POLITECHNIKA OPOLSKA
Wydział Elektrotechniki, Automatyki i Informatyki

ROZPRAWA DOKTORSKA

mgr inż. Krzysztof Smykała

Zastosowanie metod uczenia maszynowego i pomiarów hiperspektralnych w diagnostyce uprawy *Solanum lycopersicum*

Promotor:
dr hab. inż. Michał Tomaszewski, prof. uczelni

Promotor pomocniczy:
dr inż. Bogdan Ruszczak

Opole 2023

Najserdeczniejsze podziękowania kieruję w stronę
mojego promotora **Profesora Michała Tomaszewskiego**
za pokierowanie pracami nad niniejszą pracą doktorską,
za cierpliwość i nieocenioną motywację do działania.

Równie mocno pragnę podziękować
mojemu promotorowi pomocniczemu **dr inż. Bogdanowi Ruszczakowi**
za pomoc merytoryczną, cenne uwagi i zaangażowanie
w kluczowych momentach prowadzenia badań.

Szczególnie dziękuję **Zbyszkowi** i całemu zespołowi **QZ Solutions**
za wieloletnią współpracę nad projektami badawczo-rozwojowymi
i wspaniałą atmosferę pracy.

Dziękuję moim bliskim za wiarę we mnie i wsparcie,
Mojej **Ukochanej Żonie Agacie**, za zaangażowanie i poświęcenie,
bez których nie mógłbym realizować pracy naukowej i zawodowej,
Moim rodzicom, **Andrzejowi i Jolancie**, za trud
włożony w moje wychowanie i edukację
Oraz mojej siostrze, **Małgorzacie**, za inspirację
i pokazanie, że determinacja pozwala na osiąganie założonych celów.

Pracę dedykuję mojej żonie Agacie i synowi Ignacemu.

Spis treści

1. Wprowadzenie 7
2. Cel i zakres pracy 9
2.1. Zakres pracy 9
2.2. Teza pracy 10
3. Zastosowanie metrologii w diagnostyce chorób roślin 12
3.1. Obiekt badań 12
3.2. Analizowane choroby i patogeny je wywołujące 12
3.3. Diagnostyka chorób roślin na podstawie analizy sygnałów pomiarowych 16
3.4. Podsumowanie rozdziału 18
4. Zastosowanie danych hiperspektralnych w diagnostyce uprawy pomidora – stan techniki 18
4.1. Typy danych spektralnych 18
4.2. Pomiary spektralne 18
4.3. Obrazowanie wielospektralne 19
4.4. Obrazowanie hiperspektralne 20
4.5. Pomiary hiperspektralne w diagnostyce uprawy pomidora 22
4.6. Przegląd dotychczasowych rozwiązań – istotne zakresy spektralne 23
4.7. Przegląd stanu wiedzy w zakresie wczesnego wykrywania chorób 28
4.8. Podsumowanie rozdziału 30
5. Koncepcja metody diagnostyki wybranych chorób w uprawie pomidorów zwyczajnych 32
6. Eksperyment pomiarowy – obserwacja uprawy pomidora zwyczajnego w warunkach rzeczywistych i laboratoryjnych 35
6.1. Harmonogram badań 35
6.2. Przygotowanie inokulacji i materiału roślinnego do badań 37
6.3. Instrument pomiarowy 37
6.4. Pomiary hiperspektralne 44
6.5. Podsumowanie rozdziału 46
7. Wybrane metody uczenia maszynowego użyte w doktoracie 47
7.1.1. Lista wszystkich rozpatrywanych klasyfikatorów 47
7.1.2. Opis wybranych klasyfikatorów 48
7.1.3. Hiperparametry dostrajane w ramach tworzonej metody 51
7.1.4. Metryki używane w pracy doktorskiej: F1, dokładność, czułość, precyzja 53
7.2. Podsumowanie rozdziału 54
8. Zbiór danych dla potrzeb uczenia maszynowego 55
8.1. Schemat postępowania z pomiarami 55
8.2. Obróbka wstępna danych 57
8.3. Metoda łączenia danych dzień po dniu 60
8.4. Analiza danych pomiarowych – EDA 64
8.5. Wskaźnik NDVI 66
8.6. Podsumowanie rozdziału 68
9. Wykrywanie chorób z wykorzystaniem wybranych algorytmów uczenia maszynowego 69
9.1. Wstępna selekcja algorytmów klasyfikacyjnych 69
9.2. Klasyfikacja binarna 71
9.2.1. Wszystkie choroby 77
9.2.2. Anthracnose (AN) 78
9.2.3. Bacterial Speck (BS) 79

9.2.4. Early Blight (EB) ... 80
9.2.5. Late Blight (LB) ... 81
9.2.6. Septoria Leaf Spot (SL) ... 82
9.2.7. Podsumowanie wyników uzyskanych dla klasyfikatorów binarnych ... 83

9.3. Analiza udziału poszczególnych zakresów spektralnych w klasyfikacji binarnej chorób ... 86
9.3.1. Zjawisko Hughesa w danych hiperspektralnych ... 86
9.3.2. Metoda eliminacji cech ... 86

9.4. Porównanie wyników z literaturą przedmiotu ... 91
9.5. Klasyfikacja wieloklasowa ... 92
9.6. Podsumowanie rozdziału ... 94
10. Analiza skuteczności klasyfikacji chorób w kolejnych dniach od inokulacji ... 95

10.1. Wyniki w ujęciu rodzaju choroby ... 95
10.1.1. Wszystkie choroby ... 96
10.1.2. Anthracnose (AN) ... 98
10.1.3. Bacterial Speck (BS) ... 100
10.1.4. Early Blight (EB) ... 102
10.1.5. Late Blight (LB) ... 104
10.1.6. Septoria Leaf Spot (SL) ... 106
10.1.7. Wnioski dla wyników w ujęciu chorób ... 108

10.2. Wyniki w ujęciu rodzaju algorytmu klasyfikacyjnego ... 109
10.2.1. Ridge Classifier ... 109
10.2.2. LGBM Classifier ... 111
10.2.1. Random Forest Classifier ... 113
10.2.2. Wnioski dla wyników w ujęciu klasyfikatorów ... 114

10.3. Zestawienie wyników dla różnych wartości parametru T ... 114
10.3.1. Podział na poszczególne choroby ... 117
10.3.2. Połączenie wykresów ... 124
10.3.3. Wnioski dla zestawienia wyników dla różnej wartości parametru T ... 128

10.4. Podsumowanie rozdziału ... 128
11. Podsumowanie i wnioski ... 129
12. Kierunek dalszych prac ... 133
13. Bibliografia ... 134
Streszczenie ... 141
Abstract ... 142
Spis rysunków ... 143
Spis tabel ... 148
14. DODATEK A ... 150

14.1. Załączniki ... 150
14.2. Załączniki w formie skoroszytów ... 150

Wykaz ważniejszych oznaczeń i skrótów

Symbol	Nazwa	Opis	Jednostka
L	radiancja	luminacja energetyczna	[W/m^2/sr]
L_λ	radiancja widmowa	luminacja energetyczna w poszczególnych długościach fali	[W/m^2/sr/nm]
E	irradiancja	natężenie promieniowania	[W/m^2/nm]
ρ	reflektancja	stosunek mocy promieniowania elektromagnetycznego wiązki odbitej od badanej próbki do mocy promieniowania elektromagnetycznego wiązki odbitej od panelu kalibracyjnego (Spectralon)	
τ	transmitancja	stosunek natężenia wiązki promieniowania elektromagnetycznego przepuszczanego przez próbkę do natężenia wiązki padającej	
α	absorbancja	logarytm dziesiętny odwrotności transmitancji	
l	długość fali		
$E_I(l)$	energia całkowita	energia całkowita padająca na powierzchnię obiektu	[W/m^2]
$E_R(l)$	energia odbita	energia odbita od powierzchni badanego obiektu	[W/m^2]
$E_A(l)$	energia zaabsorbowana	energia zaabsorbowana przez badany obiekt	[W/m^2]
$E_T(l)$	energia transmitowana	energia transmitowana przez badany obiekt	[W/m^2]
ψ	obszar skanowania	wielkość obszaru skanowania dla pomiary	[cm^2]
FOV, γ	pole widzenia soczewki, ang. *field of view*	obszar widziany w danym momencie przez instrument pomiarowy (soczewkę kabla światłowodowego)	[°]
h	wysokość pomiarowa	odległość pomiędzy badanym obiektem, a soczewką kabla światłowodowego urządzenia pomiarowego	[cm]
NDVI	znormalizowany różnicowy wskaźnik wegetacji	wskaźnik opisujący potencjał wegetacyjny danego obszaru, będą stosunkiem różnicy reflektancji dla kanałów bliskiej podczerwieni (NIR, 790 nm) oraz czerwonego (RED, 660 nm) do ich sumy	

DPI	dni po inokulacji ang. *days post inoculum*	liczba dni, które minęły od zakażenia do dnia, w którym wykonano dany pomiar	
T	szerokość ramki danych	Szerokość podzbiorów (ramki) w zależności od dni, które minęły od inokulacji (DPI)	[dni]
AN	Antraknoza, ang. *Antracnose*	nazwy chorób zawartych w badanym zbiorze pomiarów hiperspektralnych oraz nazwy wariantów eksperymentów zawierających te pomiary	
BS	Bakteryjna cętkowatość, ang. *Bacterial Speck*		
EB	Alternarioza, ang. *Early Blight*		
LB	Zaraza ziemniaka, ang. *Late Blight*		
SL	Septorioza, ang. *Septoria Leaf Spot*		
ALL		wariant eksperymentu zawierający pomiary hiperspektralne z grupy kontrolnej oraz wszystkich grup chorobowych (AN-SL)	
F1	wskaźnik F1	metryki oceny jakościowej modelu	
Fβ	wskaźnik Fβ		
R^2	współczynnik determinacji		
UV	ultrafiolet	zakres widma ultrafioletowego (10-390 nm)	
VIS	pasma widzialne	zakres widma widzialnego (390-740 nm)	
VNIR	pasma widzialne i bliskiej podczerwieni	łączony zakres widma widzialnego oraz bliskiej podczerwieni, w literaturze oznaczający również zakres spektralny bardzo bliskiej podczerwieni (ang. *Very-Near Infrared*) (390-1000 nm)	
NIR	bliska podczerwień	zakres widma bliskiej podczerwieni (~740nm-2500 nm)	
SWIR, SWIR 1, SWIR 2	podczerwień fal krótkich	zakres widma podczerwieni fal krótkich, ze względu na wykorzystane urządzenie pomiarowe przestrzeń SWIR została rozdzielona w zależności od stosowanych detektorów – SWIR 1 (1000-1800 nm) oraz SWIR 2 (1800-2500 nm)	

UV-VNIR		łączony zakres widma ultrafioletowego, widzialnego oraz bliskiej podczerwieni (dla badanego zakresu spektralnego 350-1000 nm)	
NIR-SWIR		łączony zakres widma bliskiej podczerwieni i podczerwonych fal krótkich (1001-2500nm)	
RFE	metoda rekurencyjnej eliminacji cech	metoda pozwalająca na zmniejszenie wymiarowości zbioru danych poprzez wielokrotne przeprowadzanie procesu treningowego	
LGBM	ang. *Light Gradient Boosting Machine*, lekka maszyna wzmocnienia gradientowego	główne klasyfikatory wykorzystane w niniejszej pracy doktorskiej, szerzej opisane w rozdziale 7.1.2 „Opis wybranych klasyfikatorów”	
LSVC	ang. *Linear SVC*, liniowa klasyfikator wektorów nośnych		
LR	ang. *Logistic Regression*, regresja logistyczna		
RF	ang. *Random Forest*, las losowy		
RC	ang. *Ridge Classifier*, klasyfikator grzbietowy		
XGB	ang. *XGBoost, XGBClassifier*, ekstremalne wzmocnienie gradientowe		

1. Wprowadzenie

Poprawna diagnostyka obiektów w produkcji odgrywa istotną rolę w przedsiębiorstwach zmechanizowanych oraz zautomatyzowanych [1]. Dotyczy to również przedsiębiorców rolnych oraz producentów żywności [2].Przewiduje się, że do roku 2050 populacja ludzkości przekroczy granicę 9 miliardów, co wymusi zwiększenie produkcji żywności o 70-100% w stosunku do roku 2009 [3]. Obecnie pod uprawy wykorzystywanych jest 36% dostępnego lądu na Ziemi. Pozostałe 64% to miasta, lasy, obszary chronione, czy obszary nienadające się pod uprawę [4]. Nasuwa to wniosek, że zwiększenie produkcji żywności na tak dużą skalę, bez znaczącej szkody dla środowiska naturalnego, może odbyć się jedynie poprzez zwiększenie produktywności tej produkcji.
Zwiększenie produktywności w rolnictwie odbywa się w wielu obszarach, od tych typowo chemiczno-rolniczych, jak środki ochrony roślin, nawozy, odżywki, poprzez aspekty mechaniczno-elektryczne, jak ciągniki oraz maszyny rolnicze, aż po inteligentne systemu diagnostyki i monitoringu tak maszyn, jak i upraw, które stanowią przedmiot niniejszej rozprawy.
Podstawą w codziennej pracy rolnika są systematyczne oględziny pola i ocena uprawy pod kątem niedoborów składników odżywczych, występowania chorób, szkodników czy innych zjawisk wymagających interwencji agronomicznej. Rozwój technologiczny umożliwił usprawnienie tego procesu w ostatnich latach. Opracowanie w latach 70-tych XX w. pierwszych wskaźników wegetacyjnych, a także rozwój sektora kosmicznego i dziedziny obserwacji Ziemi w ostatnich latach doprowadziły do powstania wielu systemów teledetekcyjnych umożliwiających obserwację rozwoju wegetacyjnego uprawy z kosmosu z kilkudniową rozdzielczością czasową [5, 6].
Poziom wegetacyjny uprawy wskazuje na wystąpienie stresów biotycznych i abiotycznych roślin, jednak do tej pory nie udało się wypracować narzędzia pozwalającego na diagnozowanie chorób roślin z dużo dokładnością za pomocą danych satelitarnych. Wynika to między innymi z ograniczonej rozdzielczości spektralnej dostępnych instrumentów obrazujących (np. Sentinel-2). Dużą szansę na rozwój w tej dziedzinie może dać rozwój technologii hiperspektralnych. Obrazowanie hiperspektralne pozwala na badanie obiektów z wyższą, o rząd wielkości, rozdzielczością spektralną niż w przypadku obrazowania multispektralnego. Niestety konsekwencją wyższej rozdzielczości spektralnej jest większy wolumen badanych danych. W przypadku tego wyzwania technologicznego przychodzą narzędzia uczenia maszynowego.

Temat wykorzystania danych multispektralnych i hiperspektralnych w połączeniu z algorytmami przetwarzania sygnału cyfrowego oraz sztucznej inteligencji cieszy się coraz większą popularnością wśród badaczy [7–11]. Dane te są wykorzystywane na szeroką skalę dzięki agencjom kosmicznym (takim jak NASA czy ESA) w dziedzinie teledetekcji, m.in. w zarządzaniu kryzysowym, obserwacji zmian klimatycznych, ale również w rolnictwie do określania współczynników wegetacji roślin [12], stresu wodnego roślin [13], czy wilgotności gleby [14]. Dzięki postępowi technologicznemu i nieustannej miniaturyzacji oraz intensywnej popularyzacji dronów w ostatnich latach możliwe jest wykonywanie lotniczych zdjęć hiperspektralnych w rozdzielczości do paru milimetrów na piksel, co umożliwia dokładniejszą analizę badanego zjawiska.

Wyniki prac badawczych zawartych w rozprawie z zakresu analizy sygnałów hiperspektralnych i diagnostyki obiektów organicznych z wykorzystaniem metod uczenia

maszynowego częściowo powstały w ramach realizacji przez QZ Solutions Sp. z o.o. projektu dofinansowanego z funduszy Narodowego Centrum Badań i Rozwoju (POIR.01.01.01-00-1317/17) przy współpracy z Wydziałem Ogrodnictwa i Warzywnictwa Uniwersytetu Przyrodniczego w Poznaniu oraz Centrum Badań Kosmicznych Polskiej Akademii Nauk.

2. Cel i zakres pracy

Celem niniejszej pracy jest stworzenie metody pozwalającej na wstępną, automatyczną diagnostykę obiektu za pomocą danych hiperspektralnych oraz algorytmów uczenia maszynowego na przykładzie diagnostyki wybranych chorób w uprawach roślin.

2.1. Zakres pracy

Niniejsza praca została podzielona na 11 rozdziałów, które obejmują ogólne informacje dotyczące uprawy, opis konwencjonalnych metod diagnostycznych, przegląd literatury, opis koncepcji proponowanej metody, opis eksperymentu pomiarowego, wybór algorytmów uczenia maszynowego, propozycję metod przetwarzania danych hiperspektralnych na potrzeby diagnostyki chorób w uprawie roślin oraz opis i analizę wyników zrealizowanych badań naukowych.

Rozdział 3 przybliża podstawowe informacje na temat uprawy wykorzystanej w procesie pozyskiwania zbioru danych hiperspektralnych omawianych w niniejszej pracy, a także jej chorób. Poruszony został również temat skutków braku prawidłowej diagnostyki i reagowania na zagrożenia chorobowe. Przedstawione zostały konwencjonalne metody diagnostyczne oraz związane z nimi problemy technologiczne, które mogą być z powodzeniem rozwiązywane przez nieinwazyjne metody diagnostyki teledetekcyjnej, takie jak metody omawiane i proponowane w niniejszej pracy.

Rozdział 3.3 odpowiada na problem technologiczny związany z konwencjonalnymi metodami diagnostyki w uprawach. Przedstawione zostało rozeznanie literaturowe związane z diagnostyką chorób roślin z wykorzystaniem nieinwazyjnych metod teledetekcyjnych, takich jak obrazy w przestrzeni barwnej RGB, obrazowanie wielospektralne, obrazowanie hiperspektrlane, w tym pomiary hiperspektralne.

W **rozdziale 4** przedstawione zostały różne typy danych spektralnych wykorzystywanych w nieinwazyjnych metodach diagnostycznych. Przybliżono aktualny stan wiedzy w zakresie metod łączących zalety danych hiperspektralnych oraz uczenia maszynowego w celu prawidłowej diagnostyki chorób roślin. Zaprezentowano najważniejsze zakresy spektralne, wykorzystywane w dotychczasowych badaniach. Ze względu na charakter zbioru danych użytych w niniejszej pracy, tj. codzienne wykonywanie pomiarów, wyodrębniono i zaprezentowano również przegląd aktualnych metod wykorzystujących zbiory pomiarów wykonywanych wielokrotnie w trakcie prowadzenia obserwacji.

Rozdział 5 przedstawia w syntetyczny sposób propozycję koncepcji metody diagnostyki wybranych chorób oraz jej kolejne elementy zrealizowane w ramach niniejszej pracy doktorskiej. W rozdziale wypisano kolejne kroki badawcze prowadzące do opracowania zaproponowanej metody.

Rozdział 6 opisuje szczegóły przeprowadzonego eksperymentu pomiarowego. Przedstawiony został harmonogram przeprowadzonych badań oraz szczegóły fitopatologiczne. Istotnym elementem rozdziału jest część opisująca wykorzystane urządzenie pomiarowe oraz sam proces wykonywania pomiarów hiperspektralnych.

Rozdział 7 poświęcono na opis doboru metod uczenia maszynowego, które zastosowano w niniejszej pracy. Wspomniane zostały wszystkie klasyfikatory biorące udział we wstępnej selekcji, a także opisano w sposób bardziej szczegółowy wybrane klasyfikatory.

W **rozdziale 8** omówiono sposób postępowania z danymi od momentu wykonania pomiarów przez filtrację, wyodrębnienie metadanych, przetwarzania wstępnego danych, aż do stworzenia końcowego zbioru treningowego. Opisany został zbiór danych, objaśnione zostały metadane zawarte w zbiorze. Zaproponowana została metoda podziału zbioru z uwzględnieniem DPI (ang. *days post inoculum*) na potrzeby wykonania badań omawianych w niniejszej pracy

doktorskiej. Przedstawiono również analizę zbioru danych z podziałem na poszczególne grupy badawcze, uwzględniając najważniejsze parametry dotyczące podziału klasowego zbioru – przynależność do konkretnej jednostki chorobowej (podział na choroby, rys. 28). Przeanalizowano również parametry, które mogą mieć największy wpływ na widmo spektralne badanych obiektów, czyli odmianę rośliny. Powyższe analizy wykazują, że grupy badawcze wykorzystane w dalszych badaniach są równoliczne w ujęciu poszczególnych grup badawczych. Wyjątkiem jest podział na fitotrony ze względu na decyzje podjęte na etapie kolekcjonowania zbioru danych. Następnie przedstawiono analizę jakościową pomiarów za pomocą wskaźnika wegetacyjnego NDVI, która jasno pokazała, że wyjąwszy nieliczne wartości odstające, badane obiekty znajdowały się w dobrej lub bardzo dobrej kondycji wegetacyjnej, a różnice pomiędzy badanymi grupami nie były znaczące czy pozwalające na wnioskowanie na podstawie tego, popularnego wskaźnika.
W **rozdziale 9** przedstawiono wyniki badań mających na celu opracowanie metody postępowania z danymi hiperspektralnymi, które będą skutkowały najwyższą poprawą wyników klasyfikacji binarnej oraz zweryfikowanie hipotezy, że wybrane pasma spektralne są szczególnie istotne w procesie diagnostycznym.
W tym celu zaproponowano wykonanie eksperymentu dla różnych przypadków badawczych. Przedstawiono metodę wstępnej selekcji klasyfikatorów, na podstawie której wybrano 6 klasyfikatorów do dalszych analiz. Omówiono zaproponowaną metodę postępowania z danymi. Przedstawiono wyniki dla każdego z 6 etapów dla każdego klasyfikatora. Następnie przedstawiono wyniki metody eliminacji cech i wyznaczono zakresy spektralne najczęściej biorące udział w procesie decyzyjnym dla różnych przypadków badawczych. Dodatkowo podjęto próby stworzenia klasyfikatorów wieloklasowych.
W **rozdziale 10** przedstawiono podejście do badanego zagadnienia z podziałem zbioru uczącego w zależności od DPI, czyli liczby dni, które minęły od inokulacji. Zaproponowano podział na 3 różne wartości parametru T (1, 3 oraz 5). Badania przeprowadzono w sposób analogiczny do metodyki przedstawionej w 9. Wykonano analizę wyników w ujęciu klasyfikatorów, przypadków chorobowych, a także zastosowania różnych wartości parametru T. Uzyskane wyniki wskazują na możliwość diagnostyki badanych chorób w okresie krótszym niż 7 dni od zakażenia, co potwierdza trzecią tezę niniejszej pracy doktorskiej.

2.2. Teza pracy

Celem pracy jest wykazanie prawdziwości następującej tezy:

Istnieje możliwość wykorzystania pomiarów hiperspektralnych reflektancji z zakresu fal o długości 350-2500nm i algorytmów uczenia maszynowego w procesie automatycznej diagnostyki *Solanum lycopersicum* dla potrzeb klasyfikacji wybranych jednostek chorobowych.

Powyższą tezę uzupełniają następujące tezy pomocnicze:

1. Możliwe jest określenie algorytmów klasyfikacyjnych, które z większą efektywnością pozwalają na rozpoznanie następujących chorób pomidora zwyczajnego: alternarioza, antraknoza, bakteryjna cętkowatość pomidora, septorioza, zaraza ziemniaka w pomidorze z zastosowaniem ograniczonego zbioru danych uczących,

2. Możliwe jest określenie specyficznych długości zakresu widma odpowiadających określonym jednostkom chorobowym *Solanum lycopersicum*,
3. Hiperspektralne pomiary reflektancji pozwalają na klasyfikację jednostek chorobowych *Solanum lycopersicum* w czasie szybszym niż metoda oceny wizualnej stosowana w praktyce.

3. Zastosowanie metrologii w diagnostyce chorób roślin

Obiektem badanym był pomidor zwyczajny (*Solanum lycopersicum*) w dwóch odmianach (Benito oraz Polfast). Szczegóły przedstawiono w rozdziale 3.1, celem wykonywanych prac była klasyfikacja wybranych chorób pomidora, które to choroby przedstawiono w rozdziale 3.2.

3.1. Obiekt badań

Pomidor (łac. *Solanum lycopersicum*) jest jedną z najbardziej rozpowszechnionych roślin uprawnych na świecie. Zgodnie z Polską Klasyfikacą Wyrobów i Usług owoce pomidora zaliczają się do kategorii „01.13.3 pozostałe warzywa owocowe" [15]. Jego produkcja wynosi 160 milionów ton rocznie i jest drugim najchętniej konsumowanym warzywem na świecie [16]. Niestety roślina ta jest podatna na choroby pochodzenia grzybowego, bakteryjnego czy wirusowego, takie jak zaraza ziemniaka, bakteryjna cętkowatość, alternarioza czy mączniak prawdziwy [17]. Niektóre z tych chorób, jak na przykład alternarioza, mogą powodować redukcję plonu o 80% przy sprzyjających dla patogenu warunkach [18, 19]. Rolnicy przeciwdziałają chorobom na różne sposoby np. poprzez opryski fungicydowe czy sadząc rośliny bardziej odporne na patogeny, jednak kluczowym elementem procesu przeciwdziałania skutkom chorób jest prawidłowa diagnoza. Standardowa metoda diagnostyczna polega na eksperckiej weryfikacji symptomów choroby oraz analizie tkanek rośliny pod mikroskopem.

Konwencjonalne metody rozpoznawania chorób zakładają obserwację symptomów chorobowych roślin. Niestety symptomy różnych chorób mogą być do siebie podobne, co utrudnia prawidłowe rozpoznanie choroby nawet doświadczonym rolnikom, dlatego szczegółowe badania wymagają pobranie materiału roślinnego i przeprowadzenia badania laboratoryjnego [20]. W przypadku chorób grzybowych patogen musi zostać wyizolowany, a następnie wyhodowany na specjalnej pożywce. Finalnie jego obecność potwierdza się za pomocą obserwacji mikroskopowej, co wymaga eksperckiej wiedzy z zakresu taksonomii grzybów [20]. Choroby pochodzenia bakteryjnego rozpoznawane są najczęściej za pomocą testów biochemicznych, z kolei wirusowe – na podstawie materiału genetycznego, analizy transmisji choroby [21]. Jednocześnie, jak podaje Sakudo i in. [22], niektóre metody inwazyjne, jak m.in. test immunoenzymatyczny, reakcja łańcuchowa polimerazy czy test immunofluorescencyjny, nie są idealnymi rozwiązaniami pod kątem efektywności kosztowej, szybkości i dokładności pomiaru, natomiast spektroskopia bliskiej podczerwieni pozwala na równie dobrą, efektywną kosztowo i nieinwazyjną diagnozę roślin. Dodatkowo metody inwazyjne nie są skalowalne, natomiast metodę laboratoryjną, opartą o dane hiperspektralne można próbować z powodzeniem przenieść na platformy lotnicze (bezzałogowe statki powietrzne, załogowe statki powietrzne, satelity) [23]. Istotne jest tworzenie szybkich i możliwie niezawodnych metod diagnostyki roślin, co potwierdzają różne badania [21, 22, 24].

3.2. Analizowane choroby i patogeny je wywołujące

Choroby roślin dzielimy na grupy ze względu na ich pochodzenie. Występują zatem choroby pochodzenia wirusowego, grzybowego oraz bakteryjnego. Dodatkowo roślinom zagrażają również niedobory składników odżywczych, a także ich nadmiar, szkodniki oraz rośliny pasożytnicze. Na potrzeby stworzenia metody opisanej w niniejszej pracy i pozwalającej na automatyczną diagnostykę obiektów organicznych wykorzystane zostały pomiary

hiperspektralne uprawy *Solanum lycopersicum* w odmianach Benito, Polfast, Dyno, Guadivia pod kątem zakażenia pięcioma różnymi chorobami:

1. **Alternarioza** lub sucha plamistość liści, ang. *Early Blight*, łac. *Alternaria solani*, dalej określana skrótem **EB**

   Rodzaj choroby: choroba grzybowa
   Występowanie, szkodliwość oraz objawy [17, 25]:
   Choroba ta występuje nie tylko w uprawie pomidora, ale również w uprawie ziemniaka oraz niektórych innych roślin psiankowatych. Znaczenie ekonomiczne ma w szczególności na bardzo podatnych odmianach tych gatunków roślin. Objawia się często już na początku okresu wegetacji w postaci brunatonoczarnych, współśrodkowo strefowych plam na liściach o średnicy 0,5-2 cm. Plamy te początkowo są owalne, po czym przybierają kształt wieloboku ze względu na ograniczające je nerwy liścia. Wokół brunatnych plam pojawia się żółte zabarwienie tkanek prowadzące do całkowitego przedwczesnego zamierania liści. Cykl rozwojowy jest bardzo krótki i trwa do tygodnia.

rys. 1 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019

2. **Antraknoza**, ang. *Antracnose*, łac. *Colletotrichum coccodes*, dalej określana skrótem **AN**

   Rodzaj choroby: choroba grzybowa
   Występowanie, szkodliwość oraz objawy [17, 25]:
   Choroba ta występuje również powszechnie w uprawie ziemniaka. W uprawie pomidora powoduje często duże straty w plonie owoców. Porażone są owoce dojrzałe oraz przejrzałe. Patogen atakuje korzenie i podstawy łodyg roślin, co powoduje więdnięcie, żółknięcie i zamieranie liści. Na dojrzałych owocach pojawiają się plamy jasnobrązowe, wklęsłe, współśrodkowo strefowane z czarnymi sklerocjami (wytworami grzybni). W czasie deszczowej pogody na powierzchni porażonych owoców pojawią się różowe skupienia zarodnikowania konidialnego (powstające w procesie odcięcia komórki). Podobne objawy chorobowe niekiedy powstają również na liściach i innych nadziemnych częściach roślin.

rys. 2 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019

3. **Bakteryjna cętkowatość**, ang. *Bacterial Speck*, łac. *Pseudomonas syringae*, dalej określana skrótem **BS**

Rodzaj choroby: choroba bakteryjna
Występowanie, szkodliwość oraz objawy [17, 25]:

Choroba znana w Polsce od połowy lat 70. XX w. Od tego czasu występuje co roku, a o jej nasileniu decydują warunki atmosferyczne. Szkodliwość polega przede wszystkim na utracie wartości handlowej owoców. Porażenie roślin w okresie kwitnienia powoduje opadanie kwiatów, a także opóźnienie dojrzewania owoców, co ma bezpośredni wpływ na uzyskiwany plon. Pierwsze objawy choroby to małe, okrągłe, ciemnozielone, uwodnione plamki z/lub bez wąskiego, żółtego halo (efektu wytwarzania toksyny - koronatyny) na liścieniach i pierwszych liściach. Plamki osiągają średnicę 2-5mm, stają się szaroczarne lub brązowa, a tkanka w ich obrębie ulega nekrozie. W warunkach wysokiej wilgotności od spodniej strony blaszki liściowej widoczny jest szarawy, śluzowaty wyciek komórek bakteryjnych. W warunkach epidemicznego nasilenia choroby może dojść do przedwczesnego zamierania i opadania kwiatów oraz liści. Przy dłużej utrzymującym się zwilżeniu liści pierwsze objawy mogą wystąpić już po 3-5 dniach.

rys. 3 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019

4. **Septorioza**, ang. *Septoria Leaf Spot*, łac. *Septoria lycopersici*, dalej określana skrótem **SL**

Rodzaj choroby: choroba grzybowa
Występowanie, szkodliwość oraz objawy [17, 25]:

Występuje we wszystkich rejonach polowej uprawy pomidora. Powoduje zmniejszenie liczby zawiązywanych owoców oraz ich wielkości, jednak na samych owocach nie występują objawy choroby. Pierwsze objawy występują na najstarszych częściach roślin, a następnie przechodzą na części młodsze. Na blaszkach, rzadziej na ogonkach liściowych, działkach kielicha i szypułkach owoców występują początkowo brunatne plamy, które otoczone są wąskim, chlorotycznym (wynikającym z zakłócenia syntezy chlorofilu) lub żółtym pierścieniem. Na blaszkach liściowych plamy te są zwykle owalne o średnicy około 5 mm, a na innych częściach roślin są podłużne. Plamy te z czasem stają się prawie białe w środkowej części. Silne porażenie liści powoduje ich zamieranie, a kwiaty opadają, co przyczynia się do zmniejszenia wielkości i liczby zawiązanych owoców.

rys. 4 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019

5. **Zaraza ziemniaka**, ang. *Late Blight*, łac. *Phytophthora infestans*, dalej określana skrótem **LB**

Rodzaj choroby: choroba grzybowa
Występowanie, szkodliwość oraz objawy [17, 25]:

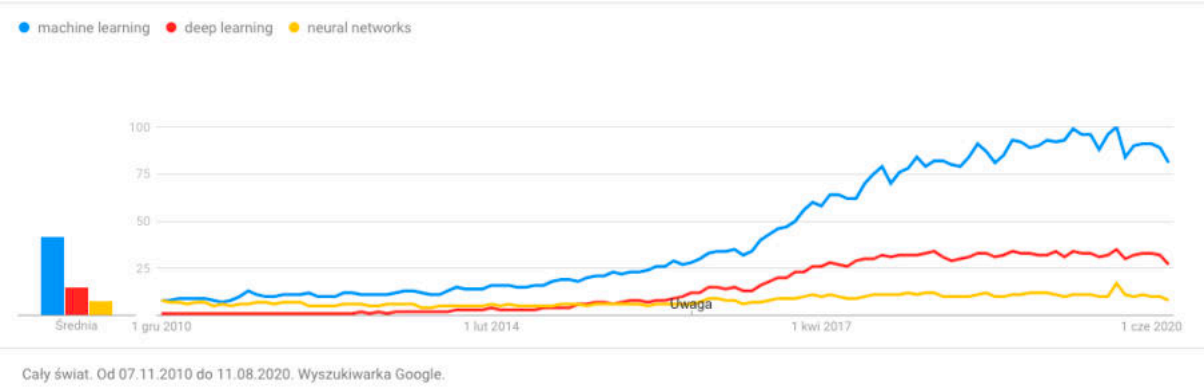
Występuje głównie w uprawie ziemniaka oraz pomidora. W warunkach optymalnych choroba ta rozprzestrzenia się nadzwyczaj szybko i przyczynia się do obniżenia plonu w granicach 20-50%, a niekiedy przy wczesnym wystąpieniu do całkowitego jego zniszczenia. Szkody często ulegają powiększeniu przez gnicie podczas przechowywania zakażonych owoców pomidora. Na pomidorze choroba ta występuje najczęściej w uprawach gruntowych i pod folią. Bardzo rzadko lub w ogóle nie występuje w uprawach szklarniowych. Choroba objawia się początkowo na nadziemnych częściach roślin (liściach, łodygach, a także owocach) w postaci brunatnych, stopniowo powiększających się, niekształtnych plam. Na pograniczu zdrowej i chorej części rośliny pojawia się biały nalot, widoczny niemal jedynie po spodniej stronie liści. Nalot ten zanika po kilkudniowej suchej pogodzie. Termin wystąpienia choroby zależy przede wszystkim od warunków atmosferycznych.

rys. 5 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019

3.3. Diagnostyka chorób roślin na podstawie analizy sygnałów pomiarowych

Diagnostyka chorób roślin jest nauką zajmującą się m. in. sposobami rozpoznawania chorób na podstawie ich charakterystycznych objawów. Do określenia tych objawów mogą służyć metody diagnostyczne oparte o dane pozyskane za pomocą przyrządów pomiarowych. W medycynie wykorzystywane są na przykład tomografy komputerowe i ekspercka wiedza lekarza, w energetyce wykorzystywany jest z kolei analizator drgań silnika elektrycznego oraz wiedza doświadczonego pracownika laboratorium diagnostycznego. Proces diagnostyczny polega zatem na pozyskaniu sygnału pomiarowego badanego obiektu, jego przetworzeniu za pomocą modelu diagnostycznego oraz finalnym określeniu diagnozy.

Coraz częściej metody eksperckie wspomagane są modelami analitycznymi zawartymi w specjalistycznym oprogramowaniu. Od połowy minionej dekady rosnącym zainteresowaniem cieszą się również algorytmy uczenia maszynowego i uczenia głębokiego oraz systemy wspomagania decyzyjnego oparte o te algorytmy.



rys. 6 Zainteresowanie technologiami a) uczenia maszynowego, b) uczenia głębokiego oraz c) sieci neuronowych w ujęciu czasowym za lata 2010-2020 [26]

Algorytmy uczenia maszynowego, uczenia głębokiego czy sieci neuronowych dobrze sprawdzają się w przypadkach diagnostyki opartej na analizie dużych zbiorów danych, jakich jak np. dane hiperspektralne.

Wykrywanie chorób roślin cieszy się dużym zainteresowaniem wśród naukowców, o czym świadczą liczne publikacje [21, 23, 24, 27–67]. Część autorów zauważa, że silnie rozwijający się przemysł agrotechniczny i przetwórstwa żywności wymaga tworzenie nieinwazyjnych metod diagnostycznych [22, 45], o czym wspomniano już w rozdziale 3.1 przy okazji omawiania konwencjonalnych metod diagnostyki chorób upraw. Istotną część publikacji dotyczących wykrywania chorób obejmują uprawy pomidorów [23, 27, 29–32, 34–36, 41, 43, 47, 48, 61, 68].
Badacze wykorzystują w tym celu zarówno kamery cyfrowe [28, 36, 41, 43, 47, 65, 69], obrazowanie wielospektralne [35, 62], jak również hiperspektralne [23, 27, 29–33], 38, 39, 44, 46, 48, 55, 57, 60, 61, 64, 66, 68, 70, 71]. Podział publikacji ze względu na wykorzystane źródło danych przedstawiono w tab. 1. Wielu autorów proponuje również rozwiązania oparte o pomiary hiperspektralne wykonane za pomocą spektrofotometru z zakresu światła widzialnego (VIS), bliskiej podczerwieni (NIR), fal krótkich (SWIR) oraz zakresu fal termalnych. Często badanym zakresem spektralnym był zakres 350-2500nm [23, 27, 29–33, 38, 39, 42, 46, 48, 54–57, 60, 61, 64, 66, 68, 70–72]. Badana była również możliwość określania chorób na bazie indeksów spektralnych [23, 27, 30, 46, 55].

tab. 1 Zestawienie publikacji dot. wykorzystania różnych źródeł danych w detekcji chorób

Źródło danych	**Publikacje**
Obrazy przestrzeni barwnej RGB	[28, 36, 41, 43, 47, 65, 69]
Dane wielospektralne	[35, 62]
Dane hiperspektralne	[23, 27, 29–33, 38, 39, 44, 46, 48, 55, 57, 60, 61, 64, 66, 68, 70, 71]

Diagnostyka roślin za pomocą analizy danych hiperspektralnych była wykonywana już między innymi dla pszenicy [7], [38], [46], [57], jęczmienia [38], [66], kukurydzy [71], buraka cukrowego[55], [56], [60], [66], ziemniaków [33], [70], pomidora [23], [27], [29]–[32], [48], [61], [68] oraz innych upraw.

tab. 2 Zestawienie publikacji nt. wykrywania choroba na podstawie danych hiperspektralnych w zależności od gatunku odmiany

Uprawa	**Publikacje**
Pszenica	[7, 38, 46, 57]
Jęczmień	[38, 66]
Kukurydza	[71]
Burak cukrowy	[55, 56, 60, 66]
Ziemniak	[33, 70]
Pomidor	[23, 27, 29–32, 48, 61, 68]

W przypadku pomidora wykorzystywane były zarówno hiperspektralne dane pomiarowe [23, 27, 31, 32, 61, 68] jak i hiperspektralne dane obrazowe [23, 29, 30, 48]. Do badań z użyciem danych obrazowych wykorzystano zakres spektralny VIS-NIR, tj. z przedziału od 400nm do 1000nm. Badania wykonane przy użyciu spektrometrów obejmowały pomiary z przedziału 200-2500nm.

3.4. Podsumowanie rozdziału

Rozdział przybliża podstawowe informacje na temat uprawy wykorzystanej w procesie pozyskiwania zbioru danych hiperspektralnych omawianych w niniejszej pracy, a także jej chorób. Poruszony został również temat skutków braku prawidłowej diagnostyki i reagowania na zagrożenia chorobowe. Przedstawione zostały konwencjonalne metody diagnostyczne oraz związane z nimi problemy technologiczne, które mogą być rozwiązywane przez nieinwazyjne metody diagnostyki teledetekcyjnej, takie jak metody omawiane i proponowane w niniejszej pracy.

