{BMC}$ values at window sizes not smaller than 64 samples, i.e., with frequency resolution of minimum 150 Hz. It could indicate, that lower frequency resolution leads to SFSs that are more variable, and their central values are less meaningful compared to features measuring their variability.

12.4 Summary

In summary, for each signal, the spectrogram was obtained, with varying parameters of window size and overlap between the windows. From each spectrogram, 21 Spectral Fluctuation Signals were obtained, each further summarized with 15 time domain features. This yielded 34 650 features per each preprocessing approach, summing up to 103 950 time-frequency features total.

The most informative features proved to be the median and the MoAD, extracted from

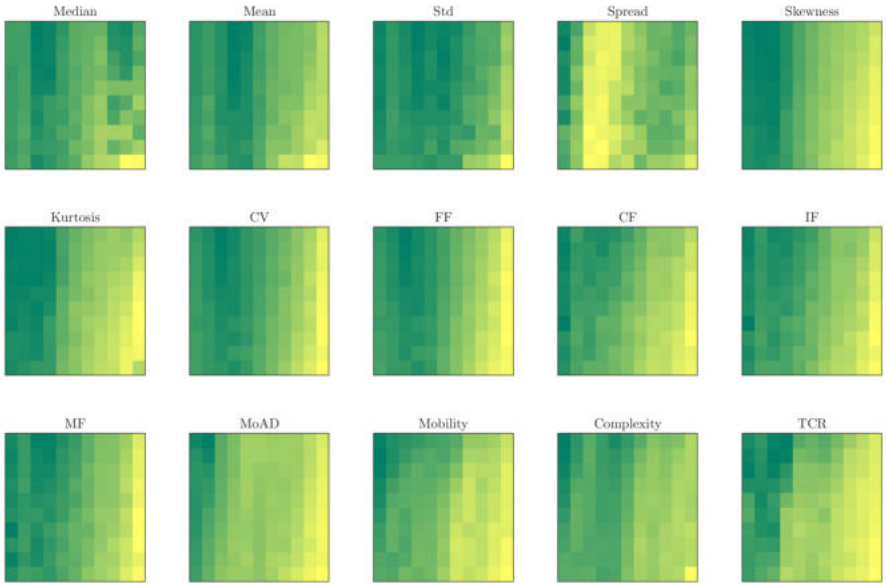

Figure 12.23: Average informativeness of specific time features extracted from all SFSs, for varying spectrogram parameters.

SFSs of the non-normalized Peak Value and the Average Power, as well as the Spectral Slope. All those features increased with progressing conditions of the knee joint, except for the median value of the Spectral Slope SFS, which decreased. It indicated, that with progressing knee joint conditions, power of the signal not only grows, but is more variable along the time axis.

For most SFSs and time domain measures, the filtering preprocessing approach allowed extraction of the most informative features. Note, that in the frequency-only analysis, this was not the case. It can be explained by the definition of each SFS, based on the *whole* short-time PSD. Perhaps extracting each SFS for some specific frequency range, as in previous chapter, would yield more informative features. That could be potentially fruitful direction for the future studies.

In general, for all SFSs and time domain features, higher overlap between windows in the spectrogram calculation led to more informative features. The other spectrogram parameter, i.e., the window size, proved to influence informativeness of the feature depending on its definition. For SFSs measuring power of the signal, shorter window sizes, i.e., of around 16 ms, allowed extraction of more information, indicating lesser importance of frequency resolution compared to time resolution. However, SFSs measuring some frequency-specific parameters, such as the Spectral Median or the F_{max}, proved to be the most informative for slightly longer window sizes, of around 64 ms. Similarly, in the time domain measures, inherently temporal features, such as the MoAD proved to be the most informative for high time resolution. On the other hand, central tendency time domain features achieved highest informativeness for somewhat longer windows, of around 128 ms.

12.5 Conclusions

Following conclusions can be derived from the conducted time-frequency analysis.

1. When the SFSs are based on the whole PSD, filtering provides the most informative features.
2. SFSs based on the power measures provide the most informative features when calculated for short windows, i.e., high time and low frequency resolution.
3. Higher frequency resolution is needed for the most informative feature extraction based on the SFSs measuring some inherently frequency-related features, such as the F_{max} or Spectral Centroid.
4. In general, time features measuring specifically temporal variability, seem to benefit from high temporal resolution.
5. Higher values of the overlap between windows in spectrogram calculation lead to more continuous SFSs, which allows extraction of more informative features.
6. Future studies could focus on conducting similar time-frequency research, but with extraction of SFSs based on some specific frequency range. It could be 10 Hz to 350 Hz suggested in the previous chapter, or preferably, the most informative frequency range per each SFS. In such case, analysis of the raw signal could provide the most informative features.

Chapter 13

Dimensionality reduction and classification

In the current chapter, results of the dimensionality reduction and classification will be provided. In the first section, two steps of the dimensionality reduction will be discussed. The second section will be devoted to the classification results and the comparison of the dimensionality reduction approaches in the classifiers' accuracy context. The last section will be include the summary and specific conclusions, which can be derived from the conducted analyzes.

13.1 Dimensionality reduction

In the first step of dimensionality reduction, all extracted features were divided into 25 groups per preprocessing. In each group, feature selection was done separately, with number of selected features determined using the elbow method. In summary, from 8 429 331 features total, 483 were selected after the first dimensionality reduction step. In Table 13.1, specific number of features selected in each group was provided.

From each feature group, between 3 and 25 features were selected. Interestingly, the most features were selected from the Frequency Range Maps. Of course, they included the most features to begin with. However, it suggests that different frequency ranges can be both informative and not necessarily redundant, even when measured by the same feature. Number of selected features did not varied much between the preprocessing approaches.

In the second dimensionality reduction step, all features selected in the first step were combined, and the dimensionality reduction was conducted again. Both MRMR and PCA were used, each providing 183 features sorted by importance. Note, that while the MRMR provided sorted list of the original features, the PCA provided new features, that were linear combinations of the original ones. Informativeness of consecutive features provided by both algorithms was provided in Figure 13.1. Complete list of all 183 features provided by the MRMR algorithm was included in Appendix B.

Clearly, the MRMR provides features that are much more informative. The lowest $coef_{BMC}$ of the PCA-provided features was 0.64, achieved for the first feature. It indicates, that most variability in the features obtained in the first step reduction provides *some* information in the analyzed context. However, consecutive features achieved much greater $coef_{BMC}$ values, indicating lower information content. Average $coef_{BMC}$ across all 183 PCA-provided features was equal to around 0.97. In the MRMR curve, much greater variability of informativeness can be seen, with much lower average $coef_{BMC}$ value, of around 0.87. Note, that features selected by the MRMR are not necessarily uncorrelated, such as in the PCA.

Table 13.1: Summary of the first dimensionality reduction step.

feature set	before reduction	after: raw	after: filtered	after: derivative
Time features	3142	5	4	3
FRM: F_{max}	125 250	14	9	3
FRM: PVn	125 250	4	4	4
FRM: PVr	125 250	3	5	3
FRM: PVs	125 250	4	5	6
FRM: MedF	125 250	14	8	6
FRM: μ_f	125 250	7	11	25
FRM: σ_f	125 250	6	14	3
FRM: $skew_f$	125 250	14	5	7
FRM: $kurt_f$	125 250	3	7	6
FRM: Pn	125 250	5	5	6
FRM: Pr	125 250	5	5	4
FRM: Gn	125 250	5	6	5
FRM: Gr	125 250	6	7	3
FRM: Gs	125 250	5	8	5
FRM: RMSn	125 250	6	7	7
FRM: RMSr	125 250	3	6	3
FRM: RMSs	125 250	7	7	13
FRM: $slope_f$	125 250	9	4	4
FRM: $decrease_f$	125 250	6	13	11
FRM: Hr	125 250	3	5	3
FRM: Hs	125 250	6	6	6
FRM: II	125 250	5	5	5
Frequency ratios and rolloffs	16 485	3	7	6
Time-frequency	34 650	14	4	7
Sum	2 809 777	162	167	154

Average absolute correlation of each feature to all previously selected features was plotted in Figure 13.2.

The highest absolute correlation, slightly above 0.3, was scored by a couple features. However, the average correlation value (indicated by dashed line) was around 0.13. Note, that the MRMR curves from Figures 13.1 and 13.2 are visibly negatively correlated ($\rho = -0.97$), which indicates *positive* correlation between informativeness and redundancy, since $coef_{BMC}$ decreases for more informative features. This correlation is caused by the definition of the MRMR score, which is a difference between informativeness and absolute correlation. In other words, a highly redundant feature can be chosen only if it is highly informative. On the other hand, if the informativeness is relatively low, the feature can be chosen if it provides relatively new information. Of course, features that are highly informative and not

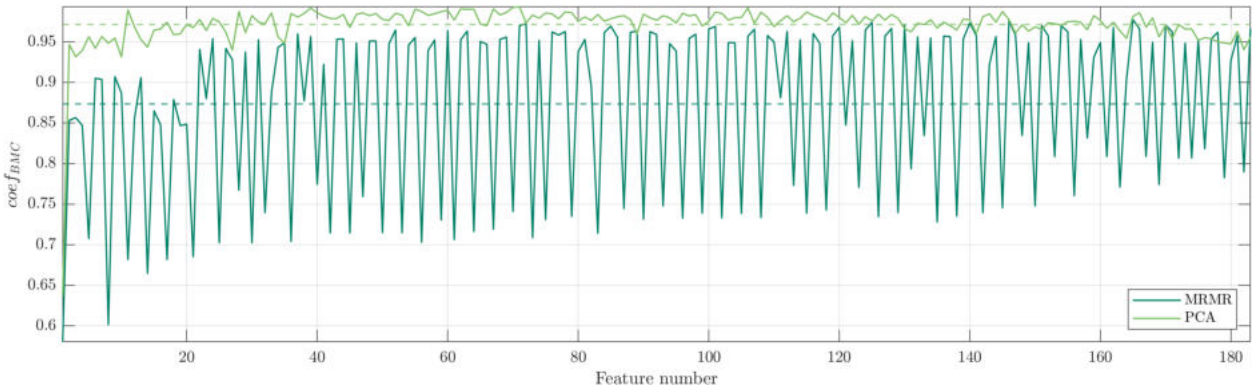

Figure 13.1: Informativeness of consecutive features obtained from the MRMR and the PCA. Average values were indicated by the dashed lines.

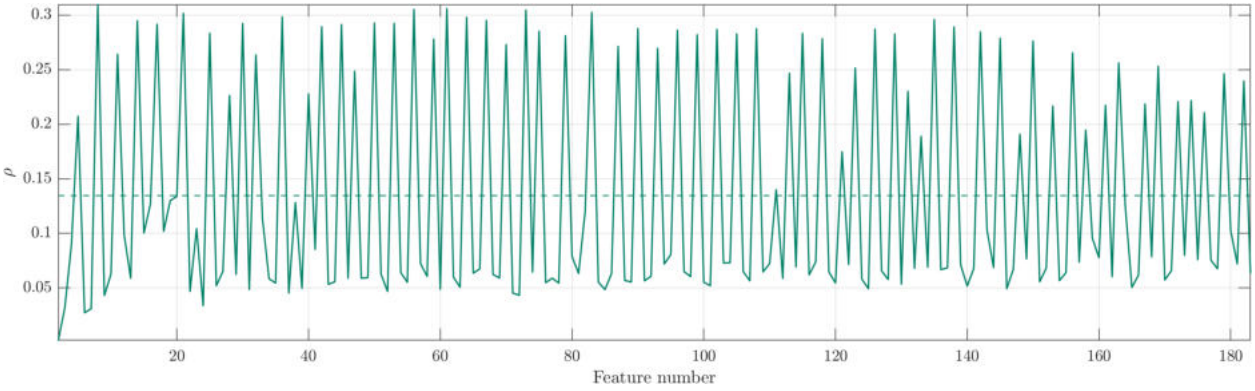

Figure 13.2: Average absolute correlation between each feature and all features previously chosen by the MRMR algorithm.

redundant are always preferred.

The periodicities visible in both plots indicate, that the MRMR first selects the feature least correlated to already selected features, even if it is relatively uninformative. Then, consecutively selected features are correlated more and more, but their informativeness also increases. After couple of features this cycle repeats, and the algorithm selects feature that is, again, relatively uninformative, but also not redundant.

13.2 Classification

After the second step of dimensionality reduction, features provided by both MRMR and PCA were used to train 11 classification algorithms. Each classifier was trained on the number of features increasing from 1 to 183. Dependence of classification accuracies on the number of features used, for each classifier, were plotted in Figure 13.3.