W rozdziale przedstawione zostało także rozeznanie literaturowe związane z diagnostyką chorób roślin z wykorzystaniem nieinwazyjnych metod teledetekcyjnych, takich jak analiza obrazy przestrzeni barwnej RGB, obrazowanie wielospektralne, obrazowanie hiperspektralne, w tym pomiary hiperspektralne. W kolejnym rozdziale autor doprecyzowuje przegląd wykorzystania danych hiperspektralnych oraz metod uczenia maszynowego.

4. Zastosowanie danych hiperspektralnych w diagnostyce uprawy pomidora – stan techniki

W niniejszym rozdziale przytoczono opracowania dotyczące diagnostyki uprawy pomidora z wykorzystaniem danych hiperspektralnych. W rozdziale 3.3. omówiono publikacje wykorzystujące tę technikę pomiarową w badaniach diagnostycznych również innych chorób.

4.1. Typy danych spektralnych

Dane spektralne, multispektralne czy hiperspektrlane pozyskuje się za pomocą urządzeń spektrometrycznych. Szczegółowy opis urządzenia pomiarowego wykorzystanego do pozyskania danych w ramach niniejszej pracy doktorskiej oraz sposobu jego działania został umieszczony w rozdziale 6.4.
W zależności od liczby rejestrowanych pasm spektralnych, pozyskane dane dzielimy na:

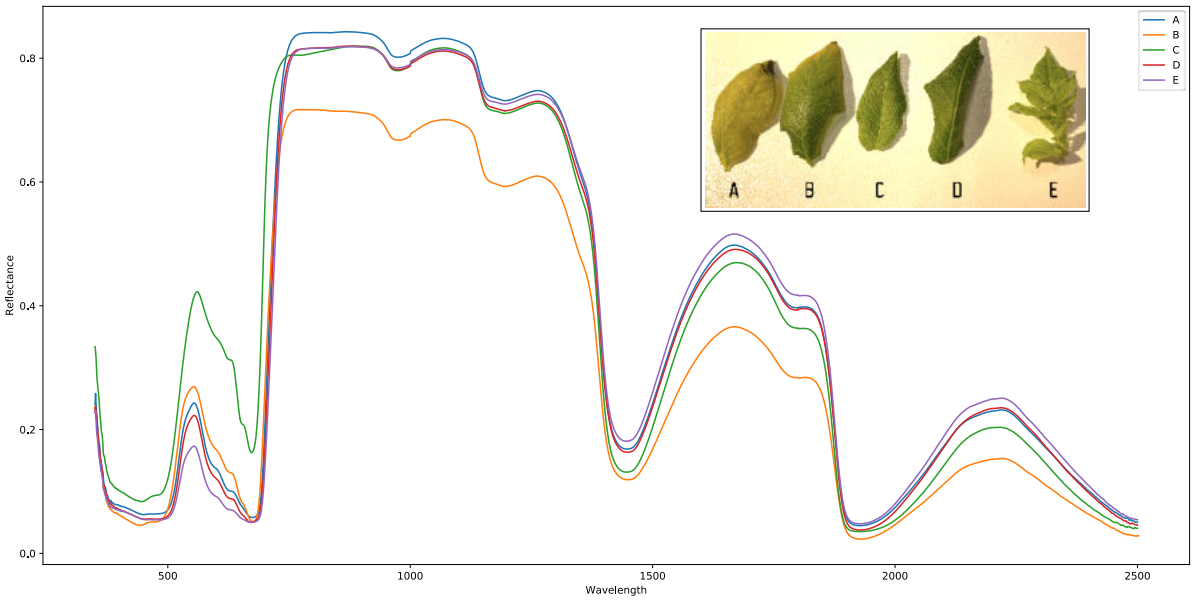
- dane spektralne (pojedyncze pasma),
- wielospektralne (kilkanaście pasm),
- hiperspektralne (kilkaset pasm),

przy czym pasmem spektralnym nazywamy zakres widma fali elektromagnetycznej o znanej długości i szerokości. W literaturze można spotkać się z wymiennym stosowaniem słów pasmo i kanał. Przyjęty podział nazewnictwa danych spektralnych pozwala zorientować się z jakiego typu wymiarowością i poziomem skomplikowania danych ma się do czynienia. Pojedyncze pasmo spektralne może zostać zarejestrowane w przestrzeni jedno- lub dwuwymiarowej. Te pierwsze rejestruje się za pomocą różnego typu spektrometrów, natomiast te drugie pozyskiwane są za pomocą kamer spektralnych.

4.2. Pomiary spektralne

W diagnostyce materiałowej oraz geologii powszechnie wykorzystywanym narzędziem jest spektroradiometr, który pozwala na wykonywanie pomiarów hiperspektralnych. Urządzenie to zazwyczaj charakteryzuje następujące parametry pomiarowe: wykonywanie pomiarów sygnału w zakresie spektralnym 350-2500nm przy rozdzielczości spektralnej 3nm oraz 8nm odpowiednio dla długości 700nm oraz 1400nm i 2100nm. Szerokość próbkowania widmowego

na poziomie 1.4nm dla zakresu UV-VNIR (350-1000nm) oraz 1.1nm dla zakresu NIR-SWIR (1001-2500nm).



rys. 7 Przykładowe pomiary spektralne wykonane za pomocą urządzenia pomiarowego ASD FieldSpec-4 Hi-Res

Przedstawiony na powyższym wykresie pomiar wykonany został za pomocą przenośnego spektrofotometru hiperspektralnego. Urządzenie to co prawda jest nazywane urządzeniem mobilnym, jednak akwizycja danych za pomocą tego urządzenia niesie ze sobą również pewne wady. Waga samego instrumentu pomiarowego to 5.44kg, jednak cały wymagany osprzęt waży ponad 10kg, co nastręcza trudności przy całodniowych pomiarach terenowych. Jednocześnie aparatura ta jest precyzyjnym sprzętem pomiarowym.

4.3. Obrazowanie wielospektralne

Obrazowanie wielospektralne pozwala na rejestrację kilku do kilkunastu obrazów reprezentujących reflektancję fali elektromagnetycznej dla konkretnej długości. Obrazy takie uzyskiwane są za pomocą kamer wielospektralnych i pozwalają m.in. na określanie poziomu wegetacji na polu. Przykładem urządzeń tego typu mogą być np. kamera Parrot Sequoia+ [73] czy DJI P4 Multispectral [74]. Z satelitarnych rozwiązań tego typu na szczególną uwagę zasługują misje kosmiczne Sentinel 2 [ESA, [75]] czy Landsat 8 [NASA, [76]]. Coraz więcej pojawia się też systemów, które pozwalają na analizę Ziemi z kosmosu za pomocą wysokorozdzielczych danych multispektralnych, jak na przykład WorldView-3 [DigitalGlobe, [77]]. Przykład obrazowania multispektralnego został przedstawiony i omówiony w rozdziale 6.3 na rysunkach rys. 18–rys. 20. Obrazowanie tego typu wymaga dodatkowych kroków pośrednich przed przystąpieniem do analizy danych, takich jak wykonanie korekcji geometrycznej, czy korekcji atmosferycznej.

4.4. Obrazowanie hiperspektralne

Obrazowanie hiperspektralne jest zbliżone do obrazowania wielospektralnego. Podstawową różnicą, jak już wspomniano, jest liczba dostępnych kanałów spektralnych. Z drugiej szerokość tych kanałów jest znacznie węższa niż w przypadku obrazowania multispektralnego, co pozwala na bardziej wnikliwą analizę. Przykładem urządzenia pozwalającego na akwizycję obrazowania hiperspektralnego jest kamera Headwall Hyperspec Co-Aligned VNIR-SWIR [78] czy Zebra X1 przedstawiona na poniższych zdjęciach.

rys. 8 Kamera hiperspektralna Zebra X1 przymocowana do bezzałogowego statku latającego DJI Matrice 600 [autor zdjęcia: Krzysztof Płatek]

rys. 9 Kamera hiperspektralna Zebra X1

Tak samo, jak w przypadku obrazowania wielospektralnego dostępne są pozyskanie obrazowania hiperspekralnego z orbity, jednak liczba satelitów wyposażonych w instrumenty służące do akwizycji obrazowania hiperspektralnego jest znacznie mniejsza niż tych wyposażonych w instrumenty do akwizycji obrazownia multispektralnego. Najważniejszym obecnie satelitą operacyjnym jest włoski satelita PRISMA [ASI, [79]]. W chwili pisania niniejszej pracy w fazie uruchomieniowej znajduje się niemiecki satelita EnMAP [DLR, [80]], a w roku 2023 ma zostać uruchomiony polski satelita hiperspektralny Intuition-I [KP Labs, [81]]. Ten ostatni jest, w przeciwieństwie do poprzedników będzie wykonywał przetwarzania danych na orbicie, co rozwiązać problem ograniczonego przesyłu dużych wolumenów danych.

To właśnie rozmiar danych jest głównym problemem, ponieważ pojedynczy zestaw danych może zajmować nawet kilkaset gigabajtów przestrzeni dyskowej, a co z tym się wiąże – dłuższym czasem przetwarzania danych niż w przypadku danych wielospektralnych. Dodatkowo, tak jak w przypadku obrazownia multispektralnego wymagane jest wykonanie odpowiednich, –odatkowych kroków pośrednich - korekcji geometryczna, czy korekcja atmosferyczna.

W niniejszej pracy Autor skupił się na opracowaniu metody opartej o precyzyjne pomiary hiperspektralne wykonane za pomocą spektroradiometru ASD FieldSpec 4 Hi-Res w środowisku kontrolowanym, a pomiary były wykonywane przy znanym oświetleniu. Urządzenie operuje w zakresie spektralnym 350-2500 nm z rozdzielczością spektralną 3nm oraz 8nm odpowiednio dla długości 700nm oraz 1400nm i 2100nm, natomiast szerokość próbkowania widmowego wynosi 1.4nm dla zakresu UV-VNIR (350-1000nm) oraz 1.1nm dla zakresu NIR-SWIR (1001-2500nm). Szczegóły dotyczące wykorzystanego sprzętu zostały opisane w rozdziale 6.3.

4.5. Pomiary hiperspektralne w diagnostyce uprawy pomidora

Do tej pory badane były Alternarioza pomidora (*ang. Early Blight*) [31, 32, 48], Zaraza ziemniaka w pomidorze (*ang. Late Blight*) [23, 27, 48, 61], Bakteryjna cętkowatość pomidora (*ang. Bacterial Spot*) [27, 68], Szara pleśń (*ang. Gray Mold*) [29] oraz Brązowa plamistość liści pomidora (*ang. Tomato Spotted Wilt Virus – TSWV*) [30].

Zhang i in. [23] w 2004 roku opublikował pracę opisującą badanie, w którym naukowcy zbadali chorą plantację pomidora na obecność Zarazy ziemniaka. W ramach badania wykonali pomiary z wykorzystaniem spektrometru GER-2600, a następnie, na podstawie uśrednionego pomiaru wybranych zakresów spektralnych (600-690, 750-930nm, 950-1030nm, 1040-1130nm, 1450-1850nm, 2000nm, 2400nm) zaproponowali 6 indeksów dedykowanych do wykrycia badanej choroby i określenia fazy rozwojowej choroby – zdrowa roślina, 4 kolejne fazy chorobowe oraz gleba. Powstałe w ten sposób wskaźniki zostały wykorzystane do analizy lotniczego obrazowa–ia hiperspektralnego (*AVIRIS - Airborne visible/infrared imaging spectrometer*).

Podobne badanie opublikował Wang i in. [61] w 2008 roku. Różnica w tym badaniu w stosunku do poprzedniego polegała na wykorzystaniu sieci neuronowej wstecznej propagacji (BPNN) opartej na algorytmie gradientu prostego. Zastosowaną architekturą sieci była struktura 3-25-9-1. Dla danych hiperspektralnych pozyskanych za pomocą spektrometru osiągnięto współczynnik korelacji R=0.99.

Opublikowane w 2010 badania [68] nt. Bakteryjnej cętkowatości pomidora wykonane były na nieco szerszym, pod względem zakresu spektralnego, zbiorze danych, ponieważ obejmowały również przestrzeń ultrafioletową (UV – 200-380nm). W badaniu tym wykorzystano regresję liniową metody cząstkowych najmniejszych kwadratów (partial least squares, PLS) oraz krokowej regresji wieloliniowej (Stepwise multilinear regression, SMLR). Badaczom udało się uzyskać współczynnik determinacji na poziomie R^2=0.82.

Xie i in. w 2 swoich pracach [29, 48] opisał wykorzystanie danych hiperspektralnych i narzędzia uczenia maszynowego w procesie wykrywania 3 chorób – Alternariozy, Zarazy ziemniaka oraz Szarej pleśni. W obu pracach wykorzystał obrazowe dane hiperspektralne z zakresu 380-1023nm. W przypadku wykrywania Alternariozy i Zarazy ziemniaka autorzy prac wykorzystali metodę Extreme Learning Machine (ELM) oraz Successive Projections Algorithm (SPA). Na podstawie ELM stworzony został model na pełnym zakresie spektralnym. Kolejnym podejściem było wykorzystanie SPA do zmniejszenia wymiarowości danych poprzez zawężenie zakresu spektralnego do kilku pasm (442nm, 508nm, 573nm, 696nm oraz 715nm), a następnie użycie ELM do ponownego wytrenowania modelu. Naukowcom udało się osiągnąć dokładność na poziomie 97.1-100% w zależności od choroby i zastosowanego podejścia. W przypadku drugiego badania dotyczącego wykrywania Szarej pleśni wykorzystano metodę KNN, C5.0 oraz FR-KNN i uzyskano dokładność na poziomie 85.71-100%.

tab. 3 Zestawienie wyników badań nad wykrywanie chorób roślin za pomocą danych hiperspektralnych

Źródło danych	**Choroba**	**Metoda detekcji**	**Metryka**	**Wyniki**	**Publikacja**
HS (380-1023 nm)	Alternariozy, Zarazy ziemniaka	Extreme Learning Machine (ELM) oraz Successive Projections Algorithm (SPA)	dokładność [%]	97.1-100	[29]

HS (380-1023 nm)	Szara pleśń	K najbliższych sąsiadów (KNN), C5.0 oraz FR-KNN	dokładność [%]	85.71-100	[48]
HS (200-380 nm)	Bakteryjnej cętkowatości	regresja liniowa metody cząstkowych najmniejszych kwadratów (PLS) oraz krokowej regresji wieloliniowej (SMLR)	R^2	0.82	[68]
HS (400-2500 nm)	Zaraza ziemniaka	stworzenie autorskich indeksów chorobowych	-	-	[23]
HS (400-2500 nm)	Zaraza ziemniaka	sieć neuronowa wstecznej propagacji (BPNN)	R^2	0.62	[61]

Powyższa analiza aktualnego stanu wiedzy, przedstawiona w tym oraz poprzednich rozdziałach, potwierdza, że cel i zakres badań wykonanych w ramach niniejszej pracy doktorskiej jest zasadny, cieszący się zainteresowaniem w świecie nauki i potrzebny z punktu widzenia rozwoju nauki oraz przemysłu.

4.6. Przegląd dotychczasowych rozwiązań – istotne zakresy spektralne

W rozdziale 9. (Wykrywanie chorób z wykorzystaniem wybranych algorytmów uczenia maszynowego) przedstawiono wyniki prac dotyczących klasyfikacji chorób na podstawie danych hiperspektralnych. Poniższy przegląd literatury uzupełnia powyższe informacje w kontekście badań zawartych we wspomnianym rozdziale.
Próby diagnostyki upraw i identyfikacji chorób *Solanum lycopersicum* na bazie pomiarów spektralnych wykonywali już np.:

- Publikacja P1 – Zhang i in. [23],
- Publikacja P2 – Wang i in. [61],
- Publikacja P3 – Jones i in. [68],
- Publikacja P4 – Moghadam P. i in [30],
- Publikacja P5 – Xie C. i in. [29, 48],
- Publikacja P6 – Lu J. i in.[27].

We wspomnianych publikacjach wykorzystano pomiary spektralne od ultrafioletu (UV) przez zakres światła widzialnego (VIS) oraz bliskiej podczerwieni (NIR) do krótkich fal podczerwonych (SWIR). W P1 i P2 wykorzystano spektroradiometr GER o zakresie spektralnym 400-2500nm, a w P3 – spektroradiometr Cary 500 o zakresie spektralnym 200-2500nm. W publikacji P4 wykorzystano zestaw 2 kamer hiperspektralnych firmy Headwall obrazujących w przestrzeniach spektralnych 400-1000nm oraz 900-2500nm. Xie w swych badaniach (P5) wykorzystuje kamerę hiperspektralną działającą w przestrzeni 380-1023nm. Z kolei w badaniach Lu (P6) wykorzystywany był ręczny spektroradiometr wykonujący pomiary w zakresie 400-1050nm.

W opracowaniach P1 i P2 wykorzystano zbiór danych spektralnych obejmujący 60 roślin oraz 1074 punkty pomiarowe, a dla P2 dodatkowo wykorzystano 195 obrazów hiperspektralnych z zakresu 700-1130nm programu AVIRIS. Wyznaczono wartości średnie dla 6 przedziałów spektralnych (600-690nm, 750-930nm, 950-1030nm, 1040-1130nm, 1450-1850nm, 2000-2400nm) i zaproponowano indeksy chorobowe będące ilorazem dwóch zakresów spektralnych.
Efektem publikacji P1 było wyodrębnienie przestrzeni spektralnej najbardziej nadającej się do teledetekcyjnego wykrywania zarazy ziemniaka w pomidorze, natomiast w publikacji P2 stworzono sztuczną sieć neuronową wstecznej propagacji do rozwiązania tego samego problemu. Autorzy badania P1 wskazują, że zakresy 750-930nm, a następnie 950-1030nm oraz 1040-1130 są szczególnie istotne w procesie wykrywania zarazy ziemniaka w pomidorze. Zwracają oni również uwagę na 5 szczególnych wąskich zakresów pasm spektralnych, w których występują wysokie różnice w zarejestrowanej reflektancji dla roślin chorych i zdrowych (625nm, 850nm, 1050nm, 1500nm i 2100nm), na podstawie których opracowali własne indeksy wegetacyjne (Vegetation Ind–ces – VI). Autorzy badania P2 wskazują zakresy 750-1350nm (dla pomiarów polowych) 700-1105nm (dla obrazowania hiperspektrlanego AVIRIS) jako obszary pozwalające na rozróżnienie obiektów chorych od zdrowych. Autorzy skupili się przede wszystkim na obiektach, w których symptomy chorobowe są już widoczne gołym okiem i można określić fazę jej rozwoju.

W opracowaniu P3 posłużono się widmowymi współczynnikami korelacji oraz metodami PLS (regresja liniowa metody cząstkowych najmniejszych kwadratów) i SMLR (krokowa regresja wielomianowa) do zdiagnozowania bakteryjnej plamistości pomidora. Badany zbiór danych obejmował 156 pomiarów hiperspektralnych. Autorzy badania P3 wskazują na istotną korelację pomiędzy absorbancją zaobserwowaną w pasmach 384nm, 626nm, 691nm oraz 761nm, a nasileniem choroby. Jako kluczowe pasma autorzy wskazują jednak zakres 750-760nm, gdyż długości fal z tego zakresu pojawiły się w każdym z trzech użytych podejść. Dodatkowo wskazują oni również jako istotne pasma 395nm, 400nm, 630nm oraz zakres 633-635nm.

Autorzy publikacji P4 zaimplementowali 3 różne podejścia do tematu rozpoznawania chorych i zdrowych roślin. Badacze postawili sobie za cel rozpoznanie roślin kontrolnych od tych zarażonych brązową plamistością liści pomidora (ang. *Tomato Spotted Wilt Virus*, TSWV) za pomocą obrazowania hiperspektralnego z zakresu 400-2500nm. Pierwszym podejściem ekstrakcji cech zastosowanym w publikacji było wykorzystanie istniejących indeksów wegetacyjnych (12 różnych indeksów), drugim – wykorzystanie całego dostępnego widma spektralnego, a trzecim – metoda probabilistycznego modelowania tematu (*Probabilistic Topic Modeling, PTM*). PTM jest metodą często wykorzystywaną w przetwarzaniu języka naturalnego (ang. *Natural Language Processing*, NLP) do określania tematu, o jakim traktuje dany dokument. Badacze jako dokumenty potraktowali piksele hiperspektralne, natomiast wartość pomiarowa dla konkretnej długości fali była traktowana jak słowo zawarte w tym dokumencie.
W eksperymencie wykorzystano 30 zakażonych i 30 kontrolnych roślin 1 odmiany, które obserwowano przez 21 dni. Wykonano 6 pomiarów w okresie od 3. Do 21. Dnia od zakażenia. Łącznie zeskanowano 133 liście zakażone oraz 103 liście kontrolne. Określono dywergencję Kullbacka-Leiblera (entropię względną) do określenia rozbieżności pomiędzy rozkładami wartości dla zakażonych i kontrolnych pomiarów.
Najlepsze rezultaty uzyskano przy wykorzystaniu pełnego zakresu spektralnego. Uzyskano wartość wskaźnika F1 na poziomie 0,92 dla pełnego zakresu SWIR oraz 0,939 dla zakresu VNIR, jednak wszystkie przedstawione metody charakteryzowały się wysoką wartością

wskaźnika F1 na poziomie >= 0,8. Autorzy nie podają konkretnych zakresów spektralnych, na które należy zwrócić uwagę tworząc rozwiązania przeznaczone do klasyfikacji chorób.

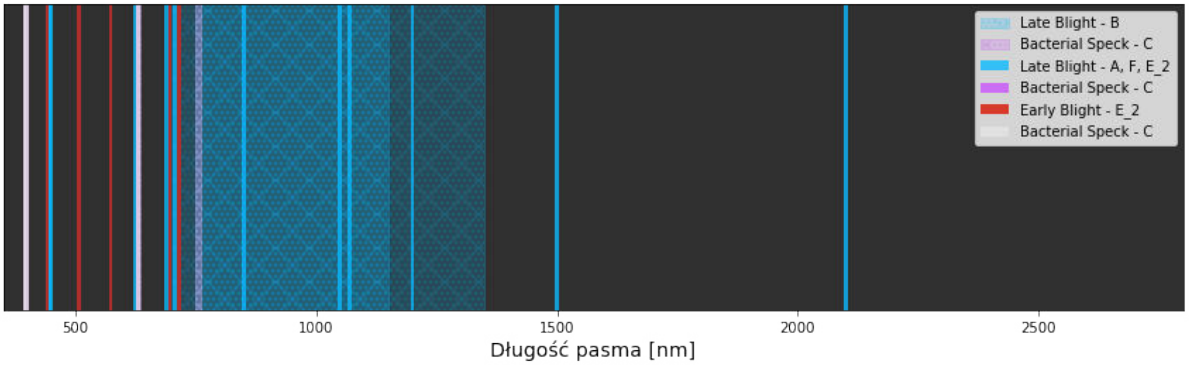
Badania opisane w podpunkcie P5 skupiają się na wykorzystaniu spektrometru pomiarowego z zakresu 380-1023nm w badaniach klasyfikacji binarnej (klasyfikacji szarej pleśni i roślin zdrowych) oraz multiklasyfikacji (Alternarioza, Zaraza ziemniaka i zdrowe rośliny).
Pierwszy z omawianych problemów w początkowej fazie badań wybrano 15 istotnych pasm (400, 402, 403, 411, 413,418, 419, 420, 422, 473, 642, 690, 722, 756 i 850 nm), a następnie spośród nich wyłoniono najważniejsze pasma (**746, 759, 760 oraz 761nm**). Uzyskano dokładność na poziomie 85,71% do 100%.
W drugim z omawianych problemów skupiono się na zakresach spektralnych **442, 508, 573, 696 i 715 nm**. W badaniu wykorzystano 310 pomiarów – 120 dla roślin kontrolnych (zdrowych), 120 roślin zakażonych Alternariozą oraz 70 zakażonych Zarazą ziemniaka. Uzyskano dokładność klasyfikacyjną na poziomie 97,1% do 100%.
Ostatni z wymienionych autorów (P6) przebadali 57 spektralnych indeksów wegetacyjnych (SVI). Na podstawie przeprowadzonych badań wyróżnili 4 najważniejsze SVI, w których wskazują na 6 konkretnych zakresów spektralnych, nad którymi warto się pochylić, tj. **445, 450, 690, 707, 1070 oraz 1200nm**. W badaniu wykorzystanych zostało 445 pomiarów. Badacze starali się zaklasyfikować pomiary do 4 różnych klas – kontrola (74 pomiary), roślina chora bez widocznych objawów (asymptotycznych, 77 pomiarów), roślina chora, z widocznymi objawami we wczesnym stadium rozwoju choroby (148 pomiarów) oraz roślina chora z widocznymi objawami w późnej fazie rozwoju patogenu (146 pomiarów). Uzyskano dokładność na poziomie 100% dla wykrywania roślin zdrowych, asymptotycznych oraz Tych w późnej fazie rozwoju.

tab. 4 Syntetyczne zestawienie charakterystyki przeprowadzonych pomiarów i badań

	Publikacja	**Źródło danych**	**Choroba**	**Istotne długości fal**	**Liczba pomiarów**	**Liczba roślin**	**Metryka**	**Wyniki**
P1 (A)	Spectral analysis of tomato late blight infections for remote sensing of tomato disease stress in California	HS (400-2500 nm)	Zaraza ziemniaka (Late Blight)	5 pasm – 625nm, 850nm, 1050nm, 1500nm i 2100nm	1074	60	-	-
P2 (B)	Spectral prediction of Phytophthora infestans infection on tomatoes using artificial neural network (ANN)	HS (400-2500 nm)	Zaraza ziemniaka (Late Blight)	pomiary polowe: zakres 750-1350nm AVIRIS: zakres 700-1105nm	1074	60	R	0.99
P3 (C)	Diagnosis of bacterial spot of tomato using spectral signatures	HS (200-2500 nm)	Bakteryjnej cętkowatości (Bacterial Speck)	5 pasm/zakresów – 395 nm, 400nm, 630nm oraz zakresy 633-635nm i 750-760nm	156 (104 trening, 52 test)	-	R^2	0.82
P4 (D)	Plant disease detection using hyperspectral imaging	HSI (400-1000nm & 900-2500nm)	Brązowa plamistość liści pomidora (Tomato Spotted Wilt Virus)	N/A	236 (133 zakażonych + 103 kontrola)	60 (30 z–każonych + 30 kontrola)	F1	0–8 - 0.939
P5 (E)	Hyperspectral imaging for classification of healthy and gray mold	HSI (380-1023nm)	Szara pleśń (Gray Mold)	4 pasm – 746, 759, 760 i 761 nm	212 (140 trening+ 72 test)	30 (25 zakażonych +	dokładność [%]	97.1-100

	diseased tomato leaves with different infection severities					5 kontrolnych)		
	Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging		Alternariozy (Early Blight), Zarazy ziemniaka (Late Blight)	5 pasm – 442, 508, 573, 696 i 715 nm	310 (120 control, 120 early bligh, 70 late blight)	-	dokładność [%]	85.71-100
P6 (F)	Detection of multi-tomato leaf diseases (late blight, target and bacterial spots) in different stages by using a spectral-based sensor	HS (400-1050nm)	Zaraza ziemniaka (Late Blight), (Target spot), Bakteryjna plamistość pomidora (Bacterial spots)	4 najistotniejsze SVI – 445, 450, 690, 707, 750, 800, 1070, 1200nm, ale 750 i 800 nie zostały wybrane przez PCA	445 (74 kontrola, 77 asympthotic late blight, 148 symptomatic early stage (?), 146 symptomatic late stage (?))	brak informacji o liczbe pomiarów, pomiary zbierane na farmie komercyjnej	dokładność [%]	100
ER	**Eksperyment opisany w rozdziale 9**	HS (350-2500 nm)	Kontrola + 5 chorób: AN, BS, EB, LB, SL	Opisane w rozdziale 9.4	11634 (dla 3 różnych wysokości) 3877 (dla wysokości 5cm)	brak informacji o liczbie pomiarów, warunki laboratoryjne (3 niezależne fitotrony)	F1	Opisane w rozdziale 9.2

Z kolei w niniejszym badaniu (ER) wykorzystany został zbiór złożony z 3877 pomiarów wykonanych za pomocą spektroradiometru ASD FieldSpec 4 Hi-Res z zakresu 350-2500nm. Zbiór danych wykorzystany w niniejszym opracowaniu zawiera również pomiary obiektów, dla których symptomy chorobowe nie są jeszcze widoczne dla człowieka, co stanowi nowość w porównaniu do dotychczas prowadzonych badań. Istotą tego problemu badawczego jest podjęcie próby klasyfikacji chorej rośliny niezależnie od stadium rozwoju choroby, tj. również na etapie, kiedy objawy nie są widoczne.



rys. 10 Pasma wyróżnione przez autorów publikacjii

Syntetyczne zestawienie charakterystyki przeprowadzonych pomiarów i badań oraz ich rezultaty przedstawiono w 9.4.

4.7. Przegląd stanu wiedzy w zakresie wczesnego wykrywania chorób

Wczesne wykrywanie ognisk chorobowych w uprawach jest istotnym zadaniem ze względu na możliwość zastosowania odpowiednich środków zaradczych w krótkim okresie od wystąpienia infekcji, ale jednocześnie nie jest to zadanie łatwe. Jak już wspomniano wcześniej, niektóre choroby przy sprzyjających warunkach pogodowych i braku zastosowania środków ochrony roślin mogą doprowadzić do obniżenia plonu nawet o 80%. Analiza stanu roślin w ujęciu czasowym (np. dzień za dniem) jest szczególnie istotna ze względu na potrzebę wczesnego wykrywania chorób w uprawach (eng. *early stage disease detection*), ponieważ pomaga w rozwiązaniu tego problemu.
W literaturze przedmiotu popularny jest temat wczesnego wykrywania chorób roślin ze względu na jego przydatność w przemyśle rolniczym, jednak w większości publikowanych badań pomiary wykonywano sporadycznie, np. 2-8 sesji pomiarowych w całym okresie wegetacyjnym [57, 68]. Niektórzy badacze podchodzą do analizy wskazując jedynie na stadium rozwoju choroby (wczesne / późne stadium rozwoju patogenu), bez uwzględnienia czasu, jaki minął od zakażenia oraz pomijając analizę rozwoju choroby w ujęciu czasowym [27]. Część badaczy wskazuje na wykonywanie pomiarów w określonym przedziale czasowym od inokulacji [46] (6-42 dni od inokulacji), jednak nie przestawiają rezultatów analiz na etapie opisywania prowadzonych prac.

Część badań została przeprowadzona z wysoką częstotliwością pomiarową podobnie, jak w niniejszej pracy. Przykładowo w [82] pomiary wykonywano co 12-24h przez 5-7 dni od zakażenia. W przypadku [83] wykonywano cotygodniowe pomiary pomiędzy 11 a 60 dniem od inokulacji, a wyniki rozpatrywano niezależnie dla zakresu VIS-NIR oraz SWIR. Dla

zakresu VIS-NIR najlepszy uzyskany wynik F1 binarnego klasyfikatora (SVM) odróżniającego rośliny chore od zdrowych wyniósł 0.79 w 32 i 39 dniu od zakażenia, natomiast dla SWIR – 0.83 w 60 dniu. Również w przypadku [84] wykonywano cotygodniowe pomiary hiperspektralne zakresu VIS-NIR pomiędzy 6 a 42 dniem od zakażenia, jednak analizę w ujęciu czasowym ograniczono jedynie do wskazania czasu pojawienia się widocznych symptomów chorobowych. Ciekawe podejście zostało zastosowane w [85], gdzie zastosowane została klasyfikacja binarna na danych multispektralnych dla dwóch odmian manioka jadalnego w 4 punktach czasowych po inokulacji, jednak wyniki były zróżnicowane w zależności od odmiany i punktu pomiarowego i wahały się od 0.50-0.92 oraz 0.50-0.85 dokładności (ang. *accuracy*, w publikacji określana jako *classification rate*) w zależności od odmiany (odpowiednio dla odmian TME204 oraz Kiroba) dla drugiego punktu pomiarowego (28 dzień od zakażenia).
Do opisywanego zagadnienia używano również narzędzi uczenia głębokiego [86]. W tym celu wykonano 40 zdjęć hiperspektralnych w 4 punktach czasowych pomiędzy 7.08.2019 a 8.10.2019. Zbiory takie zostały przetworzone z wykorzystaniem dwuwymiarowej (2D-CNN) i trójwymiarowej (3D-CNN) sieci neuronowej do wyodrębnienia cech, a następnie zostały wytrenowane klasyfikatory SVM oraz RandomForest. Random Forest dał lepsze wyniki, jednak należy zaznaczyć, że ze względu na niewielką ilość danych pomiarowych były one nieakceptowalnie niskie. Dokładność dla zbioru treningowego wynosiła odpowiednio 0.71 i 0.75 dla 2D-CNN oraz 3D-CNN, natomiast dokładność dla zbioru testowego wyniosła zaledwie 0.50. Również w przypadku [87] zastosowano pomiary spektralne 4 punktach czasowych (7.05, 20.05, 6.06 oraz 19.06 w roku 2021). W przypadku tego badania do ekstrakcji cech wykorzystano 4 metody: Principal Component Analysis (PCA), Sequential Forward Selection (SFS), Simulated Annealing (SA) oraz Genetic Algoritms (GA), zaś jako klasyfikatora użyto wektora maszyny nośnej (SVM). Połączenie GA-SVM dało najlepsze wyniki klasyfikacji na poziomie 0.91 i 0.93 odpowiednio dla liści i łodyg. Eksperyment powtórzono w 2022 roku i dla tego samego podejścia uzyskano 0.89 dokładności (ang. *accuracy*) i 0.80 wskaźnika F1 dla łodyg.
Kombinacja algorytmów genetycznych i maszyny wektorów nośnych została również wskazana w [39] jako model nadający się do rozwiązywania tego typu zagadnień. W badaniu wykonano pomiary w ciągu 5 dni pomiarowych w okresie do 15 dni od zakażenia (3, 6, 9, 12 i 15 dzień od inokulacji) i uzyskano 0.97 wskaźnika F1 dla zbioru testowego. Badanie dotyczyło zainfekowanych łodyg z widocznymi symptomami chorobowymi.
W przypadku publikacji [29] pomiary wykonywano w ciągu pierwszych 5 dni od zakażenia, wykorzystano klasyfikatory binarne takie jak KNN (K-Nearest Neighbours), C5.0 (an algorithm used to generate Decision Tree) oraz FR-KNN (Feature Ranking K-Nearest Neighbours). Uzyskano wyniki na poziomie 0.94-0.97 dokładności klasyfikacji (ang. *accuracy*), jednak wyniki dla poszczególnych dni pomiarowych wahały się od 0.10-0.58 dokładności dla FR-KNN.
Badania nad zmianami chorobowymi w czasie z wykorzystaniem danych hiperspektralnych prowadzono również w skali mikroskopijnej [88]. W przypadku tego badania wykonywano codzienne pomiary, w okresie od 3 do 14 dni od inokulacji, z wykorzystaniem urządzenia pomiarowego złożonego z kamery hiperspektralnej i mikroskopu. W badaniu wykorzystano Simplex Volume Maximization (SiVM). W tym przypadku skoncentrowano się jednak na analizie zmiany widma zainfekowanych roślin w zależności od upływu czasu od inokulacji.

tab. 5 Syntetyczne zestawienie publikacji dotyczących analizy pomiarów hiperspektralnych w celu wczesnego wykrywania chorób

ref	Częstotliwość pomiarów	Metoda, metryka i osiągnięte wyniki
[57]	2-8 razy w trakcie wegetacji	SAM uzyskano dokładność klasyfikacji równą 87%
[68]	2-8 razy w trakcie wegetacji	PLS, SMLR – uzyskano współczynnik determinacji R^2=0.82
[82]	Co 12-24h przez 5-7 dni od zakażenia	PLS-DA dla przypadków pre-symptomatycznych dokładność klasyfikacyjna wyniosła 91.3% (jedynie 64 próbki)
[83]	Co tydzień pomiędzy 11 a 60 dniem od zakażenia	SVM w 32 i 39 dniu od zakażenia osiągnął F1=0.79 dla VIS-NIR, a dla SWIR w 60 dniu uzyskano F1=0.83
[85]	4 punkty pomiarowe	Dla drugiego punktu pomiarowego (14 DPI) uzyskano 88% dokładności klasyfikacyjnej
[86]	4 punkty pomiarowe	Dokładność klasyfikacyjna na poziomie 71% oraz 75 % odpowiednio dla 2D-CNN oraz 3D-CNN
[87]	4 punkty pomiarowe	Dokładność na poziomie 0.91 i 0.93 odpowiednio dla liści i łodyg
[39]	5 dni pomiarowych w okresie do 15 dnia od zakażenia	F1=0.97, ale tylko łodygi z widocznymi symptomami chorobowymi
[29]	Pomiary w 5 dni od zakażenia	Dokładność na poziomie 94% - 0.97%
[88]	Codzienne pomiary w okresie od 3 do 14 dnia od zakażenia	-

Zgodnie z powyższą analizą literatury można zauważyć, że temat diagnostyki upraw z wykorzystaniem danych hiperspektralnych oraz metod uczenia maszynowego w ujęciu czasowym jest istotny w kontekście wczesnego wykrywania chorób roślin. Można jednak zauważyć, że żadna z przytoczonych publikacji nie podejmuje wątku w takim zakresie, jak niniejsza praca doktorska. Przede wszystkim analizowane są inne uprawy, inne choroby oraz inne metody są proponowane na potrzeby wykrywania i rozpoznawania chorób. Osiągane są wyniki na poziomie dokładności klasyfikacyjnej od 71% do 97% oraz wskaźnik F1 – od 0.79 do 0.97, ale dotyczą ograniczonych zbiorów danych, co z jednej strony wskazuje na zasadność podejścia, ale z drugiej strony wskazuje na konieczność przeprowadzenia bardziej szczegółowych badań.

Struktura danych wykorzystana w tej metodzie została opisana w rozdziale 8.3.

4.8. Podsumowanie rozdziału

W rozdziale przedstawione zostały różne typy danych spektralnych wykorzystywanych w nieinwazyjnych metodach diagnostycznych. Przybliżono aktualny stan wiedzy w zakresie

metod łączących zalety danych hiperspektralnych oraz uczenia maszynowego w celu prawidłowej diagnostyki chorób roślin. Zaprezentowano najważniejsze zakresy spektralne, wykorzystywane w dotychczasowych badaniach. Ze względu na charakter zbioru danych użytych w niniejszej pracy, zaprezentowano również przegląd aktualnych metod w ujęciu czasowym (detekcja chorób w określonym przedziale czasu od inokulacji). Przeprowadzona analiza aktualnego stanu literatury w omawianym zakresie wskazuje na zasadność podjętej tezy, jak również możliwość osiągnięcia założonych celów. Dodatkowo temat jest społecznie i gospodarczo ważny, a także cieszy się istotnym zainteresowaniem w świecie nauki. Niektórzy badacze wskazują na istotność pewnych pasm spektralnych w procesie diagnostyki roślin. Widoczne są jednak niedobór wyników dla pomiarów wykonywanych z dzienną częstotliwością czy brak badań obejmujących metodę przetwarzania danych hiperspektralnych zaproponowaną przez autora w niniejszej pracy.