Comparing curves of PCA and MRMR on the plot for the Linear Discriminant Analysis, slightly better performance was achieved using features obtained using PCA. The highest accuracy of around 0.77 was scored for 45 features, after which the accuracy decreased. This could indicate LDA's sensitivity to noisy or redundant features.

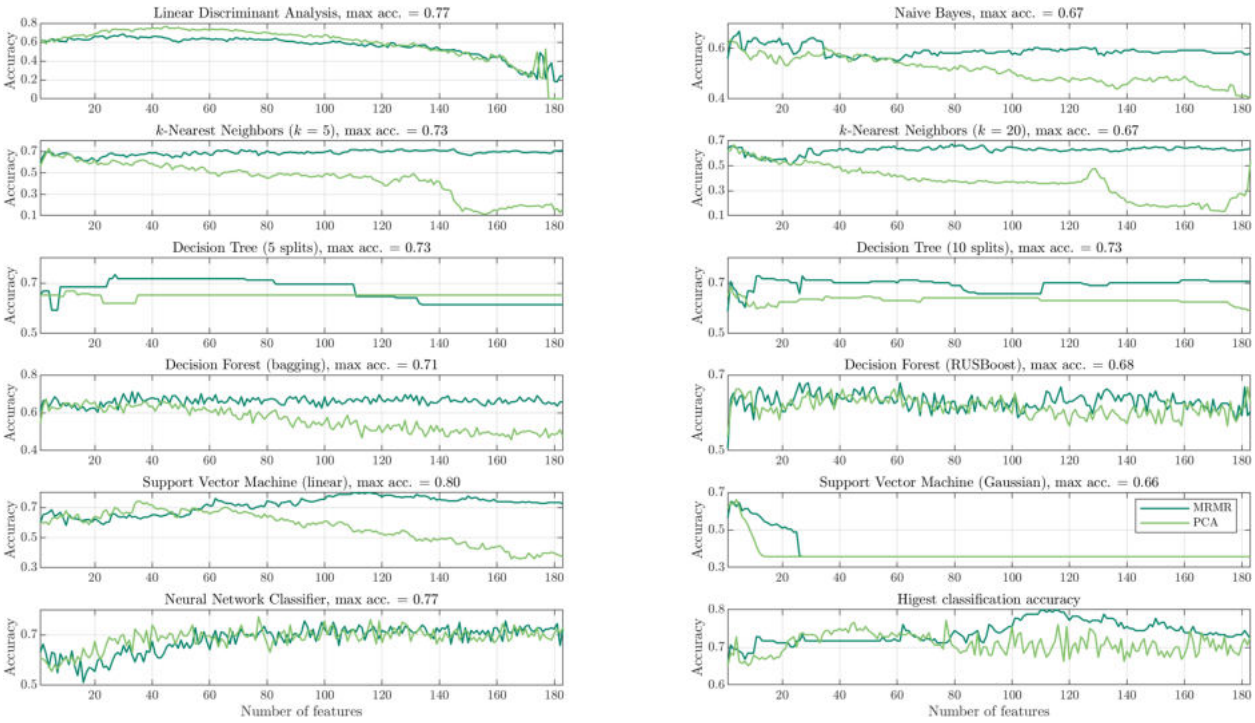

Figure 13.3: Classification results for each classifier.

The Naive Bayes classifier achieved relatively poor performance, with its peak accuracy being roughly 0.67 for 5 best features selected by MRMR algorithm. The classifier showed a decrease in accuracy with increasing number of features for both PCA and MRMR. Interestingly, better performance was achieved by the MRMR, which can be surprising considering assumption of the classifier of uncorrelated features. Since PCA creates uncorrelated features by definition, one would expect it to achieve higher accuracy. Note, however, that lack of correlation provided by the PCA does not ensure high informativeness of the features in the specific context of classification.

In both KNN classifiers, accuracies up to around 20 features were similar between the PCA and MRMR. Above those 20 features, however, performance of PCA significantly dropped, while MRMR increased. That could be caused by relatively low informativeness of features provided by the PCA. Lower neighborhood of 5 neighbors considered in the classification achieved better performance compared to 20 neighbors, indicating that the closer neighbors have a more significant influence on the high classification accuracy. That could suggest denser values distribution within the classes, leading to higher importance of close samples.

Both 5- and 10-split decision trees achieved similar results, with better performance achieved by MRMR-provided features. Note characteristic flat areas in curves. They could be caused by the fact, that despite greater number of features, decision trees had constant maximum number of splits. They were implemented in such way to avoid overfitting. Therefore, when new, less informative features were added, both models only used more informative, "old" features, achieving exact same classification accuracy.

Such flat areas are not visible in the decision forests, which curves seem much more

variable. It could be explained by multiple decision trees used within each forest, each of which was trained on different samples, achieving more diverse performances. Both bagging and boosting approaches provided relatively poor performance compared to other classifiers. For the bagging ensemble approach, MRMR provided better results compared to PCA, while in RUSboost, performances were comparable.

Linear Support Vector Machine proved to achieve highest classification accuracy of around 0.8, when trained on 110 features provided by the MRMR algorithm. The Gaussian model, however, achieved much worse performance, with maximal accuracy of 0.66. With more than 4 features, performance declined with both MRMR- and PCA-provided features. Above 27 features for MRMR and 14 for PCA, the performance dropped to around 0.36, indicating that classifier always chose first class (containing 66 from 184 samples). Much lower performance of the Gaussian SVM, compared to the linear version, could be caused by several factors. First, model complexity: the Gaussian kernel's implicit infinite-dimensional feature space allows for highly flexible decision boundaries. This flexibility, however, can capture intricate variations in the training data that do not generalize well to new data, leading to lower performance even when measured using cross-validation. Another factor could be distribution of the features. If the relationship between the features and the classes is close to linear, the kernel trick used in Gaussian SVM could introduce unnecessary complexity, trying to find nonlinear patterns that are not present.

The NN classifier achieved relatively high accuracy of 0.77. For smaller number of features, up to around 50, those provided by the PCA proved to lead to slightly better performance compared to the MRMR. For higher number of features, accuracy locally varied, but in general stayed at quasi-constant level, for both PCA and MRMR. Those variations could be caused by nonlinearity of the neural network, increasing influence of the number of features on classification performance.

In general, the MRMR provided features that led to higher accuracy for most classifiers—see lowest-right curve in Figure 13.3. It can be explained by the fact, that when selecting features, the MRMR explicitly takes into account their informativeness (or relevance). The PCA, on the other hand, transforms the features into new ones, capturing the most possible amount of variance, while keeping them orthogonal. It is done without the specific context of classification. Variability in features, that is not related to the analyzed context, can be considered noise. This noise does not greatly influence the MRMR feature selection, because it can be present in both informative and uninformative features, as well as redundant and not redundant ones. On the other hand, this noise is amplified by the PCA, potentially outweighing informative variability in features, i.e., variability that actually relate to the classified conditions.

13.3 Summary and conclusions

In summary, there were 8 429 331 features extracted in this work. Reduction of their dimensionality was conducted in two steps. In the first step, features were divided into 75 groups (25 per preprocessing), and in each group, the MRMR algorithm was used to select the feature subset. The number of selected features within groups varied between 3 and 25. In the second step, all those subsets were combined, making up a total of 483 features. To

further reduce this number, two approaches were employed: the MRMR algorithm and the PCA. Both provided 183 features, sorted by importance.

The MRMR provided features of varying informativeness ($coef_{BMC}$ ranging from 0.58 to 0.98 and an average of 0.87), and redundancy (correlation of specific features between 0.00 and 0.31, with average of 0.13). The PCA, on the other hand, provided features uncorrelated by definition, but of relatively poor information content ($coef_{BMC}$ ranging from 0.64 to 0.99, with average of 0.97). That led to generally better performance of the MRMR-provided features in the classification context.

General influence of the number of features on the classification accuracy proved to be highly dependent on the specific classification model. However, in most models trained on features provided by the MRMR, accuracy seemed to plateau above some specific number of features, and decrease for the PCA-provided ones.

Following conclusions can be drawn from the provided research.

1. Features within groups were relatively highly correlated, leading to selection of just few (3 to 25) features per group.
2. The MRMR algorithm proved to provide more informative features compared to PCA, indicated by a lower average $coef_{BMC}$ of around 0.87 vs 0.97 with PCA.
3. There proved to be a positive correlation ($\rho = 0.97$) between the informativeness and redundancy of features selected by the MRMR, indicating that there was not much informative and not-redundant features, and most features was either highly informative and redundant or less informative, but providing relatively new information.
4. There is a visible periodicity in both informativeness and redundancy of consecutive features selected by the MRMR algorithm, indicating that the method alternates between selecting less informative features that are also less redundant and more informative features which are more correlated to already selected ones.
5. For most classification algorithms, features selected by the MRMR algorithm resulted in higher classification accuracies compared to the PCA, as these features were explicitly selected for their high informativeness in the specific context of this work. Transformation of features in the PCA could potentially lead to lowered informativeness of new features, by amplifying variability in features' values that does not contribute to the classification context.

Chapter 14

Summary and conclusions

In the current chapter, a summary of the conducted research will be provided, as well as the limitations of the study and conclusions, which can be derived from it.

14.1 Summary

The conducted research allowed to prove thesis of the work: "The application of digital signal processing methods in vibroarthrographic signal analysis, specifically in the time domain, frequency domain, and time-frequency domain, will enable extraction of features allowing the classification of knee joint conditions with accuracy higher than the current state-of-the-art (0.69)". All major and minor goals were achieved.

In this work, vibroarthrograms acquired from five classes of knee joint conditions were analyzed. The largest part of this work was focused on feature extraction from signals after three preprocessing approaches: filtering, differentiation and no preprocessig. This enabled more direct comparison in terms of the information content each approach highlights or suppresses in a given signal analysis context. After extraction of the features, dimensionality reduction was conducted, followed by classification. The most accurate classifier, Linear Support Vector Machine, achieved around 0.80 accuracy, which is around 0.11 more than the current state-of-the-art achieved by Kręcisz and Bączkowicz [5].

The feature extraction was divided into three major approaches: time, frequency and time-frequency. In each approach, multiple features were extracted, and then thoroughly analyzed in terms of their informativeness. In case of parametric features, influence of the parameters on the information content was also comprehensively investigated and discussed.

In time domain, five major groups of features were extracted. First, statistical, treated the signal as an unordered group of values, whose *distribution* was assumed to be informative. Second group was extracted from a *rolling* signal, i.e., signal constructed from some parameter calculated in a sliding window. In this approach, signals were no longer assumed to be unordered groups of values, and extracted features had inherently *temporal* character. Third group included measures that more explicitly quantified signal variation in time, being calculated using the differences is the consecutive points of the signals. Forth group consisted of features measuring self-similarity of the signal, either in time, using the Autocorrelation Function, or in scale, using the Detrended Fluctuation Analysis. The last, fifth group was defined on the Reconstructed Phase Space, measuring the complexity and nonlinear dynamics of the signal. Features in all groups were extracted from all preprocessing approaches, summing up to a total of 9426 singular features.

To analyze signal in the frequency domain, first, a couple of carefully selected PSD estimation methods were compared. The method allowing extraction of the most informative

features was further used for this purpose. Majority of the frequency-domain features could be understood as being parametric, with the parameters being the lower and upper frequency of the range, on which the specific feature is calculated. Therefore, a set of 22 such features was extracted, each for a comprehensive set of different ranges. Additionally, a visualization method, called the Frequency Range Map, was proposed in order to intuitively present informativeness across various frequency ranges. Majority of features was extracted in such frequency range approach, however, a couple of different parametric features were also obtained. Their parameters were examined in terms of their influence on the feature's information content. Similarly to time-domain, frequency-domain features were extracted separately for each preprocessing approach, yielding a total of 8 315 955 features.

The last approach of the feature extraction was conducted in the time-frequency context, that is, after transforming *parts* of a signal into the frequency domain and quantifying changes of the obtained spectra across time. This approach was utilized using the spectrogram, and the influence of its specific parameters on the extracted features' informativeness were thoroughly investigated. Additional analyzes were conducted in time and frequency context separately, which indicated how specific configurations of spectrogram's parameters depend on the character of the extracted feature. For all preprocessing approaches utilized in this work, time-frequency analysis provided 103 950 separate features.