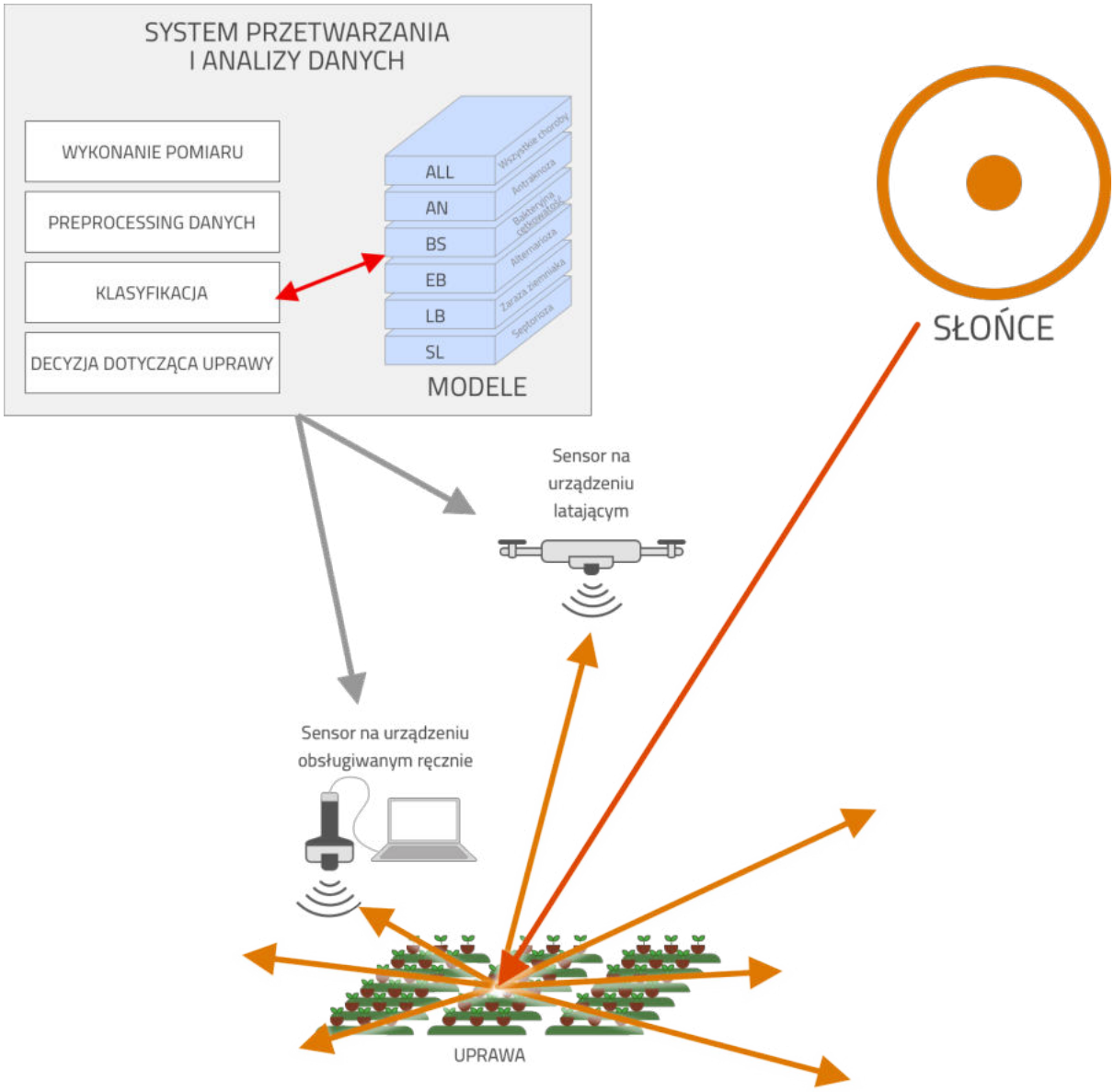
5. Koncepcja metody diagnostyki wybranych chorób w uprawie pomidorów zwyczajnych

Przedstawione w niniejszej pracy badania mają na celu opracowanie innowacyjnej koncepcji diagnostyki chorób roślin, wykorzystującej zaawansowane techniki pomiarów hiperspektralnych i uczenia maszynowego. Metoda opiera się na analizie spektralnych sygnatur roślin, które są rejestrowane przy użyciu urządzeń hiperspektralnych. Następnie, za pomocą zaawansowanych algorytmów uczenia maszynowego, tworzone są modele predykcyjne, które potrafią identyfikować i diagnozować konkretne choroby na podstawie analizy spektralnej.

rys. 11 Główne etapy realizacji proponowanej metody diagnostyki wybranych chorób w uprawie pomidorów zwyczajnych

Przyjęta koncepcja opiera się na trzech głównych fazach przedstawionych schematycznie na rysunku rys. 11. Pierwszą fazą jest zbieranie kompleksowych danych hiperspektralnych z różnych roślin dzień po dniu, zarówno zdrowych, jak i zainfekowanych chorobami. Następnie, na podstawie tych danych, tworzymy modele uczenia maszynowego, algorytmy klasyfikacji, które są w stanie rozpoznać charakterystyczne wzorce spektralne związane z konkretnymi chorobami roślin.

W fazie trzeciej, wykorzystując zbudowane modele uczenia maszynowego, można opracować system diagnostyczny, który będzie analizował spektra nowych próbek roślin i dokonywał predykcji ich stanu zdrowotnego dzień po dniu. System ten może być wykorzystywany zarówno w laboratoriach, jak i w terenie, umożliwiając szybką i nieinwazyjną diagnostykę chorób roślin. Przyjęta koncepcja diagnostyki chorób roślin oparta na pomiarach hiperspektralnych i uczeniu maszynowym ma potencjał do rewolucjonizowania dziedziny, umożliwiając szybką, precyzyjną i nieinwazyjną identyfikację chorób roślin na szeroką skalę.



rys. 12 Koncepcja przyrządu pomiarowego, który umożliwi precyzyjną diagnostykę chorób roślin przy wykorzystaniu technologii pomiarów hiperspektralnych i uczenia maszynowego

Proponowana koncepcja może pozwolić na opracowanie innowacyjnego przyrządu pomiarowego (rys. 12), który umożliwi precyzyjną diagnostykę chorób roślin przy wykorzystaniu technologii pomiarów hiperspektralnych i uczenia maszynowego. Główne cechy tego przyrządu to możliwość zbierania wysokorozdzielczych danych spektralnych z naziemnych części roślin, co pozwoli na analizę i identyfikację charakterystycznych wzorców związanych z konkretnymi chorobami. Przyrząd pomiarowy powinien być wyposażony w zaawansowany spektrometr hiperspektralny, który umożliwia rejestrowanie szerokiego zakresu spektralnego reflektancji dzień po dniu. Pozyskane dane spektralne będą następnie automatycznie przetwarzane przez opracowane wcześniej algorytmy uczenia maszynowego, które są w stanie analizować i klasyfikować spektralne sygnatury roślin i identyfikować obecność konkretnych chorób. System przetwarzania i podejmowania decyzji powinien być zintegrowany z bezzałogową platformą latającą albo urządzeniem przenośnym tak, aby od razu informował rolnika o stanie uprawy. Dzięki temu, badacze, rolnicy i specjaliści ds. ochrony roślin będą mieli możliwość szybkiego i precyzyjnego monitorowania zdrowia roślin oraz wczesnego wykrywania chorób. Zastosowanie bezzałogowych platform latających wymaga

rozwiązania szeregu problemów, m.in. związanych z wymiarami zaawansowanych instrumentów hiperspektralnych, pułapem przelotu, gdyż odległość sensora od analizowanej uprawy istotnie wpływa na dokładność wykonywanych pomiarów (problem częściowo opisano w rozdziale 6.4), zmiennymi warunkami atmosferycznymi, ograniczonymi zasobami obliczeniowymi charakteryzujących systemy wbudowane (ang. *embedded system*s) czy chociażby z problemami legislacyjnymi umożliwiającymi wykonywanie lotów automatycznych.

Przyjęta koncepcja przyrządu pomiarowego do diagnostyki chorób roślin z wykorzystaniem pomiarów hiperspektralnych i uczenia maszynowego ma potencjał do zrewolucjonizowania sposobu diagnozowania chorób roślin, umożliwiając szybką, nieinwazyjną i efektywną analizę stanu zdrowia roślin na dużą skalę.

Zaproponowana metoda została opracowana na podstawie następujących kroków badawczych opisanych w dalszej części niniejszej rozprawy:

1. Eksperyment laboratoryjny (szerzej opisany w rozdziale 6),
2. Wykonanie pomiarów hiperspektralnych (szerzej opisane w rozdziale 6.4),
3. Wstępne przetworzenie (preprocessing) danych pomiarowych i stworzenie zbioru danych (szerzej opisane w rozdziale 8.2),
4. Klasyfikacja jednostek chorobowych z wykorzystaniem metod uczenia maszynowego. Wielowariantowe trenowanie modeli uczenia maszynowego i ewaluacja na zbiorze testowym (szerzej opisana w rozdziale 9),
5. Klasyfikacja jednostek chorobowych analogicznie do punktu 4 z uwzględnieniem zmian zachodzących w czasie (szerzej opisana w rozdziale 10).

6. Eksperyment pomiarowy – obserwacja uprawy pomidora zwyczajnego w warunkach rzeczywistych i laboratoryjnych

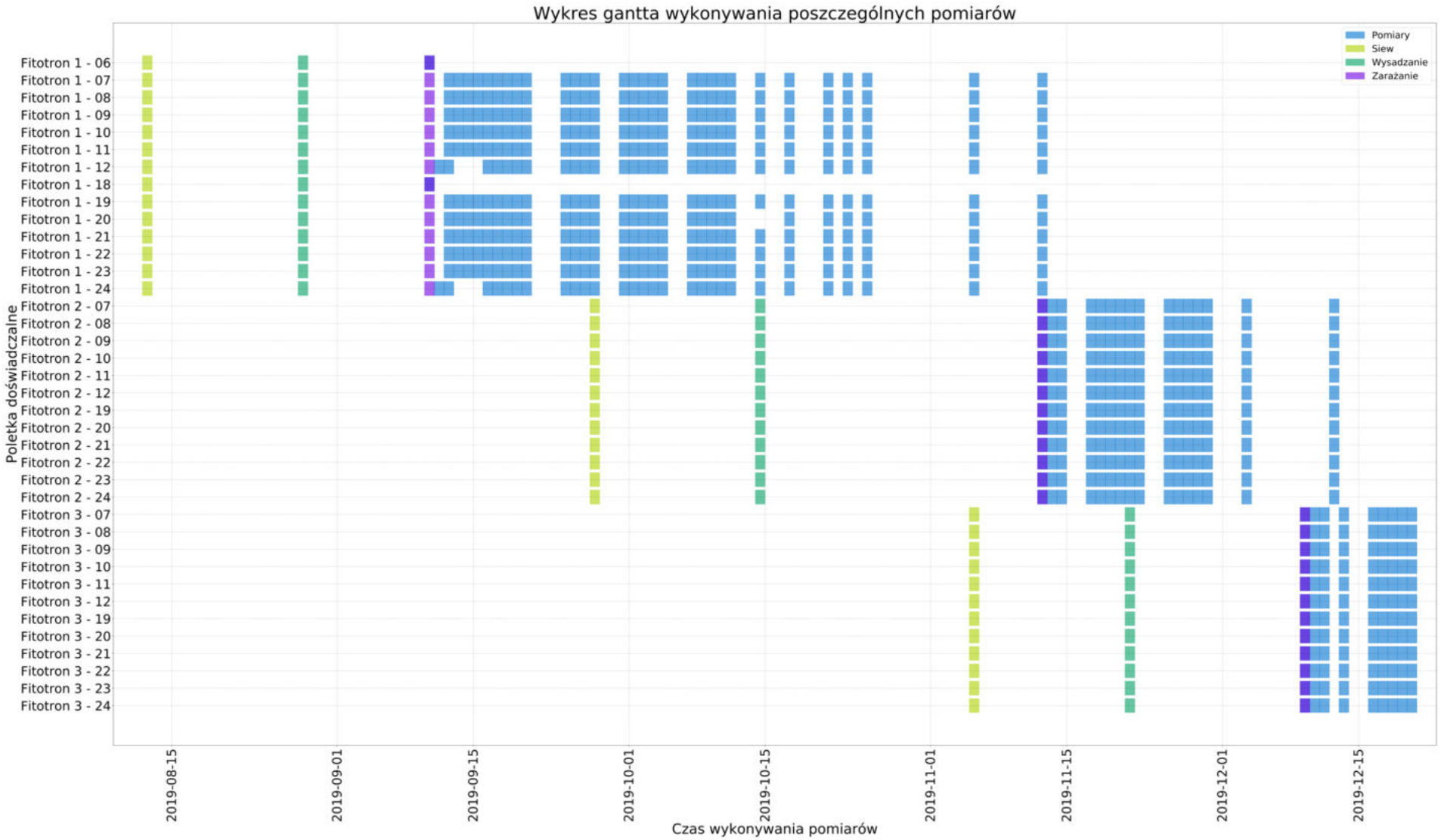
W niniejszej pracy doktorskiej Autor wykorzystał dane pochodzące z badania przeprowadzonego przez QZ Solutions Sp. z o.o. w ramach projektu POIR.01.01.01-00-1317/17. Dane zostały udostępnione bezpłatnie oraz za zgodą Prezesa Zarządu QZ Solutions [DODATEK A – Załącznik 1 - Upoważnienie do wykorzystania danych firmowych].

6.1. Harmonogram badań

Pomiary hiperspektralne stanowiące podstawę analizowanego zbioru danych wykonane były w okresie 08-12.2019. Pomiary były wykonywane w warunkach laboratoryjnych. Rośliny wyhodowane były w komorach wegetacyjnych (fitotronach) umożliwiających kontrolowanie temperatury i wilgotności powietrza.

Badanie przeprowadzono w 3 cyklach pomiarowych. Inokulacja (zarażenie) w cyklu pierwszym zostało wykonane 10.09.2019, w cyklu drugim – 12.11.2019, natomiast w cyklu trzecim – 09.12.2019. Dzień inokulacji w dalszej części badania jest określany jako DPI 0 (days-post-inoculation). Na rys. 13 w szczegółowy sposób przedstawione zostały dni siewu (kolor jasnozielony), wysadzania (zielony) i zarażania roślin (fioletowy), a także okres wykonywania pomiarów (niebieski) w 3 poszczególnych fitotronach.

Rośliny przetrzymywano w komorach wegetacyjnych, w których utrzymywano stałą temperaturę 18-20°C (z dokładnością do 0,5°C), wilgotnością 60-80% (z dokładnością do 5%) oraz doświetleniem (minimum 200 $\mu mol \cdot m^{-2} \cdot s^{-1}$) w 16-godzinnym fotoperiodzie.
Pomiary hiperspektralne oraz oznaczenie patogenów wykonano analogicznie, jak w przypadku środowiska polowego.
Pierwsze, widoczne dla człowieka objawy zarażenia, widoczne były po 3-5 dniach od zakażenia.
Pomiary wykonywano w ciągu 33 dni pomiarowych w okresie od DPI = 0 do DPI = 63, gdzie DPI (days post inoculum) oznacza liczbę dni, które upłynęły od momentu inokulacji (zakażenia roślin) zgodnie z harmonogramem przedstawionym na rys. 13.



rys. 13 Wykres Gantta przedstawiający na osi czasu dni siewu, wysadzania i zarażania roślin oraz wykonywania pomiarów; każdy pojedynczy kwadrat reprezentuje 1 dzień pomiarowy

6.2. Przygotowanie inokulacji i materiału roślinnego do badań

Zbiór danych wykorzystywanych w badaniu obejmował pomiary hiperspektralne roślin dwóch odmian uprawy *Solanum lycopersicum*: *Benito* oraz *Polfast* (*Bejo Zaden B.V.*). Dla każdej z odmian przygotowano 36 roślin doświadczalnych. Rośliny zarażono pięcioma różnymi patogenami w grupach po 6 sztuk:

1. 6 roślin zakażonych patogenem *Colletotrichum coccodes* (CBS 103.16) powodującym antraknozę (AN),
2. 6 roślin zakażonych patogenem *Alternaria solani* (CBS 142772) powodującym alternariozę (EB),
3. 6 roślin zakażonych patogenem *Pseudomonas syringe pv. Tomato* powodującym bakteryjną cętkowatość (BS),
4. 6 roślin zakażonych patogenem *Phytophthora infestans* (CBS 429.90) powodującym zarazę ziemniaka w uprawie pomidora (LB),
5. 6 roślin zakażonych patogenem *Septoria lycopersici* (CBS 354.49) powodującym septoriozę (SL).

Pozostałe 6 roślin było obiektami kontrolnymi, niezakażonymi żadnym patogenem (Control). Numer CBS odnosi się do bazy chorób pochodzenia grzybiczego, *Centraalbureau voor Schimmelcultures*. W przypadku *Pseudomonas syringe pv*. Tomato nie podano numeru CBS, ponieważ jest to patogen pochodzenia bakteryjnego. Patogen ten został wyizolowany z owoców pomidora.

Każdy z patogenów, zarówno pochodzenia bakteryjnego, jak i grzybiczego, został wyizolowany i wyhodowany na przynajmniej 6 niezależnych płytkach. Każda z płytek miała średnicę przynajmniej 9 cm. Do wyhodowania patogenów wykorzystane zostały standardowe podłoża oraz sterylne warunki. Warunki hodowli i przechowywania patogenów były stałe, tj. temperatura powietrza wynosiła 20±1°C, a wilgotność powietrza wynosiła co najmniej 70%. Rośliny zakażone zostały poprzez zanurzenie ich w inokulum rozcieńczonym do stężenia 10^5 do 10^6 zarodników na 1 ml.
Obecność patogenów była potwierdzana na liściach i/lub owocach za pomocą analizy mikroskopowej, izolacji patogenu i określeniu gatunków. Fragmenty roślin z objawami choroby wycinano z liści lub owoców. Następnie dezynfekowano powierzchniowo przez 3 min w 2% podchlorynie sodu. Fragmenty suszono, krojono na 4 mniejsze fragmenty i wykładano na pożywkę PDA. Po 7-10 dniach oceniano wyrosłe kultury i przypisywano do gatunków sprawców. Przed wyizolowaniem części roślin do badań wykonywany został pomiar za pomocą instrumentu opisanego w rozdziale 6.3.

6.3. Instrument pomiarowy

Do wykonywania pomiarów hiperspektralnych wykorzystany został spektroradiometr ASD FieldSpec 4 Hi-Res. Urządzenie wykonuje pomiar reflektancji, transmisji lub energii promieniowania elektromagnetycznego, tj. radiancji (luminacja energetyczna) lub irradiancja (natężenie promieniowania), badanej próbki w zakresie spektralnym 350-2500 nm z rozdzielczością spektralną 3nm oraz 8nm odpowiednio dla długości 700nm oraz 1400nm i 2100nm. Szerokość próbkowania widmowego wynosiła 1.4nm dla zakresu UV-VNIR (350-1000nm) oraz 1.1nm dla zakresu NIR-SWIR (1001-2500nm). Radiacja mierzona jest w watach na metr kwadratowy na kąt bryłowy (steradian) ($W/m^2/sr$), natomiast urządzenie wykonuje

pomiar radiacji spektralnej, zatem jednostką pomiarową jest wat na metr kwadrat na steradian dla poszczególnych długości fali ($W/m^2/sr/nm$).
Szczegółowe parametry urządzenia wymienione zostały w tabeli tab. 6 Parametry urządzenia pomiarowego wykorzystanego w badaniu.
Dane pozyskane na potrzeby niniejszej pracy są pomiarami reflektancji, reprezentującymi ilość światła odbitego w skali od 0 do 1 w poszczególnych kanałach spektralnych badanego zakresu w porównaniu do pomiaru kalibracyjnego, gdzie pomiar kalibracyjny jest równy 1 w całej, badanej przestrzeni spektralnej.
Podczas gdy promieniowanie pada na dowolny obiekt, cała energia jest odbijana, transmitowana lub absorbowana przez tenże obiekt. Zachowanie energii zapisać można w następujący sposób [12]:

$$E_I(l) = E_R(l) + E_A(l) + E_T(l) \tag{7.1}$$

gdzie
l - długość fali,
$E_I(l)$ – energia padająca na powierzchnię,
$E_R(l)$ – energia odbita od powierzchni,
$E_A(l)$ – energia zaabsorbowana przez obiekt,
$E_T(l)$ – energia transmitowana przez obiekt.

Powyższy wzór można zapisać również w następujący sposób:

$$\frac{E_R(l)}{E_I(l)} + \frac{E_A(l)}{E_I(l)} + \frac{E_T(l)}{E_I(l)} = 1 \tag{7.2}$$

A to z kolei można zapisać w następujący sposób:

$$\rho + \alpha + \tau = 1 \tag{7.3}$$

gdzie:
ρ – reflektancja,
α – absorbancja,
τ – transmitancja.

Dodatkowo można zaznaczyć, że absorbancja oraz transmitancja powiązane są ze sobą w następujący sposób:

$$\alpha = \log\frac{1}{\tau} \tag{7.4}$$

tab. 6 Parametry urządzenia pomiarowego wykorzystanego w badaniu

Parametr	**Wartość**
Zakres spektralny	350 nm – 2500 nm
Rozdzielczość spektralna VNIR @ 700nm	3 nm
Rozdzielczość spektralna SWIR @ 1400 oraz 2100nm	8 nm
Próbkowanie spektralne (szerokość pasma) VNIR	2.2 nm

Próbkowanie spektralne (szerokość pasma) SWIR	2.2 nm
Czas skanowania	200 ms
NEdL (Noise Equivalent Radiance) – VNIR @ 700nm	1.0 x 10^{-9} W/cm^2/nm/sr
NEdL (Noise Equivalent Radiance) – SWIR @ 1400nm	1.4 x 10^{-9} W/cm^2/nm/sr
NEdL (Noise Equivalent Radiance) – SWIR @ 2100nm	2.2 x 10^{-9} W/cm^2/nm/sr
Odtwarzalność długości fali	0.1 nm
Niepewność pomiarowa	Średni błąd pomiarowy na poziomie 0.5 nm; długość fali +/- 1nm
Maksymalna radiacja – VNIR	2x radiacja Słońca
Maksymalna radiacja – SWIR	10x radiacja Słońca
Prędkość zapisu danych	2 pomiary na sekundę
Liczba kanałów	2151
Detektor VNIR (350-1000 nm)	512-elementowa rozszerzona macierz krzemowa bliskiej podczerwieni
Detektor SWIR 1 (1000-1800 nm)	Graded Index InGaAs Photodiode, 2 Stage TE Cooled Graded
Detektor SWIR 2 (1800-2500 nm)	Graded Index InGaAs Photodiode, 2 Stage TE Cooled Graded
Wejście	Światłowód 1.5 m, kąt widzenia 25°
Waga	5,44 kg
Kalibracja	Długość fali, absolutna reflektancja

Wielkość skanowanego obszaru próbki stanowi okrąg o wielkości zależnej od wysokości, na której umieszona jest soczewka urządzenia pomiarowego i wynosi:

$$\psi = \pi \left(h \cdot tg \left(\frac{\frac{\gamma}{2} \cdot \pi}{180°} \right) \right)^2 \tag{7.5}$$

gdzie:
- ψ to skanowany obszar,
- h to wysokość, z jakiej wykonywany jest pomiar,
- γ to pole widzenia soczewki (FOV).

Dla pomiarów z wykorzystaniem domyślnego osprzętu optycznego pole widzenia soczewki (FOV) wynosi 25°. Zatem dla domyślnego FOV skanowany obszar wynosi:

$$\psi = \pi \left(h \cdot tg \left(\frac{12{,}5° \, \pi}{180°} \right) \right)^2 \tag{7.6}$$

$$r = \sqrt{\frac{\psi}{\pi}} \tag{7.7}$$

Dla powyższego wzoru otrzymujemy następujące wartości skanowanego obszaru ψ oraz r (promień skanowanego obszaru):

tab. 7 Obszar skanowania próbki w zależności od wysokości, na której umieszczona była soczewka urządzenia skanującego przy zastosowaniu domyślnego osprzętu (brak dodatkowych soczewek korygujących FOV)

$\boldsymbol{h[cm]}$	5	10	20	30	40	50	60
$\boldsymbol{\psi[cm^2]}$	3,86	15,44	61,76	138,96	247,05	386,01	555,86
$\boldsymbol{r[cm]}$	1,11	2,22	4,43	6,65	8,87	11,08	13,30

a) b)

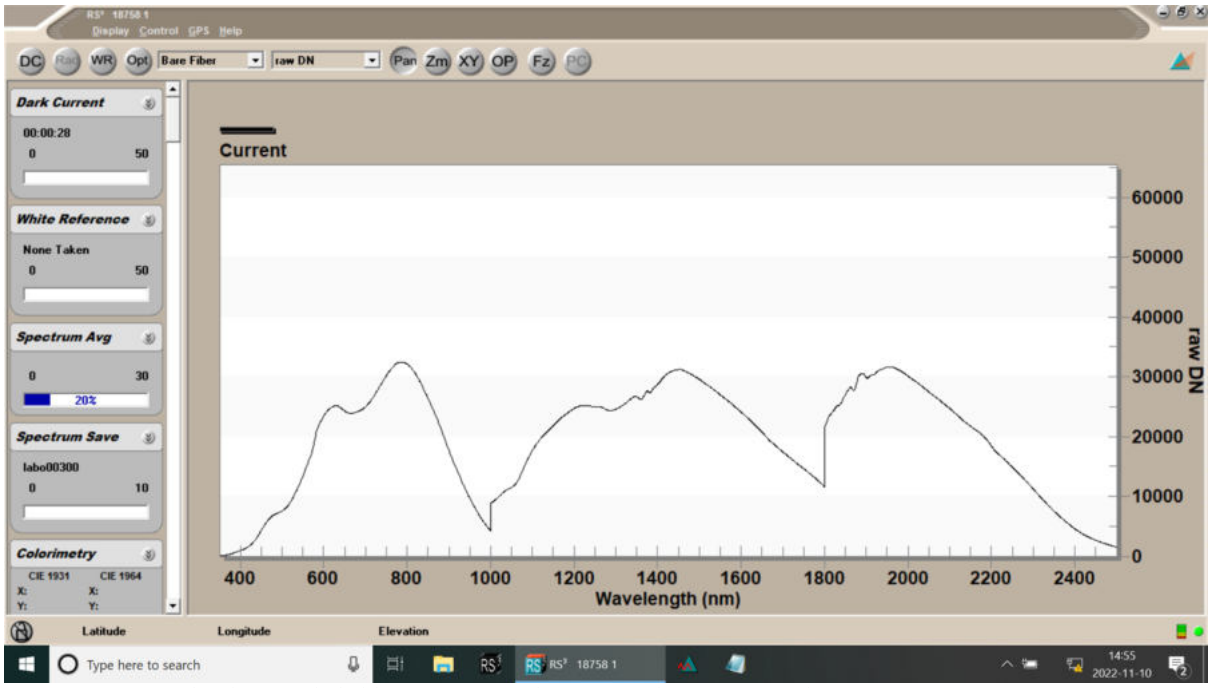
rys. 14 Spektroradiometr ASD FieldSpec Hi-Res wykonujący pomiary hiperspektralne w zakresie 350-2500nm

Prawidłowo wykonany pomiar spektralny z wykorzystaniem spektrometru ASD FieldSpec 4 Hi-Res musi zostać poprzedzony odpowiednią procedurą kalibracyjną. Kalibracja jest szczególnie istotna w sytuacji, w której warunki atmosferyczne ulegają zmianie, np. zmieni się zachmurzenie, zmieni się źródło światła, np. badanie próbki w różnych pomieszczeniach laboratoryjnych, następuje saturacja urządzenia, tzn. pomiary nie są spójne (widoczny uskok dla pomiarów panelu kalibracyjnego).

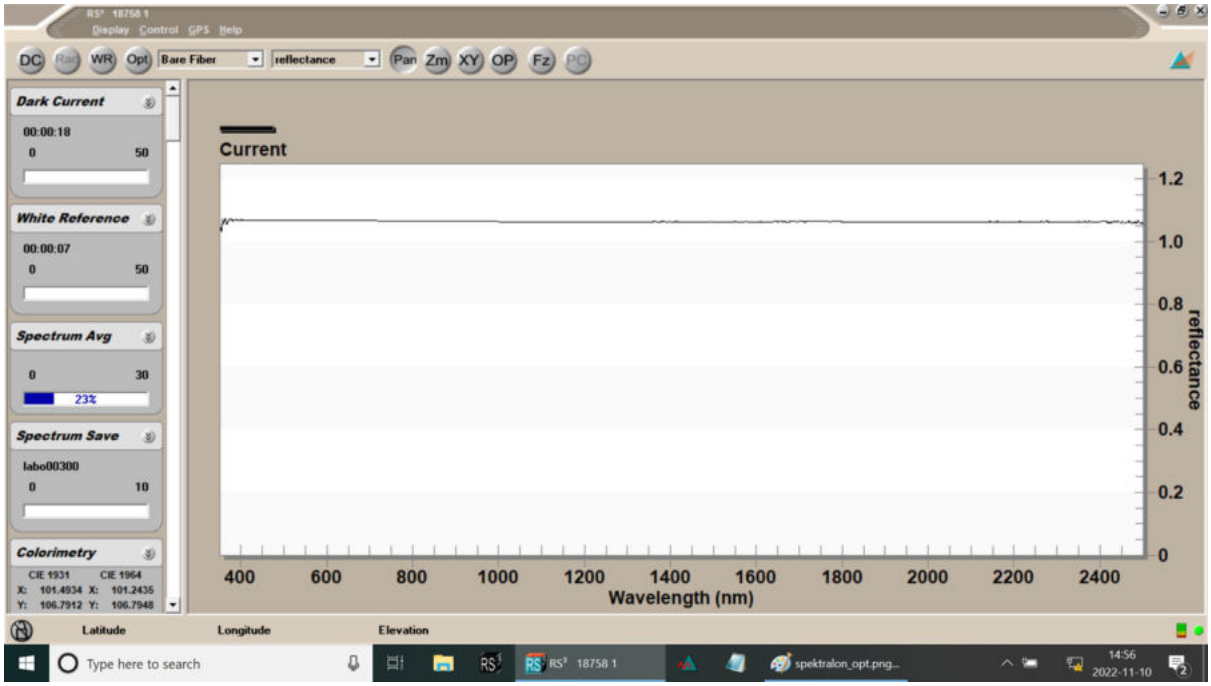
Kalibracja urządzenia polega na jego ustabilizowaniu. W praktyce oznacza to konieczność uruchomienia urządzenia odpowiednio wcześnie przed pomiarami, tj. przynajmniej około 15 do 20 minut przed przystąpieniem do pomiarów. W przypadku pomiarów wykonywanych w opisywanej pracy spektrometr uruchomiono przynajmniej 30 min przed rozpoczęciem procedury optymalizacji pomiarowej. Optymalizacja jest procesem regulacji urządzenia elektronicznego w celu właściwego przetworzenia sygnału wejściowego. Proces ten polega na

a) wykonaniu pomiaru panelu bieli referencyjnej (Spectralon) w celu wykonania optymalizacji,

b) wykonania pomiaru za pomocą prądu ciemnego (ang. *dark current*), co w praktyce oznacza odcięcie źródła światła od pomiarowego przewodu światłowodowego poprzez jego szczelne zakrycie jego końcówki,
c) ponownego zapisania pomiaru wzornika bieli jako sygnału bieli referencyjnej.



rys. 15 Krzywa spektralna panelu kalibracyjnego przed przeprowadzeniem optymalizacji



rys. 16 Krzywa spektralna reflektancji panelu kalibracyjnego po przeprowadzeniu optymalizacji i kalibracji wzornika bieli

Po tak przeprowadzonej kalibracji można przystąpić do wykonywania pomiarów. Należy jednak od czasu do czasu upewniać się, czy urządzenia nie ulega saturacji, co charakteryzuje

się m.in. nagły wzrost lub spadek wartości w zakresie około 1000nm i/lub 1400nm zarejestrowanej krzywej spektralnej, gdyż wtedy konieczne jest powtórzenie procesu optymalizacyjnego i ponowne zapisanie sygnału bieli referencyjnej. Stanowisko pomiarowe podczas wykonywania kalibracji zostało zaprezentowane na rys. 17.

rys. 17 Stanowisko pomiarowe w trakcie wykonywania procesu kalibracji instrumentu pomiarowego ASD FieldSpec 4 HiRes wraz z oświetleniem halogenowym

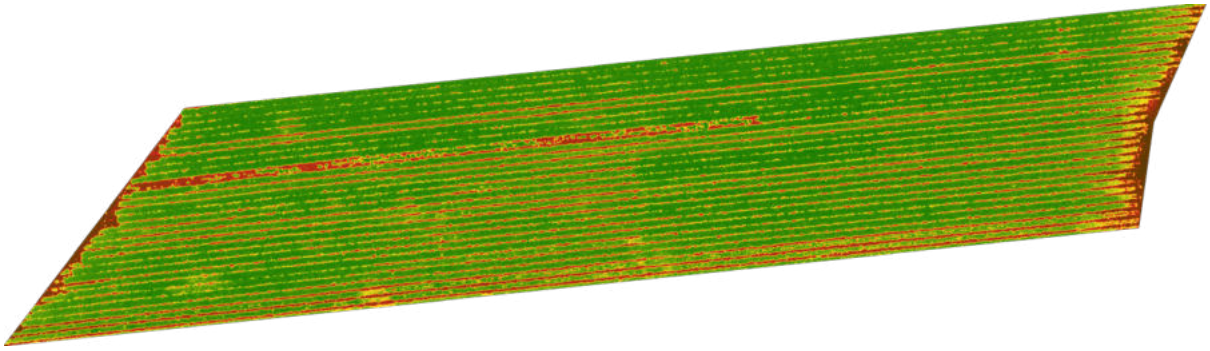
Oświetlenie widoczne na rys. 17, tj. lampa halogenowa Eurostar Reflektor MR16 firmy Ushio [89], jest oświetleniem zarekomendowanym przez dystrybutora urządzenia do wykonywania pomiarów laboratoryjnych. W celu zapewnienia równomiernego oświetlenia badanych obiektów stosowany był zestaw oświetleniowy złożony z dwóch lamp halogenowych. Natężenie strumienia pojedynczego źródła światła wynosiła 11 tys. kandeli [cd], temperatura barwowa światła wynosiła 3000 kelwinów [K].
Biel referencyjna wykorzystana w badaniu to Spectralon, czyli flouropolimeru o najwyższym współczynniku rozproszenia odbicia w zakresie od ultrafioletu do bliskiej podczerwieni widma ze wszystkich znanych materiałów. Reflektancja w zakresie 400-1500nm wykorzystanego wzornika bieli wynosiła powyżej 99% w zakresie 400-1500nm oraz powyżej 95% w zakresie 250-2500nm [90].

Ręczne wykonywanie pomiarów w przedstawiony powyżej sposób jest żmudne i mało efektywne. Najnowszym trendem jest wykorzystanie bezzałogowych statków powietrznych (BSP, ang. UAV) do wykonywania takich pomiarów. Autor brał udział w badaniach, w których wykorzystywano BSP do podobnych zastosowań zarówno w aspekcie wykrywania chorób roślin, jak i innych aspektów upraw – zapotrzebowanie na wodę, zasobność gleby w makroelementy. Zebrany materiał pomiarowy nie pozwalał jednak na wykonanie tak

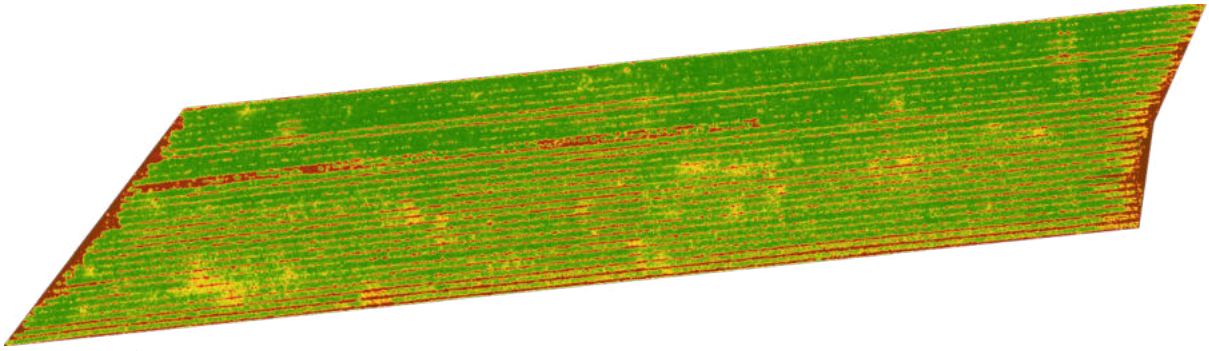
wnikliwych analiz jak te przedstawione w niniejszej pracy doktorskiej, jednak celowe jest wykonanie tego typu badań w przyszłości. Fakt ten wynikał przede wszystkim z zastosowania kamery spektralnej pozwalającej na akwizycję dużo węższego zakresu reflektancji o dużo niższej rozdzielczości spektralnej niż jest to możliwe przy wykorzystaniu opisanego powyżej urządzenia pomiarowego.
Na rysunkach rys. 18–rys. 20 przedstawiono ortofotomapy znormalizowanego, różnicowego wskaźnika wegetacyjnego (NDVI) obliczone na podstawie obrazów uzyskanych za pomocą kamery multispectralnej (Parrot Sequoia+) z poziomu BSP.
Przedstawiony wskaźnik wegetacyjny można przedstawić wzorem:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{7.8}$$

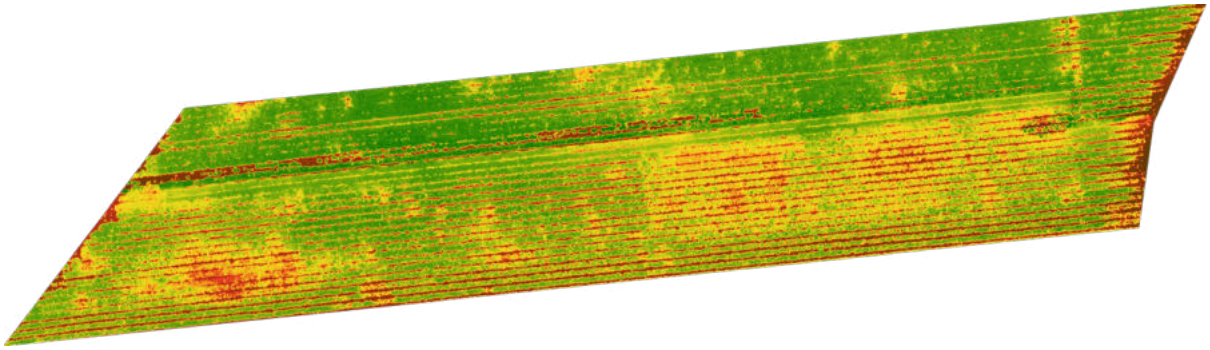
gdzie:
NIR – kanał bliskiej podczerwieni, 790 nm z rozdzielczością spektralną na poziomie ± 40 nm,
RED – kanał czerwony, 660 nm z rozdzielczością spektralną na poziomie ± 40 nm.



rys. 18 Ortofotomapa wskaźnika wegetacyjnego (NDVI) pola doświadczalnego w miejscowości Ernestynów z ledwie zauważalnymi lub niewidocznymi ogniskami chorobowymi (26.07.2019)



rys. 19 Ortofotomapa wskaźnika wegetacyjnego (NDVI) pola doświadczalnego w miejscowości Ernestynów z widocznymi ogniskami chorobowymi (02.08.2019)

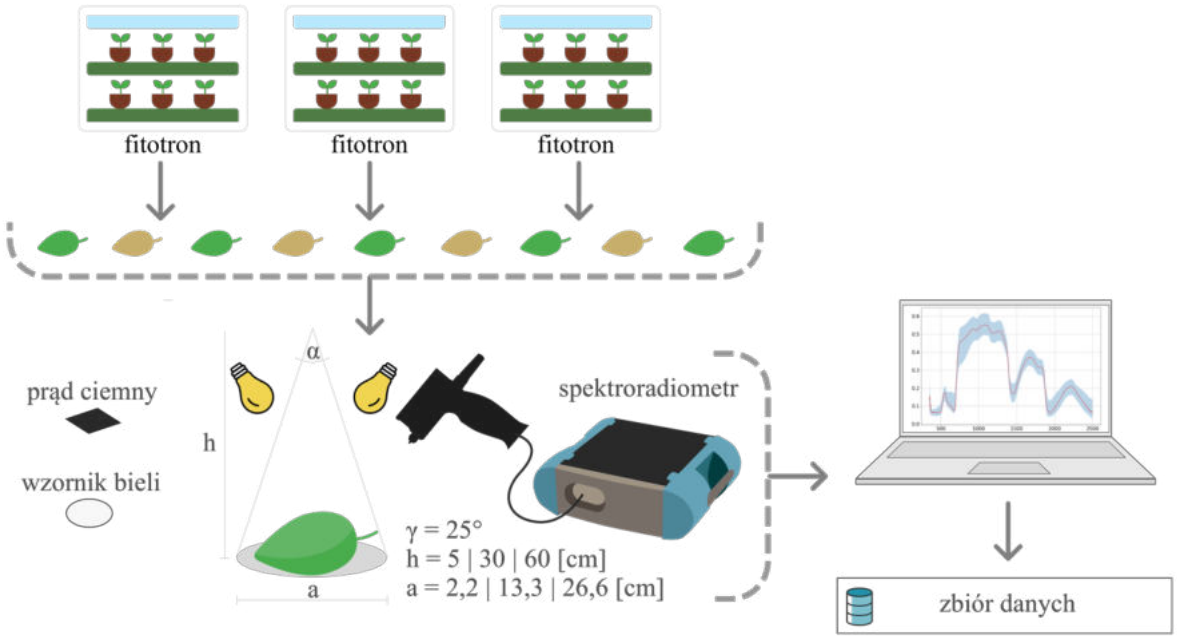


rys. 20 Ortofotomapa wskaźnika wegetacyjnego (NDVI) pola doświadczalnego w miejscowości Ernestynów z widocznymi, obszernymi zmianami chorobowymi sprzyjających dla patogenu warunkach atmosferycznych i po zastosowaniu nieadekwatnych środków ochrony roślin (05.08.2019)

W związku z powyższym nie zdecydowano się na wykorzystanie zebranych danych do przeprowadzenia analiz przedstawionych w dalszej części pracy.
Jednocześnie należy zauważyć, że wykorzystanie metod opracowanych w ramach niniejszej pracy doktorskiej w środowisku produkcyjnym dla pola przedstawionego powyżej (rys. 18–rys. 20), w dniach pomiędzy 26.07.2019 a 02.08.2019, mogłoby pomóc z minimalizacji strat widocznych na rys. 20. Właśnie dlatego podejście wykorzystujące obrazowanie hiperspektralne oraz multispektralne pozyskane z poziomu bezzałogowych statków powietrznych będzie przedmiotem dalszych badań Autora w przyszłości.

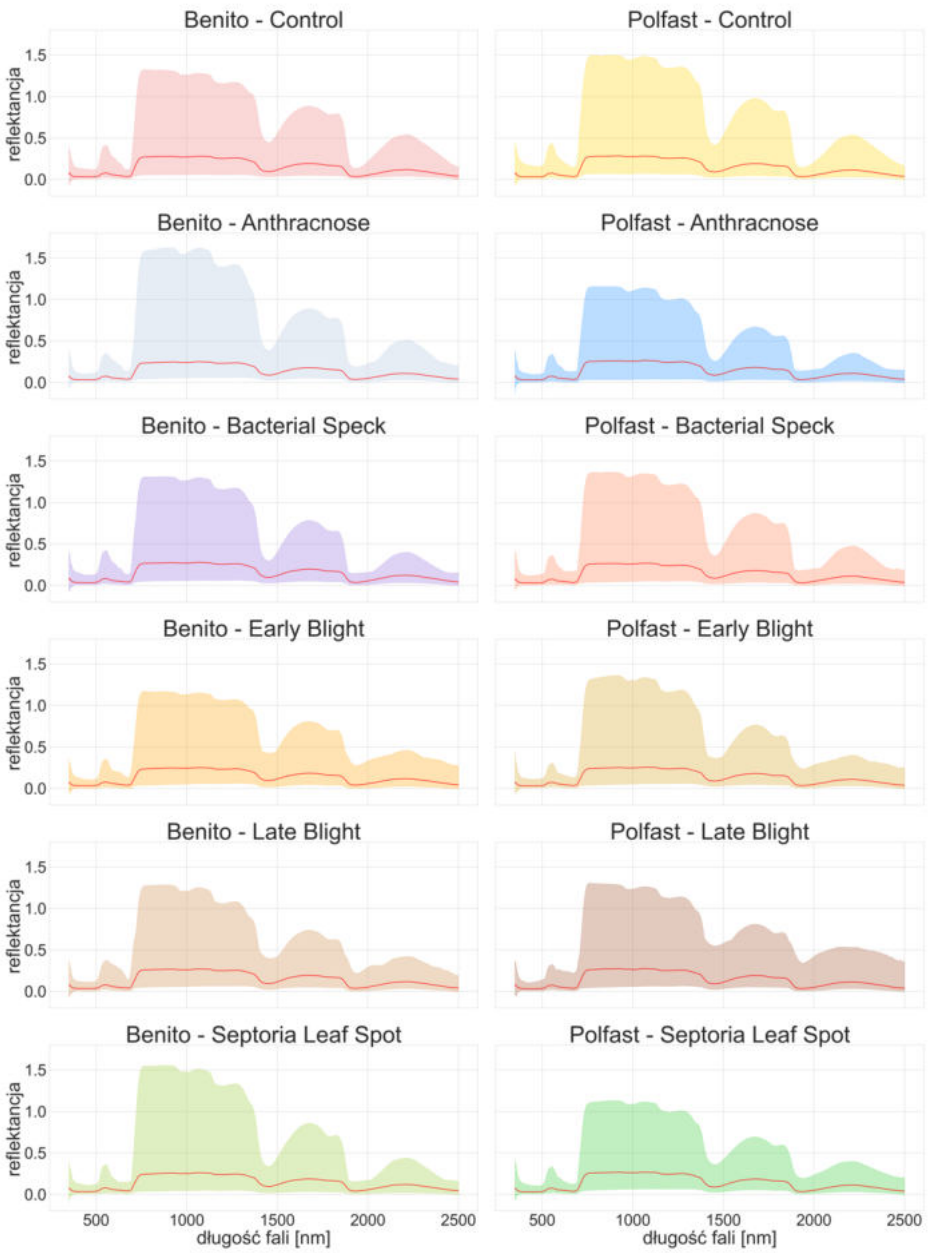
6.4. Pomiary hiperspektralne

W 3 rundach pomiarowych przeprowadzonych w środowiskach kontrolowanych (phytotron 1-3) wykonano pomiary hiperspektralne reflektancji badanych obiektów. Badane były liście pomidora zarażonego różnymi szczepami chorobowymi. Pomiar został wykonany poprzez odczyt reflektancji za pomocą spektroradiometru ASD FieldSpec 4 Hi-Res w zakresie 350-2500nm, opisanym w rozdziale 6.3. Do prawidłowego doświetlenia badanych obiektów wykorzystano sztuczne źródło światła w postaci dwóch identycznych lamp halogenowych, widocznych na rys. 20. Każdorazowo przed przystąpieniem do wykonywania serii pomiarowej wykonywano kalibrację spektroradiometru za pomocą bieli referencyjnej oraz kalibrację prądu czarnego, co szerzej opisane zostało w rozdziale 6.3.



rys. 21 Schemat obrazujący proces zbierania danych pomiarowych w tunelach foliowych

Pomiar badanych obiektów był wykonywany z 3 wysokości – 5 cm, 30 cm oraz 60 cm. Wiązka światła odbitego była przenoszona do spektroradiometru za pomocą przewodu optycznego o polu widzenia równym 25°. Koniec przewodu światłowodowego był skierowany prostopadle do badanego obiektu. Dla wysokości pomiarowej równej 5 cm średnica skanowanego obszaru wynosiła około 2,2 cm, dla 30 cm – 13,3 cm, a dla 60 cm – 26,6 cm, co znacząco przekracza przeciętne wymiary liści uprawy *Solanum lycopersicum*. Między innymi z tego powodu zdecydowano się na wykorzystanie danych pomiarowych wykonanych z wysokości 5cm od badanych obiektów na etapie przetwarzania danych, ich wstępnej analizy i tworzeniu zbioru danych treningowych. Dokładne wyliczenia w temacie zakresu badanej powierzchni przedstawiono w rozdziale 6.3 i podsumowano w tabeli tab. 7 Obszar skanowania próbki w zależności od wysokości, na której umieszczona była soczewka urządzenia skanującego przy zastosowaniu domyślnego osprzętu (brak dodatkowych soczewek korygujących FOV). Pomiary reflektancji każdorazowo zapisywane były do formatu .asd.
Na rysunku rys. 22 przedstawiono graficzne podsumowanie wszystkich, surowych pomiarów hiperspektralnych z podziałem na poszczególne grupy badawcze. Kolorowe tło przedstawia minimalne i maksymalne zakresy otrzymywanych wartości. Czerwoną linią przedstawiono medianę otrzymywanych wartości pomiarowych.



rys. 22 Rozkład pomiarów hiperspektralnych dla poszczególnych odmian i chorób; kolorowe tło stanowi jest wizualizacją zakresu min-max wszystkich wykonanych pomiarów; linią czerwoną oznaczono medianę wszystkich pomiarów

Każdy pomiar odbywał się w identycznych warunkach oświetleniowych, w kontrolowanych warunkach laboratoryjnych. Pojedynczy pomiar był uśrednieniem pięciu zarejestrowanych krzywych spektralnych w celu zapewnienia poprawności i wiarygodności pomiarowej.

6.5. Podsumowanie rozdziału

Rozdział opisuje szczegóły przeprowadzonego eksperymentu pomiarowego. Przedstawiony został harmonogram przeprowadzonych badań oraz szczegóły fitopatologiczne. Istotnym elementem rozdziału jest część opisująca wykorzystane urządzenie pomiarowe oraz sam

proces wykonywania pomiarów hiperspektralnych, które następnie zostały zastosowane do budowy zbioru danych przetwarzanych przez różne algorytmy uczenia maszynowego.

7. Wybrane metody uczenia maszynowego użyte w doktoracie

Ze względu na ilość danych zarejestrowanych podczas wykonywanego badania, tj. 2151 pasm spektralnych z zakresu 350-2500nm, zarejestrowanych w ponad 72 tys. pomiarów, 8zdecydowano się na wykorzystanie metod uczenia maszynowego, jako podstawowego narzędzia służącego do stworzenia nieinwazyjnej metody diagnostyki badanych obiektów. Przy wyborze kierowano się doświadczeniami innych badaczy i specjalistów, którzy z sukcesem przeprowadzili liczne badania w zakresie wykorzystania narzędzi uczenia maszynowego przy tworzeniu nieinwazyjnych, teledetekcyjnych metod diagnostycznych upraw, a które to badania przedstawiono w rozdziałach 3.3 i 4, a w szczególności w podrozdziale 4.5.

7.1.1. Lista wszystkich rozpatrywanych klasyfikatorów

W ramach przeprowadzonych prac badawczych wykonano serię eksperymentów w poszukiwaniu odpowiedniego rozwiązania pozwalającego na rozwiązanie problemu klasyfikacji binarnej, tj. określenie, czy badany obiekt jest obiektem zdrowym, czy chorym. Szerzej badanie opisane jest w załączniku (14.1).

W eksperymencie pod uwagę były brane następujące klasyfikatory:

- Ada Boost,
- Bagging Classifier,
- Calibrated Classifier,
- Decision Tree,
- Extra Tree,
- K-Nearest Neighbors,
- Label Propagation,
- Label Spreading,
- LGBM (Light Gradient Boosting Machines),
- Linear Discriminant Analysis,
- Linear SVC (Linear Support Vector Classification),
- Logistic Regression,
- Naive Bayes,
- Nearest Centroid,
- Nu SVC (Nu-Support Vector Classification),
- Passive Aggressive,
- Perceptron,
- Quadratic Discriminant Analysis,
- Random Forest,
- Ridge Classifier,
- SGD (Stochastic Gradient Descent),
- SVC (C-Support Vector Classification),
- XGBoost (eXtreme Gradient Boosting).

7.1.2. Opis wybranych klasyfikatorów

Po wykonaniu prac badawczych, szerzej opisanych w podrozdziale 9.2, do dalszych badań zdecydowano się na zawężenie wyboru wybór następujących klasyfikatorów:

1. Linear SVC,
2. Logistic Regression,
3. Ridge Classifier,
4. Random Forest,
5. LGBMClassifier,
6. XGBClassifier.

Klasfykatory 1-3 zostały wybrane ze względu na osiągnięcie najlepszego wyniku wśród klasyfikatorów nie ulegających zjawisku przetrenowania, natomiast klasyfikatory 4-6 osiągnęły najlepszy wynik wśród klasyfikatorów ulegających przetrenowaniu. Wykorzystane zostały implementacje klasyfikatorów dostępne w pakietach XGBoost, LightGBM oraz scikit-learn dla języka Python, a poniższy opis został opracowany w oparciu o dostępne w ich dokumentacji informacje [91–93].

Linear SVC

Liniowy klasyfikator wektorów nośnych (Linear Support Vector Machine) jest odmianą klasyfikatora wektorów nośnych o liniowym parametrze jądra. Domyślną normą funkcji kary jest dla tego klasyfikatora L2, a domyślnym strategią funkcji kary jest funkcja zawiasowa (*hinge loss*). Problem można sformułować następującą formułą:

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \max\left(0, 1 - y_i(w^T \phi(x_i) + b)\right) \tag{7.9}$$

gdzie:

- w – wektor wag,
- w^T – transponowana macierz współczynników,
- b – współczynnik biasu (uchybu),
- i – indeks,
- n – wielkość zbioru,
- y_i – wartość docelowa i-tego elementu zbioru treningowego (etykieta),
- x_i – i-ty element zbioru treningowego X o wymiarach (liczba próbek, liczba cech),
- C – parametr regularyzacji,
- ϕ – funkcja tożsamości (*identity function*) – ze względu na użytą funkcję liniową jądra zwraca wartości takimi, jakie są, tzn. $\langle x, x' \rangle$.

Wektorem wag nazywamy macierz wartości numerycznych określających istotność poszczególnych cech w procesie predykcyjnym. Skalarny składnik odchylenia dodawany jest do wag modelu, co pozwala na jego lepsze dopasowanie. Parametr regularyzacji jest wykorzystywany do uniknięcia efektu przetrenowania modelu poprzez uzyskanie kompromisu pomiędzy maksymalizacją odległości pomiędzy granicą decyzyjną, a najbliższym elementem zbioru treningowego oraz minimalizacją sumy kwadratów wektora wag. Im mniejszy parametr C, tym większa odległość pomiędzy granicami decyzyjnymi, ale tym samym większa suma kwadratów wektora wag.

Logistic Regression
Regresja logistyczna, zwana też w literaturze klasyfikacją maksymalnej entropii czy klasyfikatorem liniowo-logarytmicznym, jest liniową funkcją służącą do klasyfikacji. W tym modelu prawdopodobieństwo możliwego wyniku dla pojedynczej próby jest modelowe za pomocą funkcji logistycznej.
W wykorzystanej w niniejszym doktoracie klasyfikacji binarnej wykorzystano binarną implementację tego klasyfikatora.
Domyślnym parametrem funkcji kary $r(w)$ jest $l_2 = \frac{1}{2}\|w\|_2^2 = \frac{1}{2}w^T w$. Zatem w ramach zadania stara się minimalizować następującą funkcję kosztu dla przyjętych klas $\{0,1\}$:

$$\min_w C \sum_{i=1}^{n} \left(-y_i \log\left(\hat{p}(x_i)\right) - (1 - y_i) \log\left(1 - \hat{p}(x_i)\right)\right) + \frac{1}{2} w^T w \tag{7.10}$$

dla predykcji prawdopodobieństwa wystąpienia klasy 1, $P(y_i = 1|x_i)$, jako:

$$\hat{p}(x_i) = expit(x_i w + b) = \frac{1}{1 + \exp\left(-x_i w - b\right)} \tag{7.11}$$

gdzie:
- w – wektor wag,
- w_0 – macierz współczynników,
- y_i – wartość docelowa i-tego elementu zbioru treningowego (etykieta),
- x_i – i-ty element zbioru treningowego,
- C – parametr regularyzacji,
- b – współczynnik biasu (uchybu),
- $expit$ – funkcja sigmoidalna $\frac{1}{1+e^{-\beta x}}$, przy czym $\beta = 1$.

Ridge Classifier
Klasyfikator grzbietowy w pierwszej fazie zamienia klasy binarne na wartości $\{-1,1\}$, a następnie traktuje zadanie jak problem regresji brzegowej, tzn.:

$$\min_w \|Xw - y\|_2^2 + \alpha\|w\|_2^2 \tag{7.12}$$

gdzie:
- w – wektor wag,
- X – macierz danych wejściowych,
- y – wektor wartości docelowych,
- α – parametr regularyzacji.

Pierwsza część równania funkcji celu jest sumą różnic pomiędzy wartością przewidywaną, a wartością prawdziwą. Druga część formuły odpowiada za regularyzację poprzez dodanie kary proporcjonalnej do wielkości współczynników wag. Dzięki temu zapobiega się przetrenowaniu modelu, czyli nadmiernemu dopasowaniu do wartości treningowych.

Random Forest
Klasyfikator lasu losowego pochodzi z rodziny metod zespołowych, które łączą wyniki kilku klasyfikatorów bazowych w celu usprawnienia generalizacji. W tym przypadku budowanych jest wiele niezależnych klasyfikatorów, które w losowy, nieznaczny sposób różnią się od siebie, a następnie wynik jest uśredniany. Klasyfikatorem bazowym w lasach losowych jest algorytm drzewa decyzyjnego. Użyte w doktoracie podejście różni się od oryginalnego [94] w ten sposób, że w uśrednieniu podlega predykcja probabilistyczna pojedynczego klasyfikatora bazowego, a nie głos oddany na pojedynczą klasę. Zakładając, że dany jest wektor treningowy x_i oraz wektor etykiet y_i, pojedyncze drzewo decyzyjne rekurencyjnie dzieli przestrzeń cech w taki sposób, że przypadki z tymi samymi etykietami lub podobnymi wartościami docelowymi są grupowane razem. Niech dane w węźle m będą reprezentowane przez Q_m z n_m przykładami. Dla każdego podziału $\theta = (j, t_m)$ złożonego z cechy j oraz wartości brzegowej t_m zostaje uwzględniony następujący podział danych na podzbiory $Q_m^l(\theta)$ oraz $Q_m^p(\theta)$ w następującej postaci:

$$Q_m^l(\theta) = \{(x, y) | x_j \leq t_m\} \tag{7.13}$$

$$Q_m^p(\theta) = Q_m \setminus Q_m^l(\theta) \tag{7.14}$$

Jakość podziału w węźle m jest obliczana za pomocą funkcji zanieczyszczenia oraz straty $H()$, której wybór zależy od rozpatrywanego zadania. W badaniach, dla drzew decyzyjnych wykorzystywanych w modelach lasów losowych, zastosowano domyślne kryterium oceny, tj. indeks *Gini*. Zatem uwzględniając, że klasy wyjściowe przyjmują wartości *0, 1, ..., K-1* dla węzła m, należy przyjąć, że:

$$p_{mk} = \frac{1}{n_m} \sum_{y \in Q_m} I(y = k) \tag{7.15}$$

Wtedy funkcja straty przy zastosowanym indeksie Gini można opisać równaniem:

$$H(Q_m) = \sum_k p_{mk}(1 - p_{mk}) \tag{7.16}$$

Z kolei miarę jakości podziału można opisać równaniem:

$$G(Q_m, \theta) = \frac{n_m^l}{n_m} H\left(Q_m^l(\theta)\right) + \frac{n_m^p}{n_m} H\left(Q_m^p(\theta)\right) \tag{7.17}$$

Następnie wybierany jest parametr, który minimalizuje zanieczyszczenie

$$\theta^* = argmin_\theta G(Q_m, \theta) \tag{7.18}$$

Podział na podzbiory $Q_m^l(\theta^*)$ oraz $Q_m^p(\theta^*)$ powtarzany jest rekurencyjnie aż do osiągnięcia maksymalnej, dopuszczalnej głębokości drzewa.
Jak już zostało wspomniane, w podejściu lasów losowych tworzonych jest wiele klasyfikatorów bazowych, których wynik jest następnie uśredniany (ang. bagging). W

przypadku kolejnych klasyfikatorów, LGBM oraz XGB, wykorzystane jest podejście ze wzmocnieniem, co oznacza, że klasyfikatory bazowe budowane są w sposób sekwencyjny.

LGBMClassifier

Klasyfikator Light Gradient Boosting Machine zaimplementowany został w ramach pakietu LightGBM stworzonego przez firmę Microsoft. Geneza tego algorytmu dotyczy wysoko efektywnych drzew decyzyjnych wzmocnionych gradientem [95]. Podstawowa różnica tego rozwiązania w stosunku do innych rozwiązań opartych o algorytmy drzew decyzyjnych wzmocnionych gradientem (np. wspomniany niżej XGBoost) jest taka, że LightGBM nie wykorzystuje algorytmów sortowania wstępnego, a zamiast tego korzysta z algorytmów opartych na histogramie. Polega to na podziale ciągłych wartości cech na przedziały dyskretne, co znacząco przyspiesza trening i redukuje wykorzystanie pamięci.
Większość algorytmów wykorzystujących drzewa decyzyjne korzysta z podejścia zorientowanego na poziome budowanie drzew. W przypadku LightGBM wykorzystane jest podejście zorientowane na tworzenie liści (podjeście *leaf-wise*) promując najlepsze węzły w pierwszej kolejności.

XGBClassifier
XGBoost, a właściwie eXtreme Gradient Boosting, czyli klasyfikator ekstremalnego wzmocnienia gradientowego, podobnie jak LGBM wykorzystuje rodzinę metod wzmocnienia gradientowego. Algorytm korzysta z sortowania wstępnego. Struktura drzew wykorzystywanych przez XGBoost jest budowana poprzez dokładanie kolejnych poziomów drzewa (podejście *level-wise*). XGBoost wspiera metodę wzmocnienia gradientowego, stochastycznego wzmocnienia gradientowego, jak również regulowanego wzmocnienia gradientowego.

7.1.3. Hiperparametry dostrajane w ramach tworzonej metody

W ramach prac omawianych w rozdziale 9.2 w ramach działań przeprowadzonych dla etapu 6 (hypertuning) przeprowadzono proces dostrajania hiperparametrów trenowanych klasyfikatorów poprzez wykorzystane przeszukania losowego arbitralnie zadeklarowanej macierzy wartości tych parametrów. Poniżej umieszczono zestawienie parametrów z ich opisem oraz dopuszczalnymi wartościami tych parametrów.
W zależności od wykorzystanego modelu zakres wykorzystanych hiperparametrów różnił się, a w poniższych punktach przedstawiono zastosowane wartości parametrów dostrajanych w ramach etapu 6 zawartego w rozdziale 9.2.

Linear SVC – Liniowy klasyfikator wektorów nośnych

Liniowy klasyfikator wektorów nośnych (*ang. Linear Support Vector Classification*).

Wybrane zostały następujące wartości hiperparametrów do procesu losowego doboru (ang. Random Search):

- ‘C’: [0.1, 1, 10, 100, 1000],
- ‘class_weight’: [None, ‘balanced’],
- ‘dual’: [True, False],
- ‘fit_intercept’: [True, False],

- ‘intercept_scaling’: [0.1, 1, 10, 100],
- ‘loss’: [‘hinge’, ‘squared_hinge’],
- ‘max_iter’: [10, 100, 1000, 2000],
- ‘multi_class’: [‘ovr’, ‘crammer_singer’],
- ‘penalty’: [‘l1’, ‘l2’],
- ‘tol’: [1, 0.1, 0.01, 0.001, 0.0001].

Proces doboru hiperparametrów został przeprowadzony analogicznie dla każdego klasyfikatora.

Logistic Regression – Regresja logistyczna

- ‘C’: [0.1, 1, 10, 100, 1000],
- ‘class_weight’: [None, ‘balanced’],
- ‘dual’: [True, False],
- ‘fit_intercept’: [True, False],
- ‘intercept_scaling’: [0.1, 1, 10, 100],
- ‘l1_ratio’: [0.1, 0.25, 0.5, 0.75, 0.9],
- ‘max_iter’: [10, 100, 1000, 2000],
- ‘multi_class’: [‘auto’, ‘ovr’, ‘multinomial’],
- ‘penalty’: [‘l1’, ‘l2’, ‘elasticnet’, ‘none’],
- ‘solver’: [‘newton-cg’, ‘lbfgs’, ‘liblinear’, ‘sag’, ‘saga’],
- ‘tol’: [1, 0.1, 0.01, 0.001, 0.0001],
- ‘warm_start’: [True, False].

Ridge

- ‘alpha’: [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0],
- ‘class_weight’: [None, ‘balanced’],
- ‘copy_X’: [True, False],
- ‘fit_intercept’: [True, False],
- ‘max_iter’: [10, 100, 1000, 2000],
- ‘normalize’: [True, False],
- ‘solver’: [‘auto’, ‘svd’, ‘cholesky’, ‘lsqr’, ‘sparse_cg’, ‘sag’, ‘saga’].

Random Forest

- ‘n_estimators’: [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000],
- ‘max_features’: [‘auto’, ‘sqrt’],
- ‘max_depth’: [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
- ‘min_samples_split’: [2, 5, 10],
- ‘min_samples_leaf’: [1, 2, 4],
- ‘bootstrap’: [True, False].

LGBM

- ‘num_leaves’: [20, 40, 60, 80, 100],
- ‘min_child_samples’: [5, 10, 15],
- ‘max_depth’: [-1, 5, 10, 20],
- ‘learning_rate’: [0.05, 0.1, 0.2],
- ‘reg_alpha’: [0, 0.01, 0.03].

XGBoost

- ‘min_child_weight’: [1, 5, 10],
- ‘gamma’: [0.5, 1, 1.5, 2, 5],
- ‘subsample’: [0.6, 0.8, 1.0],
- ‘colsample_bytree’: [0.6, 0.8, 1.0],
- ‘max_depth’: [3, 4, 5].

Ze względu na objętość rozprawy nie zdecydowano się na przytaczanie opisów używanych hiperparametrów poszczególnych modeli uczenia maszynowego. Szczegółowe opisy znajdują się w dokumentacji technicznej znajdującej się pod następującymi odnośnikami internetowymi [91–93].

7.1.4. Metryki używane w pracy doktorskiej: F1, dokładność, czułość, precyzja

W badaniach przeprowadzonych w ramach niniejszej pracy doktorskiej wykorzystano miarę F1 dla określenia jakości dopasowania modelu. W literaturze pojawia się również miara dokładności (ang. accuracy), którą określamy sumę wszystkich poprawnie zaklasyfikowanych przypadków w odniesieniu do wszystkich przypadków.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7.19}$$

gdzie:

- TP – przypadki prawdziwie pozytywne,
- TN – przypadki prawdziwie negatywne,
- FP – przypadki fałszywie pozytywne,
- FN – przypadki fałszywie negatywne.

Za pomocą precyzji (ang. *precision*) określamy jaka część wyników wskazana przez klasyfikator jako przypadki zarażone faktycznie była przypadkami chorobowymi.

$$precision = \frac{TP}{TP + FP} \tag{7.20}$$

Czułość (ang. *recall*) określa jaka część przypadków chorobowych została wykrywa przez klasyfikator.

$$recall = \frac{TP}{TP + FN} \tag{7.21}$$

Miara F_1 pozwala uwzględnić zarówno precyzję, jak i czułość. Miara F_1 jest podstawową odmianą współczynnika F_β, gdzie β przyjmuje zazwyczaj wartość z zakresu $[0, 2]$ w zależności od tego, czy większą wagę przyjmuje czułość ($\beta = 2$) czy precyzja ($\beta = 0{,}5$).

$$F\beta = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision + recall)} \tag{7.22}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{7.23}$$

Podstawowa metryka wykorzystana w pracy doktorskiej ma za zadanie ocenić w sposób obiektywny i rzetelny poziom dopasowania modelu. W ramach prowadzonych działań badawczych jako zadowalającą wartość metryki F1 przyjęto wartość *0.8*.

7.2. Podsumowanie rozdziału

W niniejszym rozdziale przedstawiono opis procesu wyboru klasyfikatorów wykorzystanych w proponowanej, nieinwazyjnej metodzie diagnostyki badanych obiektów. Wyszczególnione zostały wszystkie algorytmy uczenia maszynowego, które brały udział we wstępnej selekcji. Opisano również szczegóły dotyczące poszczególnych klasyfikatorów wybranych do dalszych analiz.

8. Zbiór danych dla potrzeb uczenia maszynowego

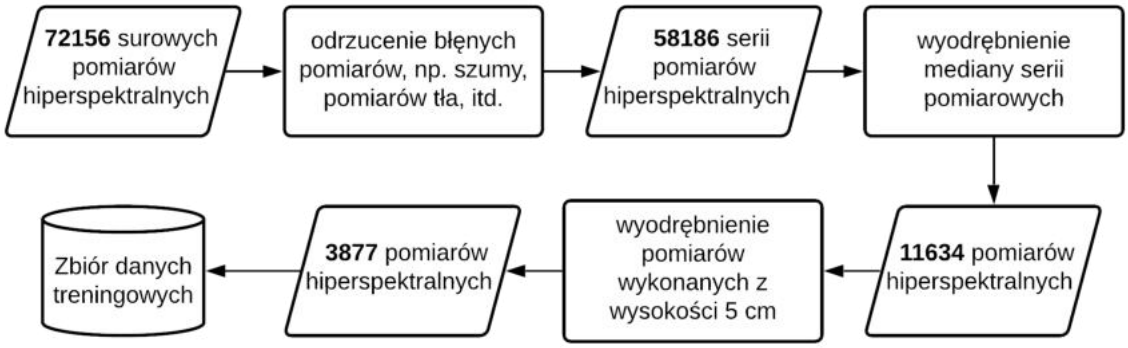
Dane pomiarowe zostały zebrane za pomocą spektroradiometru ASD FieldSpec 4 Hi-Res, szerzej opisanego w rozdziale 6.3.Sprzęt pomiarowy wykorzystany do stworzenia zbioru danych pozwala na analizę:

- reflektancji (ilości energii elektromagnetycznej odbitej od badanego obiektu),
- transmitancji (ilości energii elektromagnetycznej, która przechodzi przez badany obiekt),
- absorpcji (ilości energii pochłanianej przez badany obiekt).

W realizowanym badaniu wykorzystane zostały pomiary reflektancji.

8.1. Schemat postępowania z pomiarami

Wstępne przetworzenie danych polegało na odizolowaniu danych niepoprawnych, takich jak błędne pomiary kalibracyjne, pomiary obiektów nie będących przedmiotem projektu (np. tło badanego obiektu) od pomiarów wykonanych poprawnie. Następnie wyekstrahowano metadane zakodowane w numerach referencyjnych pomiarów i dodano je do zbioru danych tworząc macierz modelową (*ang. tidy dataset*)[96]. Obliczono znormalizowany, różnicowy współczynnik wegetacyjny (NDVI).
W efekcie zredukowano liczbę pomiarów pierwotnych z **72156** do **58186** odsiewając błędy pomiarowe, następnie do **11634** wyznaczając mediany serii pomiarowych, a ostatecznie wybrano pomiary wykonane z wysokości 5 cm jako najmniej zaszumione obrazem tła, a jednocześnie najbardziej wiarygodne w procesie badawczym, co poskutkowało uzyskaniem **3877** wiarygodnych, w pełni opisanych metadanymi unikatowych pomiarów.



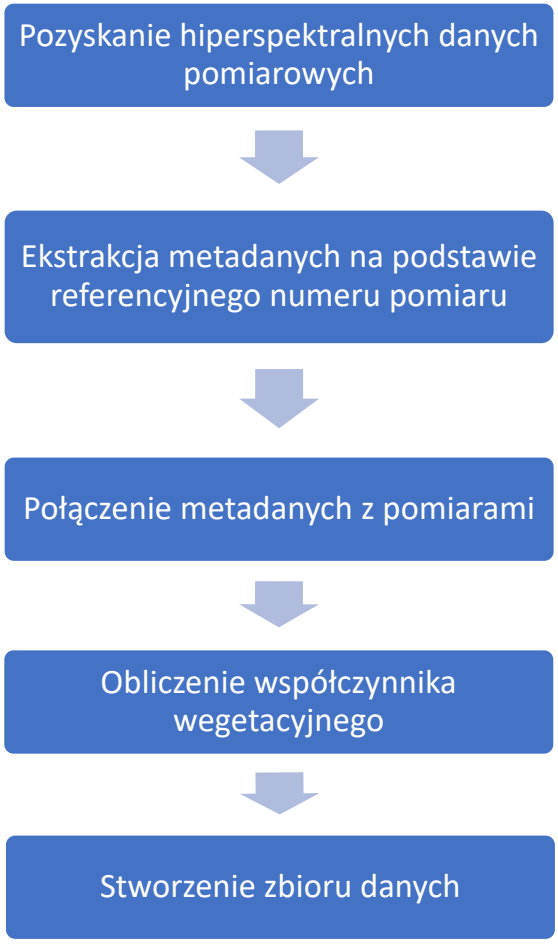
rys. 23 Proces wstępnego przetwarzania danych i zmiana ilości danych podczas wykonywania tego procesu

W procesie wstępnego przetwarzania danych, ilość pomiarów uległa znacznej redukcji, jednak w zbiorze treningowym pozostawiono jedynie wiarygodne dane.
Przez surowe pomiary hiperspektralne rozumie się pomiary wykonane za pomocą urządzenia ASD FieldSpec 4 Hi-Res wg procedury przedstawionej w rozdziale 6. Wśród tych danych znalazły się pomiary nieużyteczne dla stworzenia metod tworzonych w ramach niniejszej pracy, jednak potrzebnych w innych działania potrzebnych w ramach projektu, na potrzeby którego wykonywane były niniejsze pomiary. Jak zostało już wspomniane w procedurze opisującej proces wykonywania pomiarów, pojedynczy surowy pomiar hiperspektralny był uśrednieniem 5 zarejestrowanych krzywych spektralnych w celu zapewnienia poprawności i wiarygodności pomiarowej. Pomiar dla pojedynczego obiektu badawczego był wykonywany 5-krotnie, dalej określanych jako seria pomiarowa. W procesie wstępnego przetwarzania

zbioru danych obliczono medianę serii pomiarowej, aby wyeliminować potencjalnie niepoprawne pomiary oraz powtórzenia pomiarów tych samych obiektów.
Na podstawie referencyjnego numeru pomiarowego ekstrahowane były metadane poszczególnych pomiarów. Metadane zawierały następujące informacje:

- datę wykonania pomiaru,
- grupę badawczą,
- numer cyklu eksperymentalnego (fitotron 1-3),
- odmianę rośliny,
- odmianę patogenu,
- datę wysiania,
- datę sadzenia,
- datę inokulacji,
- liczbę dni, które minęły od momentu inokulacji.

Metadane stanowiły uzupełnienie zbioru danych, na potrzeby grupowania i filtracji zbioru w procesie uczenia, jednak nie brały one bezpośredniego udziału w procesie treningu modeli.



rys. 24 Uproszczony schemat przetwarzania danych do postaci czystego zbioru danych

Dodatkowo, na podstawie danych z pomiarów spektralnych obliczono znormalizowany, różnicowy indeks wegetacyjny (równanie (7.8)) przyjmując za RED długość fali 670nm, natomiast za NIR przyjęto długość fali 800nm. Uproszczony schemat opisanego powyżej postępowania przestawiono na rys. 24.

8.2. Obróbka wstępna danych

Pomiary wykonane za pomocą spektrometru należało poddać obróbce wstępnej. Pliki wejściowe krzywych spektralnych zostały zapisane w formacie '.asd'. Należało wykonać uporządkowanie plików i przyporządkowanie do miejsca cyklu eksperymentalnego (fitotron 1-3), daty wykonania pomiaru, odmiany rośliny (*Benito*, *Polfast*) oraz patogenu, za pomocą którego została wykonana inokulacja rośliny (*Alternaria solani, Colletotrichum coccodes, Pseudomonas syringe pv. Tomato, Septoria lycopersici, Phytophthora infestans*).

tab. 8 Wybrane wiersze z tabeli mapującej służącej do wstępnego przetwarzania danych (pozostawiono oryginalne nazwy kolumn odpowiadające kolumną znajdującym się w przetwarzanym zbiorze danych – opis znajduje się na końcu podrozdziału)

field_no	variety	pathogen	disease	disease_PL	type
23	Polfast	*Phytophthora infestans*	Late Blight	Zaraza ziemniaka	phytotron 1
256	Polfast	Control	Control	Kontrola	phytotron 3
12	Benito	Control	Control	Kontrola	phytotron 2
20	Polfast	*Colletotrichum coccodes*	Anthracnose	Antraknoza pomidora	phytotron 3
3	Benito	*Pseudomonas syringe pv. Tomato*	Bacterial Speck	Bakteryjne cętkowatość	phytotron 2
230	Benito	*Colletotrichum coccodes*	Anthracnose	Antraknoza	phytotron 2
237	Polfast	*Phytophthora infestans*	Late Blight	Zaraza ziemniaka	phytotron 3
229	Polfast	*Colletotrichum coccodes*	Anthracnose	Antraknoza	phytotron 3
240	Polfast	*Septoria lycopersici*	Septoria Leaf Spot	Septorioza	phytotron 1
235	Polfast	*Pseudomonas syringe pv. Tomato*	Bacterial Speck	Bakteryjne cętkowatość	phytotron 1
27	Polbig	*Pseudomonas syringe pv. Tomato*	Bacterial Speck	Bakteryjne cętkowatość	phytotron 3

Do odczytu danych z formatu .asd posłużyła biblioteka zewnętrzna specdal. Jest to pakiet przeznaczony do odczytu, przetwarzania oraz agregowania danych pozyskanych za pomocą spektroskopów polowych firm ASD, SVC czy PSR.
Losowo wybrane elementy gotowego zbioru danych zostały przedstawione w tab. 9. W poniższej tabeli kolumny 350.0 i 2500.0 reprezentują wartości pomiarowe dla długości fali elektromagnetycznej odpowiednio 350 i 2500 nm.

tab. 9 Przykładowe rekordy zawarte w zbiorze danych wykorzystanych w prowadzonym badaniu. Kolumny "..." reprezentują brakujące kolumny opisane poniżej tabeli, m.in. istotne daty zabiegów agrotechnicznych (siew, rozsada, inokulacja), wartość wskaźnika wegetacyjnego dla wykonywanego pomiaru, ilość dni, które minęły od inokulacji do dnia wykonywania pomiaru oraz wartości pomiarowych dla zakresu spektralnego 351-2499[nm] (pozostawiono oryginalne nazwy kolumn odpowiadające kolumną znajdującym się w przetwarzanym zbiorze danych – opis znajduje się na końcu podrozdziału)

measurement_id	**group_id**	**date**	**type**	**variety**	**pathogen**	**disease**	**…**	**350.0**	**…**	**2500.0**
kom2_2019-11-13_1900047	19	2019-11-13	phytotron_2	Polfast	*Alternaria solani*	Early Blight	…	0,070228	…	0,013017
kom1_2019-09-17_2200007	22	2019-09-17	phytotron_1	Polfast	*Septoria lycopersici*	Septoria Leaf Spot	…	-0,02374	…	0,012925
kom3_2019-12-20_1200085	12	2019-12-20	phytotron_3	Benito	Control	Control	…	0,095769	…	0,087443
kom1_2019-10-17_2000076	20	2019-10-17	phytotron_1	Polfast	*Colletotrichum coccodes*	Anthracnose	…	0,215997	…	0,036504
kom1_2019-10-14_2300069	23	2019-10-14	phytotron_1	Polfast	*Phytophthora infestans*	Late Blight	…	0,032333	…	0,027258
kom1_2019-10-03_0700014	07	2019-10-03	phytotron_1	Benito	*Alternaria solani*	Early Blight	…	0,128032	…	0,096656
kom2_2919-12-12_2100009	21	2019-12-12	phytotron_2	Polfast	*Pseudomonas syringe pv. Tomato*	Bacterial Speck	…	0,118505	…	0,0314
kom3_2019-12-09_2400055	24	2019-12-09	phytotron_3	Polfast	Control	Control	…	-0,02624	…	0,068639
kom1_2019-09-18_2000052	20	2019-09-18	phytotron_1	Polfast	*Colletotrichum coccodes*	Anthracnose	…	0,020275	…	0,009961
kom3_2019-12-19_1000063	10	2019-12-19	phytotron_3	Benito	*Septoria lycopersici*	Septoria Leaf Spot	…	0,155491	…	0,061995
kom2_2019-11-17_0700019	07	2019-11-17	phytotron_2	Benito	*Alternaria solani*	Early Blight	…	0,065724	…	0,040945

tab. 10 Tabela mapująca zawierająca informacje o badanym obiekcie w zależności od numeru referencyjnego (field_no, pozostawiono oryginalne nazwy kolumn odpowiadające kolumną znajdującym się w przetwarzanym zbiorze danych – opis znajduje się na końcu podrozdziału)

field_no	variety	pathogen	disease	disease_PL	type
7	Benito	*Alternaria solani*	Early Blight	Alternarioza	phytotron
8	Benito	*Colletotrichum coccodes*	Anthracnose	Antraknoza	phytotron
9	Benito	*Pseudomonas syringe pv. Tomato*	Bacterial Speck	Bakteryjne cętkowatość	phytotron
10	Benito	*Septoria lycopersici*	Septoria Leaf Spot	Septorioza	phytotron
11	Benito	*Phytophthora infestans*	Late Blight	Zaraza ziemniaka	phytotron
12	Benito	Control	Control	Kontrola	phytotron
19	Polfast	*Alternaria solani*	Early Blight	Alternarioza pomidora	phytotron
20	Polfast	*Colletotrichum coccodes*	Anthracnose	Antraknoza pomidora	phytotron
21	Polfast	*Pseudomonas syringe pv. Tomato*	Bacterial Speck	Bakteryjne cętkowatość	phytotron
22	Polfast	*Septoria lycopersici*	Septoria Leaf Spot	Septorioza	phytotron
23	Polfast	*Phytophthora infestans*	Late Blight	Zaraza ziemniaka	phytotron
24	Polfast	Control	Control	Kontrola	phytotron

tab. 11 Tabela mapująca dla dat siania, rozsady oraz inokulacji obiektów badawczych (pozostawiono oryginalne nazwy kolumn odpowiadające kolumną znajdującym się w przetwarzanym zbiorze danych – opis znajduje się na końcu podrozdziału)

type	seeding	planting	inoculum
phytotron_1	12.08.2019	28.08.2019	10.09.2019
phytotron_2	27.09.2019	14.10.2019	12.11.2019
phytotron_3	05.11.2019	21.11.2019	09.12.2019

Poniżej przedstawiono opis poszczególnych kolumn zbioru danych:

- measurement_id – unikatowy kod pomiaru zawierający informacje odseparowane znakiem podkreślenia; na podstawie danych zawartych w tej kolumnie oraz danych z tabel mapujących (tab. 10, tab. 11) przyporządkowano pozostałe dane zbioru do danych pomiarowych (kolumny 350-2500),
- group_id – numer grupy, określający numer referencyjny obiektu pomiarowego (field_no w tab. 10),
- date – data wykonania pomiaru,
- type – rodzaj cyklu pomiarowego – kom1 to phytotron_1, kom2 to phytotron_2, natomiast kom_3 to phytotron_3,
- variety – odmiana rośliny (Benito lub Polfast),
- pathogen – patogen, którym zakażano rośliny,
- disease – choroba zaobserwowana na roślinie,
- days_from – liczba dni, które upłynęły od momentu inokulacji do dnia wykonania pomiaru,
- kolumn 350 do 2500 – wartości pomiaru reflektancji, nazwa kolumny jest jednocześnie reprezentacja długości fali podanej w nanometrach [nm].