Combining all features yielded a total of 8 429 331. Extraction of such a great number of features was done primarily to analyze informativeness of parametric features in the context of their parameters. In order to do so, features were extracted for a number of parameters' values with relatively high resolution, or small differences in those values. Such approach enabled intuitive visualization, however, also rendered majority of the obtained features highly correlated. To effectively utilize features for the classification task, they should contain as little redundant information as possible. Therefore, after the feature extraction, the dimensionality reduction was conducted in two stages. In the first stage, the most informative and lest redundant features were selected using the MRMR algorithm, reducing number of features to 483. In the second step, this number was further reduced using both MRMR and PCA algorithms, which allowed direct comparison of those two approaches in the context of classification accuracy. Since distribution of the most informative, selected features was unknown, 11 different classification algorithms were implemented. Finally, influence of the number of features on the classification accuracy was analyzed and discussed.

14.2 Limitations

Although the conducted study was rather comprehensive, some limitations could be identified. First, utilized preprocessing methods were relatively simple, which potentially reduced amount of information extracted from the signal, especially in time domain. Perhaps less frequency-specific approaches, such as the Empirical or Variational Mode Decompositions, already used in VAG-related studies (see Chapter 3), could provide more informative signals in general. Such approaches could influence frequency domain in nonlinear ways, potentially improving also informativeness of the frequency domain features.

Second, informativeness obtained for each feature, measured by the $coef_{BMC}$, was calculated using *all* signals in the signal base. Therefore, in the feature selection step, features that

describe *all* signals in the best possible way were selected. That could potentially create a sort of information leakage, that could artificially increase classification accuracy, even when evaluated using cross-validation. To avoid this limitation, the parameter selection of each parametric feature could also be done in a cross-validation manner. However, two problems would arise in such approach. The first, more practical issue, would be high computational cost. Even training one classifier would be more computationally expensive compared to the computation of $coef_{BMC}$, and training multiple ones needed for cross-validation would rise computational cost much higher. Second, more conceptual issue, is related to the choice of specific classifier. As discussed, different classification algorithms have different strengths and weaknesses in terms of the data distribution. Therefore, selecting one specific classification algorithm would render analyzes less general. That is the exact reason for using $coef_{BMC}$ as an informativeness measure instead of a accuracy of some classifier.

Another potential limitation is the exclusive usage of $coef_{BMC}$ in the MRMR algorithm. Since $coef_{BMC}$ is defined as the weighted average of the $coef_B$ values, MRMR selects features that are well suited to differentiate *all* classes, not their specific pairs. Results of specific features indicated, that some pairs of classes, such as the ctrl and cmp1, or cmp3 and oa are much more difficult to differentiate compared to other pairs. It is not unexpected, since signals in those classes are acquired from the knee joints with similar qualities of movement. However, some features could potentially be better suited to differentiate them than others. Therefore, maybe MRMR could be implemented in a way to measure relevancy not as a singular $coef_{BMC}$, but to successively select features, that are more informative only in the class pair context, that is not well differentiated by already selected features. Such implementation, though, was kept for the future research.

In order to keep the analyzes relatively simple, the time-frequency features were defined on the whole spectrum. More information, however, could potentially be extracted by features defined on a spectrum limited to some specific frequency range, potentially first chosen by the frequency analysis. Such analyzes were also kept for future research, since spectral fluctuations would not necessarily be the most informative in a frequency range chosen in the FRM. Perhaps more general approach would be to choose the frequency range on a basis of a curve similar to one on Figure 11.44. That is, on the average informativeness of some frequency bins in *all* frequency ranges containing this bin.

Finally, a more technical limitation, is a lack of information about the knee joint movement phase synchronized with the VAG signal. Results of the time analysis indicated, that such information is already embedded in a signal, in a form of low-frequency component. More explicit information, however, would allow more direct analyzes of different movement phases. In some publications (see Chapter 3), segmentation of the signal proved to result in more informative features. Therefore, this lack of explicit information about the knee joint movement phase, could also be consider a limitation of this study.

14.3 Conclusions

Conclusions derived from the conducted research could be divided into three categories, regarding: general analysis of the vibroarthrogram, changes in the signal is progressing knee joint conditions and directions for future studies.

14.3.1 General analysis of the vibroarthrogram

Conclusions regarding general analysis of the vibroarthrogram are as follows.

1. In general, most time-domain features proved to be more informative when calculated on the signal filtered in specific frequency range. One exception, however, turned out to be features based on proximity of points (or sub-signals). They seem to benefit from low-frequency movement phase information embedded in the raw signal. Embedding this information in the filtered signal could potentially increase their informativeness even more.
2. The Welch method with window size of 2^{14} samples provided the most informative estimate of the Power Spectral Density.
3. Frequency analysis indicated that various frequency ranges can be informative depending on specific feature calculated. In general, however, frequency range from 10 Hz to 350 Hz proved to be the most informative across all studied features.
4. Analysis of power in the frequency domain provided the most informative features when conducted on the non-normalized spectrum.
5. Moreover, frequency analysis proved to provide the most informative features for the raw signal.
6. Time-frequency analysis, on the other hand, provided the most information when conducted on the filtered signal.
7. In time-frequency analysis, the most informative features proved to be those related to the signal's power, rather than the shape of the PSD.
8. Spectrogram's parameters that provide the most informative features seem to depend on the specific features being extracted. However, in most features greater overlap provided features of greater informativeness. In terms of time and frequency resolution tradeoff, features related specifically to time (like MoAD) seem to benefit from better time resolution, while features measuring central tendencies of the PSD provided the best results for better frequency resolution.
9. The dimensionality reduction part of the study indicated that the MRMR provided more informative features compared to the PCA.
10. From all utilized classification methods, the Linear Support Vector Machine provided the highest classification accuracy, when trained on 110 features.

14.3.2 Changes in the signal with progressing knee joint conditions

Conducted research provided following conclusions regarding VAG signal characteristics in the progressing knee joint conditions.

1. Power of the signal generally increases, which, in time domain, can be quantified by for example ASP or RMS features.
2. Variability also increases, which can be measured by variance, standard deviation, etc.
3. Complexity measures, such as entropy features or LLE also increase.
4. On the other hand, features measuring long-range dependencies of the signal decrease.
5. Power of the signal increases in wide frequency range, yielding increased values of measures such as Spectral Spread or Entropy.
6. However, lower frequencies tend to increase more compared to higher ones, yielding

decreased values of Spectral Slope or Skewness features.
7. Variability of the signal's spectral content increase, which can be measured, for example, by MoAD feature of the Spectral Slope SFS.

14.3.3 Directions for the future VAG-related studies

Future VAG-related research could potentially benefit from implementing the following approaches.

1. Lowered sampling frequency. Conducted study indicated that the most informative frequency range is around from 10 Hz to 350 Hz. Therefore, sampling frequencies as low as 700 Hz would potentially be sufficient for extraction of relatively informative features.
2. Moreover, this frequency range could be used for signal filtering.
3. Augmenting the signal by adding movement phase information could increase informativeness of features based on proximity of sub-signals or samples.
4. Instead of full frequency range, time-frequency analysis conducted in this work could be implemented on a specific frequency range based on the FRMs.

14.3.4 Novelty of the research

The novelty and value of this research lies in the following list of original achievements.

1. Comprehensive examination of VAG features in various domains, including time, frequency, and time-frequency, with a specific focus on their information content. That provides a useful reference for the choice of signals characteristics in the future studies.
2. Explicit calculation of the feature's informativeness, that is, its ability to differentiate between the analyzed classes. It enabled all analyzes to be conducted with the specific focus on the classification context.
3. Comparison of multiple signal and data processing methods, including: different preprocessing approaches, spectrum estimation methods, time-frequency resolution tradeoffs, as well as dimensionality reduction and classification algorithms. This comparison serves as a useful resource for making informed decisions in a general biomedical signal processing chain.
4. Thorough analysis of the parametric features. That is, investigation of all parameters and their influence on the features' informativeness. That enabled optimization of the parameters, as well as better understanding of the changes in the vibroarthrograms in the analyzed context.
5. Dimensionality reduction of the feature space, providing a list of features in the most informative subset (see Appendix B), and improving upon current stat-of-the-art classification accuracy.
6. The research provided conclusions that were specifically grouped into three major categories: the general approach to VAG signal analysis, changes in the signal with progressing degradation of the knee joint, and general directions for future VAG-related studies. This allows for easy comprehension and application of findings.

Appendix A

Detailed results of the PSD estimation comparison

Table A.1: Detailed results of the most informative features per PSD estimation method, per FRM feature, for the raw signal.

feature	period	w512	w1k	w2k	w4k	w8k	w16k	w32k	mt3	mt7	mt15
F_{max}	0.935	0.925	0.927	0.926	0.920	0.921	0.921	0.916	0.917	0.930	0.916
PVn	0.773	0.773	0.768	0.774	0.760	0.770	0.723	0.753	0.754	0.746	0.755
PVr	0.815	0.822	0.821	0.817	0.816	0.829	0.802	0.801	0.801	0.826	0.822
PVs (Crest)	0.887	0.878	0.881	0.873	0.876	0.880	0.888	0.904	0.921	0.909	0.893
Median	0.897	0.897	0.898	0.898	0.900	0.908	0.896	0.913	0.904	0.905	0.902
Centroid	0.812	0.809	0.809	0.810	0.810	0.823	0.816	0.808	0.813	0.813	0.815
Spread	0.733	0.749	0.736	0.737	0.730	0.741	0.736	0.730	0.728	0.733	0.735
Skewness	0.747	0.739	0.737	0.749	0.755	0.765	0.759	0.759	0.761	0.765	0.765
Kurtosis	0.790	0.763	0.772	0.779	0.792	0.800	0.811	0.811	0.804	0.805	0.801
Pn	0.742	0.748	0.743	0.738	0.736	0.755	0.715	0.739	0.730	0.738	0.742
Pr	0.810	0.817	0.812	0.809	0.808	0.818	0.808	0.800	0.812	0.813	0.812
Gn	0.706	0.696	0.700	0.701	0.697	0.720	0.684	0.712	0.700	0.703	0.704
Gr	0.778	0.770	0.773	0.775	0.775	0.788	0.783	0.784	0.779	0.782	0.779
Gs (Flatness)	0.780	0.772	0.775	0.776	0.776	0.789	0.785	0.784	0.781	0.783	0.780
RMSn	0.747	0.751	0.743	0.742	0.739	0.755	0.732	0.745	0.739	0.740	0.744
RMSr	0.807	0.816	0.812	0.812	0.808	0.818	0.818	0.803	0.812	0.815	0.813
RMSs (FF)	0.874	0.873	0.878	0.877	0.877	0.883	0.883	0.890	0.897	0.899	0.894
Slope	0.741	0.746	0.743	0.738	0.735	0.757	0.715	0.735	0.728	0.736	0.741
Decrease	0.940	0.887	0.895	0.891	0.906	0.914	0.910	0.926	0.917	0.926	0.922
Hr	0.795	0.784	0.785	0.788	0.788	0.800	0.790	0.784	0.794	0.797	0.797
Hs	0.869	0.784	0.830	0.875	0.877	0.882	0.878	0.888	0.887	0.897	0.890
Irregularity Index	0.777	0.760	0.754	0.759	0.756	0.769	0.765	0.760	0.771	0.777	0.782
average	0.807	0.798	0.800	0.802	0.802	0.813	0.801	0.807	0.807	0.811	0.809

Table A.2: Detailed results of the most informative features per PSD estimation method, per FRM feature, for the filtered signal.

feature	period	w512	w1k	w2k	w4k	w8k	w16k	w32k	mt3	mt7	mt15
F_{max}	0.856	0.893	0.887	0.883	0.896	0.896	0.864	0.874	0.873	0.891	0.895
PVn	0.775	0.773	0.768	0.775	0.761	0.771	0.723	0.754	0.753	0.746	0.756
PVr	0.903	0.924	0.924	0.918	0.913	0.901	0.907	0.910	0.910	0.906	0.906
PVs (Crest)	0.919	0.919	0.930	0.927	0.919	0.914	0.914	0.921	0.914	0.916	0.914
Median	0.849	0.807	0.864	0.854	0.851	0.846	0.843	0.852	0.847	0.848	0.848
Centroid	0.839	0.831	0.835	0.840	0.845	0.842	0.835	0.845	0.839	0.839	0.840
Spread	0.898	0.882	0.888	0.896	0.897	0.905	0.899	0.913	0.914	0.911	0.904
Skewness	0.871	0.850	0.860	0.869	0.873	0.872	0.870	0.871	0.872	0.872	0.872
Kurtosis	0.934	0.916	0.934	0.935	0.934	0.936	0.930	0.934	0.925	0.934	0.930
Pn	0.731	0.731	0.731	0.728	0.725	0.752	0.715	0.743	0.726	0.731	0.733
Pr	0.876	0.883	0.879	0.877	0.875	0.873	0.867	0.875	0.873	0.874	0.875
Gn	0.705	0.706	0.707	0.703	0.699	0.723	0.687	0.708	0.697	0.703	0.704
Gr	0.823	0.822	0.816	0.822	0.820	0.822	0.816	0.814	0.803	0.797	0.796
Gs (Flatness)	0.886	0.822	0.825	0.828	0.825	0.829	0.829	0.833	0.816	0.793	0.788
RMSn	0.750	0.753	0.746	0.745	0.742	0.757	0.735	0.749	0.742	0.743	0.746
RMSr	0.868	0.871	0.873	0.873	0.870	0.867	0.871	0.884	0.871	0.870	0.871
RMSs (FF)	0.919	0.873	0.901	0.905	0.915	0.918	0.914	0.916	0.915	0.919	0.920
Slope	0.729	0.729	0.728	0.726	0.723	0.750	0.712	0.739	0.723	0.728	0.730
Decrease	0.947	0.867	0.873	0.935	0.924	0.932	0.948	0.952	0.953	0.942	0.911
Hr	0.877	0.881	0.881	0.880	0.878	0.875	0.869	0.875	0.875	0.875	0.876
Hs	0.902	0.850	0.885	0.890	0.898	0.899	0.908	0.919	0.901	0.900	0.897
Irregularity Index	0.875	0.878	0.874	0.873	0.872	0.871	0.866	0.873	0.872	0.872	0.873
average	0.851	0.839	0.846	0.849	0.848	0.852	0.842	0.852	0.846	0.846	0.845

Table A.3: Detailed results of the most informative features per PSD estimation method, per FRM feature, for the derivative signal.

feature	period	w512	w1k	w2k	w4k	w8k	w16k	w32k	mt3	mt7	mt15
F_{max}	0.805	0.752	0.822	0.808	0.810	0.810	0.798	0.795	0.799	0.806	0.806
PVn	0.761	0.764	0.760	0.761	0.760	0.763	0.738	0.758	0.757	0.746	0.755
PVr	0.884	0.894	0.889	0.887	0.886	0.883	0.881	0.890	0.882	0.881	0.884
PVs (Crest)	0.853	0.885	0.864	0.848	0.850	0.852	0.850	0.867	0.850	0.851	0.845
Median	0.788	0.799	0.777	0.791	0.791	0.797	0.787	0.809	0.792	0.789	0.791
Centroid	0.759	0.742	0.756	0.765	0.763	0.773	0.762	0.782	0.765	0.761	0.762
Spread	0.892	0.893	0.892	0.895	0.894	0.895	0.893	0.901	0.891	0.894	0.894
Skewness	0.771	0.751	0.765	0.776	0.778	0.785	0.772	0.793	0.774	0.772	0.774
Kurtosis	0.876	0.878	0.876	0.878	0.877	0.876	0.874	0.885	0.881	0.878	0.880
Pn	0.742	0.744	0.739	0.732	0.737	0.760	0.723	0.740	0.732	0.739	0.740
Pr	0.873	0.874	0.873	0.874	0.875	0.872	0.863	0.869	0.868	0.872	0.873
Gn	0.706	0.701	0.701	0.700	0.701	0.724	0.684	0.712	0.699	0.704	0.704
Gr	0.882	0.877	0.877	0.880	0.883	0.884	0.882	0.897	0.883	0.883	0.882
Gs (Flatness)	0.911	0.918	0.917	0.919	0.914	0.914	0.910	0.915	0.918	0.919	0.918
RMSn	0.753	0.755	0.752	0.746	0.750	0.765	0.738	0.750	0.746	0.749	0.752
RMSr	0.863	0.873	0.872	0.872	0.868	0.873	0.865	0.876	0.869	0.871	0.872
RMSs (FF)	0.871	0.894	0.881	0.874	0.876	0.873	0.873	0.877	0.875	0.873	0.874
Slope	0.783	0.768	0.772	0.769	0.772	0.774	0.766	0.768	0.774	0.776	0.780
Decrease	0.890	0.766	0.784	0.833	0.845	0.858	0.894	0.918	0.913	0.872	0.862
Hr	0.876	0.874	0.874	0.875	0.878	0.879	0.873	0.878	0.875	0.877	0.877
Hs	0.895	0.898	0.896	0.896	0.895	0.891	0.888	0.894	0.895	0.894	0.894
Irregularity Index	0.871	0.868	0.868	0.869	0.869	0.867	0.860	0.867	0.866	0.869	0.870
average	0.832	0.826	0.828	0.830	0.831	0.835	0.826	0.838	0.832	0.831	0.831

Appendix B

List of the most informative features

Table B.1: List of the most informative features selected by the MRMR algorithm.

no.	pp	feature name	feature parameters	$coef_{BMC}$
1	F	RAV	-	0.580
2	F	Slope	$f_L = 0$ Hz, $f_U = 300$ Hz	0.853
3	D	Slope	$f_L = 270$ Hz, $f_U = 300$ Hz	0.857
4	D	Slope	$f_L = 190$ Hz, $f_U = 200$ Hz	0.847
5	R	Singularity	m = 1, q = 2	0.708
6	D	SampEn	m = 3, r = max	0.905
7	D	ApEn	m = 4, r = 0.2	0.904
8	F	MoAD	-	0.601
9	R	Slope	$f_L = 490$ Hz, $f_U = 510$ Hz	0.907
10	R	Slope	$f_L = 110$ Hz, $f_U = 120$ Hz	0.887
11	D	Gn	$f_L = 0$ Hz, $f_U = 350$ Hz	0.682
12	R	Slope	$f_L = 140$ Hz, $f_U = 160$ Hz	0.856
13	R	Slope	$f_L = 640$ Hz, $f_U = 670$ Hz	0.906
14	R	LLE_{zc}	-	0.665
15	R	Slope	$f_L = 290$ Hz, $f_U = 300$ Hz	0.865
16	R	Slope	$f_L = 280$ Hz, $f_U = 320$ Hz	0.848
17	R	Gn	$f_L = 0$ Hz, $f_U = 350$ Hz	0.682
18	F	Slope	$f_L = 390$ Hz, $f_U = 400$ Hz	0.879
19	R	Ratio L2H	$f_{th} = 150$ Hz	0.847
20	R	Slope	$f_L = 240$ Hz, $f_U = 250$ Hz	0.849
21	F	Gn	$f_L = 0$ Hz, $f_U = 360$ Hz	0.685
22	R	m	-	0.941
23	F	Slope	$f_L = 210$ Hz, $f_U = 230$ Hz	0.880
24	D	Spectral Slope SFS—CV	ws = 16384, ovl = 0.1	0.954
25	R	Slope	$f_L = 80$ Hz, $f_U = 3550$ Hz	0.702
26	F	Spectral Slope SFS—MF	ws = 32, ovl = 0.4	0.942
27	D	Spectral Skewness SFS—CV	ws = 2048, ovl = 0.6	0.928
28	R	Skewness	$f_L = 0$ Hz, $f_U = 3940$ Hz	0.767
29	R	Spread	$f_L = 4550$ Hz, $f_U = 4760$ Hz	0.937
30	R	Slope	$f_L = 80$ Hz, $f_U = 3390$ Hz	0.702
31	R	Spread	$f_L = 2400$ Hz, $f_U = 2510$ Hz	0.953
32	D	Spectral Decrease SFS—TCR	ws = 128, ovl = 0.9	0.740

Continued on next page

Table B.1: List of the most informative features selected by the MRMR algorithm. (Continued)

no.	pp	feature name	feature parameters	$coef_{BMC}$
33	D	Spread	$f_L = 0$ Hz, $f_U = 230$ Hz	0.888
34	D	Gn SFS—complexity	ws = 4096, ovl = 0.2	0.943
35	F	Hs	$f_L = 4520$ Hz, $f_U = 4530$ Hz	0.949
36	R	Pn	$f_L = 80$ Hz, $f_U = 1780$ Hz	0.704
37	D	Gn SFS—MF	ws = 2048, ovl = 0.3	0.960
38	R	Gr	$f_L = 420$ Hz, $f_U = 430$ Hz	0.877
39	F	Decrease	$f_L = 3530$ Hz, $f_U = 3600$ Hz	0.956
40	R	Spectral Decrease SFS—skewness	ws = 64, ovl = 0.6	0.775
41	F	Median	$f_L = 3610$ Hz, $f_U = 3660$ Hz	0.922
42	R	Spectral Slope SFS—MoAD	ws = 16, ovl = 0.9	0.714
43	D	Decrease	$f_L = 800$ Hz, $f_U = 820$ Hz	0.953
44	R	Pn SFS—kurtosis	ws = 16384, ovl = 0.0	0.954
45	F	PVn	$f_L = 230$ Hz, $f_U = 270$ Hz	0.714
46	D	Spread	$f_L = 2330$ Hz, $f_U = 2590$ Hz	0.949
47	R	Skewness	$f_L = 0$ Hz, $f_U = 2840$ Hz	0.759
48	F	Median	$f_L = 1530$ Hz, $f_U = 4220$ Hz	0.951
49	R	Skewness	$f_L = 1560$ Hz, $f_U = 1650$ Hz	0.951
50	R	Pn SFS—MoAD	ws = 16, ovl = 0.9	0.715
51	R	Spread	$f_L = 1390$ Hz, $f_U = 1530$ Hz	0.947
52	D	Kurtosis	$f_L = 4710$ Hz, $f_U = 4900$ Hz	0.965
53	R	PVn	$f_L = 230$ Hz, $f_U = 270$ Hz	0.715
54	F	ApEn	r_{fixed}, m10, r0.18	0.946
55	R	Median	$f_L = 4190$ Hz, $f_U = 4500$ Hz	0.955
56	F	Spectral Slope SFS—MoAD	ws = 16, ovl = 0.6	0.703
57	F	F_{max}	$f_L = 4760$ Hz, $f_U = 4770$ Hz	0.940
58	R	Median	$f_L = 3780$ Hz, $f_U = 3820$ Hz	0.952
59	D	Centroid	$f_L = 0$ Hz, $f_U = 130$ Hz	0.731
60	R	Median	$f_L = 690$ Hz, $f_U = 710$ Hz	0.964
61	F	Slope	$f_L = 80$ Hz, $f_U = 970$ Hz	0.706
62	R	Skewness	$f_L = 2700$ Hz, $f_U = 2740$ Hz	0.953
63	F	Centroid	$f_L = 1630$ Hz, $f_U = 1710$ Hz	0.963
64	D	PVn	$f_L = 220$ Hz, $f_U = 270$ Hz	0.716
65	R	Skewness	$f_L = 3210$ Hz, $f_U = 3340$ Hz	0.951
66	F	Gs (Flatness)	$f_L = 4750$ Hz, $f_U = 4760$ Hz	0.947
67	R	RMSn SFS—MoAD	ws = 16, ovl = 0.9	0.719
68	F	Decrease	$f_L = 610$ Hz, $f_U = 750$ Hz	0.953
69	F	Gs (Flatness)	$f_L = 50$ Hz, $f_U = 60$ Hz	0.956
70	D	Hurst	m = 1, q = 2	0.741
71	D	Centroid	$f_L = 4410$ Hz, $f_U = 4450$ Hz	0.970
72	F	Centroid	$f_L = 3350$ Hz, $f_U = 3370$ Hz	0.972