Dodatkowe kolumny niezawarte w tab. 9 ze względu na ograniczoną szerokość strony:

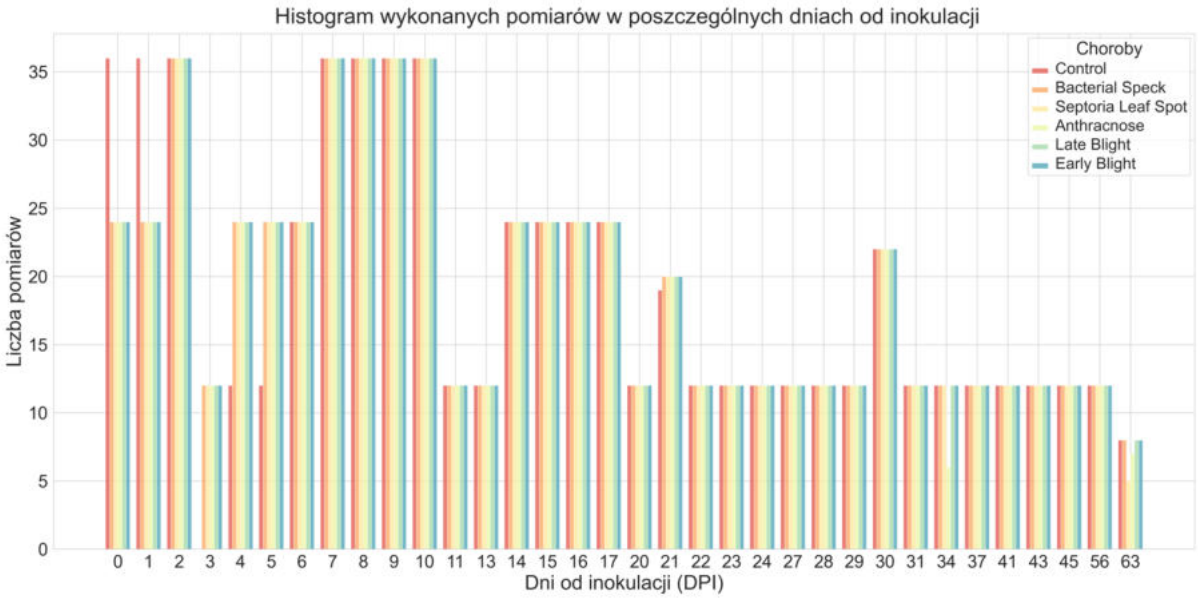
- seeding – data siewu,
- planting – data rozsady,
- inoculum – data zakażenia patogenem (inokulacji),
- days_from_inoculum – liczba dni, które upłynęły od inokulacji do daty wykonania pomiaru,
- ndvi – wartość znormalizowanego, różnicowego wskaźnika wegetacyjnego,
- height – wysokość wykonania pomiaru, czyli odległość pomiędzy badanym obiektem, a końcem przewodu światłowodowego,
- iteration_id – numer referencyjny,
- ref_img_path – ścieżka do folderu zawierającego referencyjne zdjęcia roślin.

8.3. Metoda łączenia danych dzień po dniu

Metoda utworzona w ramach niniejszej pracy doktorskiej powstawała w sposób sekwencyjny. Pierwszym podejściem do zagadnienia rozpoznawania chorób na podstawie danych hiperspektralnych z wykorzystaniem algorytmów uczenia maszynowego była praca mająca na celu rozpoznanie konkretnej choroby na ograniczonym zbiorze danych [31]. Zbiór ten obejmował pomiary wykonane jedynie jednego dnia, jednak wyniki skłoniły zespół badawczy do wykonania szerszego zbioru pomiarów. Przedstawione w powyższym zestawieniu dane zostały zebrane w trzech cyklach pomiarowych, wykonywanych systematycznie prawie każdego dnia od inokulacji tak, aby ująć dynamikę zmian spektralnych zachodzących w badanych obiektach zarówno przed, jak i po wystąpieniu objawów widocznych gołym okiem i rozpoznawalnych przez osoby wykwalifikowane w zakresie taksonomii patogenów roślin.
Pomiary wykonywano w okresie od inokulacji patogenu aż do 63 dnia od zakażenia. W początkowej fazie pomiary wykonywane były codziennie, w fazie pełnego rozwoju patogenu pomiary wykonywano z niższą częstotliwością. Pełne zestawienie dni pomiarowych przestawiono w opisie harmonogramu badań pomiarowych na rys. 13. Całość badań nad stworzeniem metody pozwalającej na diagnostykę chorobową badanych obiektów podzielono na część opartą o pełne zestawienie danych przedstawionych w niniejszym rozdziale, ale

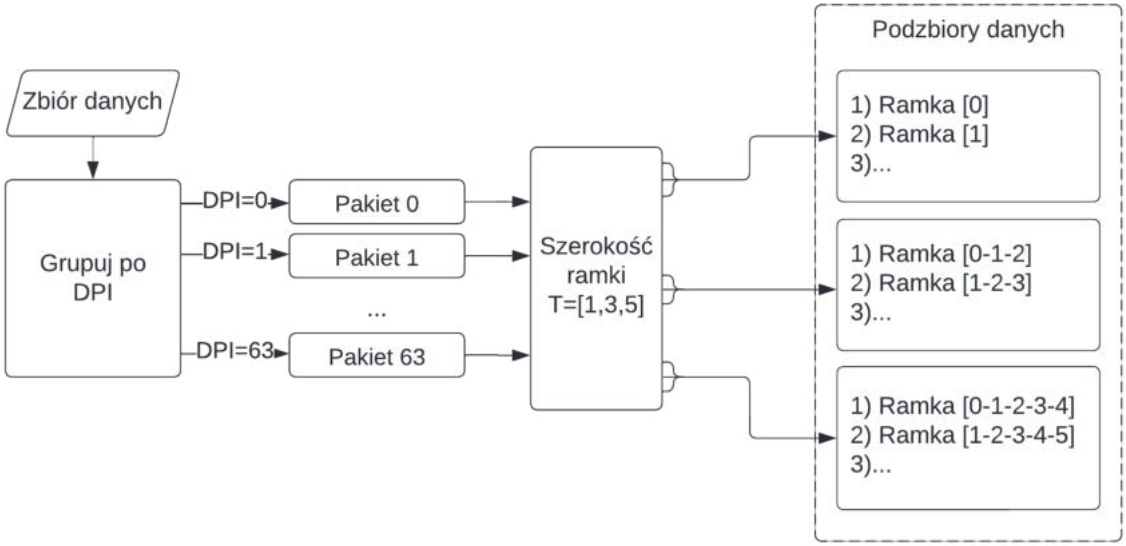
również na część opracowaną na podstawie zbiorów cząstkowych w celu określenia najwcześniejszego momentu, w którym możliwa jest diagnoza chorób. Jak wykazano na podstawie ortofotomap rzeczywistego pola badawczego przedstawiających rozprzestrzenianie się ognisk chorobowych w uprawach (rys. 18–rys. 20) wczesna i prawidłowa diagnoza, choć nie jest zadaniem łatwym, jest kluczowym czynnikiem w minimalizacji strat poprzez umożliwienie zastosowania odpowiednich środków zaradczych w krótkim czasie od wystąpienia infekcji.
Główny zbiór danych (3877 pomiarów) był dzielony na podzbiory na podstawie dni, które minęły od inokulacji (DPI, *ang. days post inoculum*).

Systematyczne zbieranie danych w warunkach laboratoryjnych, zbliżonych do rzeczywistych (fitotrony) jest utrudnione m.in. ze względu na występowanie błędów pomiarowych, błędów kalibracji urządzeń oraz problemy związane z poprawnym zapisem pomiarów w bazie danych. Z tych powodów, pomimo odpowiedniego zaplanowania eksperymentu i dołożenia wszelkiej staranności podczas wykonywania pomiarów, uzyskano różne pod względem liczebności zbiory pomiarowe w kolejnych dniach od inokulacji. Ze względu na niską reprezentację zbiorów pomiarowych w niektórych dniach (rys. 3), np. dzień 3, 11, 13, 20, itd. zdecydowano się na zastosowanie autorskiej metody łączenia pomiarów z kilku dni we wspólne zbiory treningowe.



rys. 25 Liczba pomiarów w zależności od dni, które minęły od inokulacji w środowisku kontrolowanym z podziałem na analizowane choroby

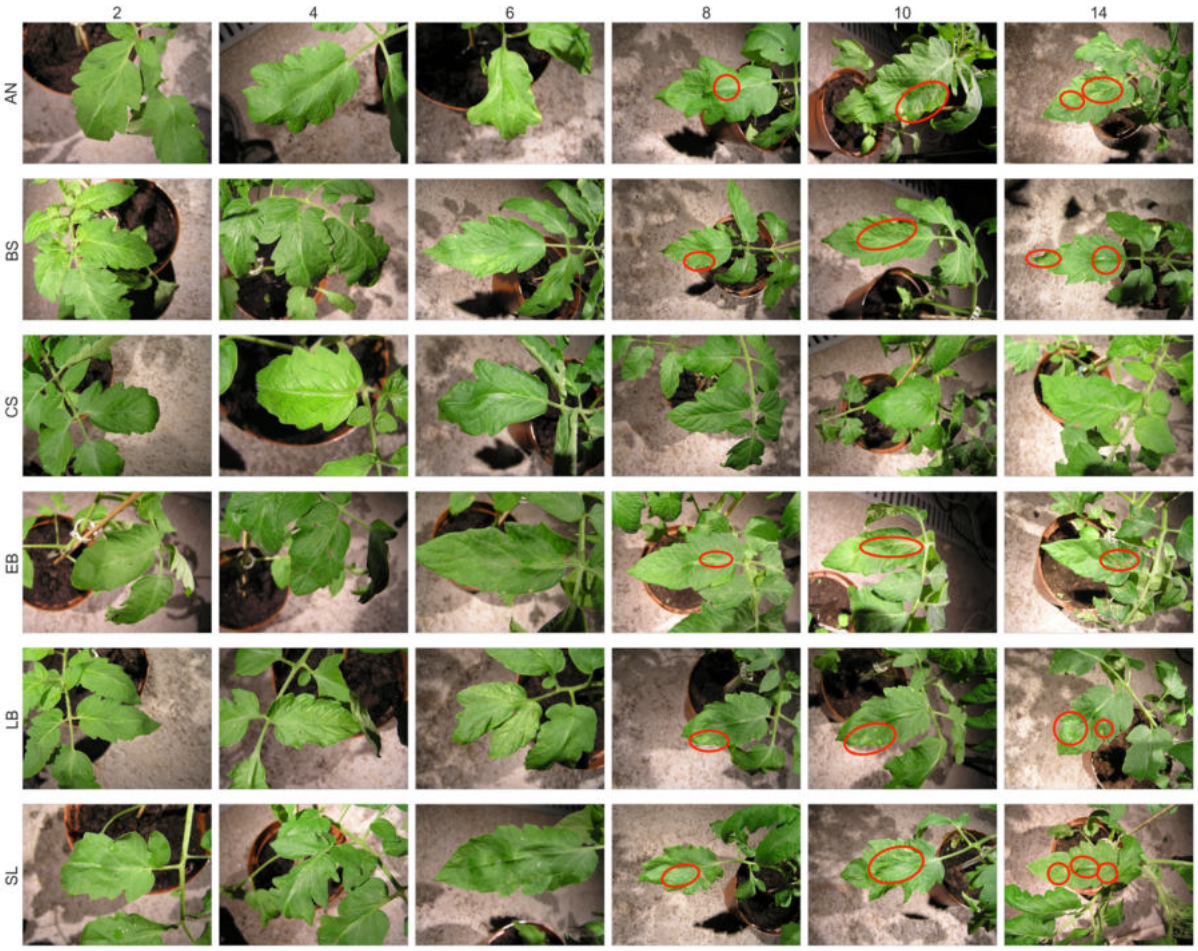
Zbiór danych podzielony został na pakiety pomiarów w zależności od liczby dni, które minęły od zakażenia (days post-inoculum, DPI = [0, 1, 2, …, 63]), zgodnie ze schematem przedstawionym na rysunku rys. 25. Pakiet, to zbiór pomiarów z danego dnia zawierający wszystkie zakresy spektralne dla wszystkich roślin z obserwowanych fitotronów. Następnie ustalono szerokość ramki danych (T = [1, 3, 5]), czyli przedział czasu obejmujący określoną liczbę kolejnych dni pomiarowych.



rys. 26 Schemat tworzenia podzbiorów danych do analizy przedziałów czasowych, DPI (days post inoculum) liczba dni, które minęły od inokulacji, źródło: [97]

Szerokości ramki zostały wyznaczone tak, by jak najwcześniej wykryć pojawienie się choroby, tzn. na etapie, kiedy pojawiają się dopiero pierwsze oznaki chorobowe lub nie są one widoczne jeszcze gołym okiem [17]. W praktyce oznacza to czas poniżej 7 dni. Zbiór główny uszeregowany rosnąco na podstawie wartości DPI podzielono na podzbiory (ramki) o szerokości *T*. Dla każdego podzbioru przeprowadzono trening klasyfikatorów binarnych zgodnie z metodyką zaprezentowaną w rozdziale 9.2. Klasyfikacja binarna. Wyniki badania przeprowadzonych na podzbiorach zaprezentowano w rozdziale 10. Analiza skuteczności klasyfikacji chorób w kolejnych dniach od inokulacji

Ze względu na metodykę pozyskiwania danych pomiarowych dla *T*=1 nie było możliwe przeprowadzenie klasyfikacji binarnej dla każdej wartości DPI z zakresu 1-63 ze względu na braki pomiarów. Przykładowo dla DPI=3 nie zostały wykonane żadne pomiary kontrolne.
W przedstawionych obliczeniach zsumowana wartość przypadków biorących udział w tworzeniu modeli jest większa od liczby wykonanych pomiarów ze względu na udział pomiarów kontrolnych w treningu dla różnych chorób oraz łączenie ze sobą pomiarów z sąsiadujących dni pomiarowych dla *T*>1. Całkowita liczba pomiarów wyniosła: 7384, 19960 oraz 33016 odpowiednio dla *T*=1, *T*=3 oraz *T*=5.



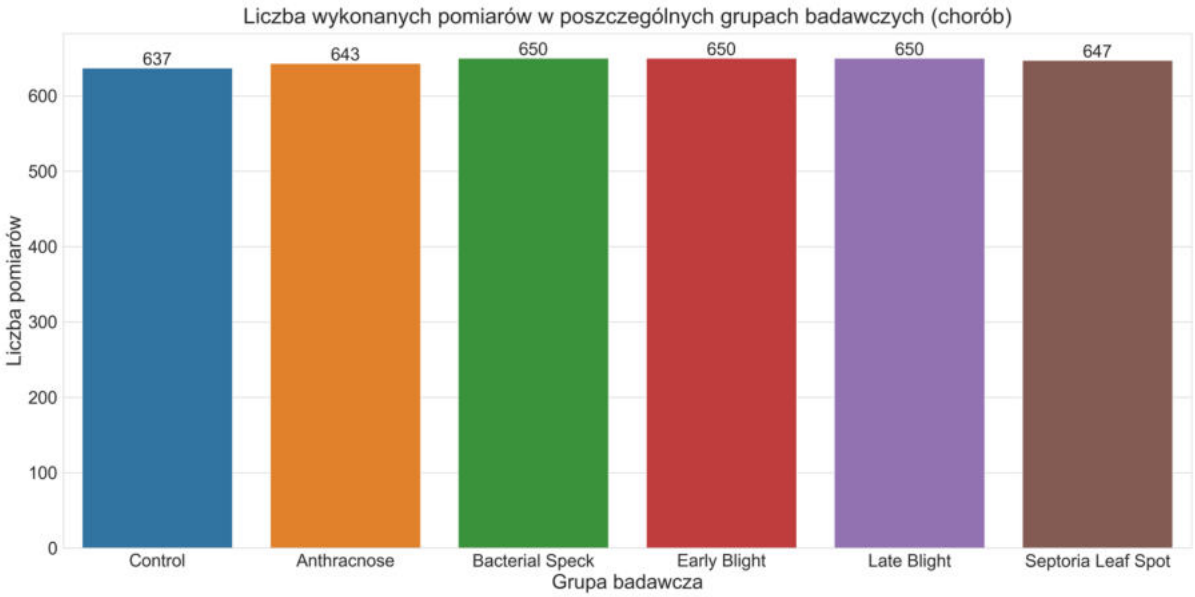
rys. 27 Zestawienie zdjęć referencyjnych badanych roślin w czasie rozwoju chorób w pierwszych dwóch tygodniach od inokulacji, czerwonym obwodem zaznaczono widoczne zmiany chorobowe, numeracja kolumn oznacza liczbę dni, które minęły od inokulacji, wiersze AN, BS, CS, EB, LB i SL oznaczają przypadki chorobowe opisane w rozdziale 9 „Wykrywanie chorób z wykorzystaniem wybranych algorytmów uczenia maszynowego", źródło: [97]

Na rys. 27 przedstawiono zdjęcia referencyjne badanych roślin, na których można zauważyć rozwój poszczególnych chorób w czasie, który minął od inokulacji. Niewielka liczba zmian chorobowych widocznych gołym okiem w pierwszych dniach od zakażenia obrazuje jak trudnym, a jednocześnie istotnym zadaniem jest tworzenie nieinwazyjnych systemów wykrywania choroby we wczesnym stadium jej rozwoju. Na podstawie oceny wizualnej bardzo trudno jest ocenić rolnikowi stan danej uprawy lub rozpoznać rodzaj patogenu, który istotnie może wpłynąć na przyszłe zbiory.

8.4. Analiza danych pomiarowych – EDA

Przed przystąpieniem do prac badawczych przeanalizowany został zbiór wejściowy poprzez wykonanie eksploracyjnej analizy danych (Exploration Data Analysis, EDA). W niniejszym rozdziale ujęto najważniejsze informacje dotyczące przeprowadzonej analizy.

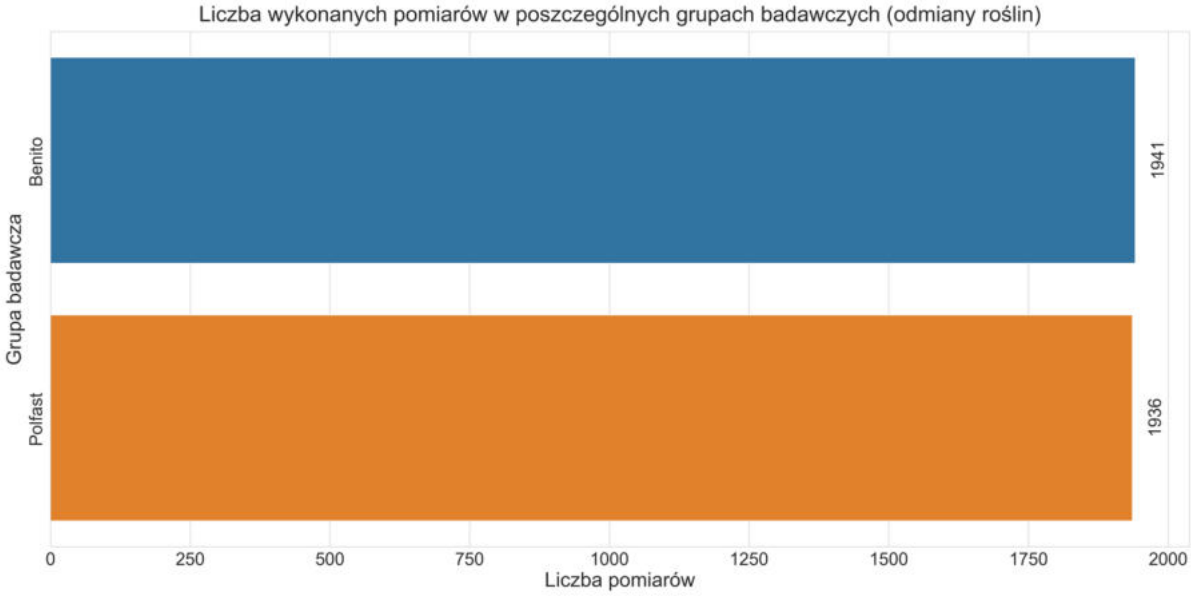
W rozdziale 8.2 zamieszczono tabele oraz listy kolumn opisujących wykonane pomiary. W pierwszej kolejności sprawdzono, czy podział danych ze względu na wykorzystany patogen jest równy. W każdej z klas chorobowych wykonano 637 do 650 pomiarów w trzech cyklach pomiarowych. Rozkład liczby reprezentantów klas w tym podziale przedstawiono na rysunku 27.



rys. 28 Liczba wykonanych pomiarów w zależności od użytego patogenu

Zbliżona liczba pomiarów w każdej grupie (klasie) świadczy o prawidłowo zaplanowanej i wykonanej części pomiarowej eksperymentu. Zbiór posiadającą równą liczbę reprezentantów w każdej z klas nazywamy zbiorem zbalansowanym. Taki zbiór pozwala na prawidłowe przeprowadzenie procesu trening uczenia maszynowego ze względu na redukcję stronniczości modelu.

W badanym zbiorze wykorzystano dwie odmiany roślin, aby umożliwić lepszą generalizację modelu w kontekście odmian tzn., aby model nie był nakierowany jedynie na jedną określoną grupę.



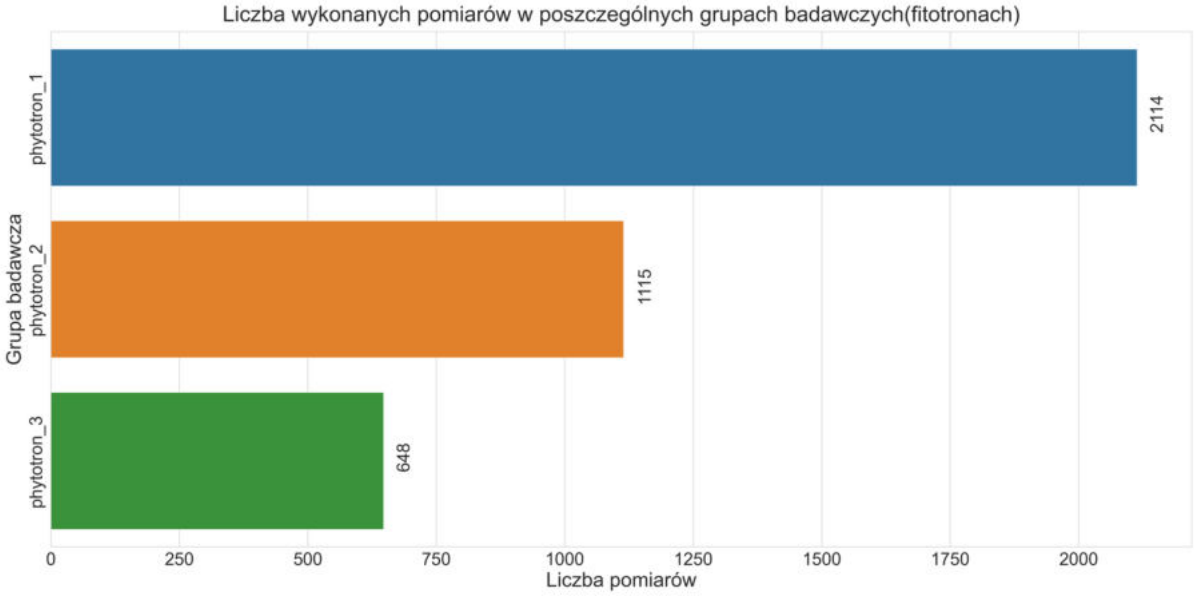
rys. 29 Liczba wykonanych pomiarów w zależności od odmiany rośliny

Na rys. 29 można zauważyć, że niemal równą liczbę pomiarów wykonano dla obu odmian rośliny – 1936 dla odmiany Polfast oraz 1941 dla odmiany Benito. W tym miejscu warto również przypomnieć rysunek rys. 22 reprezentujący krzywe spektralne dla danych podzielonych pod kątem odmian i chorób. Dla ułatwienia poniżej przedstawiono fragment wykresu w powiększeniu.

rys. 30 Fragment wykresu rys. 22; Rozkład pomiarów hiperspektralnych dla poszczególnych odmian i chorób; kolorowe tło stanowi jest wizualizacją zakresu min-max wszystkich wykonanych pomiarów; linią czerwoną oznaczono medianę wszystkich pomiarów

Pomimo różnic w poszczególnych pomiarach, w tym zakresie od minimalnych do maksymalnych wartości pomiarowych (kolorowe tło wykresu) pomiędzy odmianami, czy nawet między chorobami medianowa krzywa spektralna (oznaczona czerwoną linią) we wszystkich grupach ma podobny przebieg.

Na kolejnym rysunku (rys. 31) przedstawiono podział na cykle badawcze (phytotron_1-3). W tym przypadku grupy nie są równoliczne.



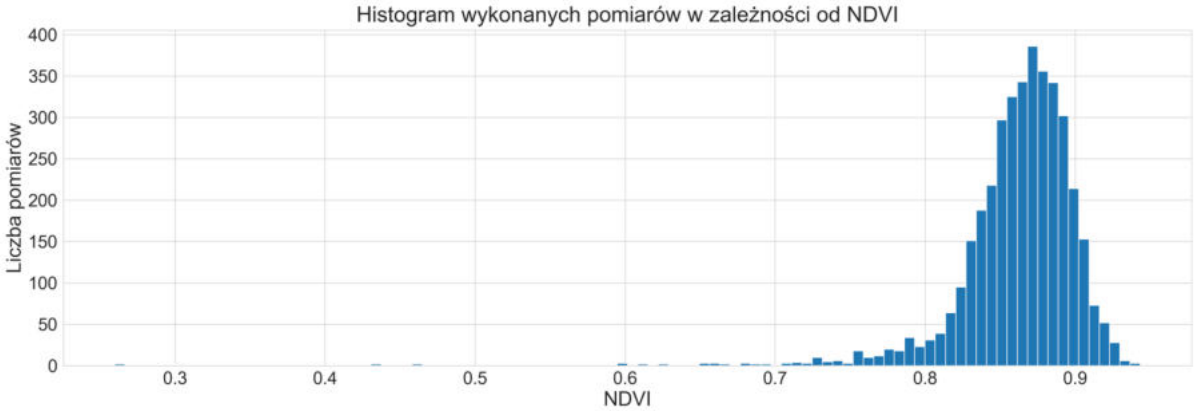
rys. 31 Liczba wykonanych pomiarów w zależności od fitotronu

Zabieg ten jest celowy. Cykle nie były prowadzone równolegle i już po pierwszej iteracji badawczej, w której wykonano 2114 pomiarów, wyciągnięto wniosek, że kluczowych jest kilka pierwszych tygodni rozwoju roślin. W drugim cyklu wykonano 1115 pomiarów. Przed przystąpieniem do badań w trzecim, a zarazem ostatnim cyklu podjęto decyzję o zawężeniu tego czasu do kluczowych, pierwszych 11 dni rozwoju rośliny. Wykonano wtedy 648 pomiarów.

Zbiorczą strukturę wykonanych we wszystkich cyklach pomiarów przedstawiono już wcześniej na rys. 25, na którym zobrazowano liczbę wykonanych pomiarów w zależności od badanej choroby i dni, które minęły od inokulacji. Z kolei na rys. 13 przedstawiono strukturę wykonanych pomiarów na osi czasu z uwzględnieniem podziału na poszczególne cykle i dni pomiarowe. Łącznie wykonano 3877 unikatowych pomiarów, które wykorzystano w dalszych pracach, co szerzej omówiono w rozdziale 8.

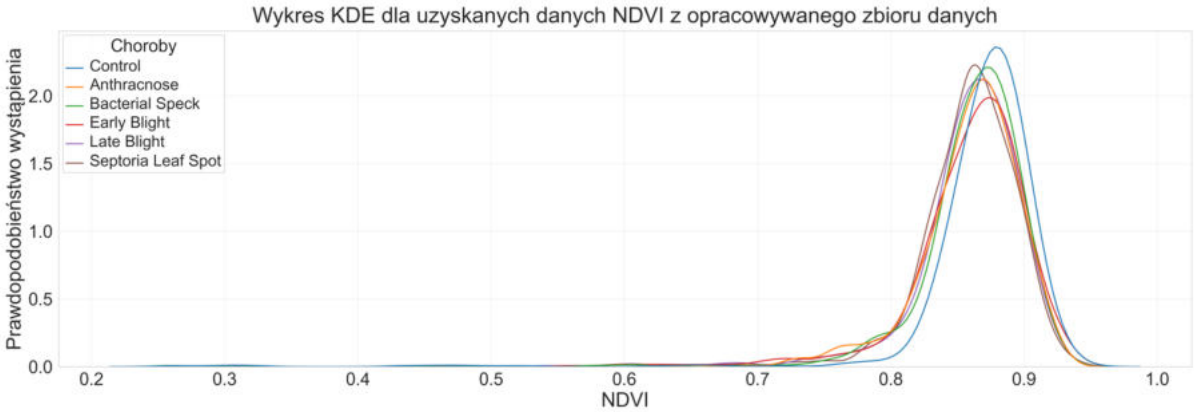
8.5. Wskaźnik NDVI

Istotnym parametrem oceny stanu roślin jest znormalizowany, różnicowy wskaźnik wegetacyjny (NDVI), o którym wspominano już w rozdziale 6.3 (równanie (7.8)). Swoje zastosowanie wskaźnik ten znajduje przede wszystkim w satelitarnych rozwiązaniach wielospektralnych obrazujących przynajmniej w zakresie światła czerwonego i bliskiej podczerwieni. Przyjmuje się, że od poziomu około 0.6 tego wskaźnika mówimy o dobrej, poprawnej wegetacji upraw. Poziom 0.8 oznacza bardzo dobrą wegetację.



rys. 32 Histogram wskaźnika wegetacyjnego wykonanych pomiarów hiperspektralnych

W prowadzonych badaniach starano się uchwycić infekcję rośliny pomimo wysokiego poziomu wskaźnika NDVI, czyli w podstawowych badaniach wielospektralnych roślina wydaje się zdrowa, na co wskazuje m.in. rys. 32.



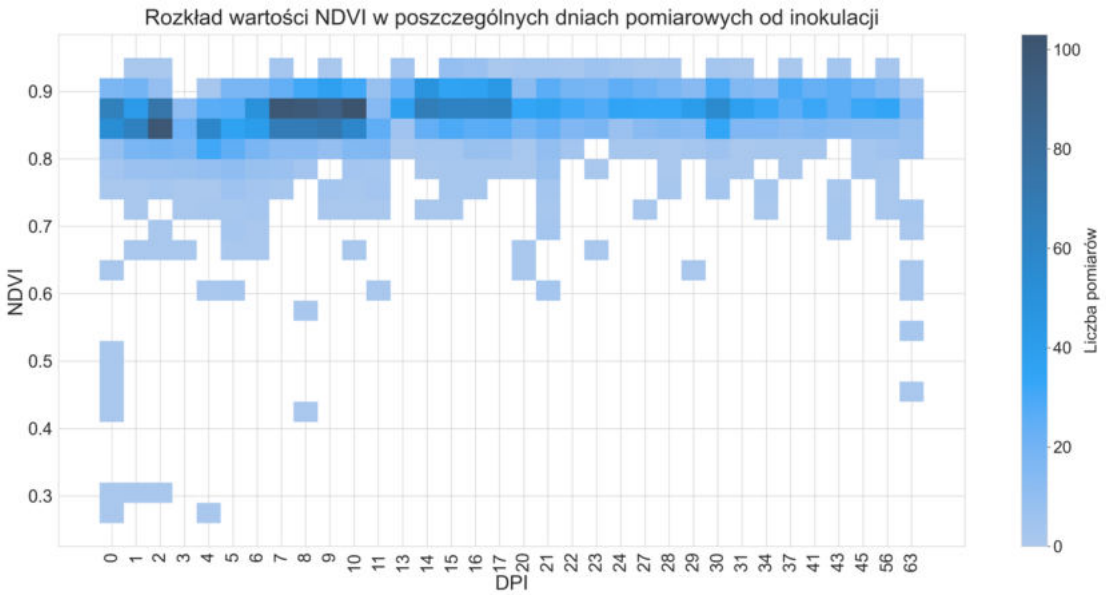
rys. 33 Wykres KDE dla danych NDVI zbioru, czyli rozkładu prawdopodobieństwa wystąpienia NDVI dla losowej wartości w przedziale 0-1 na podstawie wartości NDVI otrzymanej ze zbioru

Po rozbiciu pomiarów na poszczególne grupy badawcze w zależności od choroby, można zauważyć nieznacznie przesunięcie pomiarów kontrolnych (zaznaczonych niebieską linią na rys. 33) w stosunku do jednostek chorobowych, jednak różnica pomiędzy miarami tendencji centralnych tych podzbiorów jest nieznaczna, rzędu od 0,009-0,014 dla mediany oraz 0,009-0,012 dla średniej, co przedstawiono w tabeli tab. 12.

tab. 12 Miary tendencji centralnej wskaźnika NDVI dla poszczególnych grup badawczych, podzielonych wg badanych chorób

Choroba	**Mediana NDVI**	**Średnia NDVI**
Control	0,877	0,869
Anthracnose	0,866	0,860
Bacterial Speck	0,868	0,861
Early Blight	0,867	0,859
Late Blight	0,863	0,857
Septoria Leaf Spot	0,863	0,860

Wyjąwszy nieliczne pomiary odstające, wartości wskaźnika w poszczególnych dniach pomiarowych zawierały się w przeważającej części w zakresie 0.75 do 0.95, co przedstawiono na poniższym wykresie (rys. 34).



rys. 34 Rozkład wartości wskaźnika wegetacyjnego w poszczególnych dniach od inokulacji z uwzględnieniem liczby pomiarów w poszczególnych obszarach

Powyższe wyniki analizy potwierdzają, że rośliny pomimo zakażenia wykazywały dobre lub bardzo dobre zdolności wegetacyjne, co może oznacza, że wskaźnik NDVI nie umożliwia na tym etapie określenia, czy roślina była zakażona.

8.6. Podsumowanie rozdziału

W rozdziale omówiono sposób postępowania z danymi od momentu wykonania pomiarów przez filtrację, wyodrębnienie metadanych, przetwarzania wstępnego danych, aż do stworzenia końcowego zbioru poddanego opisywanym analizom. Przybliżona została struktura zestawu danych oraz objaśnione zostały metadane zawarte dodatkowo w zbiorze. Zaproponowana została metoda podziału zbioru z uwzględnieniem wymiaru DPI (ang. *days post inoculum*) na potrzeby wykonania badań omawianych w niniejszej pracy doktorskiej.

Dodatkowo w rozdziale przedstawiono analizę zbioru danych z podziałem na poszczególne grupy badawcze, uwzględniając najważniejsze parametry dotyczące podziału klasowego zbioru – przynależność do konkretnej jednostki chorobowej (podział na choroby, rys. 28). Przeanalizowano również parametry, które mogą mieć największy wpływ na widmo spektralne badanych obiektów. Powyższe analizy wykazują, że grupy badawcze wykorzystane w dalszych badaniach są równoliczne w ujęciu poszczególnych grup badawczych. Wyjątkiem jest podział na fitotrony ze względu na decyzje podjęte na etapie kolekcjonowania zbioru danych.
Następnie przedstawiono analizę jakościową pomiarów za pomocą wskaźnika wegetacyjnego NDVI, która pokazała, że wyjąwszy nieliczne wartości odstające, badane obiekty znajdowały się w dobrej lub bardzo dobrej kondycji wegetacyjnej, a różnice pomiędzy badanymi grupami nie były znaczące czy pozwalające na wnioskowanie na podstawie tego, popularnego wskaźnika.

9. Wykrywanie chorób z wykorzystaniem wybranych algorytmów uczenia maszynowego

Celem przeprowadzonego eksperymentu badawczego było:

1. Znalezienie najlepszego klasyfikatora pozwalającego na rozróżnieniu obiektów chorych od obiektów zdrowych na podstawie pomiarów hiperspektralnych w 6 wariantach eksperymentu:
    a. obiekty kontrolne vs. obiekty zakażone dowolną z chorób (**AN-SL** z rozdziału 3.1),
    b. obiekty kontrolne – obiekty zakażone chorobą **AN** (*Anthracnose*),
    c. obiekty kontrolne – obiekty zakażone chorobą **BS** (*Bacterial Speck*),
    d. obiekty kontrolne – obiekty zakażone chorobą **EB** (*Early Blight*),
    e. obiekty kontrolne – obiekty zakażone chorobą **LB** (*Late Blight*),
    f. obiekty kontrolne – obiekty zakażone chorobą **SL** (*Septoria Leaf Spot*).
2. Opracowanie metody postępowania z danymi, tzn. wskazanie takiej procedury przetwarzania danych, która skutkuje najwyższą poprawą wyników klasyfikacji binarnej,
3. Weryfikacja hipotezy, iż wybrane pasma spektralne są szczególnie istotne w procesie wykrywaniu analizowanych chorób.

Przegląd literatury dotyczący omawianych w niniejszym rozdziale zagadnień, a przede wszystkim wybranie najistotniejszych zakresów spektralnych wynikających z literatury przedstawiono w rozdziale 4.6.

9.1. Wstępna selekcja algorytmów klasyfikacyjnych

W ramach przeprowadzonych prac badawczych wykonano serię eksperymentów w poszukiwaniu rozwiązania pozwalającego na rozwiązanie problemu klasyfikacji binarnej, tj. określenia, czy badany obiekt jest obiektem zdrowym, czy chorym. Szerzej badanie opisane zostało w materiałach dodatkowych zawartych w Dodatku A – Dziennik eksperymentów (rozdział 14.1, pkt. 2). Pełna lista klasyfikatorów (algorytmów uczenia maszynowego) wykorzystanych w badaniu została przedstawiona w rozdziale 7.1.1.