Continued on next page

Table B.1: List of the most informative features selected by the MRMR algorithm. (Continued)

no.	pp	feature name	feature parameters	$coef_{BMC}$
73	F	Pn	$f_L = 80$ Hz, $f_U = 1810$ Hz	0.709
74	F	Gs (Flatness)	$f_L = 790$ Hz, $f_U = 830$ Hz	0.952
75	D	Centroid	$f_L = 0$ Hz, $f_U = 120$ Hz	0.731
76	F	Spread	$f_L = 1250$ Hz, $f_U = 1370$ Hz	0.962
77	R	F_{max}	$f_L = 250$ Hz, $f_U = 600$ Hz	0.958
78	R	Skewness	$f_L = 1750$ Hz, $f_U = 1770$ Hz	0.963
79	D	Centroid	$f_L = 0$ Hz, $f_U = 140$ Hz	0.735
80	R	Spectral Flux Non-normalized SFS—kurtosis	ws = 32, ovl = 0.6	0.938
81	D	Skewness	$f_L = 1560$ Hz, $f_U = 1650$ Hz	0.953
82	R	PVr	$f_L = 540$ Hz, $f_U = 550$ Hz	0.895
83	D	Pn	$f_L = 40$ Hz, $f_U = 360$ Hz	0.714
84	R	Skewness	$f_L = 2910$ Hz, $f_U = 3010$ Hz	0.962
85	F	Decrease	$f_L = 1440$ Hz, $f_U = 1500$ Hz	0.969
86	D	Skewness	$f_L = 4180$ Hz, $f_U = 4470$ Hz	0.956
87	D	Skewness	$f_L = 0$ Hz, $f_U = 120$ Hz	0.745
88	F	Median	$f_L = 690$ Hz, $f_U = 710$ Hz	0.962
89	D	Centroid	$f_L = 3100$ Hz, $f_U = 3460$ Hz	0.963
90	R	RMSn	$f_L = 210$ Hz, $f_U = 1780$ Hz	0.731
91	D	Centroid	$f_L = 1220$ Hz, $f_U = 1240$ Hz	0.963
92	D	Centroid	$f_L = 2410$ Hz, $f_U = 2480$ Hz	0.959
93	D	Skewness	$f_L = 0$ Hz, $f_U = 130$ Hz	0.748
94	R	Skewness	$f_L = 790$ Hz, $f_U = 820$ Hz	0.948
95	F	Median	$f_L = 980$ Hz, $f_U = 1030$ Hz	0.939
96	R	Hs	$f_L = 10$ Hz, $f_U = 2000$ Hz	0.733
97	R	Skewness	$f_L = 4770$ Hz, $f_U = 4800$ Hz	0.954
98	R	Median	$f_L = 1450$ Hz, $f_U = 1480$ Hz	0.959
99	R	RMSn	$f_L = 240$ Hz, $f_U = 270$ Hz	0.739
100	F	Kurtosis	$f_L = 3510$ Hz, $f_U = 3600$ Hz	0.966
101	R	Skewness	$f_L = 4840$ Hz, $f_U = 4860$ Hz	0.969
102	R	Hs	$f_L = 10$ Hz, $f_U = 1500$ Hz	0.733
103	D	Skewness	$f_L = 3220$ Hz, $f_U = 3340$ Hz	0.949
104	D	Decrease	$f_L = 0$ Hz, $f_U = 90$ Hz	0.949
105	D	Centroid	$f_L = 0$ Hz, $f_U = 110$ Hz	0.738
106	D	Centroid	$f_L = 3750$ Hz, $f_U = 3850$ Hz	0.957
107	D	Median	$f_L = 1810$ Hz, $f_U = 1950$ Hz	0.966
108	F	RMSn	$f_L = 210$ Hz, $f_U = 330$ Hz	0.734
109	R	Decrease	$f_L = 2320$ Hz, $f_U = 2530$ Hz	0.958
110	F	Median	$f_L = 3750$ Hz, $f_U = 3780$ Hz	0.950

Continued on next page

Table B.1: List of the most informative features selected by the MRMR algorithm. (Continued)

no.	pp	feature name	feature parameters	$coef_{BMC}$
111	R	Spectral Flux Normalized SFS—median	ws = 64, ovl = 0.9	0.881
112	D	Median	f_L = 2930 Hz, f_U = 3000 Hz	0.963
113	D	PVn	f_L = 100 Hz, f_U = 130 Hz	0.773
114	R	Median	f_L = 1550 Hz, f_U = 1670 Hz	0.953
115	F	Pn	f_L = 250 Hz, f_U = 260 Hz	0.739
116	F	Decrease	f_L = 4040 Hz, f_U = 4050 Hz	0.960
117	R	Centroid	f_L = 4270 Hz, f_U = 4460 Hz	0.948
118	D	Median	f_L = 0 Hz, f_U = 120 Hz	0.743
119	R	Centroid	f_L = 3850 Hz, f_U = 3890 Hz	0.957
120	F	Decrease	f_L = 3390 Hz, f_U = 3410 Hz	0.968
121	R	Spectral Kurtosis SFS—mean	ws = 16384, ovl = 0.1	0.847
122	F	F_{max}	f_L = 3780 Hz, f_U = 3870 Hz	0.952
123	F	PVn	f_L = 110 Hz, f_U = 130 Hz	0.771
124	D	Decrease	f_L = 4790 Hz, f_U = 4950 Hz	0.965
125	D	Decrease	f_L = 1140 Hz, f_U = 1150 Hz	0.974
126	R	Spread	f_L = 0 Hz, f_U = 360 Hz	0.734
127	R	Decrease	f_L = 3140 Hz, f_U = 3390 Hz	0.957
128	D	Centroid	f_L = 3950 Hz, f_U = 4030 Hz	0.966
129	R	Pn	f_L = 220 Hz, f_U = 270 Hz	0.740
130	F	Gr	f_L = 2870 Hz, f_U = 3900 Hz	0.971
131	D	Spectral Slope SFS—median	ws = 128, ovl = 0.0	0.794
132	D	Centroid	f_L = 1330 Hz, f_U = 1380 Hz	0.956
133	F	PVr	f_L = 0 Hz, f_U = 10 Hz	0.834
134	F	Ratio L2H	f_{th} = 505 Hz	0.955
135	D	RMSn	f_L = 70 Hz, f_U = 330 Hz	0.728
136	F	Decrease	f_L = 4030 Hz, f_U = 4060 Hz	0.957
137	D	Centroid	f_L = 1540 Hz, f_U = 1640 Hz	0.956
138	R	Spread	f_L = 0 Hz, f_U = 350 Hz	0.735
139	F	Kurtosis	f_L = 50 Hz, f_U = 60 Hz	0.954
140	D	Decrease	f_L = 1470 Hz, f_U = 1480 Hz	0.973
141	D	Centroid	f_L = 3850 Hz, f_U = 3890 Hz	0.958
142	R	Irregularity Index	f_L = 0 Hz, f_U = 40 Hz	0.739
143	R	SampEn	m = 9, r = 0.0	0.922
144	D	Centroid	f_L = 270 Hz, f_U = 300 Hz	0.957
145	D	Centroid	f_L = 0 Hz, f_U = 150 Hz	0.746
146	D	Centroid	f_L = 2730 Hz, f_U = 2750 Hz	0.976
147	R	Spread	f_L = 3080 Hz, f_U = 3200 Hz	0.958
148	D	Slope	f_L = 60 Hz, f_U = 330 Hz	0.835
149	D	Centroid	f_L = 4280 Hz, f_U = 4450 Hz	0.949

Continued on next page

Table B.1: List of the most informative features selected by the MRMR algorithm. (Continued)

no.	pp	feature name	feature parameters	$coef_{BMC}$
150	D	Median	$f_L = 0$ Hz, $f_U = 130$ Hz	0.748
151	R	Centroid	$f_L = 2530$ Hz, $f_U = 2670$ Hz	0.970
152	F	Spread	$f_L = 3180$ Hz, $f_U = 3220$ Hz	0.958
153	R	Gn	$f_L = 340$ Hz, $f_U = 350$ Hz	0.809
154	D	Decrease	$f_L = 4740$ Hz, $f_U = 4990$ Hz	0.970
155	R	Median	$f_L = 3020$ Hz, $f_U = 3140$ Hz	0.963
156	D	Slope	$f_L = 0$ Hz, $f_U = 160$ Hz	0.760
157	D	Skewness	$f_L = 3770$ Hz, $f_U = 3850$ Hz	0.953
158	F	RMSr	$f_L = 0$ Hz, $f_U = 10$ Hz	0.832
159	D	Kurtosis	$f_L = 1220$ Hz, $f_U = 1260$ Hz	0.931
160	F	Centroid	$f_L = 790$ Hz, $f_U = 820$ Hz	0.949
161	F	Gn	$f_L = 340$ Hz, $f_U = 350$ Hz	0.809
162	D	Hs	$f_L = 30$ Hz, $f_U = 40$ Hz	0.968
163	D	Pn	$f_L = 100$ Hz, $f_U = 160$ Hz	0.771
164	D	Ratio L2H	$f_{th} = 32$ Hz	0.903
165	D	Centroid	$f_L = 3440$ Hz, $f_U = 3600$ Hz	0.977
166	F	Spread	$f_L = 2300$ Hz, $f_U = 2360$ Hz	0.966
167	D	Gn	$f_L = 340$ Hz, $f_U = 350$ Hz	0.809
168	F	Spread	$f_L = 4290$ Hz, $f_U = 4310$ Hz	0.949
169	D	F_{max}	$f_L = 0$ Hz, $f_U = 130$ Hz	0.774
170	F	Gr	$f_L = 2820$ Hz, $f_U = 4100$ Hz	0.971
171	D	Gs (Flatness)	$f_L = 800$ Hz, $f_U = 840$ Hz	0.963
172	R	Kurtosis	$f_L = 0$ Hz, $f_U = 330$ Hz	0.807
173	R	Kurtosis	$f_L = 3850$ Hz, $f_U = 3930$ Hz	0.948
174	D	Pn	$f_L = 340$ Hz, $f_U = 350$ Hz	0.807
175	D	Median	$f_L = 4680$ Hz, $f_U = 4780$ Hz	0.953
176	D	Decrease	$f_L = 10$ Hz, $f_U = 180$ Hz	0.818
177	F	Skewness	$f_L = 1560$ Hz, $f_U = 1660$ Hz	0.954
178	F	F_{max}	$f_L = 2680$ Hz, $f_U = 2830$ Hz	0.962
179	R	Skewness	$f_L = 0$ Hz, $f_U = 760$ Hz	0.783
180	F	Kurtosis	$f_L = 1220$ Hz, $f_U = 1260$ Hz	0.927
181	F	Spread	$f_L = 3370$ Hz, $f_U = 3380$ Hz	0.958
182	R	PVn	$f_L = 120$ Hz, $f_U = 130$ Hz	0.790
183	D	Centroid	$f_L = 4770$ Hz, $f_U = 4890$ Hz	0.967

Bibliography

[1] J. H. Park, J. H. Moon, H. J. Kim, M. H. Kong, and Y. H. Oh. Sedentary Lifestyle: Overview of Updated Evidence of Potential Health Risks. *Korean Journal of Family Medicine*, 41(6):365–373, 2020.

[2] P. K. Levangie and C. C. Norkin, editors. *Joint Structure and Function: A Comprehensive Analysis*. F.A. Davis Co, Philadelphia, 5th edition, 2011.

[3] S. E. Putnis, J. P. Linklater, and B. A. Fritsch. Plain Radiography of the Knee. In R. F. LaPrade and J. Chahla, editors, *Evidence-Based Management of Complex Knee Injuries*, pages 6–28. Elsevier, Philadelphia, 2022.

[4] E. C. Argentieri, K. C. Zochowski, T. W. Braun, A. J. Burge, and H. G. Potter. Magnetic Resonance Imaging of the Knee: Conventional and Novel Techniques. In R. F. LaPrade and J. Chahla, editors, *Evidence-Based Management of Complex Knee Injuries*, pages 29–49. Elsevier, Philadelphia, 2022.

[5] K. Kręcisz and D. Bączkowicz. Analysis and multiclass classification of pathological knee joints using vibroarthrographic signals. *Computer Methods and Programs in Biomedicine*, 154:37–44, 2017.

[6] I. A. Kapandji. *The Physiology of the Joints*. Churchill Livingstone, Edinburgh ; New York, 6th edition, 2007.

[7] A. Trepczynski, I. Kutzner, E. Kornaropoulos, W. R. Taylor, G. N. Duda, G. Bergmann, and M. O. Heller. Patellofemoral joint contact forces during activities with high knee flexion. *Journal of Orthopaedic Research: Official Publication of the Orthopaedic Research Society*, 30(3):408–415, 2012.

[8] K. S. Halonen, M. E. Mononen, J. S. Jurvelin, J. Töyräs, J. Salo, and R. K. Korhonen. Deformation of articular cartilage during static loading of a knee joint–experimental and finite element analysis. *Journal of Biomechanics*, 47(10):2467–2474, 2014.

[9] Y. Wu. *Knee Joint Vibroarthrographic Signal Processing and Analysis*. SpringerBriefs in Bioengineering. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[10] R. M. Rangayyan. *Biomedical Signal Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2nd edition, 2015.

[11] S. F. Habusta, R. Coffey, S. Ponnarasu, A. Mabrouk, and E. E. Griffin. Chondromalacia Patella. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2023.

[12] R. E. Outerbridge. The etiology of chondromalacia patellae. *The Journal of Bone and Joint Surgery. British Volume*, 43-B:752–757, 1961.

[13] L. Sharma. Osteoarthritis of the Knee. *New England Journal of Medicine*, 384(1):51–59, 2021.

[14] R. Karpiński, P. Krakowski, J. Jonak, A. Machrowska, M. Maciejewski, and A. Nogalski. Diagnostics of Articular Cartilage Damage Based on Generated Acoustic Signals Using ANN—Part I: Femoral-Tibial Joint. *Sensors*, 22(6):2176, 2022.