Każdy z modeli został wytrenowany na tym samym zestawie danych, omówionym wcześniej w rozdziale 8.
Poniżej przedstawiono wyniki dla trzech modeli, które osiągnęły najlepsze wyniki, a jednocześnie nie uległy zjawisku przetrenowania (*ang. overfitting*) dla każdej z wysokości pomiarowej przed zbalansowaniem zbioru treningowego i po jego zbalansowaniu.

tab. 13 Zestawienie wyników najlepszych klasyfikatorów (nieprzetrenowanych) w zależności od wysokości wykonywanych pomiarów oraz poziomu zbalansowania zbioru

Wysokość wykonania pomiaru	**Wyniki dla zbioru niezbalansowanego**		**Wyniki dla zbioru zbalansowanego**	
5 cm	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	LinearSVC	0.934	LinearSVC	0.848
	LogisticRegerssion	0.913	CalibratedClassifierCV	0.848
	CalibratedClassiefierCV	0.908	LogisticRegression	0.843
30 cm	**klasyfikator**	**F1**	**klasyfikator**	**F1**

	RidgeClassifier	0.881	LinearSVC	0.816
	LinearSVC	0.877	CalibratedClassifierCV	0.806
	LogisticRegression	0.870	LogisticRegression	0.806
60 cm	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	RidgeClassifier	0.835	CalibratedClassifierCV	0.738
	LogisticRegression	0.825	RidgeClassifier	0.714
	LinearSVC	0.817	LinearSVC	0.711
uśredniony wynik dla wszystkich wysokości	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	LinearSVC	0.876	CalibratedClassifierCV	0.797
	RidgeClassifier	0.872	LinearSVC	0.792
	LogisticRegression	0.869	LogisticRegression	0.784
wszystkie wysokości jednocześnie	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	LinearSVC	0.875	CalibratedClassifierCV	0.823
	RidgeClassifierCV	0.862	RidgeClassifier	0.804
	LinearDiscriminantAnalysis	0.859	RidgeClassifierCV	0.800

Poniżej przedstawiono wyniki dla trzech modeli, które osiągnęły najlepsze wyniki, a jednocześnie uległy zjawisku przetrenowania dla każdej z wysokości pomiarowej przed zbalansowaniem zbioru treningowego i po jego zbalansowaniu.

tab. 14 Zestawienie wyników najlepszych klasyfikatorów (przetrenowanych) w zależności od wysokości wykonywanych pomiarów oraz poziomu zbalansowania zbioru

Wysokość wykonania pomiaru	**Wyniki dla zbioru niezbalansowanego**		**Wyniki dla zbioru zbalansowanego**	
5 cm	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	XGBClassifier	0.820	RandomForestClassifier	0.730
	LGBMClassifier	0.817	LGBMClassifier	0.716
	ExtraTreesClassifier	0.805	ExtraTreesClassifier	0.712
30 cm	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	XGBClassifier	0.802	LGBMClassifier	0.691
	LGBMClassifier	0.792	XGBClassifier	0.687
	ExtraTreesClassifier	0.787	RandomForestClassifier	0.653
60 cm	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	XGBClassifier	0.781	XGBClassifier	0.686
	LGBMClassifier	0.778	LGBMClassifier	0.654
	ExtraTreesClassifier	0.777	RandomForestClassifier	0.619
uśredniony wynik dla wszystkich wysokości	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	XGBClassifier	0.801	LGBMClassifier	0.687
	LGBMClassifier	0.796	XGBClassifier	0.685
	ExtraTreesClassifier	0.790	RandomForestClassifier	0.667
wszystkie wysokości jednocześnie	**klasyfikator**	**F1**	**klasyfikator**	**F1**
	XGBClassifier	0.802	XGBClassifier	0.676
	ExtraTreesClassifier	0.784	ExtraTreesClassifier	0.637
	RandomForestClassifier	0.778	RandomForestClassifier	0.634

Wartości pomiarowe reflektancji obiektów różnią się od siebie w zależności od wysokości wykonanego pomiaru, dlatego do dalszych badań wykorzystane zostały pomiary wykonane z wysokości 5 cm nad badanym obiektem, o czym wspomniano również przy okazji omawiania zbiorów danych w rozdziale 8.1.

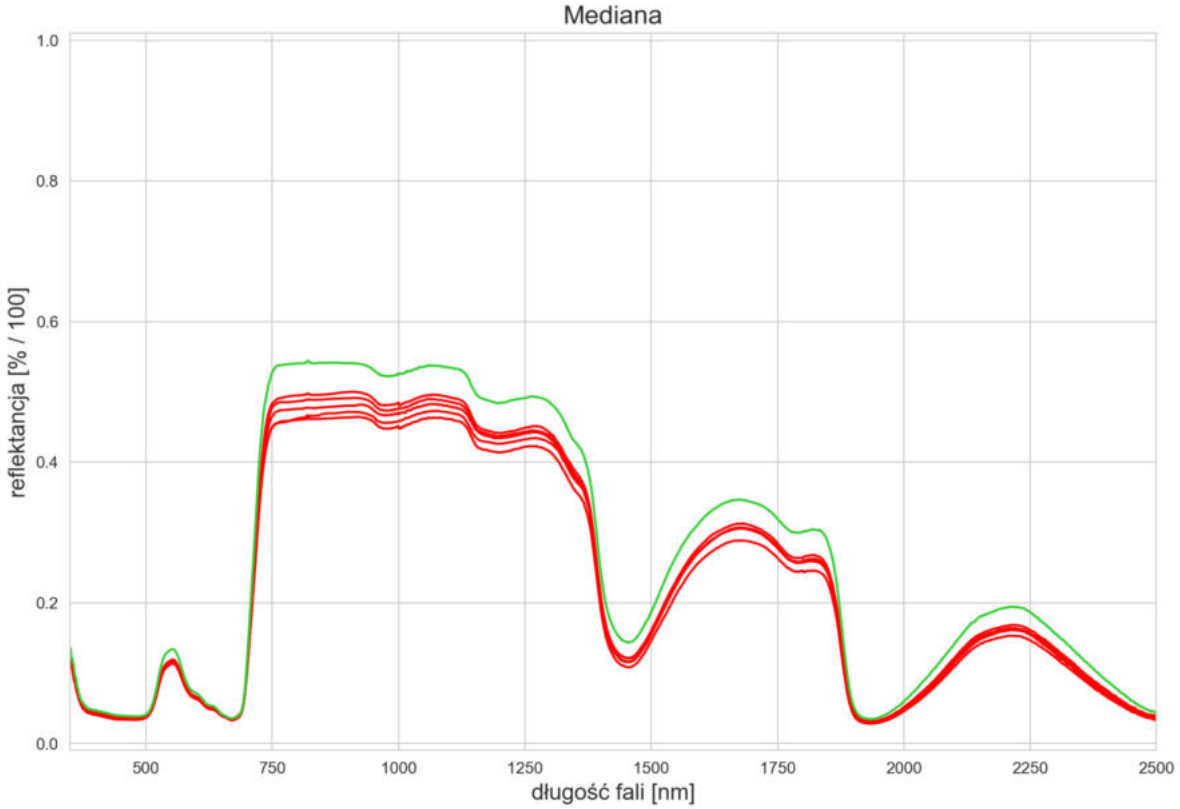
W ramach badania wykazano, że modele liniowe (m.n. Linear SVC, Logistic Regression.itp.) nie ulegają efektowi przetrenowania, a modele bardziej złożone (np. XGBoost, RandomForest, etc.) ulegają wspomnianemu zjawisku.
Do dalszych badań zdecydowano się na wybór następujących klasyfikatorów:

1. Linear SVC,
2. Logistic Regression,
3. Ridge Classifier,
4. Random Forest,
5. LGBMClassifier,
6. XGBClassifier.

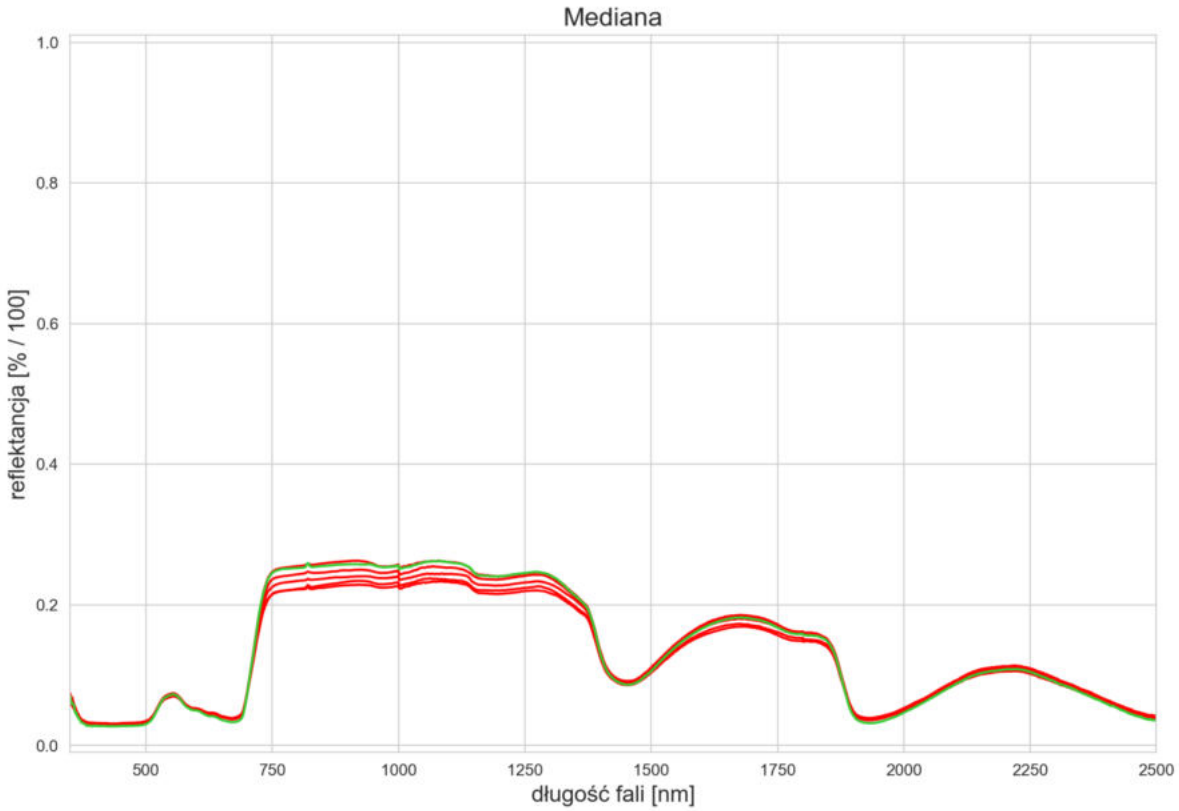
Klasfykatory 1-3 zostały wybrane ze względu na osiągnięcie najlepszego wyniku wśród algorytmów nie ulegających zjawisku przetrenowania, natomiast klasyfikatory 4-6 osiągnęły najlepszy wynik wśród algorytmów ulegających przetrenowaniu.

9.2. Klasyfikacja binarna

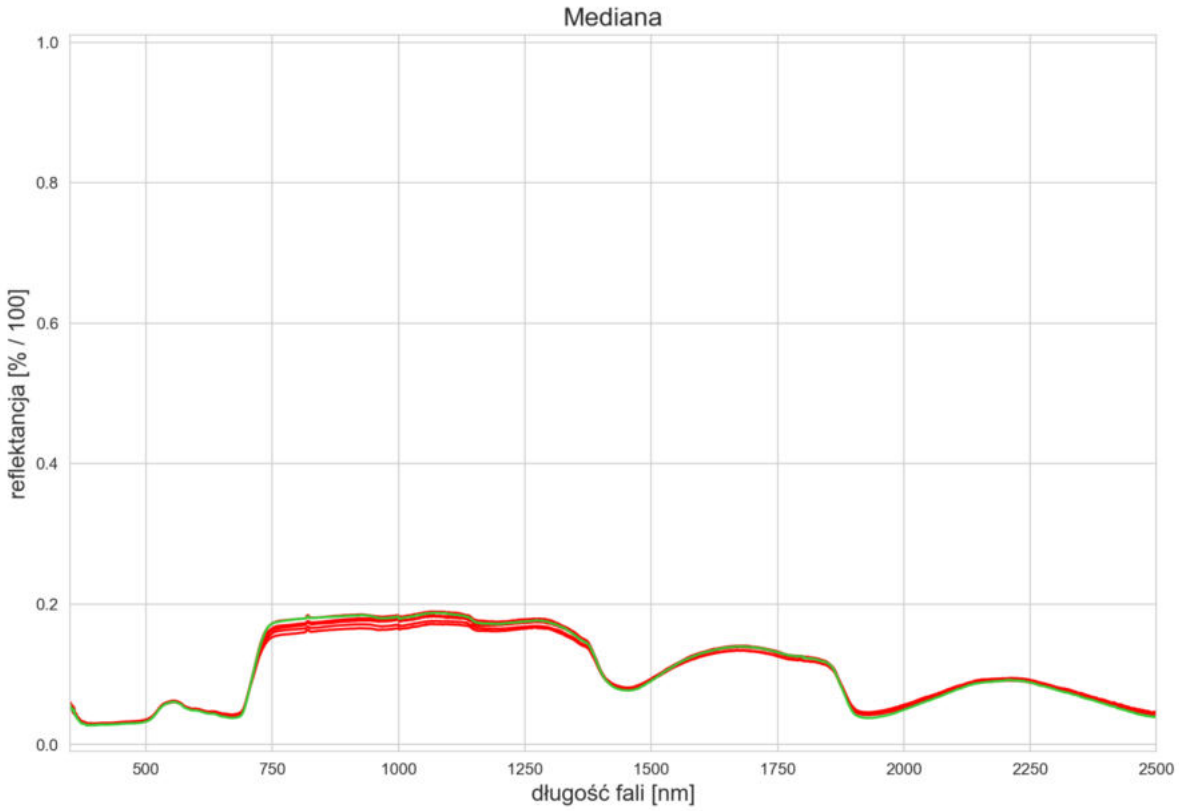
Pierwszym krokiem w klasyfikacji binarnej było przetworzenie zbioru danych, co opisane zostało w rozdziale 8. Zbiór źródłowy zawierał pomiary hiperspektralne wykonane za pomocą spektrometru ASD FieldSpeck 4 Hi-Res. Pomiary te określają reflektancję badanego obiektu, czyli stosunek mocy światła odbitego do mocy światła padającego na badany obiekt, w tym przypadku liść. Reflektancja jest wyrażana w wartościach od 0.0 do 1.0, gdzie 1.0 stanowi pełną reflektancję określoną w procesie kalibracji spektrometru [rozdział 6.4 Pomiary hiperspektralne].



rys. 35 Mediana widma pomiarów dla wysokości 5 cm

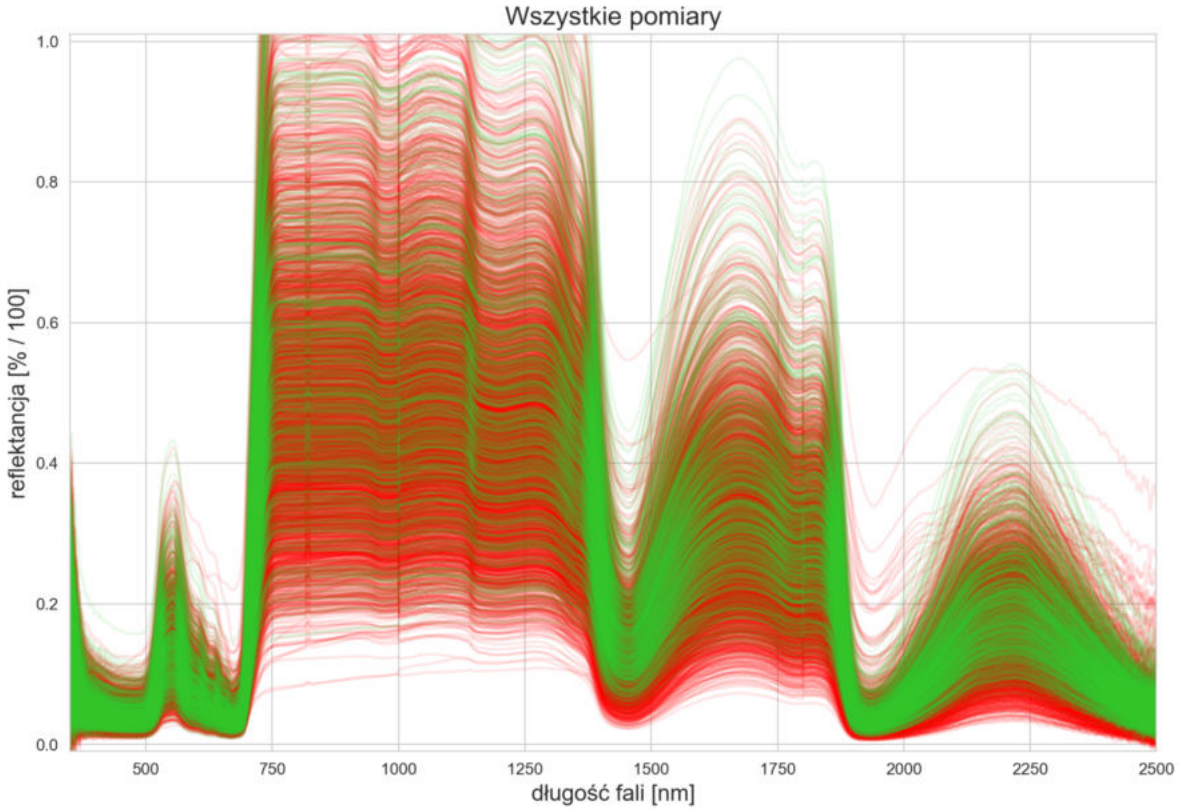


rys. 36 Mediana widma pomiarów dla wysokości 30 cm

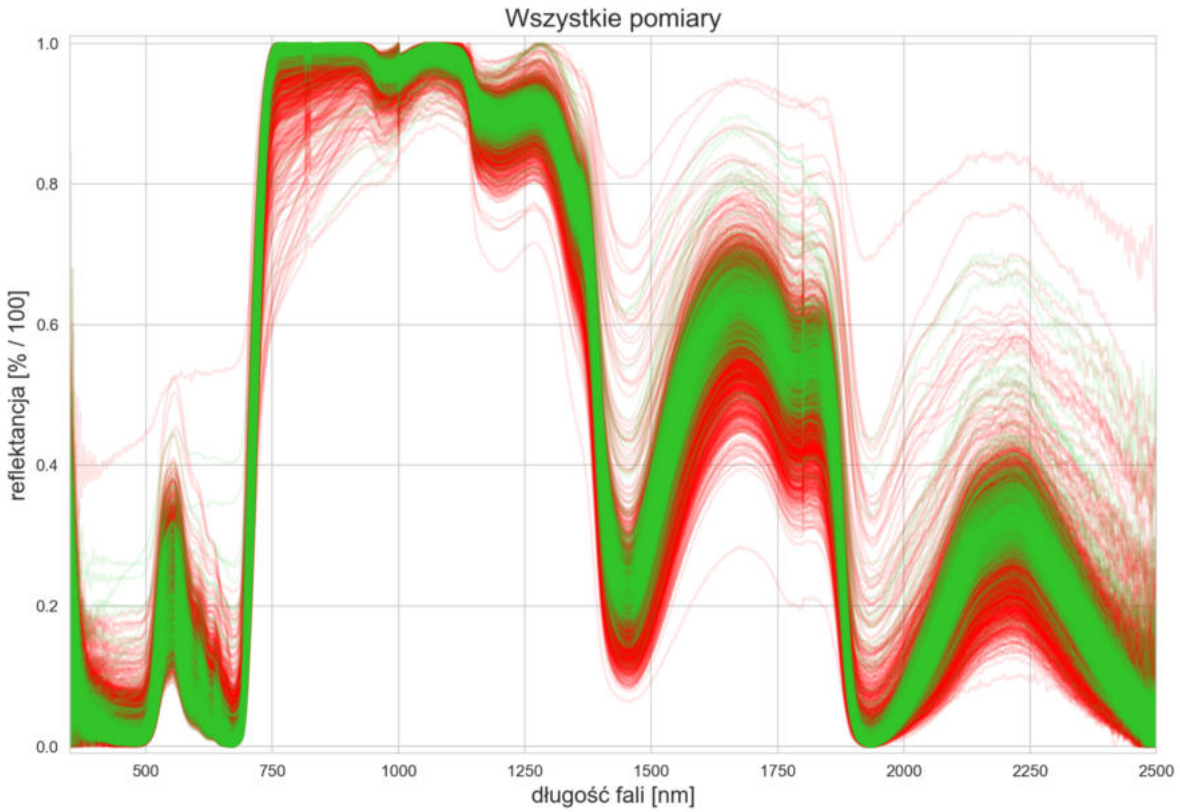


rys. 37 Mediana widma pomiarów dla wysokości 60 cm

Rozproszenie światła, a także parametry światłowodu spektrometru powodują, że pomiary tych samych obiektów różnią się od siebie w zależności od odległości dzielącej koniec światłowodu od obiektu badanego [rys. 35, rys. 37]. Dodatkowo, ponieważ wartość 1.0 reprezentuje pełną reflektancję panelu kalibracyjnego, możliwe jest uzyskanie wartości pomiarowych z zakresu przekraczającego wartość 1.0 [rys. 38].



rys. 38 Widmo wszystkich pomiarów dla wysokości 5 cm. Zielonym kolorem oznaczono pomiary kontrolny, czerwonym - obiekty zakażone



rys. 39 Widmo wszystkich pomiarów dla wysokości 5 cm poddanych normalizacji. Zielonym kolorem oznaczono pomiary kontrolny, czewonym - obiekty zakażone

Zdecydowano się na kilkuetapowe podejście do analizowanych danych. W pierwszej kolejności wykonano klasyfikację dla zbioru bazowego, czyli pomiarów uzyskanych za pomocą spektrometru bez dodatkowego przetworzenia [ETAP 1 na rys. 40]. W kolejnym kroku wykonano normalizację zbioru i ponownie wykonano klasyfikację z wykorzystaniem uprzednio wybranych metod [ETAP 2, rys. 40].

Następnie, z uwagi na wskazanie przez dystrybutora sprzętu pomiarowego wysokiej, trudnej do określenia niepewności pomiaru w początkowym (350-475 nm) i końcowym (2400-2500nm) zakresie spektrum pomiarowego, zdecydowano się na usunięcie tego zakresu ze zbioru danych i ponownie przeprowadzono proces klasyfikacji [ETAP 3, rys. 40].

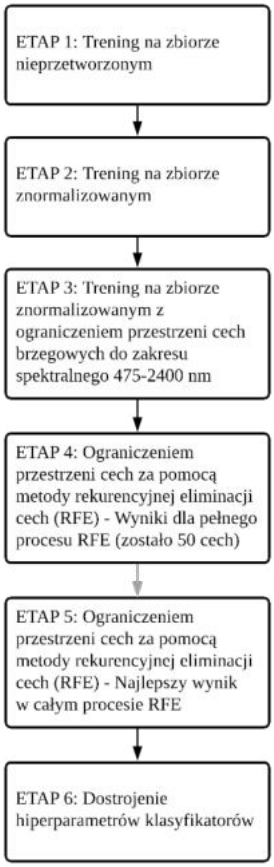
Kolejnym krokiem było ograniczenie przestrzeni cech potrzebnych w procesie klasyfikacji. W tym celu zastosowano metodę rekurencyjnej eliminacji cech (ang. *Recursive Feature Elimination*, RFE). Krok wykonano w dwóch wariantach. Wariant pierwsze [ETAP 4, rys. 40] polegał na wykonaniu RFE do ograniczenia liczby cech do 50 pasm z eliminacją 50 cech w każdym kolejnym kroku eliminacyjnym. W efekcie wybranych zostało 50 cech spektralnych (pasm), które wzięły udział w końcowej klasyfikacji, dla której obliczono miarę F1.

Wariant drugi [ETAP 5, rys. 40] polegał na eliminacji cech w sposób dynamiczny, tzn. przeprowadzono pełny proces RFE, jak w etapie 4, jednak dla każdego z kroków eliminacji określano miarę F1. Na podstawie miary F1 określono najlepszy etap eliminacji dla każdego z rozważanych przypadków klasyfikacyjnych. Podczas wyznaczania miary F1 wykorzystano metodę walidacji krzyżowej.

Szczegóły nt. sposobu wyboru cech opisano w rozdziale 9.3 Analiza udziału poszczególnych zakresów spektralnych w klasyfikacji binarnej chorób.

Ostatni etap [ETAP 6, rys. 40] polegał na dostrojeniu hiperparametrów poszczególnych klasyfikatorów dla najlepszego kroku z etapu 5 poprzez zastosowanie metody losowego wyszukania optymalnych hiperparametrów z zastosowaniem walidacji krzyżowej.

Schemat postępowania w poniższym eksperymencie:

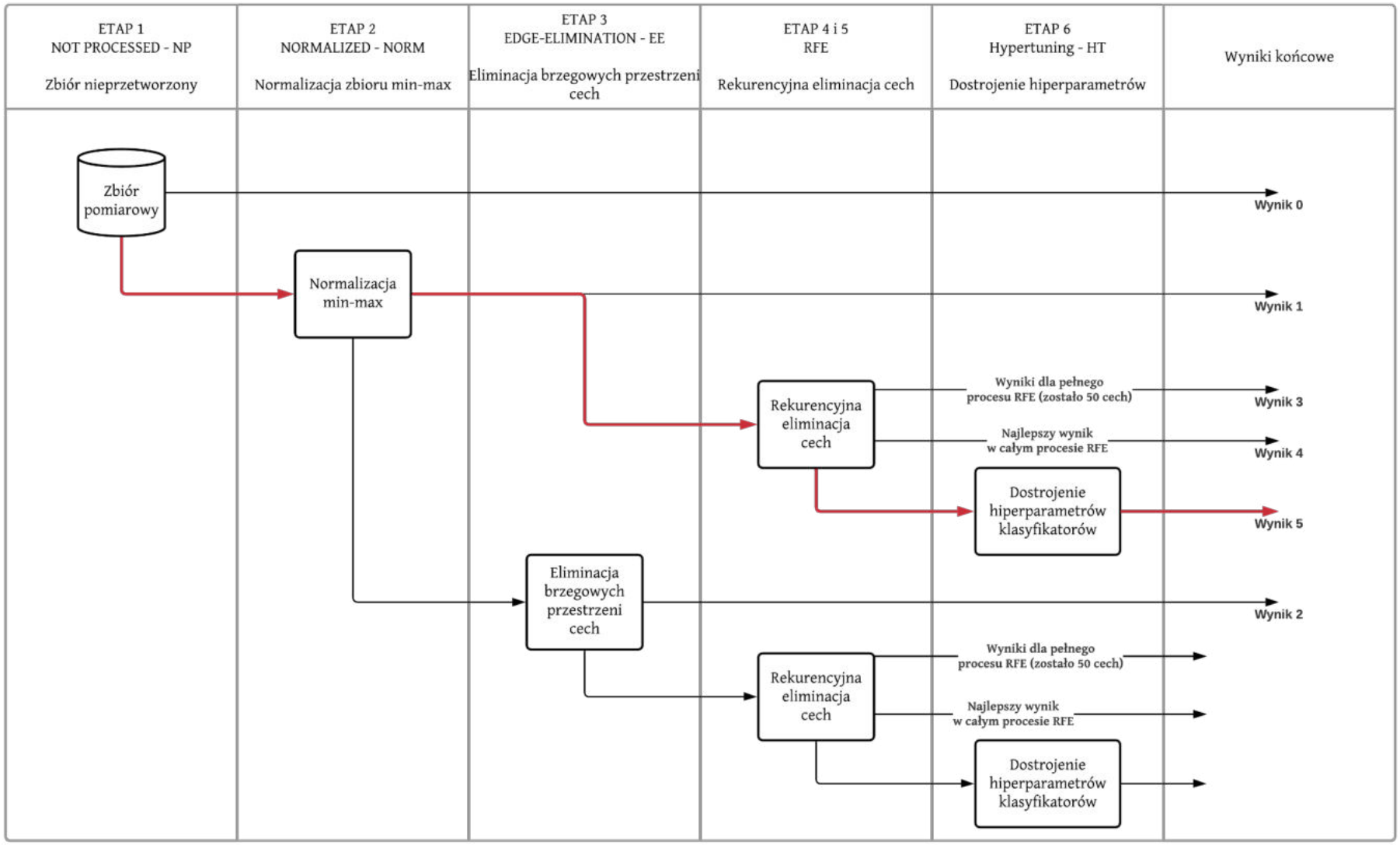


rys. 40 Schemat planowanych etapów eksperymentu

Prace badawcze w poszczególnych etapach obejmowały:

1. dla etapu 1: trening na surowym zbiorze danych, bez dodatkowego przetworzenia pomiarów,
2. dla etapu 2: trening na zbiorze znormalizowanym, dla którego zastosowano normalizację min-max dla każdego z pomiarów,
3. dla etapu 3: trening na zbiorze znormalizowanym, z usunięciem skrajnych przestrzeni spektralnych, które w przypadku pomiarów za pomocą wykorzystanego spektroradiometru wykazują się największym zaszumieniem sygnału,
4. dla etapu 4 i 5: wykorzystanie metody rekurencyjnej eliminacji cech w celu ograniczenia badanej przestrzeni spektralnej przy jednoczesnym uzyskaniu możliwie najwyższego wyniku – w etapie 4 przeprowadzano pełny proces eliminacji cech, natomiast w etapie 5 wybierano poziom eliminacyjny, dla którego uzyskano najlepsze wyniki,
5. dla etapu 6: dostrojenie hiperparametrów poszczególnych klasyfikatorów z kroku 5 poprzez zaimplementowanie metody przeszukania losowego z wykorzystaniem walidacji krzyżowej.

Szczegółowo działania wykonane w ramach etapu 4 i 5 opisano w rozdziale 9.3 Analiza udziału poszczególnych zakresów spektralnych w klasyfikacji binarnej chorób.
Opisane powyżej etapy zaimplementowano w środowisku PyCharm 2021.2.2 (Professional Edition). Opracowane oprogramowanie pozwoliło na wygenerowanie serii wyników dla wielu różnych wariantów. Schemat postępowania w opisanym eksperymencie przedstawiono na rys. 41.



rys. 41 Kolejne etapy eksperymentu badawczego. Kolorem czerwonym oznaczono rekomendowaną procedurę przetwarzania danych

Zdecydowano się na analizę wyników osiąganych dla wszystkich przedstawionych na diagramie wariantów, a uzyskane w efekcie końcowym wyniki przedstawiono w dodatku A. Duża liczba wariantów wynika z faktu, że zdecydowano się na analizę wyników dla każdego opisanego etapu dla różnych klasyfikatorów oraz różnych chorób.

W kolejnym rozdziale przedstawiono wybrane wyniki (Wynik 0 – Wynik 5), które doprowadziły do wyznaczenia procedury, która skutkowała osiąganiem najlepszych rezultatów przez trenowane modele. Ta procedura, na diagramie rys. 41, została oznaczona kolorem czerwonym.

W tabelach prezentujących wyniki badań w kolejnych rozdziałach, zastosowano następujące skróty dla nazw klasyfikatorów:

1. LGBM – Light Gradient Boosting Machine,
2. LSVC – Linear SVC – Linear Support Vector Classification,
3. LR – Logistic Regression – Regresja Liniowa,
4. RF – Random Forest – Las losowy,
5. RC – Ridge Classifier – Klasyfikator Grzbietowy,
6. XGB – XGBoost – Wzmocnienie Gradientowe.

Poszczególne wyniki przedstawione w tabelach odpowiadają czynnościom, które wykonywane były z poszczególnych ścieżek z diagramu rys. 41:

- Wynik 0 – czysty zbiór danych,
- Wynik 1 – czysty zbiór danych > normalizacja min-max,
- Wynik 2 – czysty zbiór danych > eliminacja brzegowych reflektancji,
- Wynik 3 – czysty zbiór danych > normalizacja min-max > wynik dla pełnego procesu rekurencyjnej eliminacji cech,
- Wynik 4 – czysty zbiór danych > normalizacja min-max > najlepszy wynik uzyskany w procesie rekurencyjnej eliminacji cech,
- Wynik 5 – czysty zbiór danych > normalizacja min-max > najlepszy wynik uzyskany w procesie rekurencyjnej eliminacji cech > dostrojenie hiperparametrów klasyfikatora.

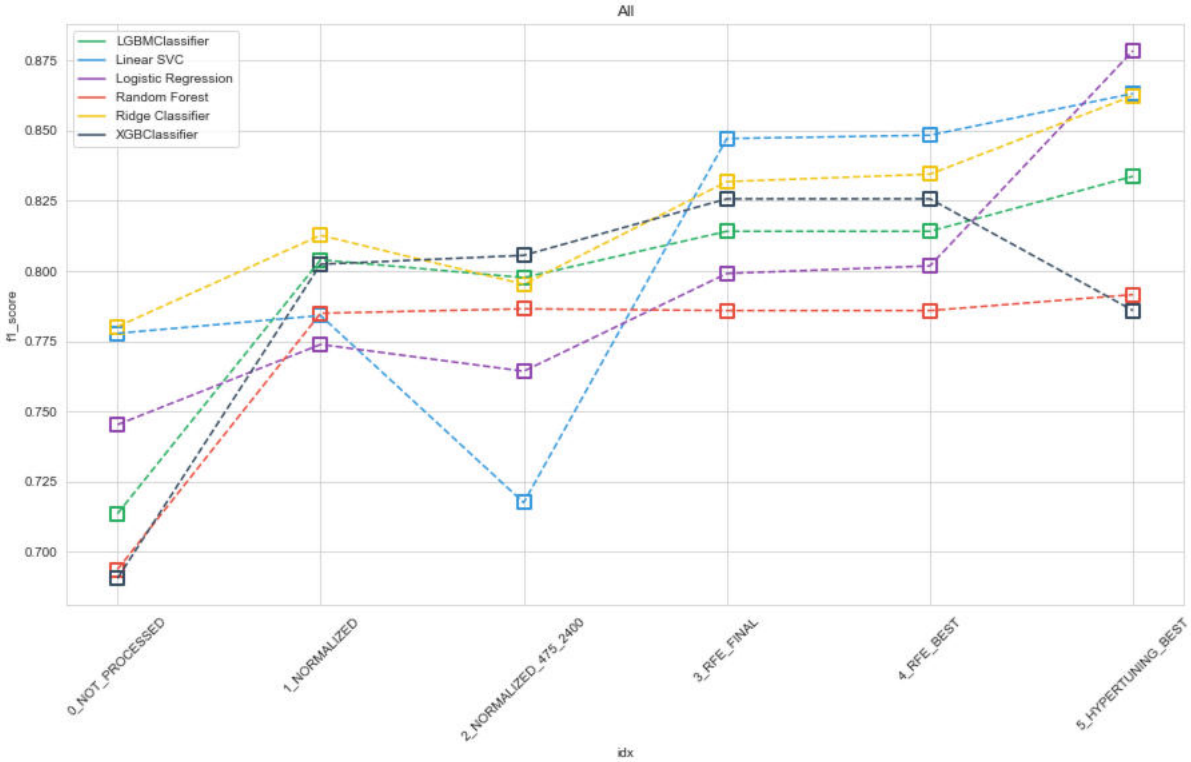
Wyniki przedstawione w tabelach to wartości wskaźnika F1 uzyskane podczas badań. Wskaźnik ten omówiony został w rozdziale 7.1.4.

9.2.1. Wszystkie choroby

W niniejszym podpunkcie opisano wyniki klasyfikacji binarnej rozróżniającej rośliny zdrowe od chorych niezależnie od choroby.

tab. 15 Wyniki treningu klasyfikatorów na różnych etapach metody dla pełnego zbioru danych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

	LGBM	LSVC	LR	RF	RC	XGB
Wynik 0	0,713	0,778	0,745	0,694	0,78	0,69
Wynik 1	0,804	0,784	0,774	0,785	0,813	0,802
Wynik 2	0,798	0,717	0,764	0,787	0,795	0,806
Wynik 3	0,814	0,847	0,799	0,786	0,832	0,826
Wynik 4	0,814	0,848	0,802	0,786	0,834	0,826
Wynik 5	0,834	0,863	0,879	0,792	0,862	0,786



rys. 42 Wyniki treningu klasyfikatorów na różnych etapach metody dla pełnego zbioru danych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

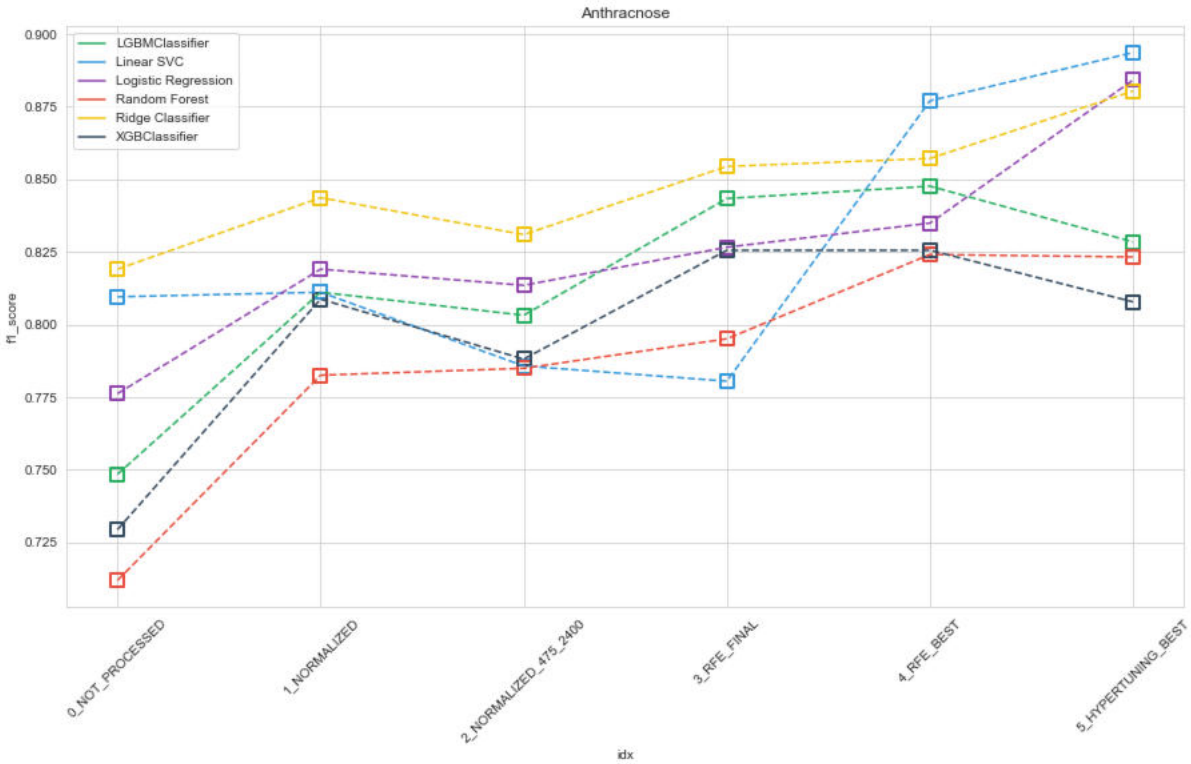
Najniższy wynik wskaźnika F1 uzyskał klasyfikator XGB wytrenowany na zbiorze nieprzetworzonym (0,690).
Najwyższy wynik wskaźnika F1 uzyskał klasyfikator Logistic Regression po dostrojeniu hiperparametrów (0,879).
W 5 przypadkach na 6 wykorzystanie zbioru nieprzetworzonego dało najgorsze rezultaty (0,69-0,78). W 6 przypadkach na 6 zastosowanie zaproponowanego procesu przetwarzania dało wynik bardzo zbliżony lub wyższy od 0,8 (0,799-0,883).