[15] R. Karpiński, P. Krakowski, J. Jonak, A. Machrowska, M. Maciejewski, and A. Nogalski. Diagnostics of Articular Cartilage Damage Based on Generated Acoustic Signals Using ANN—Part II: Patellofemoral Joint. *Sensors*, 22(10):3765, 2022.

[16] E. Hengeveld and K. Banks, editors. *Maitland's Peripheral Manipulation: Management of Neuromusculoskeletal Disorders*. Elsevier, Edinburgh, 5th edition, 2013.

[17] C. C. Norkin and D. J. White. *Measurement of Joint Motion: A Guide to Goniometry*. F.A. Davis Company, Philadelphia, 5th edition, 2016.

[18] A. M. Davies and M. E. Abd-El-Bagi, editors. *Imaging of the Knee: Techniques and Applications*. Medical Radiology / Ed. Board L. W. Brady. Springer, Berlin Heidelberg, 2002.

[19] I. Damasena and T. Spalding. Computed Tomography of the Knee Joint: Indications and Significance. In R. F. LaPrade and J. Chahla, editors, *Evidence-Based Management of Complex Knee Injuries*, pages 50–65. Elsevier, Philadelphia, 2022.

[20] A. M. Ulasli, L. Ozcakar, and W. D. Murrel. Ultrasound imaging and guidance in the management of knee osteoarthritis in regenerative medicine field. *Journal of Clinical Orthopaedics and Trauma*, 10(1):24–31, 2019.

[21] A. G. Culvenor, C. N. Engen, B. E. Øiestad, L. Engebretsen, and M. A. Risberg. Defining the presence of radiographic knee osteoarthritis: A comparison between the Kellgren and Lawrence system and OARSI atlas criteria. *Knee Surgery, Sports Traumatology, Arthroscopy*, 23(12):3532–3539, 2015.

[22] J. Petch, S. Di, and W. Nelson. Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Canadian Journal of Cardiology*, 38(2):204–213, 2022.

[23] W. E. Blodgett. Auscultation of the Knee Joint. *The Boston Medical and Surgical Journal*, 146(3):63–66, 1902.

[24] M. L. Chu, I. A. Gradisar, M. R. Railey, and G. F. Bowling. An electro-acoustical technique for the detection of knee joint noise. *Medical Research Engineering*, 12(1):18–20, 1976.

[25] S. C. Abbott and M. D. Cole. Vibration Arthrometry: A Critical Review. *Critical Reviews in Biomedical Engineering*, 41(3):223–242, 2013.

[26] R. A. Mollan, G. C. McCullagh, and R. I. Wilson. A critical appraisal of auscultation of human joints. *Clinical Orthopaedics and Related Research*, (170):231–237, 1982.

[27] W. G. Kernohan, D. E. Beverland, G. F. Mccoy, S. N. Shaw, R. G. H. Wallace, G. C. Mccullagh, and R. Mollan. The Diagnostic Potential of Vibration Arthrography. *Clinical Orthopaedics and Related Research*, 210:106–112, 1986.

[28] W. G. Kernohan, D. E. Beverland, G. F. McCoy, A. Hamilton, P. Watson, and R. Mollan. Vibration arthrometry. *Acta Orthopaedica Scandinavica*, 61(1):70–79, 1990.

[29] Z. Q. Liu, Y. T. Zhang, K. Ladly, C. B. Frank, R. M. Rangayyan, and G. D. Bell. Reduction of interference in knee sound signals by adaptive filtering. In *[1990] Proceedings. Third Annual IEEE Symposium on Computer-Based Medical Systems*, pages 389–396, 1990.

[30] Y. T. Zhang, R. M. Rangayyan, C. B. Frank, G. D. Bell, K. O. Ladly, and Z. Q. Liu. Classification of knee sound signals using neural networks: A preliminary study. In *Proceedings of IASTED Int Symp Expert Systems*, volume 60 of *2*, Honolulu, 1990.

[31] S. Tavathia, R. M. Rangayyan, C. B. Frank, G. D. Bell, K. O. Ladly, and Y. T. Zhang. Analysis of knee vibration signals using linear prediction. *IEEE Transactions on Biomedical Engineering*, 39(9):959–970, 1992.

[32] A. Peylan. Direct auscultation of the joints; preliminary clinical observations. *Rheumatism*, 9(4):77–81, 1953.

[33] Y. T. Zhang, C. B. Frank, R. M. Rangayyan, and G. D. Bell. Mathematical modeling and spectrum analysis of the physiological patello-femoral pulse train produced by slow knee movement. *IEEE Transactions on Biomedical Engineering*, 39(9):971–979, 1992.

[34] K. Kalo, D. Niederer, F. Stief, L. Würzberger, S. van Drongelen, A. Meurer, and L. Vogt. Validity of and recommendations for knee joint acoustic assessments during different movement conditions. *Journal of Biomechanics*, 109:109939, 2020.

[35] D. Bączkowicz, G. Skiba, K. Falkowski, P. Domaszewski, and N. Selkow. Effects of Immobilization and Re-Mobilization on Knee Joint Arthrokinematic Motion Quality. *Journal of Clinical Medicine*, 9(2):451, 2020.

[36] Y. T. Zhang, K. O. Ladly, R. M. Rangayyan, C. B. Frank, G. D. Bell, and Z. Q. Liu. Muscle Contraction Interference In Acceleration Vibroarthrography. In *[1990] Proceedings of the Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2150–2151, Philadelphia, PA, USA, 1990. IEEE.

[37] R. E. Andersen, L. Arendt-Nielsen, and P. Madeleine. Knee joint vibroarthrography of asymptomatic subjects during loaded flexion-extension movements. *Medical & Biological Engineering & Computing*, 56(12):2301–2312, 2018.

[38] K. Kalo, D. Niederer, R. Sus, K. Sohrabi, W. Banzer, V. Groß, and L. Vogt. Reprint of "The detection of knee joint sounds at defined loads by means of vibroarthrography". *Clinical Biomechanics*, 79:105175, 2020.

[39] N. B. Bolus, H. K. Jeong, D. C. Whittingslow, and O. T. Inan. A Glove-Based Form Factor for Collecting Joint Acoustic Emissions: Design and Validation. *Sensors*, 19(12):2683, 2019.

[40] E. Ołowiana, N. Selkow, K. Laudner, D. Puciato, and D. Bączkowicz. Vibroarthrographic analysis of patellofemoral joint arthrokinematics during squats with increasing external loads. *BMC Sports Science, Medicine and Rehabilitation*, 12(1):51, 2020.

[41] R. Gong, H. Ohtsu, K. Hase, and S. Ota. Vibroarthrographic signals for the low-cost and computationally efficient classification of aging and healthy knees. *Biomedical Signal Processing and Control*, 70:103003, 2021.

[42] D. Bączkowicz, E. Majorczyk, and K. Kręcisz. Age-Related Impairment of Quality of Joint Motion in Vibroarthrographic Signal Analysis. *BioMed Research International*, 2015:1–7, 2015.

[43] K. Kręcisz, D. Bączkowicz, and A. Kawala-Sterniuk. Using Nonlinear Vibroartrographic Parameters for Age-Related Changes Assessment in Knee Arthrokinematics. *Sensors*, 22(15):5549, 2022.

[44] R. M. Rangayyan, S. Krishnan, G. D. Bell, C. B. Frank, and K. O. Ladly. Parametric representation and screening of knee joint vibroarthrographic signals. *IEEE Transactions on Biomedical Engineering*, 44(11):1068–1074, 1997.

[45] S. Krishnan, R. M. Rangayyan, G. D. Bell, C. B. Frank, and K. O. Ladly. Adaptive filtering, modelling and classification of knee joint vibroarthrographic signals for non-invasive diagnosis of articular cartilage pathology. *Medical & Biological Engineering & Computing*, 35(6):677–684, 1997.

[46] Y. Wu, S. Krishnan, and R. M. Rangayyan. Computer-Aided Diagnosis of Knee-Joint Disorders via Vibroarthrographic Signal Analysis: A Review. *Critical Reviews™ in Biomedical Engineering*, 38(2):201–224, 2010.

[47] S. Krishnan, R. M. Rangayyan, G. D. Bell, C. B. Frank, and K. O. Ladly. Screening of knee joint vibroarthrographic signals by statistical pattern analysis of dominant poles. In *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 3, pages 968–969, Amsterdam, Netherlands, 1997. IEEE.

[48] S. Krishnan and R. M. Rangayyan. Denoising knee joint vibration signals using adaptive time-frequency representations. In *Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No.99TH8411)*, volume 3, pages 1495–1500, Edmonton, Alta., Canada, 1999. IEEE.

[49] S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank. Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology. *IEEE Transactions on Biomedical Engineering*, 47(6):773–783, 2000.

[50] K. Umapathy and S. Krishnan. Modified Local Discriminant Bases Algorithm and Its Application in Analysis of Human Knee Joint Vibration Signals. *IEEE Transactions on Biomedical Engineering*, 53(3):517–523, 2006.

[51] T. Mu, A. K. Nandi, and R. M. Rangayyan. Screening of knee-joint vibroarthrographic signals using the strict 2-surface proximal classifier and genetic algorithm. *Computers in Biology and Medicine*, 38(10):1103–1111, 2008.

[52] R. M. Rangayyan and Y. Wu. Screening of knee-joint vibroarthrographic signals using statistical parameters and radial basis functions. *Medical & Biological Engineering & Computing*, 46(3):223–232, 2008.

[53] R. M. Rangayyan and Y. Wu. Screening of knee-joint vibroarthrographic signals using parameters of activity and radial-basis functions. In *2008 Canadian Conference on Electrical and Computer Engineering*, pages 000057–000060, Niagara Falls, ON, Canada, 2008. IEEE.

[54] R. M. Rangayyan and Y. Wu. Analysis of Vibroarthrographic Signals with Features Related to Signal Variability and Radial-Basis Functions. *Annals of Biomedical Engineering*, 37(1):156–163, 2009.

[55] Y. Wu and S. Krishnan. Classification of knee-joint vibroarthrographic signals using time-domain and time-frequency domain features and least-squares support vector machine. In *2009 16th International Conference on Digital Signal Processing*, pages 1–6, Santorini, Greece, 2009. IEEE.

[56] R. M. Rangayyan and Y. Wu. Screening of knee-joint vibroarthrographic signals using probability density functions estimated with Parzen windows. *Biomedical Signal Processing and Control*, 5(1):53–58, 2010.

[57] Y. Wu, S. Cai, M. Lu, and S. Krishnan. An Artificial-Neural-Network-Based Multiple Classifier System for Knee-Joint Vibration Signal Classification. In Y. Wu, editor, *Advances in Computer, Communication, Control and Automation*, volume 121, pages 235–242. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[58] Y. Wu and S. Krishnan. Combining least-squares support vector machines for classification of biomedical signals: A case study with knee-joint vibroarthrographic signals. *Journal of Experimental & Theoretical Artificial Intelligence*, 23(1):63–77, 2011.

[59] S. Cai, Y. Wu, N. Xiang, Z. Zhong, J. He, L. Shi, and F. Xu. Detrending knee joint vibration signals with a cascade moving average filter. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4357–4360, San Diego, CA, 2012. IEEE.

[60] S. Cai, S. Yang, F. Zheng, M. Lu, Y. Wu, and S. Krishnan. Knee Joint Vibration Signal Analysis with Matching Pursuit Decomposition and Dynamic Weighted Classifier Fusion. *Computational and Mathematical Methods in Medicine*, 2013:1–11, 2013.

[61] R. M. Rangayyan, F. Oloumi, Y. Wu, and S. Cai. Fractal analysis of knee-joint vibroarthrographic signals via power spectral analysis. *Biomedical Signal Processing and Control*, 8(1):23–29, 2013.

[62] A. Sundar, V. Pahwa, and C. Das. A new method for denoising knee joint vibroarthrographic signals. In *2015 Annual IEEE India Conference (INDICON)*, pages 1–5, New Delhi, India, 2015. IEEE.

[63] S. Nalband, R. R. Sreekrishna, and A. A. Prince. Analysis of Knee Joint Vibration Signals Using Ensemble Empirical Mode Decomposition. *Procedia Computer Science*, 89:820–827, 2016.

[64] A. Sundar, C. Das, and V. Pahwa. Denoising Knee Joint Vibration Signals Using Variational Mode Decomposition. In S. C. Satapathy, J. K. Mandal, S. K. Udgata, and V. Bhateja, editors, *Information Systems Design and Intelligent Applications*, volume 433, pages 719–729, New Delhi, 2016. Springer India.