9.2.2. Anthracnose (AN)

Poniższy punkt przedstawia wyniki dla klasyfikacji binarnej odróżniającej rośliny zdrowe od choroby Anthracnose.

tab. 16 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Antracnose" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

	LGBM	LSVC	LR	RF	RC	XGB
Wynik 0	0,748	0,81	0,776	0,712	0,819	0,729
Wynik 1	0,811	0,811	0,819	0,783	0,844	0,809
Wynik 2	0,803	0,786	0,813	0,785	0,831	0,788
Wynik 3	0,843	0,78	0,827	0,795	0,854	0,826
Wynik 4	0,848	0,877	0,835	0,824	0,857	0,826
Wynik 5	0,828	0,894	0,884	0,823	0,88	0,808



rys. 43 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Antracnose" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

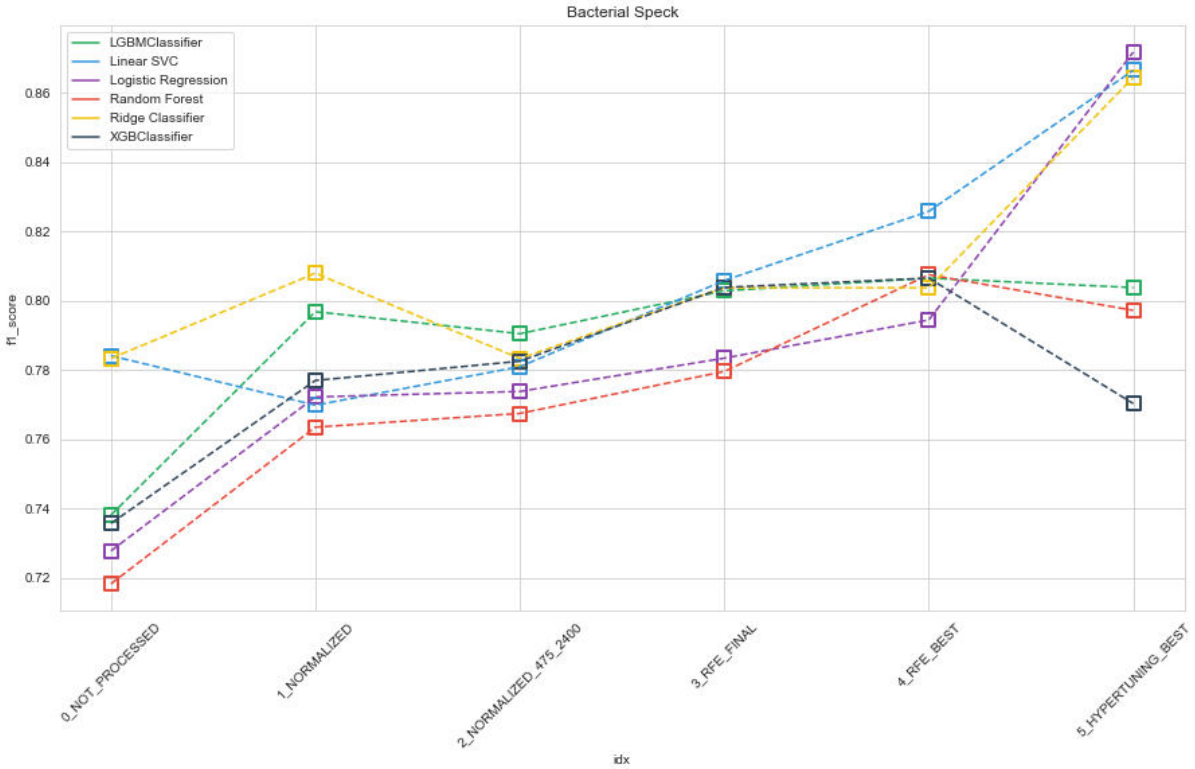
Najniższy wynik wskaźnika F1 uzyskał klasyfikator Random Forest wytrenowany na zbiorze nieprzetworzonym (0,712).
Najwyższy wynik wskaźnika F1 uzyskał klasyfikator Linear SVC po dostrojeniu hiperparametrów (0,894).
W 5 przypadkach na 6 wykorzystanie zbioru nieprzetworzonego dało najgorsze rezultaty (0,712-0,819). W 6 przypadkach na 6 zastosowanie zaproponowanego procesu przetwarzania dało wynik wyższy od 0,8 (0,808-0,894).

9.2.3. Bacterial Speck (BS)

Poniższy punkt przedstawia wyniki dla klasyfikacji binarnej odróżniającej rośliny zdrowe od choroby Bacterial Speck.

tab. 17 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Bacterial Speck" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

	LGBM	LSVC	LR	RF	RC	XGB
Wynik 0	0,738	0,784	0,728	0,718	0,783	0,736
Wynik 1	0,797	0,77	0,772	0,763	0,808	0,777
Wynik 2	0,79	0,781	0,774	0,767	0,783	0,783
Wynik 3	0,803	0,806	0,783	0,78	0,804	0,804
Wynik 4	0,807	0,826	0,794	0,808	0,804	0,807
Wynik 5	0,804	0,867	0,872	0,797	0,864	0,77



rys. 44 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Bacterial Speck" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

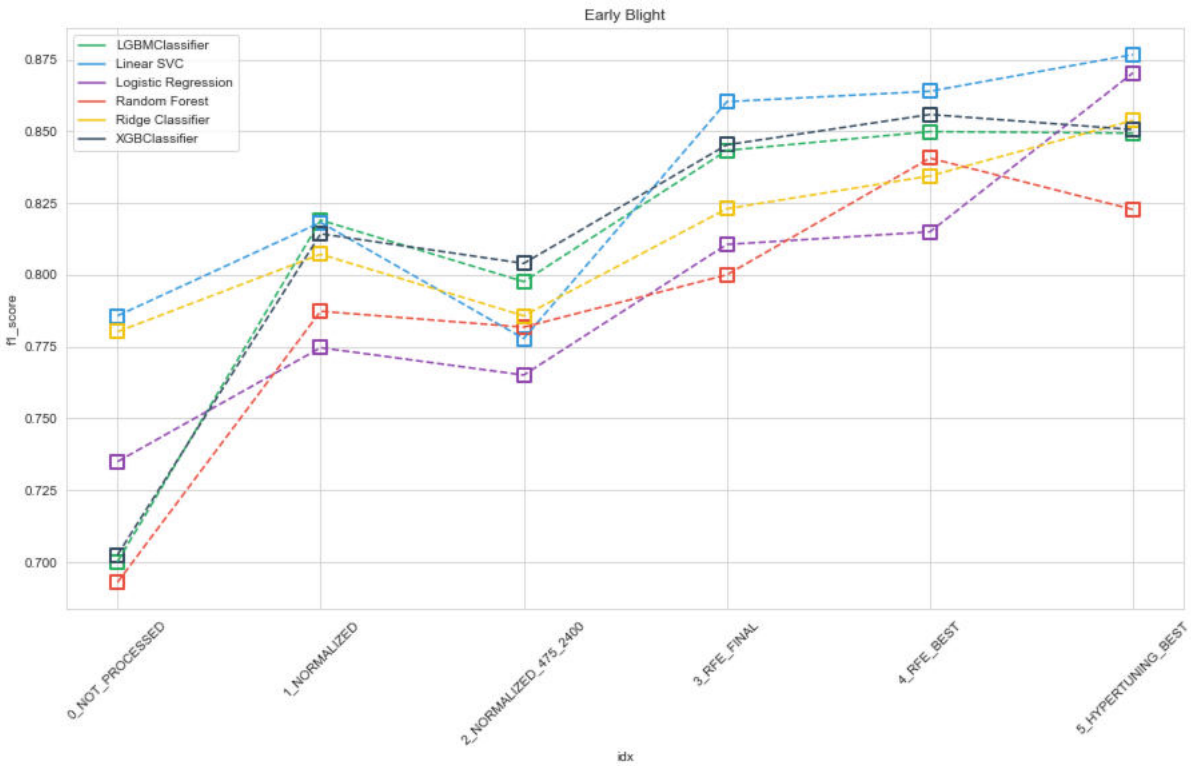
Najniższy wynik wskaźnika F1 uzyskał klasyfikator Random Forest wytrenowany na zbiorze nieprzetworzonym (0,718).
Najwyższy wynik wskaźnika F1 uzyskał klasyfikator Logistic Regression po dostrojeniu hiperparametrów (0,872).
W 5 przypadkach na 6 wykorzystanie zbioru nieprzetworzonego dało najgorsze rezultaty (0,718-0,783). W przypadku klasyfikatora XGBoost dostrojenie hiperparametrów (Etap 5) dało znacznie gorsze rezultaty niż w Etapie 4 (spadek z 0,807 do 0,77). W 5 przypadkach na 6 zastosowanie zaproponowanego procesu przetwarzania dało wynik bardzo zbliżony lub wyższy od 0,8 (0,797-0,872).

9.2.4. Early Blight (EB)

Poniższy punkt przedstawia wyniki dla klasyfikacji binarnej odróżniającej rośliny zdrowe od choroby Early Blight.

tab. 18 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Early Blight" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

	LGBM	LSVC	LR	RF	RC	XGB
Wynik 0	0,7	0,786	0,735	0,693	0,78	0,702
Wynik 1	0,819	0,818	0,775	0,787	0,807	0,814
Wynik 2	0,798	0,778	0,765	0,782	0,786	0,804
Wynik 3	0,843	0,86	0,811	0,8	0,823	0,845
Wynik 4	0,85	0,864	0,815	0,841	0,834	0,856
Wynik 5	0,849	0,877	0,87	0,823	0,854	0,851



rys. 45 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Early Blight" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

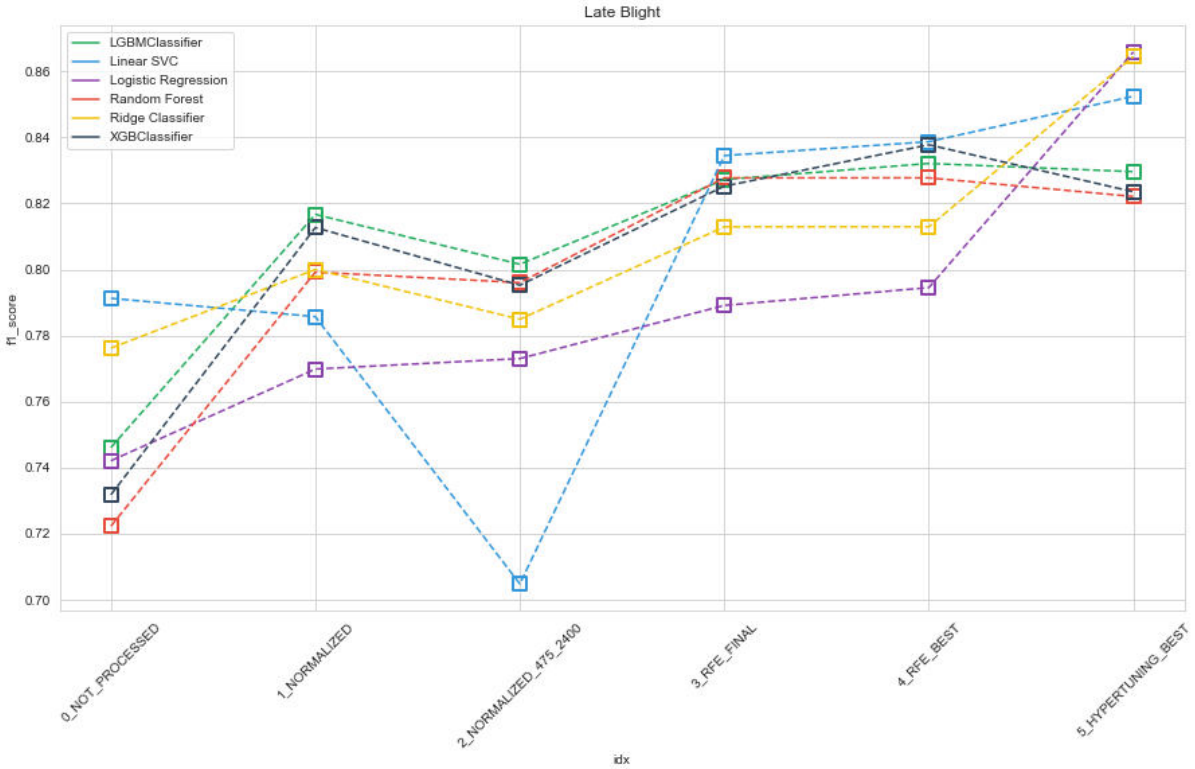
Najniższy wynik wskaźnika F1 uzyskał klasyfikator Random Forest wytrenowany na zbiorze nieprzetworzonym (0,693).
Najwyższy wynik wskaźnika F1 uzyskał klasyfikator Linear SVC po dostrojeniu hiperparametrów (0,877).
W 5 przypadkach na 6 wykorzystanie zbioru nieprzetworzonego dało najgorsze rezultaty (0,693-0,786). W 6 przypadkach na 6 zastosowanie zaproponowanego procesu przetwarzania dało wynik wyższy od 0,8 (0,823-0,877).

9.2.5. Late Blight (LB)

Poniższy punkt przedstawia wyniki dla klasyfikacji binarnej odróżniającej rośliny zdrowe od choroby Late Blight.

tab. 19 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Late Blight" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

	LGBM	LSVC	LR	RF	RC	XGB
Wynik 0	0,746	0,791	0,742	0,722	0,776	0,732
Wynik 1	0,817	0,786	0,77	0,799	0,8	0,813
Wynik 2	0,802	0,705	0,773	0,796	0,785	0,795
Wynik 3	0,827	0,834	0,789	0,828	0,813	0,825
Wynik 4	0,832	0,839	0,795	0,828	0,813	0,838
Wynik 5	0,83	0,852	0,866	0,822	0,864	0,824



rys. 46 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Late Blight" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

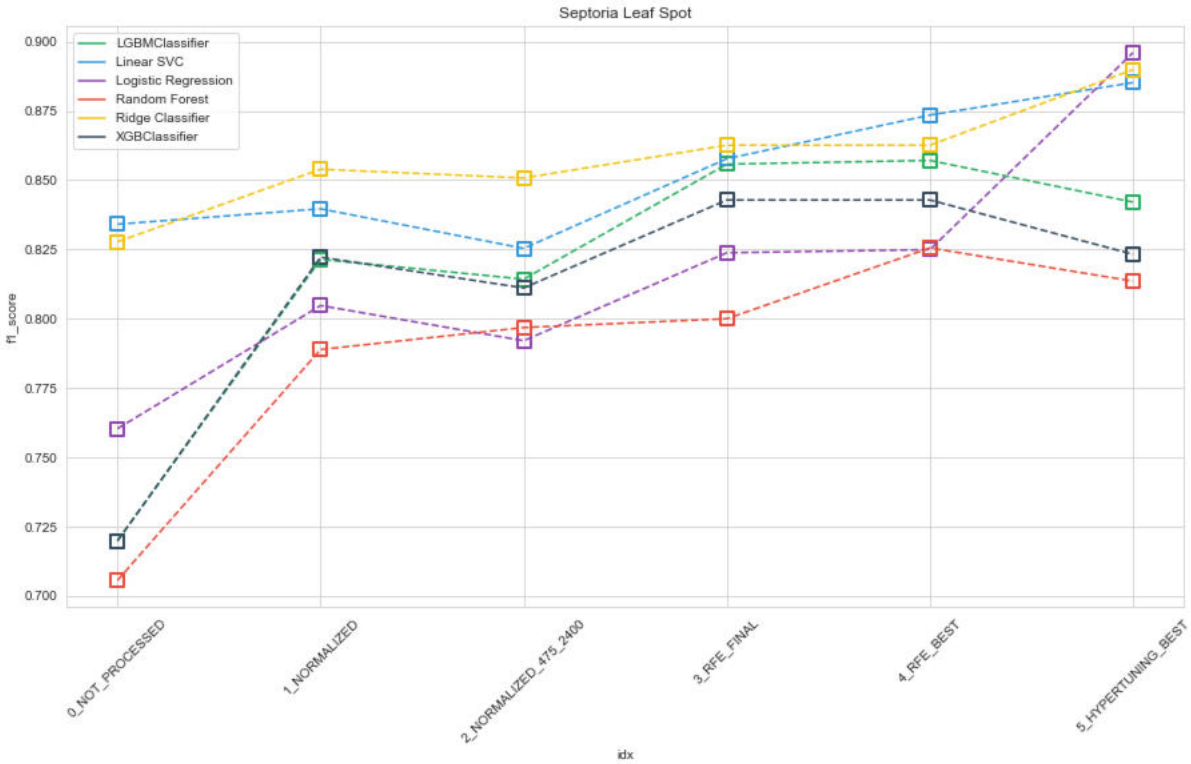
Najniższy wynik wskaźnika F1 uzyskał klasyfikator Linear SVC wytrenowany na zbiorze o skróconym zakresie spektralnym 475-2400 (0,705).
Najwyższy wynik wskaźnika F1 uzyskał klasyfikator Logistic Regression po dostrojeniu hiperparametrów (0,866).
W 5 przypadkach na 6 wykorzystanie zbioru nieprzetworzonego dało najgorsze rezultaty (0,722-0,776). W przypadku klasyfikatora Linear SVC zaobserwowano znaczy spadek metryki po wykonaniu normalizacji i ograniczeniu zakresu spektralnego – spadek z F1=0,791 dla Etapu 1 do F1=0,705 dla Etapu 3. W 6 przypadkach na 6 zastosowanie zaproponowanego procesu przetwarzania dało wynik wyższy od 0,8 (0,822-0,866).

9.2.6. Septoria Leaf Spot (SL)

Poniższy punkt przedstawia wyniki dla klasyfikacji binarnej odróżniającej rośliny zdrowe od choroby Septoria Leaf Spot.

tab. 20 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Septoria Leaf Spot" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

	LGBM	LSVC	LR	RF	RC	XGB
Wynik 0	0,720	0,834	0,760	0,706	0,828	0,720
Wynik 1	0,821	0,840	0,805	0,789	0,854	0,822
Wynik 2	0,814	0,825	0,792	0,797	0,851	0,811
Wynik 3	0,856	0,858	0,824	0,800	0,863	0,843
Wynik 4	0,857	0,874	0,825	0,826	0,863	0,843
Wynik 5	0,842	0,885	0,896	0,814	0,890	0,823



rys. 47 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Septoria Leaf Spot" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym

Najniższy wynik wskaźnika F1 uzyskał klasyfikator Random Forest wytrenowany na zbiorze nieprzetworzonym (0,706).
Najwyższy wynik wskaźnika F1 uzyskał klasyfikator Logistic Regression po dostrojeniu hiperparametrów (0,896).
W 5 przypadkach na 6 wykorzystanie zbioru nieprzetworzonego dało najgorsze rezultaty (0,706-0,828), jednak w przypadku tej choroby 2 spośród 6 klasyfikatorów dały wynik na poziomie wyższym od 0,8 (0,828 dla Ridge Classifier oraz 0,834 dla Linear SVC). Należy zauważyć, że najlepszy wynik uzyskany za pomocą klasyfikatora Random Forest nawet po dostrojeniu hiperparametrów dawał wynik niższy (F1=0,826 dla Etapu 5). W 6 przypadkach na 6 zastosowanie zaproponowanego procesu przetwarzania dało wynik wyższy od 0,8 (0,814-0,896).

9.2.7. Podsumowanie wyników uzyskanych dla klasyfikatorów binarnych

Podsumowując powyższe wyniki można zauważyć, że zaproponowany sposób postępowania z tego typu danymi skutkuje wzrostem skuteczności klasyfikatorów.

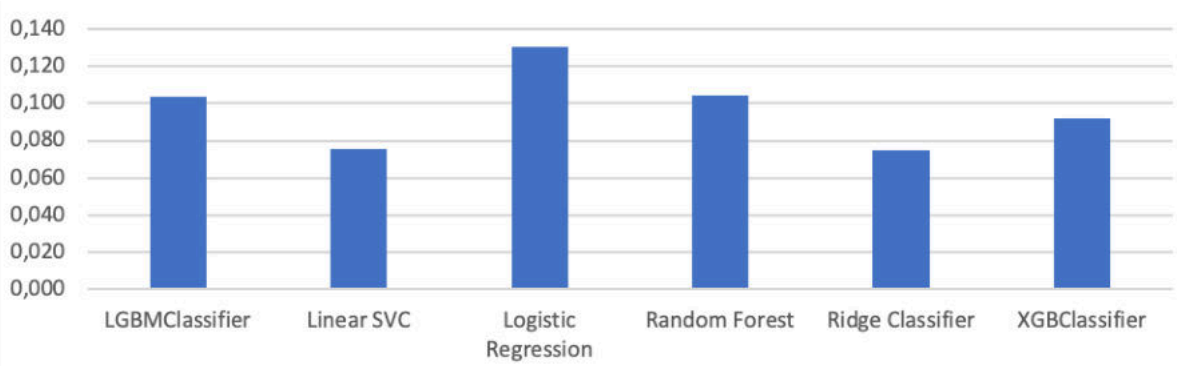
tab. 21 Wyniki treningu najlepszych klasyfikatorów na różnych etapach metody dla poszczególnych chorób

	ALL - LR	AN - LSVC	BS - LR	EB - LSVC	LB - LR	SL - LR
Wynik 0	0,745	0,810	0,728	0,786	0,742	0,760
Wynik 1	0,774	0,811	0,772	0,818	0,770	0,805
Wynik 2	0,764	0,786	0,774	0,778	0,773	0,792

Wynik 3	0,799	0,780	0,783	0,860	0,789	0,824
Wynik 4	0,802	0,877	0,794	0,864	0,795	0,825
Wynik 5	0,879	0,894	0,872	0,877	0,866	0,896

Najniższy wynik wskaźnika F1 uzyskał klasyfikator Logistic Regression dla choroby BS wytrenowany na zbiorze nieprzetworzonym (0,728).
Najwyższy wynik wskaźnika F1 uzyskał klasyfikator Logistic Regression dla choroby SL po dostrojeniu hiperparametrów (0,896).
W 4 przypadkach na 6 wykorzystanie zbioru nieprzetworzonego dało najgorsze rezultaty (0,728-0,760). W 6 przypadkach na 6 zastosowanie zaproponowanego procesu przetwarzania dało wynik wyższy od 0,8 (0,866-0,896).

W najkorzystniejszym przypadku wynik poprawił się o **0,149** pomiędzy Etapem 1 (zbiór nieprzetworzony) a Etapem 6 (dostrojenie hiperparametrów) w przypadku klasyfikatorów LGBM oraz XGBoost dla klasyfikacji binarnej przeprowadzonej dla choroby Early Blight. Średni uzysk pomiędzy Etapem 1 a Etapem 6 wyniósł **0,097** (z poziomu 0.749 do 0.846). Należy jednak odnotować, że w pojedynczym przypadku uzyskano lepszy wynik, tzn. pomiędzy Etapem 1, a Etapem 5 (najlepszy wynik w całym procesie rekurencyjnej eliminacji cech) klasyfikator XGBoost dla choroby Early Blight uzysk poprawę na poziomie 0,154.
Poszczególne etapy postępowania z danymi (Etapy 1-6) powodują osiąganie coraz lepszych wyników, nie licząc redukcji zakresów brzegowych wskazanych przez dystrybutora sprzętu. Modele klasyfikacyjne uznają te zakresy (350-475nm oraz 2400-2500nm) za obszary spektralne wnoszące wartość w procesie decyzyjnych i w 27 przypadkach na 36 ich usunięcie (Etap 3) spowodował obniżenie metryki F1 w porównaniu do Etapu 2. Zredukowanie przestrzeni cech poprzez usunięcie wspomnianych zakresów brzegowych spowodowało obniżenie metryki F1 średnio o -0,013 przy zmniejszeniu liczby cech z 2151 do 1926 (redukcja o ~10,5%).
Dla większości eksperymentów najlepszy wynik uzyskiwał klasyfikator Logistic Regression (przeciętnie 0,878). Najgorzej z problemem poradził sobie klasyfikator Random Forest (przeciętnie 0,708).



rys. 48 Średnia poprawa wyników dla poszczególnych klasyfikatorów

Realizacja przedstawionej na rys. 41 procedury (Etap 1 – Etap 6) daje porównywalną poprawę wyników zarówno dla klasyfikatora odróżniającego rośliny zdrowe od chorych niezależnie od rodzaju choroby (średni uzysk 0,103), jak również podczas rozpatrywania klasyfikacji binarnej dla każdej choroby osobno (średnia poprawa wyników w zakresie od 0,081 do 0,121).

Wyniki te potwierdzają tezę pierwszą niniejszej pracy doktorskiej („Możliwe jest określenie algorytmów klasyfikacyjnych, które z większą efektywnością pozwalają na rozpoznanie następujących chorób pomidora zwyczajnego: alternarioza, antraknoza, bakteryjna cętkowatość pomidora, septorioza, zaraza ziemniaka w pomidorze z zastosowaniem ograniczonego zbioru danych uczących”) **oraz świadczą o zasadności stosowania zaproponowanej w niniejszym rozdziale metody przetwarzania danych hiperspektralnych**.

9.3. Analiza udziału poszczególnych zakresów spektralnych w klasyfikacji binarnej chorób

Kolejnym etapem prowadzonego badania było określenie udział poszczególnych pasm spektralnych (cech) w procesie decyzyjnym. W zależności od wykorzystywanego klasyfikatora określono wagi cech (atrybut *coef_*) lub współczynniki istotności cech (atrybut *feature_importances_*). Jak podaje J. Brownlee [98] współczynniki mogą być traktowane jako wskaźniki istotności cech dla takich klasyfikator.in.ak m.in. regresja liniowa czy klasyfikator grzbietowy (Ridge).

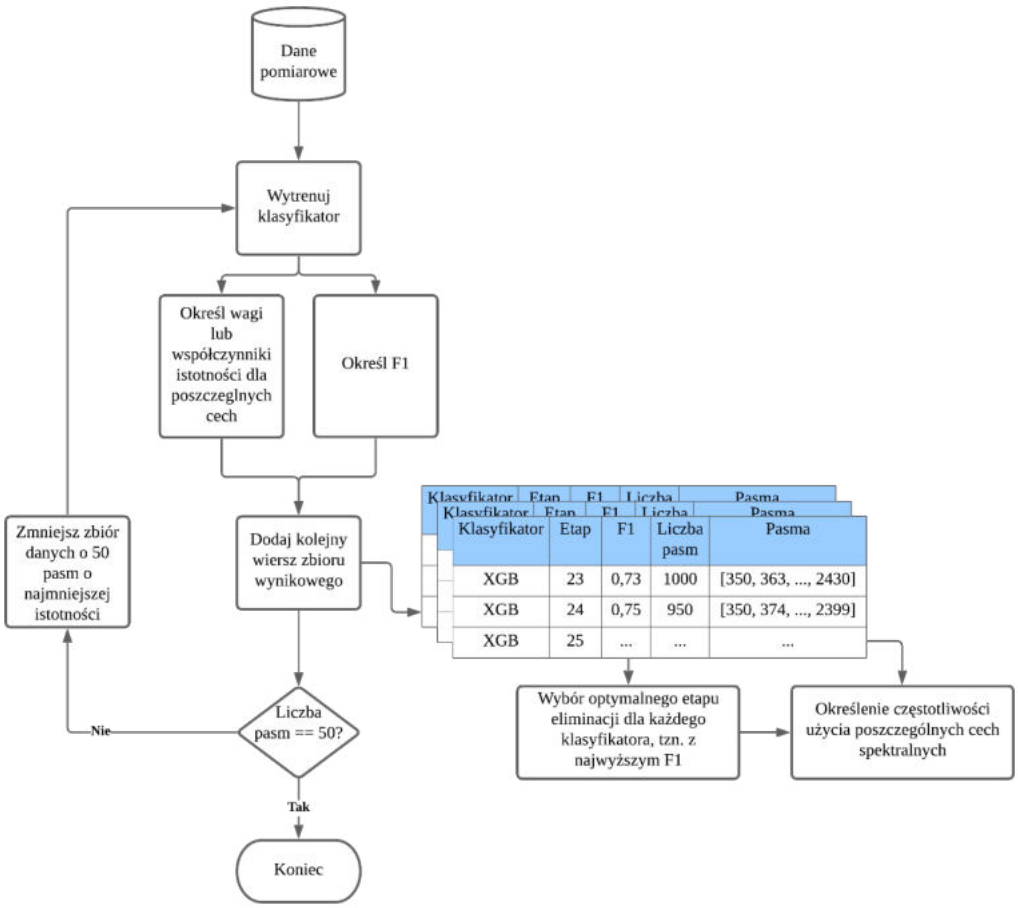
9.3.1. Zjawisko Hughesa w danych hiperspektralnych

Dane hiperspektralne składają się z setek kanałów spektralnych o wysokiej rozdzielczości spektralną, dzięki czemu pozyskiwana jest niemal ciągła charakterystyka spektralnej badanego obiektu i dlatego znajdują zastosowanie m.in. w klasyfikacji wegetacyjnej. Trudność wykorzystania danych hiperspektralnych polega na tym, że zazwyczaj są one nadmiarowe, silnie skorelowane z sobą i podlegające zjawisku Hughesa. W zjawisku tym miara dokładności klasyfikacji wzrasta stopniowo wraz ze wzrostem liczby badanych pasm spektralnych (zwiększenie wymiarowości), ale znacznie spada po przekroczeniu pewnej granicy [99, 100]. Istotne zatem jest znalezienie odpowiedniej liczby wymiarów wykorzystywanych w procesie uczenia maszynowego.

9.3.2. Metoda eliminacji cech

Problem, o którym wspomniano w punkcie 9.3.1 można rozwiązać za pomocą metod stopniowej eliminacji cech poprzez usunięcie cech o najmniejszej istotności. W tym celu dla klasyfikatorów RidgeClassifier, Linear SVC oraz Logistic Regression wykorzystano wyznaczone wagi cech, natomiast współczynniki istotności cech wykorzystano w przypadku klasyfikatorów Random Forest, XGBClassifier oraz LGBMClassifier.

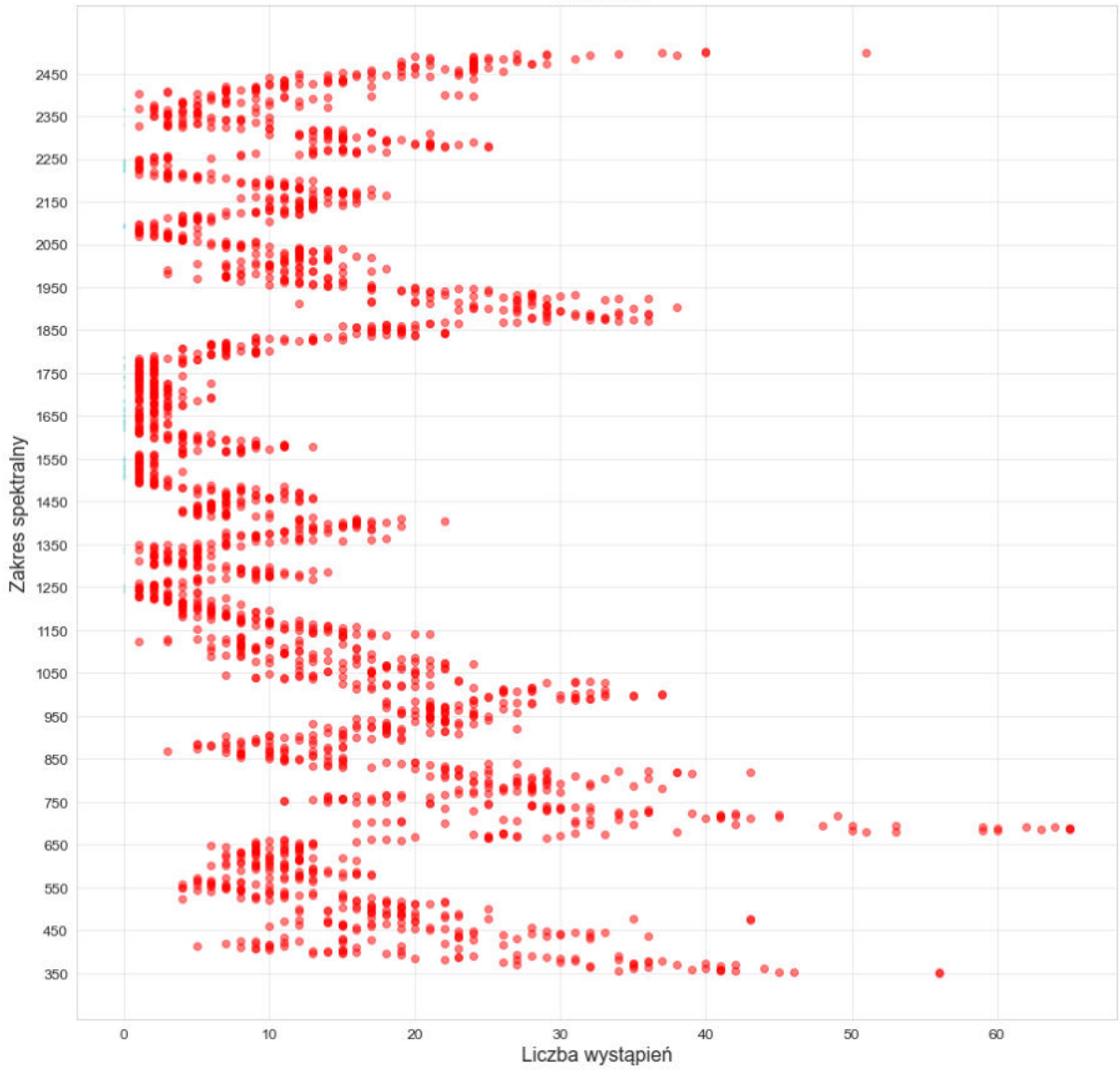
Zastosowano metodę rekurencyjnej eliminacji cech (ang. *Recursive Featrue Elimination*, RFE) [101, 102] wykorzystującej wspomniane wyżej poziomy istotności cech oraz współczynniki cech. Metoda polega na iteracyjnej eliminacji kolejnych, najmniej istotnych cech zbioru danych. W przypadku rozpatrywanego zbioru danych były to kolumny danych reprezentujące kolejne długości badanej. W każdej kolejnej iteracji ze zbioru usuwanych było 50 cech, które wykazywały najmniejszą istotność w procesie decyzyjnym. Następnie klasyfikator był ponownie trenowany na mniejszym zbiorze danych. Proces eliminacji dobiegał końca, gdy ograniczono zbiór do 50 cech. W zależności, od rozpatrywanego zbioru danych, cały proces składał się z 39 lub 44 kroków eliminacji. Każdej z cech przypisany został wskaźnik rankingowy, który pozwolił określić na jakim etapie cecha został wyeliminowana. Do każdego z etapów określono również wartość miary F1.



rys. 49 Proces rekurencyjnej eliminacji cech i wyznaczania optymalnego etapu eliminacji

Na podstawie miary F1 określono optymalny etap eliminacji, tzn. krok, dla którego uzyskano najwyższą wartość wspomnianego wskaźnika. W ten sposób wyszczególniono wszystkie pasma biorące udział w procesie decyzyjnym. Następnie określono jak często poszczególne cechy były uwzględniane przez klasyfikator. Na tej podstawie stworzono wykres udziału cech spektralnych w procesie decyzyjnym.

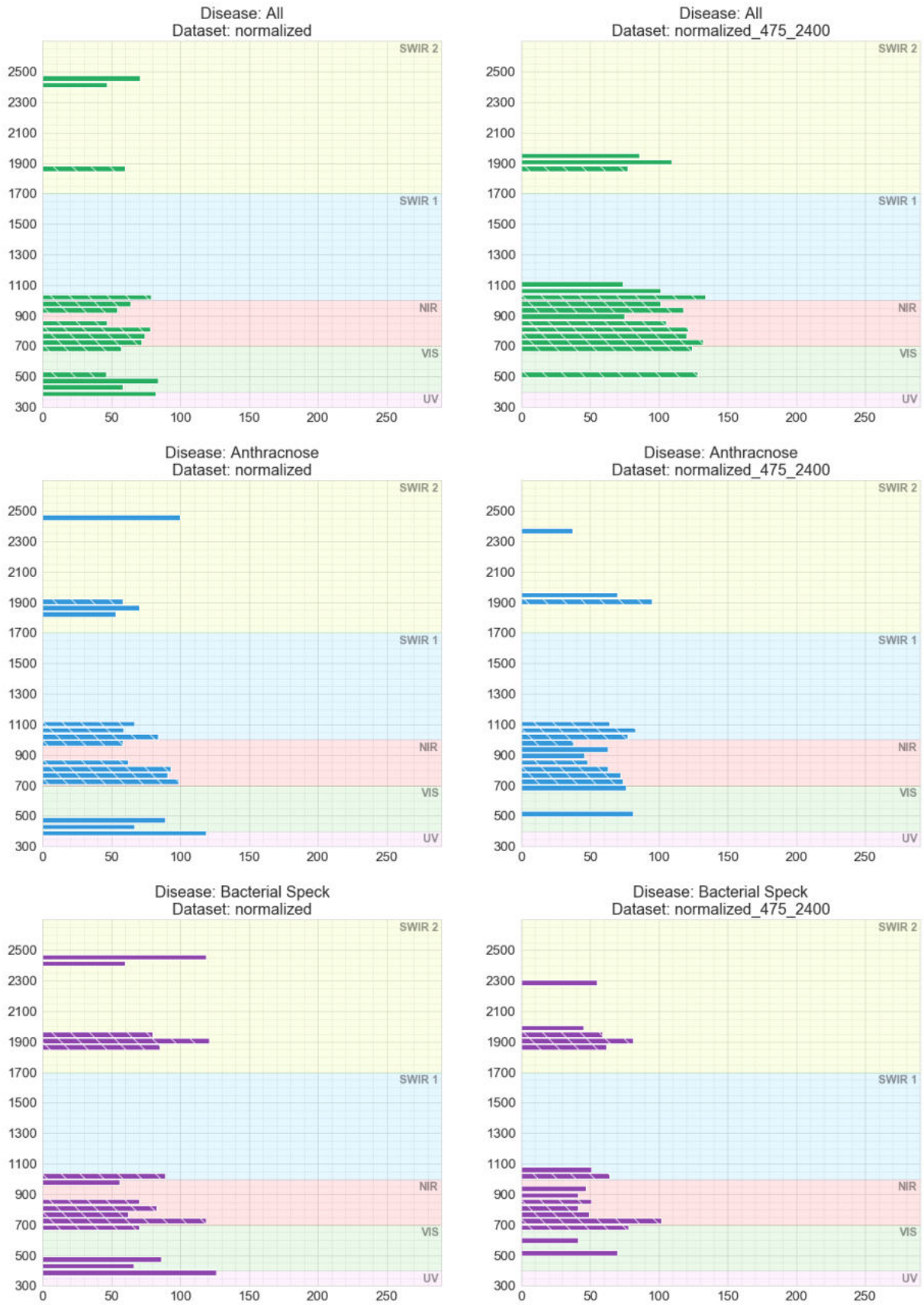
Na rysunku (rys. 50) przedstawiono zsumowany udział poszczególnych pasm spektralnych dla wszystkich badanych klasyfikatorów oraz dla wszystkich analizowanych chorób.



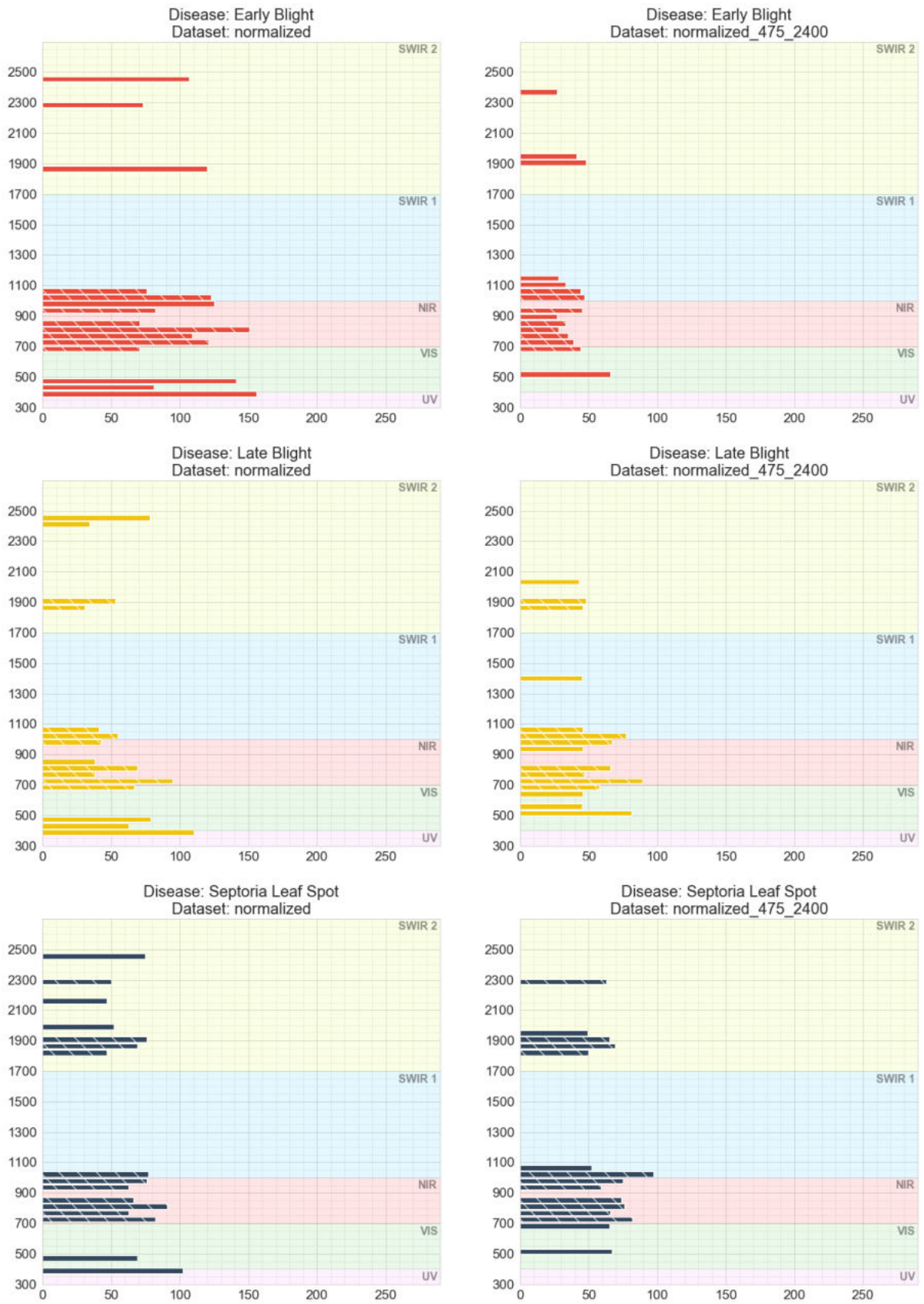
rys. 50 Udział poszczególnych pasm spektralnych w procesie decyzyjnym

W ramach pogłębionej analizy zdecydowano się na zgrupowanie sąsiednich zakresów widmowych (50 pasm) oraz określenie najczęściej powtarzających się zakresów spektralnych (usunięto 70% zakresów pasm występujących najrzadziej).