[65] A. H. Al-Timemy. Boosting-based decision tree for improved screening of vibroarthrographic signals. In *2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME)*, pages 1–4, Beirut, 2017. IEEE.

[66] S. Nalband, C. A. Valliappan, R. Gupta, A. A. Prince, and A. Agrawal. Feature extraction and classification of knee joint disorders using Hilbert Huang transform. In *2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 266–269, Phuket, 2017. IEEE.

[67] M. Sharma and U. R. Acharya. Analysis of knee-joint vibroarthographic signals using bandwidth-duration localized three-channel filter bank. *Computers & Electrical Engineering*, 72:191–202, 2018.

[68] Y. Athavale and S. Krishnan. A telehealth system framework for assessing knee-joint conditions using vibroarthrographic signals. *Biomedical Signal Processing and Control*, 55:101580, 2020.

[69] D. Kraft and G. Bieber. Vibroarthrography using convolutional neural networks. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–6, Corfu Greece, 2020. ACM.

[70] E. Mascarenhas, S. Nalband, A. R. J. Fredo, and A. A. Prince. Analysis and Classification of Vibroarthrographic Signals using Tuneable 'Q' Wavelet Transform. In *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 65–70, Noida, India, 2020. IEEE.

[71] M. M. Shidore, S. S. Athreya, S. Deshpande, and R. Jalnekar. Screening of knee-joint vibroarthrographic signals using time and spectral domain features. *Biomedical Signal Processing and Control*, 68:102808, 2021.

[72] S. Hersek, M. Baran Pouyan, C. Teague, M. N. Sawka, M. L. Millard-Stafford, G. F. Kogler, P. Wolkoff, and O. T. Inan. Acoustical Emission Analysis by Unsupervised Graph Mining: A Novel Biomarker of Knee Health Status. *IEEE Transactions on Biomedical Engineering*, 65(6):1291–1300, 2018.

[73] G. C. Ozmen, A. H. Gazi, S. Gharehbaghi, K. L. Richardson, M. Safaei, D. C. Whittingslow, S. Prahalad, J. L. Hunnicutt, J. W. Xerogeanes, T. K. Snow, and O. T. Inan. An Interpretable Experimental Data Augmentation Method to Improve Knee Health Classification Using Joint Acoustic Emissions. *Annals of Biomedical Engineering*, 2021.

[74] Y. Zheng, Y. Wang, J. Liu, H. Jiang, and Q. Yue. Knee joint vibration signal classification algorithm based on machine learning. *Neural Computing and Applications*, 33(3):985–995, 2021.

[75] N. Befrui, J. Elsner, A. Flesser, J. Huvanandana, O. Jarrousse, T. N. Le, M. Müller, W. H. W. Schulze, S. Taing, and S. Weidert. Vibroarthrography for early detection of knee osteoarthritis using normalized frequency features. *Medical & Biological Engineering & Computing*, 56(8):1499–1514, 2018.

[76] N. Tanaka and M. Hoshiyama. Articular Sound and Clinical Stages in Knee Arthropathy. *Journal of Musculoskeletal Research*, 14(01):1150006, 2011.

[77] N. Tanaka and M. Hoshiyama. Vibroarthrography in patients with knee arthropathy. *Journal of Back and Musculoskeletal Rehabilitation*, 25(2):117–122, 2012.

[78] K. S. Kim, C. G. Song, J. H. Seo, K. H. Park, J. Y. La, and J. C. Kim. An efficient algorithm to improve feature extraction and classification of knee joint sound. In *2009 ICME International Conference on Complex Medical Engineering*, pages 1–6, Tempe, AZ, USA, 2009. IEEE.

[79] K. S. Kim, J. H. Seo, J. U. Kang, and C. G. Song. An enhanced algorithm for knee joint sound classification using feature extraction based on time-frequency analysis. *Computer Methods and Programs in Biomedicine*, 94(2):198–206, 2009.

[80] B. Mascaro, J. Prior, L. K. Shark, J. Selfe, P. Cole, and J. Goodacre. Exploratory study of a non-invasive method based on acoustic emission for assessing the dynamic integrity of knee joints. *Medical Engineering & Physics*, 31(8):1013–1022, 2009.

[81] T. Lee, W. Lin, L. Wu, and H. Wang. Analysis of Vibroarthrographic Signals for Knee Osteoarthritis Diagnosis. In *2012 Sixth International Conference on Genetic and Evolutionary Computing*, pages 223–228, 2012.

[82] S. Ota, A. Ando, Y. Tozawa, T. Nakamura, S. Okamoto, T. Sakai, and K. Hase. Preliminary study of optimal measurement location on vibroarthrography for classification

of patients with knee osteoarthritis. *Journal of Physical Therapy Science*, 28(10):2904–2908, 2016.

[83] F. Golshan, Y. C. Hum, B. Pingguan-Murphy, and K. W. Lai. Vibroarthrography Difference Between Left and Right Knee for Osteoarthritis Detection. In F. Ibrahim, J. Usman, M. Y. Ahmad, N. Hamzah, and S. J. Teh, editors, *2nd International Conference for Innovation in Biomedical Engineering and Life Sciences*, volume 67, pages 289–294. Springer Singapore, Singapore, 2018.

[84] P. Madeleine, R. E. Andersen, J. B. Larsen, L. Arendt-Nielsen, and A. Samani. Wireless multichannel vibroarthrographic recordings for the assessment of knee osteoarthritis during three activities of daily living. *Clinical Biomechanics*, 72:16–23, 2020.

[85] Y. Wang, T. Zheng, J. Song, and W. Gao. A novel automatic Knee Osteoarthritis detection method based on vibroarthrographic signals. *Biomedical Signal Processing and Control*, 68:102796, 2021.

[86] R. Karpiński, P. Krakowski, J. Jonak, A. Machrowska, M. Maciejewski, and A. Nogalski. Analysis of differences in vibroacoustic signals between healthy and osteoarthritic knees using EMD algorithm and statistical analysis. *Journal of Physics: Conference Series*, 2130(1):012010, 2021.

[87] R. Karpiński, P. Krakowski, J. Jonak, A. Machrowska, M. Maciejewski, and A. Nogalski. Estimation of differences in selected indices of vibroacoustic signals between healthy and osteoarthritic patellofemoral joints as a potential non-invasive diagnostic tool. *Journal of Physics: Conference Series*, 2130(1):012009, 2021.

[88] R. Karpiński. Knee joint osteoarthritis diagnosis based on selected acoustic signal discriminants using machine learning. *Applied Computer Science*, 18(2):71–85, 2022.

[89] D. K. Verma, P. Kumari, and S. Kanagaraj. Engineering Aspects of Incidence, Prevalence, and Management of Osteoarthritis: A Review. *Annals of Biomedical Engineering*, 50(3):237–252, 2022.

[90] K. Falkowski, G. Skiba, M. Czerner, M. Szmajda, and D. Bączkowicz. Effects of Viscosupplementation on Knee Joint Arthrokinematics – Pilot Study. *Ortopedia Traumatologia Rehabilitacja*, 20(5):409–419, 2018.

[91] D. Bączkowicz, G. Skiba, M. Szmajda, I. Vařeka, K. Falkowski, and K. Laudner. Effects of Viscosupplementation on Quality of Knee Joint Arthrokinematic Motion Analyzed by Vibroarthrography. *CARTILAGE*, 12(4):438–447, 2021.

[92] K. Kalo, D. Niederer, M. Schmitt, and L. Vogt. Acute effects of a single bout of exercise therapy on knee acoustic emissions in patients with osteoarthritis: A double-blinded, randomized controlled crossover trial. *BMC Musculoskeletal Disorders*, 23(1):657, 2022.

[93] D. Bączkowicz and E. Majorczyk. Joint motion quality in vibroacoustic signal analysis for patients with patellofemoral joint disorders. *BMC Musculoskeletal Disorders*, 15(1):426, 2014.

[94] Y. Wu, S. Cai, F. Xu, L. Shi, and S. Krishnan. Chondromalacia Patellae Detection by Analysis of Intrinsic Mode Functions in Knee Joint Vibration Signals. In M. Long, editor, *World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China*, volume 39, pages 493–496. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[95] D. Bączkowicz, K. Kręcisz, and Z. Borysiuk. Analysis of patellofemoral arthrokinematic motion quality in open and closed kinetic chains using vibroarthrography. *BMC Musculoskeletal Disorders*, 20(1):48, 2019.

[96] J. Jonak, R. Karpiński, A. Machrowska, P. Krakowski, and M. Maciejewski. A preliminary study on the use of EEMD-RQA algorithms in the detection of degenerative changes in knee joints. *IOP Conference Series: Materials Science and Engineering*, 710(1):012037, 2019.

[97] D. Bączkowicz and E. Majorczyk. Joint Motion Quality in Chondromalacia Progression Assessed by Vibroacoustic Signal Analysis. *PM&R*, 8(11):1065–1071, 2016.

[98] B. Semiz, S. Hersek, D. C. Whittingslow, L. A. Ponder, S. Prahalad, and O. T. Inan. Using Knee Acoustical Emissions for Sensing Joint Health in Patients With Juvenile Idiopathic Arthritis: A Pilot Study. *IEEE Sensors Journal*, 18(22):9128–9136, 2018.

[99] D. C. Whittingslow, J. Zia, S. Gharehbaghi, T. Gergely, L. A. Ponder, S. Prahalad, and O. T. Inan. Knee Acoustic Emissions as a Digital Biomarker of Disease Status in Juvenile Idiopathic Arthritis. *Frontiers in Digital Health*, 2:571839, 2020.

[100] S. Gharehbaghi, D. C. Whittingslow, L. A. Ponder, S. Prahalad, and O. T. Inan. Acoustic Emissions From Loaded and Unloaded Knees to Assess Joint Health in Patients With Juvenile Idiopathic Arthritis. *IEEE Journal of Biomedical and Health Informatics*, 25(9):3618–3626, 2021.

[101] N. P. Reddy, B. M. Rothschild, E. Verrall, and A. Joshi. Noninvasive Measurement of Acceleration at the Knee Joint in Patients with Rheumatoid Arthritis and Spondyloarthropathy of the Knee. *Annals of Biomedical Engineering*, 29(12):1106–1111, 2001.

[102] A. C. D. Faria, G. R. C. Pinheiro, J. Neri, and P. L. Melo. Instrumentation for the analysis of changes in the knee joint of patients with rheumatoid arthritis: Focus on low-frequency vibrations. *Journal of Physics: Conference Series*, 1044:012038, 2018.

[103] K. L. Richardson, S. Gharehbaghi, G. C. Ozmen, M. M. Safaei, and O. T. Inan. Quantifying Signal Quality for Joint Acoustic Emissions Using Graph-Based Spectral Embedding. *IEEE Sensors Journal*, 21(12):13676–13684, 2021.

[104] G. F. McCoy, J. D. McCrea, D. E. Beverland, W. G. Kernohan, and R. A. Mollan. Vibration arthrography as a diagnostic aid in diseases of the knee. A preliminary report. *The Journal of Bone & Joint Surgery British Volume*, 69-B(2):288–293, 1987.

[105] Y. Shen, R. M. Rangayyan, G. D. Bell, C. B. Frank, Y. T. Zhang, and K. O. Ladly. Localization of knee joint cartilage pathology by multichannel vibroarthrography. *Medical Engineering & Physics*, 17(8):583–594, 1995.

[106] S. Krishnan and R. M. Rangayyan. Automatic de-noising of knee-joint vibration signals using adaptive time-frequency representations. *Medical & Biological Engineering & Computing*, 38(1):2–8, 2000.

[107] S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank. Sonification of knee-joint vibration signals. In *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.00CH37143)*, volume 3, pages 1995–1998, Chicago, IL, USA, 2000. IEEE.

[108] S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank. Auditory display of knee-joint vibration signals. *The Journal of the Acoustical Society of America*, 110(6):3292–3304, 2001.

[109] Y. Wu, S. Cai, S. Yang, F. Zheng, and N. Xiang. Classification of Knee Joint Vibration Signals Using Bivariate Feature Distribution Estimation and Maximal Posterior Probability Decision Criterion. *Entropy*, 15(12):1375–1387, 2013.

[110] Y. Wu, S. Yang, F. Zheng, S. Cai, M. Lu, and M. Wu. Removal of artifacts in knee joint vibroarthrographic signals using ensemble empirical mode decomposition and detrended fluctuation analysis. *Physiological Measurement*, 35(3):429–439, 2014.

[111] S. Yang, S. Cai, F. Zheng, Y. Wu, K. Liu, M. Wu, Q. Zou, and J. Chen. Representation of fluctuation features in pathological knee joint vibroarthrographic signals using kernel density modeling method. *Medical Engineering & Physics*, 36(10):1305–1311, 2014.

[112] J. C. Chen, P. C. Tung, S. F. Huang, S. W. Wu, S. L. Lin, and K. L. Tu. Extraction and screening of knee joint vibroarthrographic signals using the empirical mode decomposition method. *International Journal of Innovative Computing, Information and Control*, 9(6):2689–2700, 2013.

[113] C. Teague, S. Hersek, H. Toreyin, M. L. Millard-Stafford, M. L. Jones, G. F. Kogler, M. N. Sawka, and O. T. Inan. Novel approaches to measure acoustic emissions as biomarkers for joint health assessment. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 1–6, Cambridge, MA, USA, 2015. IEEE.

[114] Y. Wu, P. Chen, X. Luo, H. Huang, L. Liao, Y. Yao, M. Wu, and R. M. Rangayyan. Quantification of knee vibroarthrographic signal irregularity associated with patellofemoral joint cartilage pathology based on entropy and envelope amplitude measures. *Computer Methods and Programs in Biomedicine*, 130:1–12, 2016.

[115] A. Łysiak, A. Froń, D. Bączkowicz, and M. Szmajda. The new descriptor in processing of vibroacoustic signal of knee joint. *IFAC-PapersOnLine*, 52(27):335–340, 2019.

[116] A. Łysiak, A. Froń, D. Bączkowicz, and M. Szmajda. Vibroarthrographic Signal Spectral Features in 5-Class Knee Joint Classification. *Sensors*, 20(17):5015, 2020.

[117] K. Kalo, D. Niederer, R. Sus, K. Sohrabi, W. Banzer, V. Groß, and L. Vogt. The detection of knee joint sounds at defined loads by means of vibroarthrography. *Clinical Biomechanics*, 74:1–7, 2020.

[118] K. O. Ladly, C. B. Frank, G. D. Bell, Y. T. Zhang, and R. M. Rangayyan. The Effect of External Loads and Cyclic Loading on Normal Patellofemoral Joint Signals. *Defence Science Journal*, 43(3):201–210, 1993.

[119] Y. T. Zhang and R. M. Rangayyan. Adaptive cancellation of muscle contraction interference in vibroarthrographic signals. *IEEE Transactions on Biomedical Engineering*, 41(2):181–191, 1994.

[120] A. D. Wiens, S. Prahalad, and O. T. Inan. VibroCV: A computer vision-based vibroarthrography platform with possible application to Juvenile idiopathic arthritis. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4431–4434, Orlando, FL, USA, 2016. IEEE.

[121] K. L. Scherpereel, N. B. Bolus, H. K. Jeong, O. T. Inan, and A. J. Young. Estimating Knee Joint Load Using Acoustic Emissions During Ambulation. *Annals of Biomedical Engineering*, 49(3):1000–1011, 2021.

[122] D. M. Hochman, S. Gharehbaghi, D. C. Whittingslow, and O. T. Inan. A Pilot Study to Assess the Reliability of Sensing Joint Acoustic Emissions of the Wrist. *Sensors*, 20(15):4240, 2020.

[123] D. Bączkowicz and K. Kręcisz. Vibroarthrography in the evaluation of musculoskeletal system – a pilot study. *Ortopedia Traumatologia Rehabilitacja*, 15(5):407–416, 2013.

[124] K. Kalo, D. Niederer, R. Sus, K. Sohrabi, V. Groß, and L. Vogt. Reliability of Vibroarthrography to Assess Knee Joint Sounds in Motion. *Sensors*, 20(7):1998, 2020.

[125] A. Machrowska, R. Karpiński, P. Krakowski, and J. Jonak. Diagnostic Factors for Opened and Closed Kinematic Chain of Vibroarthrography Signals. *Applied Computer Science*, 15(3):34–44, 2019.

[126] D. A. Barr, W. G. Kernohan, and R. A. B. Mollan. An investigation of the short-term effect of exercise on the knee joint by means of computer-assisted auscultation. *Advances in Engineering Software*, 21(1):27–35, 1994.

[127] C. N. Teague, S. Hersek, H. Toreyin, M. L. Millard-Stafford, M. L. Jones, G. F. Kogler, M. N. Sawka, and O. T. Inan. Novel Methods for Sensing Acoustical Emissions From the Knee for Wearable Joint Health Assessment. *IEEE Transactions on Biomedical Engineering*, 63(8):1581–1590, 2016.

[128] W. G. Kernohan, D. A. Barr, G. F. McCoy, and R. A. Mollan. Vibration arthrometry in assessment of knee disorders: The problem of angular velocity. *Journal of Biomedical Engineering*, 13(1):35–38, 1991.

[129] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.

[130] N. E. Huang and Z. Wu. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46, 2008.

[131] Z. Wu and N. E. Huang. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, 01(01):1–41, 2009.

[132] K. Dragomiretskiy and D. Zosso. Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, 62(3):531–544, 2014.

[133] A. Łysiak and M. Szmajda. Empirical Comparison of the Feature Evaluation Methods Based on Statistical Measures. *IEEE Access*, 9:27868–27883, 2021.

[134] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri. Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia*, 22(1):67–72, 2019.

[135] F. B. Oppong and S. Y. Agbedra. Assessing Univariate and Multivariate Normality, A Guide For Non-Statisticians. *Mathematical Theory and Modeling*, 6(2):26, 2016.

[136] H. Zhivomirov and I. Nedelchev. A Method for Signal Stationarity Estimation. *Romanian Journal of Acoustics and Vibration*, 17(2):149–155, 2020.

[137] E. Herranz. Unit root tests. *WIREs Computational Statistics*, 9(3):e1396, 2017.

[138] P. Lovie. Coefficient of Variation. In B. S. Everitt and D. C. Howell, editors, *Encyclopedia of Statistics in Behavioral Science*, page bsa107. John Wiley & Sons, Ltd, Chichester, UK, 2005.

[139] P. Hall and S. C. Morton. On the estimation of entropy. *Annals of the Institute of Statistical Mathematics*, 45(1):69–88, 1993.

[140] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz. *Handbook of Differential Entropy*. Chapman and Hall/CRC, 1st edition, 2013.

[141] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.

[142] A. B. Hassanat, M. A. Abbadi, G. A. Altarawneh, and A. A. Alhasanat. Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach. *International Journal of Computer Science and Information Security*, 12(8):33–39, 2014.

[143] M. Sharma, P. Sharma, R. Bilas, and V. M. Gadre. Double Density Dual-Tree Complex Wavelet Transform-Based Features for Automated Screening of Knee-Joint Vibroarthrographic Signals. In *Machine Intelligence and Signal Analysis*, volume 748 of *Advances in Intelligent Systems and Computing*. Springer Singapore, Singapore, 2019.

[144] R. M. Aarts. Tracking and Estimation of Frequency, Amplitude, and Form Factor of a Harmonic Time Series [Lecture Notes]. *IEEE Signal Processing Magazine*, 38(5):86–91, 2021.

[145] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Systems with Applications*, 38(3):2888–2911, 2011.

[146] W. Caesarendra and T. Tjahjowidodo. A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing. *Machines*, 5(4):21, 2017.

[147] B. Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1970.

[148] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc, Hoboken, New Jersey, 5th edition, 2016.

[149] D. Grech and Z. Mazur. On the scaling ranges of detrended fluctuation analysis for long-term memory correlated short series of data. *Physica A: Statistical Mechanics and its Applications*, 392(10):2384–2397, 2013.

[150] E. A. F. Ihlen. Introduction to Multifractal Detrended Fluctuation Analysis in Matlab. *Frontiers in Physiology*, 3, 2012.

[151] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, 2002.

[152] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, UK ; New York, 2nd edition, 2004.

[153] M. T. Rosenstein, J. J. Collins, and C. J. De Luca. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1):117–134, 1993.

[154] W. Caesarendra, B. Kosasih, K. Tieu, and C. A. S. Moodie. An application of nonlinear feature extraction - A case study for low speed slewing bearing condition monitoring and prognosis. In *2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pages 1713–1718, Wollongong, NSW, 2013. IEEE.

[155] N. Marwan and K. H. Kraemer. Trends in recurrence analysis of dynamical systems. *The European Physical Journal Special Topics*, 232(1):5–27, 2023.

[156] S. Spiegel, D. Schultz, and N. Marwan. Approximate Recurrence Quantification Analysis (aRQA) in Code of Best Practice. In C. L. Webber, C. Ioana, and N. Marwan, editors, *Recurrence Plots and Their Quantifications: Expanding Horizons*, volume 180, pages 113–136. Springer International Publishing, Cham, 2016.

[157] N. Marwan, M. Carmenromano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5-6):237–329, 2007.

[158] S. Nalband, A. Sundar, A. A. Prince, and A. Agarwal. Feature selection and classification methodology for the detection of knee-joint disorders. *Computer Methods and Programs in Biomedicine*, 127:94–104, 2016.

[159] S. M. Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6):2297–2301, 1991.

[160] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology. Heart and Circulatory Physiology*, 278(6):H2039–2049, 2000.

[161] A. Delgado-Bonal and A. Marshak. Approximate Entropy and Sample Entropy: A Comprehensive Tutorial. *Entropy*, 21(6):541, 2019.

[162] D. Bajić and N. Japundžić-Žigon. On Quantization Errors in Approximate and Sample Entropy. *Entropy*, 24(1):73, 2022.

[163] S. Lu, X. Chen, J. K. Kanters, I. C. Solomon, and K. H. Chon. Automatic selection of the threshold value R for approximate entropy. *IEEE transactions on bio-medical engineering*, 55(8):1966–1972, 2008.

[164] P. Stoica and R. L. Moses. *Spectral Analysis of Signals*. Pearson/Prentice Hall, Upper Saddle River, N.J, 2005.

[165] L. Hu and Z. Zhang, editors. *EEG Signal Processing and Feature Extraction*. Springer Singapore, Singapore, 2019.

[166] G. Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, Icram, 2004.

[167] T. Inouye, K. Shinosaki, H. Sakamoto, S. Toi, S. Ukai, A. Iyama, Y. Katsuda, and M. Hirano. Quantification of EEG irregularity by use of the entropy of the power spectrum. *Electroencephalography and Clinical Neurophysiology*, 79(3):204–210, 1991.

[168] A. Łysiak, T. Marciniak, and D. Bączkowicz. Repeatability of the Vibroarthrogram in the Temporomandibular Joints. *Sensors*, 22(23):9542, 2022.

[169] L. Cohen. *Time-Frequency Analysis*. Prentice Hall Signal Processing Series. Prentice Hall PTR, Englewood Cliffs, N.J, 1995.

[170] R. G. Lyons. *Understanding Digital Signal Processing*. Prentice Hall, Upper Saddle River, NJ, 3rd edition, 2011.

[171] E. Scheirer and M. Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 1331–1334, Munich, Germany, 1997. IEEE Comput. Soc. Press.

[172] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. *Feature Extraction*. Number 207 in Studies in Fuzziness and Soft Computing. Springer-Verlag, Heidelberg, 1 edition, 2006.

[173] D. Petri. Big data, dataism and measurement. *IEEE Instrumentation & Measurement Magazine*, 23(3):32–34, 2020.

[174] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 1 edition, 2000.

[175] A. Djouadi, O. Snorrason, and F. D. Garber. The quality of training sample estimates of the Bhattacharyya coefficient. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):92–97, 1990.

[176] S. Bi, M. Broggi, and M. Beer. The role of the Bhattacharyya distance in stochastic model updating. *Mechanical Systems and Signal Processing*, 117:437–452, 2019.

[177] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[178] Z. Zhao, R. Anand, and M. Wang. Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 442–452, Washington, DC, USA, 2019. IEEE.

[179] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal Detection of Changepoints With a Linear Computational Cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

[180] S. L. Brunton and J. N. Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2 edition, 2022.

[181] S. Balakrishnama and A. Ganapathiraju. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, 18(1998):1–8, 1998.

[182] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, NY, 2009.

[183] W. Y. Loh and Y. S. Shih. Split Selection Methods for Classification Trees. *Statistica Sinica*, 7(4):815–840, 1997.

[184] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano. RUSBoost: Improving classification performance when training data is skewed. In *2008 19th International Conference on Pattern Recognition*, pages 1–4, 2008.

[185] P. Refaeilzadeh, L. Tang, and H. Liu. Cross-Validation. In L. Liu and M. T. Özsu, editors, *Encyclopedia of Database Systems*, pages 532–538. Springer US, Boston, MA, 2009.

[186] J. B. W. Webber. A bi-symmetric log transformation for wide-range data. *Measurement Science and Technology*, 24(2):027001, 2012.