Na podstawie powyższych założeń na kolejnych rysunkach dla poszczególnych chorób przedstawiono najczęściej powtarzające się zakresy spektralne.



rys. 51 Istotne przedziały spektralne w procesie klasyfikacji. Kreskowanie białą linią wyznacza przedziały pasm istotne również w przypadku poszczególnych chorób. Kreskowanie jest niezależne od zbioru.



rys. 52 Istotne przedziały spektralne w procesie klasyfikacji. Kreskowanie białą linią wyznacza wspólne przedziały spektralne dla wszystkich chorób. Kreskowanie jest niezależne od zbioru.

Zarówno w przypadku zbioru danych znormalizowanych, jak i znormalizowanych z zawężonego zakresu (475-2400nm) istotne znaczenie w rozpoznaniu choroby ogrywa całe pasmo bliskiej podczerwieni. W badaniach przyjęto szerokość pasma na poziomie 43nm, co oznacza wyznaczenie 50 pasm w zakresie 350-2500 nm, a jednocześnie wpisuje się w standardy spektralnej teledetekcji satelitarnej. Dla porównania, szerokości spektralne w instrumencie obrazującym satelity Sentinel-2 wahają się od 15 do 185nm (mediana 30.5 nm, średnia 50.42nm).

Wspólne przedziały spektralne uzyskane dla zbioru znormalizowanego: 371, 455, 708, 751, 793, 835, 961, 1004, 1847, 2438 nm.
Wspólne przedziały spektralne uzyskane dla zbioru znormalizowanego, ograniczonego do przestrzeni 475-2400nm: 498, 666, 708, 751, 793, 919, 1004, 1046, 1889 nm.

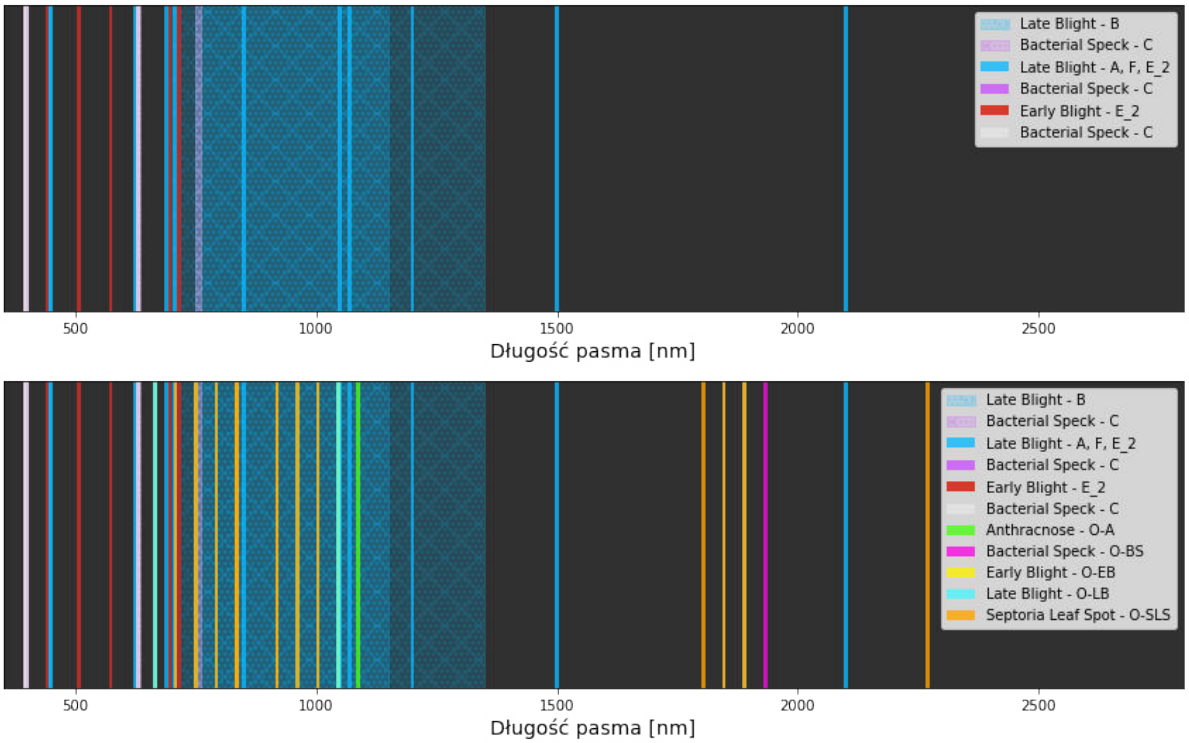
Wspólne przedziały spektralne uzyskane dla chorób, niezależnie od zbioru (szerokość pasma równa 43nm):

1. Anthracnose: 708, 751, 793, 835, 961, 1004, 1046, 1088, 1889,
2. Bacterial Speck: 666, 708, 751, 793, 835, 1004, 1847, 1889, 1932,
3. Early Blight: 666, 708, 751, 793, 835, 919, 1004, 1046,
4. Late Blight: 666, 708, 751, 793, 961, 1004, 1046, 1847, 1889,
5. Septoria Leaf Spot: 708, 751, 793, 835, 919, 961, 1004, 1805, 1847, 1889, 2269,
6. All: 498, 666, 708, 751, 793, 835, 919, 961, 1004, 1847.

Zaobserwować można, że spośród 50 pasm zbiorczych niektóre przedziały spektralne pojawiają się we wszystkich przypadkach niezależnie od zastosowanego podejścia.

9.4. Porównanie wyników z literaturą przedmiotu

Wykonano porównanie zakresów spektralnych wymienionych na początku rozdziału z otrzymanymi wynikami. Na wykres zaprezentowany w ramach przeglądu literatury (rys. 10 w rozdziale 4.6) naniesiono pasma otrzymane w ramach przeprowadzonych badań.



rys. 53 Zestawienie porównawcze z odniesieniem do pasm spektralnych zaobserwowanych w literaturze (góra – najistotniejsze pasma spektralne wg literatury, dolny – najistotniejsze pasma spektralne wg literatury i przeprowadzonych badań)

Zakres światła widzialnego oraz bliskiej podczerwieni również pojawił się w otrzymanych badaniach, jednak dodatkowo przeprowadzone badania wykazały istotność zakresów SWIR w procesie klasyfikacyjnym.

Na podstawie przedstawionych wyników można potwierdzić tezę drugą niniejszej pracy doktorskiej („Możliwe jest określenie specyficznych długości zakresu widma odpowiadających określonym jednostkom chorobowym *Solanum lycopersicum*"), **a dodatkowo wskazany został obszar SWIR, w szczególności w zakresach 1800-1890 nm oraz 2269 nm, jako istotny w badanym zagadnieniu, co stanowi nowość w ujęciu badań naukowych.**

9.5. Klasyfikacja wieloklasowa

W ramach prac badawczych rozpatrzono również przypadek multiklasyfikacji. Proces uczenia powtórzono w sposób analogicznych, jak w przypadku klasyfikacji binarnej. Zbiór danych podzielono na część treningową i testową z uwzględnieniem pomiarów tych samych obiektów tak, aby nie dopuścić do wycieku danych w procesie treningowym. W podzielonych zbiorach zapewniono równą liczbę reprezentantów każdej klasy poprzez redukcję liczby nadmiarowych reprezentantów w poszczególnych klasach.
Wykorzystano te same klasyfikatory, jak w przypadku klasyfikacji binarnej. Trening powtórzono dla wszystkich trzech zbiorów. Na każdym wykonano proces rekurencyjnej eliminacji cech (RFE) oraz dostrojenia hiperparametrów.

Dla zbioru nieprzetworzonego z wykorzystaniem tuningu hiperparametrów uzyskano następujące wyniki:

tab. 22 Wyniki multiklasyfikacji dla nieprzetworzonego zbioru danych z wykorzystaniem metody dotrenowania hiperparametrów

	AN	BS	ALL	EB	LB	SL
RF	0,238	0,214	0,502	0,286	0,234	0,257
LGBM	0,275	0,273	0,553	0,287	0,248	0,225
RC	0,382	0,313	0,693	0,350	0,342	0,369
LR	0,311	0,293	0,722	0,380	0,300	0,370
LSVC	0,309	0,299	0,706	0,376	0,328	0,364
XGB	0,226	0,230	0,543	0,325	0,282	0,223

Dla zbioru znormalizowanego z ograniczonym zakresem (475-2400nm):

tab. 23 Wyniki multiklasyfikacji dla znormalizowanego zbioru danych z ograniczonym zakresem spektralnym (475-2400nm) z wykorzystaniem metody dotrenowania hiperparametrów

	AN	BS	ALL	EB	LB	SL
RF	0,225	0,221	0,646	0,335	0,269	0,230
LGBM	0,225	0,287	0,646	0,329	0,258	0,234
RC	0,342	0,236	0,657	0,335	0,372	0,404
LR	0,322	0,313	0,715	0,376	0,350	0,396
LSVC	0,358	0,243	0,674	0,335	0,369	0,333
XGB	0,226	0,234	0,630	0,325	0,282	0,268

Dla zbioru znormalizowanego w pełnym zakresie (350-2500nm):

tab. 24 Wyniki multiklasyfikacji dla znormalizowanego zbioru danych z pełnym zakresem spektralnym (350-2500nm) z wykorzystaniem metody dotrenowania hiperparametrów

	AN	BS	ALL	EB	LB	SL
RF	0,238	0,212	0,608	0,342	0,321	0,261
LGBM	0,291	0,296	0,678	0,394	0,340	0,216
RC	0,392	0,308	0,698	0,361	0,385	0,377
LR	0,328	0,317	0,738	0,359	0,358	0,363
LSVC	0,408	0,293	0,735	0,368	0,388	0,351
XGB	0,242	0,279	0,676	0,342	0,337	0,256

Najlepsze wyniki uzyskano dla roślin kontrolnych przy użyciu regresji logistycznej (0.715-0.738), natomiast najgorsze dla BS przy zastosowaniu lasów losowych (0.212-0.221).

Uzyskane wyniki nie pozwoliły na kontynuowanie prac w zakresie multiklasyfikacji.

9.6. Podsumowanie rozdziału

W rozdziale przedstawiono wyniki badań mających na celu opracowanie metody postępowania z danymi hiperspektralnymi, które będą skutkowały najwyższą poprawą wyników klasyfikacji binarnej oraz zweryfikowanie hipotezy, że wybrane pasma spektralne są szczególnie istotne w procesie diagnostycznym.
W tym celu zaproponowano wykonanie eksperymentu dla różnych przypadków badawczych. Przedstawiono metodę wstępnej selekcji klasyfikatorów, na podstawie której wybrano 6 klasyfikatorów do dalszych analiz. Omówiono zaproponowaną metodę postępowania z danymi. Przedstawiono wyniki dla każdego z 6 etapów dla każdego klasyfikatora. Następnie przedstawiono wyniki metody eliminacji cech i wyznaczono zakresy spektralne najczęściej biorące udział w procesie decyzyjnym dla różnych przypadków badawczych. Dodatkowo podjęto próby stworzenia klasyfikatorów wieloklasowych, jednak uzyskane rezultaty nie pozwalały na kontynuowanie prac w tym zakresie.
Celem tego rozdziału była prezentacja wyników dla zaproponowanej metody postępowania z danymi hiperspektralnymi, a także potwierdzenie pierwszej oraz drugiej tezy niniejszej pracy doktorskiej. Dodatkowo wyznaczono obszar SWIR (1800-1890 nm oraz 2269 nm) jako istotne dla badanego zagadnienia obszary spektralne.

10. Analiza skuteczności klasyfikacji chorób w kolejnych dniach od inokulacji

Metodyka prowadzenia badań i pozyskiwania pomiarów spektralnych przedstawiona w rozdziale 6. „Eksperyment pomiarowy – obserwacja uprawy pomidora zwyczajnego w warunkach rzeczywistych i laboratoryjnych" pozwala na analizę danych w ujęciu czasowym. Wykrywanie ognisk chorobowych w uprawie stanowi istotny element nowoczesnej agronomii, a wczesne wykrywanie i rozpoznawanie chorób roślin jest przedmiotem zainteresowania wielu badaczy, co szerzej opisano w rozdziale 4.7. W sprzyjających dla patogenu warunkach choroba może rozprzestrzenić się w ciągu kilku tygodni z jednej rośliny na powierzchnię całego pola, jak przedstawiono w rozdziale 6.3, rys. 18–rys. 20, dlatego tak ważnym jest prawidłowe zdiagnozowanie choroby w jak najkrótszym czasie od zakażenia. Wyniki zawarte w niniejszym rozdziale zostały w części zaprezentowane w czasopismie Scientific Reports [97].

10.1. Wyniki w ujęciu rodzaju choroby

W poniższym podrozdziale zaprezentowano otrzymane rezultaty analiz w 6 wariantach eksperymentu, analogicznie do eksperymentu przedstawionego w rozdziale 9:

a) obiekty kontrolne – obiekty zakażone dowolną z chorób (AN-SL z rozdziału 3.1),
b) obiekty kontrolne – obiekty zakażone chorobą **AN** (*Anthracnose*),
c) obiekty kontrolne – obiekty zakażone chorobą **BS** (*Bacterial Speck*),
d) obiekty kontrolne – obiekty zakażone chorobą **EB** (*Early Blight*),
e) obiekty kontrolne – obiekty zakażone chorobą **LB** (*Late Blight*),
f) obiekty kontrolne – obiekty zakażone chorobą **SL** (*Septoria Leaf Spot*).

Takie podejście pozwoliło na porównanie wyników zbudowanych klasyfikatorów pomiędzy sobą oraz na zachowanie spójności z metodyką zastosowaną we wcześniej przedstawionych badaniach. W badaniu uwzględniono klasyfikatory uzyskujące najlepsze rezultaty, tj. Random Forest, LGBMClassifier oraz Ridge Classifier. Metoda prezentowana w niniejszym rozdziale pracy doktorskiej dotyczy analizy pomiarów w kolejnych dniach pomiarowych, w ujęciu czasowym (DPI, ang. *days post-inoculum*). Zbiór pomiarowy, przedstawiony w rozdziale 8.4 (Analiza danych pomiarowych – EDA), zawierający dane pozyskane w środowisku kontrolowanym w kolejnych dniach od inokulacji oraz przeprowadzony w trzech seriach pomiarowych, został podzielony na podzbiory zgodnie z metodą zaprezentowaną w rozdziale 8.3, (rys. 26).

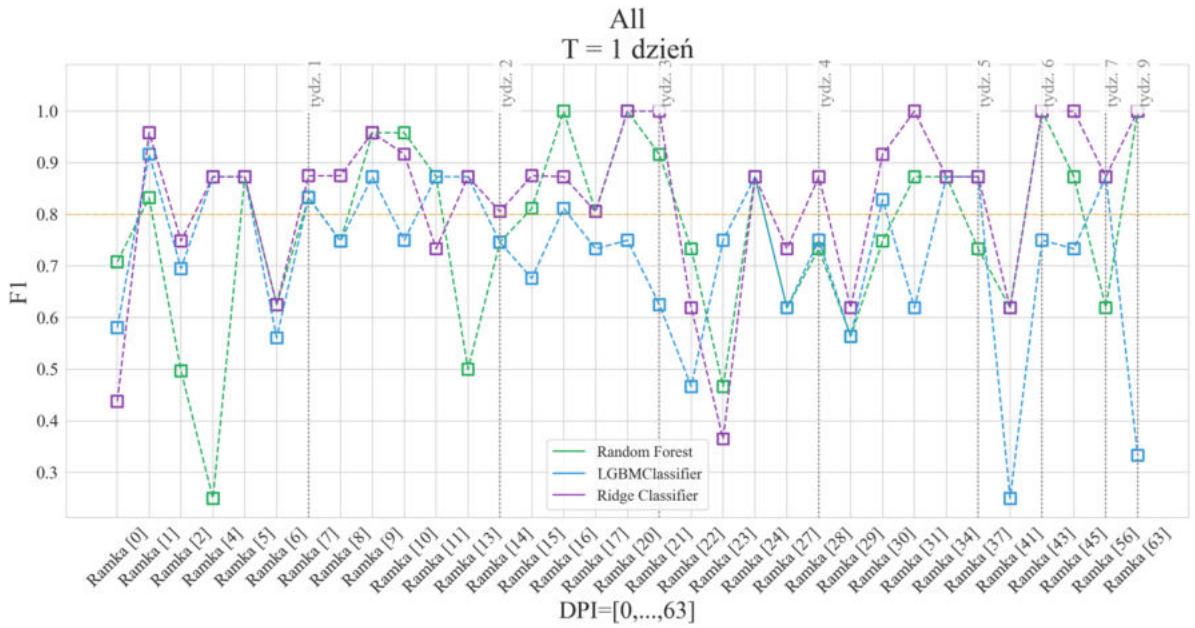
Na wynikowych wykresach zaznaczono przerywaną, pomarańczową linią poziom wskaźnika F1 jako poziom oczekiwanej miary skuteczności modelu klasyfikacji. Arbitralne przyjęcie wskaźnika jakości modelu konieczne było do analizy możliwości rozpoznania wybranych chorób badanej uprawy w ujęciu czasowym. Dobranie właściwego poziomu wskaźnika F1 czy innych metryk modeli klasyfikacyjnych zależy przede wszystkim od kontekstu, w jakim dany model będzie funkcjonował. Jednocześnie wskazuje się poziom F1=0.8 jako wartość zadowalającą [103].

10.1.1. Wszystkie choroby

W niniejszym podrozdziale opisano wyniki klasyfikacji binarnej rozróżniającej rośliny zdrowe od chorych niezależnie od choroby. Analizę wykonano w ujęciu T = [1, 3 oraz 5].

Pomiary z podziałem na poszczególne dni pomiarowe (T = 1)

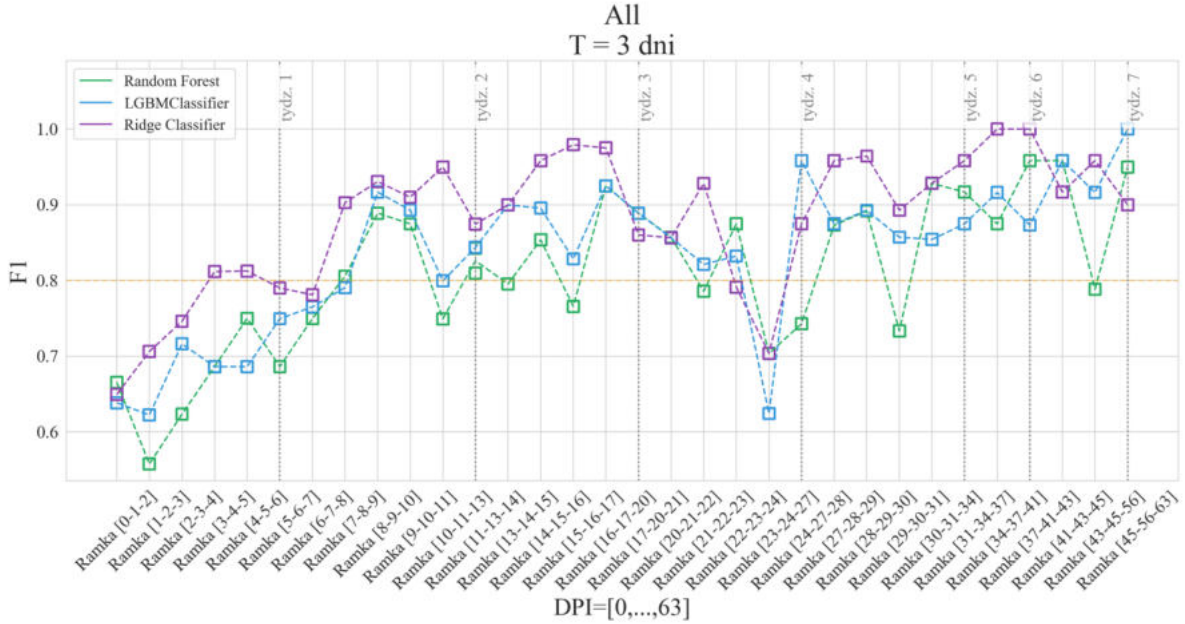
Na przedstawionym wykresie dla T = 1 przedstawiono wyniki treningu klasyfikatorów dla poszczególnych dni pomiarowych (DPI).



rys. 54 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI) w scenariuszu a. pomiary kontrolne kontra pomiary obiektów zakażonych (AN-SL), T=1

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.75 (LGBM) do 0.87 (Ridge) przy odchyleniu standardowym na poziomie 0.16-0.179.

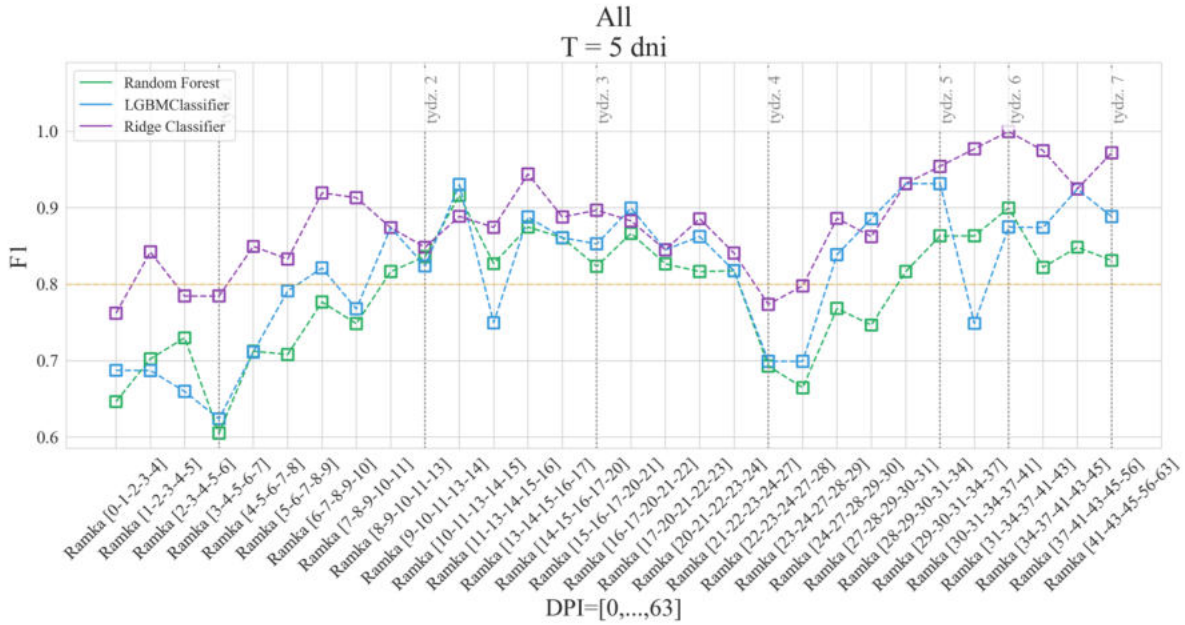
Pomiary z 3 kolejnych dni (T =3)



rys. 55 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu a. pomiary kontrolne kontra pomiary obiektów zakażonych (AN-SL), T=3

Zauważalny jest istotny wzrost wyników dla zaprezentowanego podejścia z poziomu 0.6 do 0.9 w ciągu pierwszych 12 iteracji (2 tygodnie od inokulacji). Zauważalny jest spadek wartości wskaźnika F1 w okolicach DPI= 23.
Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.83 (Random Forest) do 0.9 (Ridge) przy odchyleniu standardowym na poziomie 0.09-0.1. Widać zatem znaczną poprawę wyników w stosunku do podejścia wykorzystującego próbki jedynie z pojedynczego dnia pomiarowego.

Pomiary z 5 kolejnych dni (T =5)



rys. 56 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu a. pomiary kontrolne kontra pomiary obiektów zakażonych (AN-SL), T=5

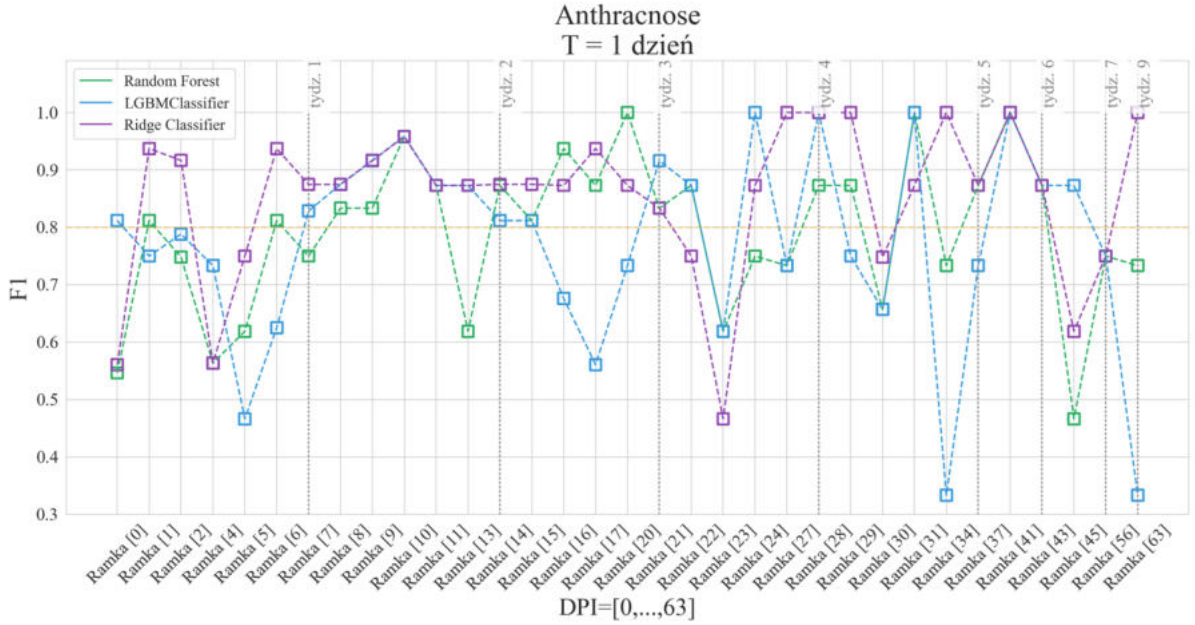
Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.82 (Random Forest) do 0.88 (Ridge) przy odchyleniu standardowym na poziomie 0.064-0.09. Uzyskano zatem nieco niższe, jednak bardziej powtarzalne wyniki.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (*T*=1), a pozostałymi (*T*=3, *T*=5) wyniosła:

- Dla *T*=1 vs. *T*=3: 0.0748
- Dla *T*=1 vs. *T*=5: 0.0492

10.1.2. Anthracnose (AN)

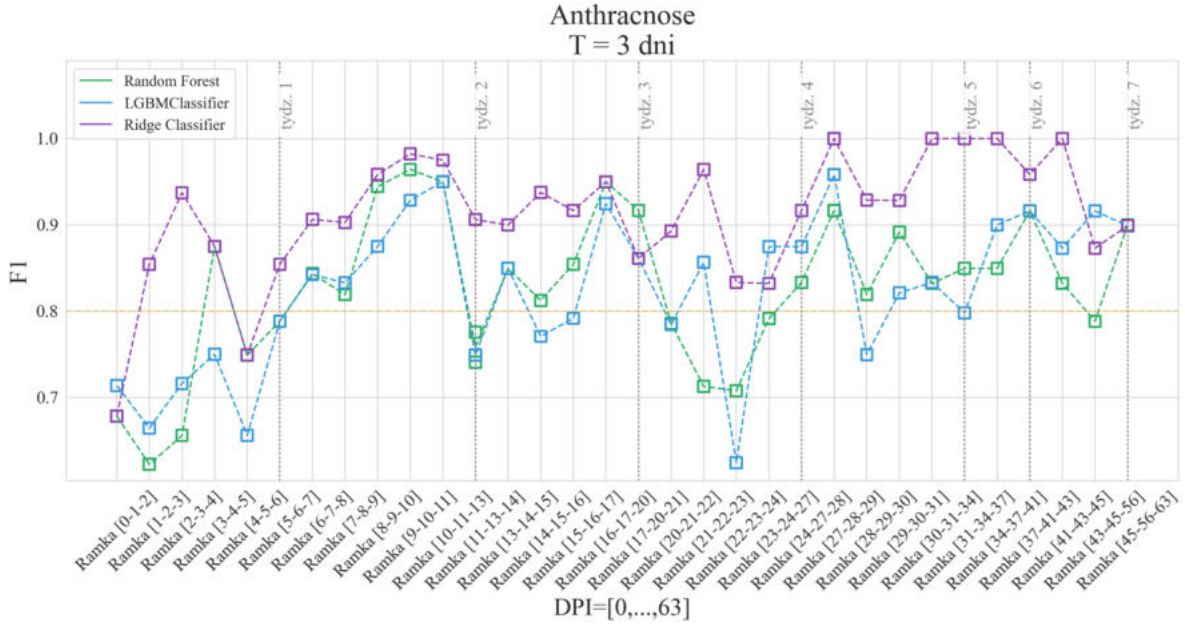
Pomiary z podziałem na poszczególne dni pomiarowe (T =1)



rys. 57 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę AN, T=1

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.81 (Random Forest, LGBM) do 0.87 (Ridge) przy odchyleniu standardowym na poziomie 0.135-0.172.

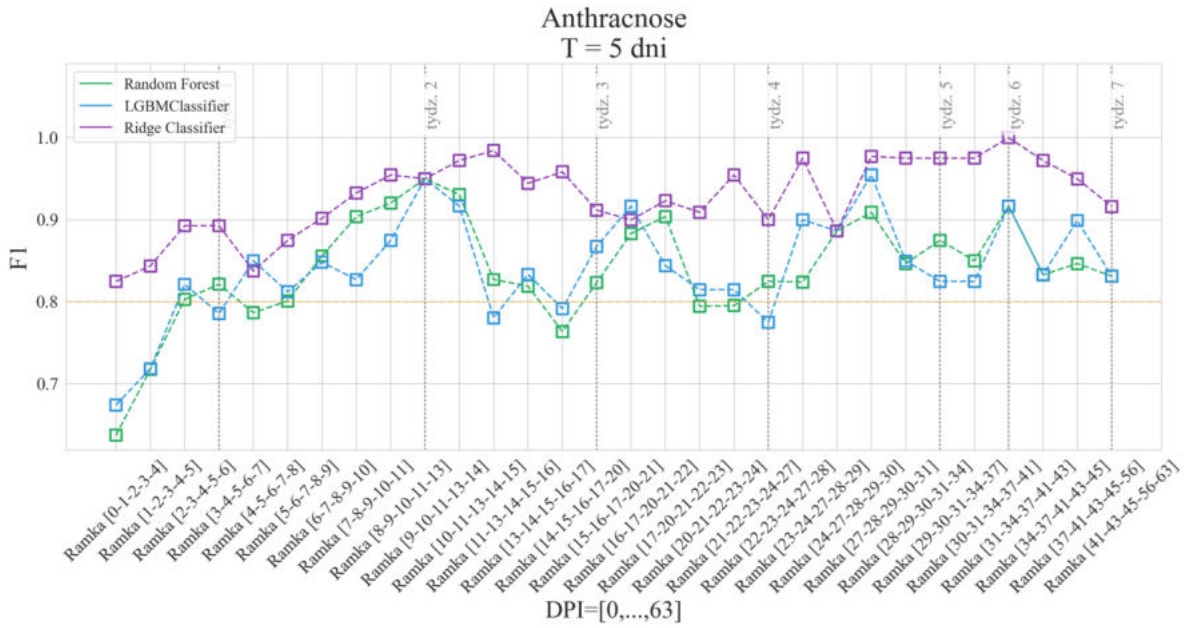
Pomiary z 3 kolejnych dni (T =3)



rys. 58 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę AN, T=3

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.83 (Random Forest) do 0.92 (Ridge) przy odchyleniu standardowym na poziomie 0.072-0.092.

Pomiary z 5 kolejnych dni (T =5)



rys. 59 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę AN, T=5

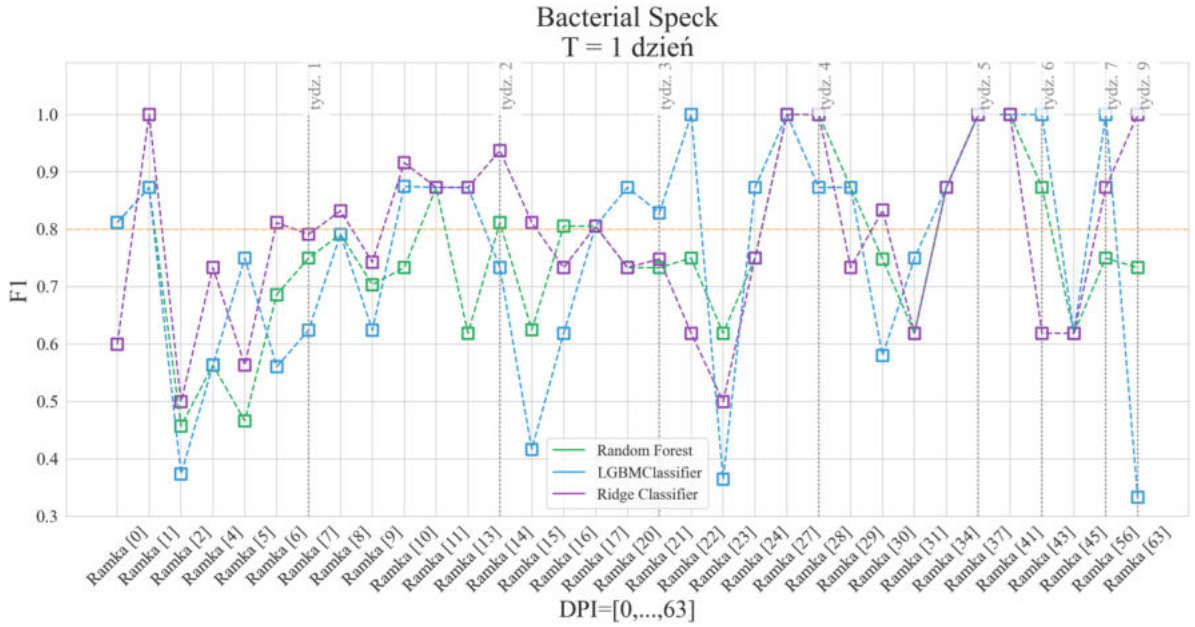
Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.83 (Random Forest, LGBM) do 0.94 (Ridge) przy odchyleniu standardowym na poziomie 0.047-0.066.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (T=1), a pozostałymi (T=3, T=5) wynosiła:

- Dla T=1 vs. T=3: 0.0540
- Dla T=1 vs. T=5: 0.0598

10.1.3. Bacterial Speck (BS)

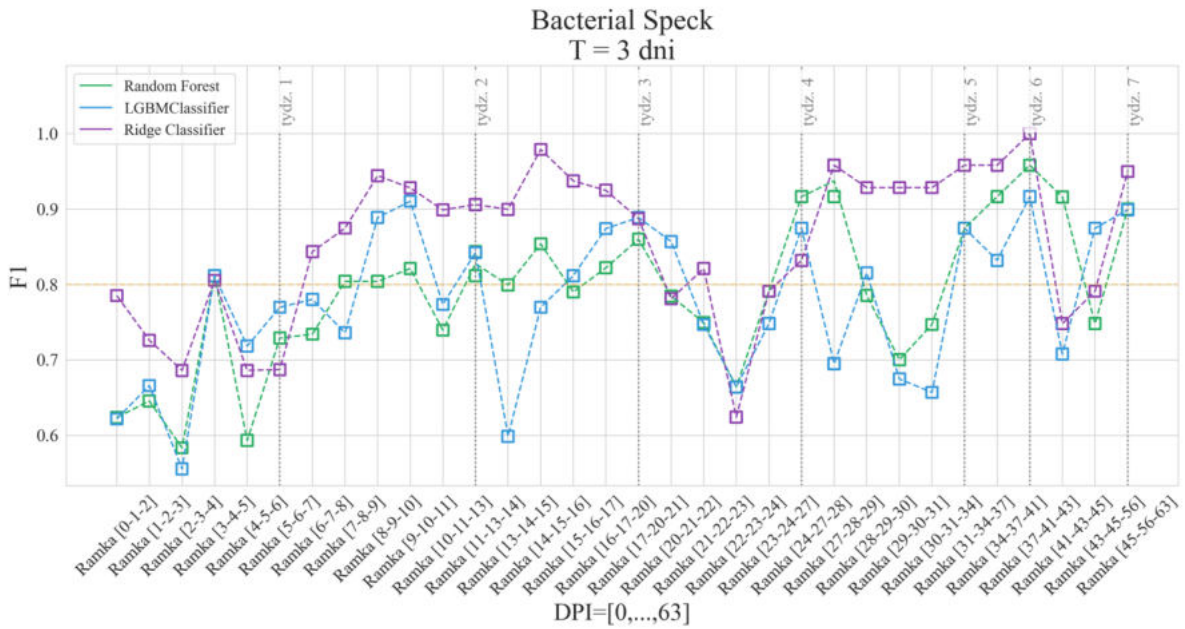
Pomiary z podziałem na poszczególne dni pomiarowe (T =1)



rys. 60 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę BS, T=1

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.75 (Random Forest) do 0.81 (LGBM, Ridge) przy odchyleniu standardowym na poziomie 0.151-0.199.

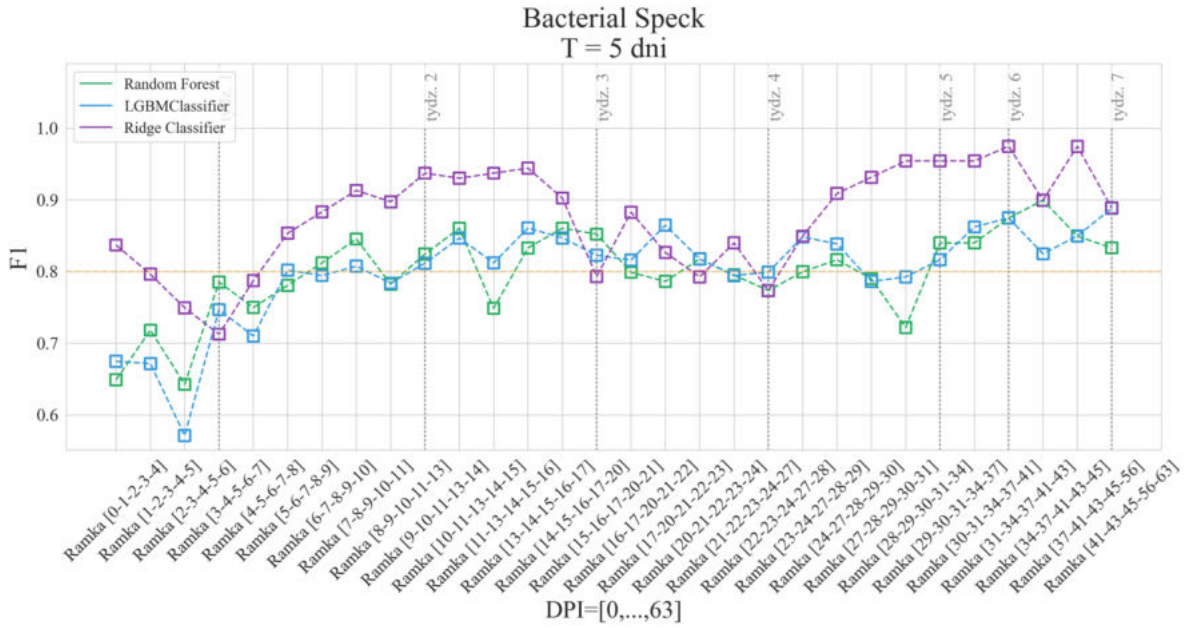
Pomiary z 3 kolejnych dni (T =3)



rys. 61 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę BS, T=3

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.78 (LGBM) do 0.9 (Ridge) przy odchyleniu standardowym na poziomie 0.098-0.1.

Pomiary z 5 kolejnych dni (T =5)



rys. 62 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę BS, T=5

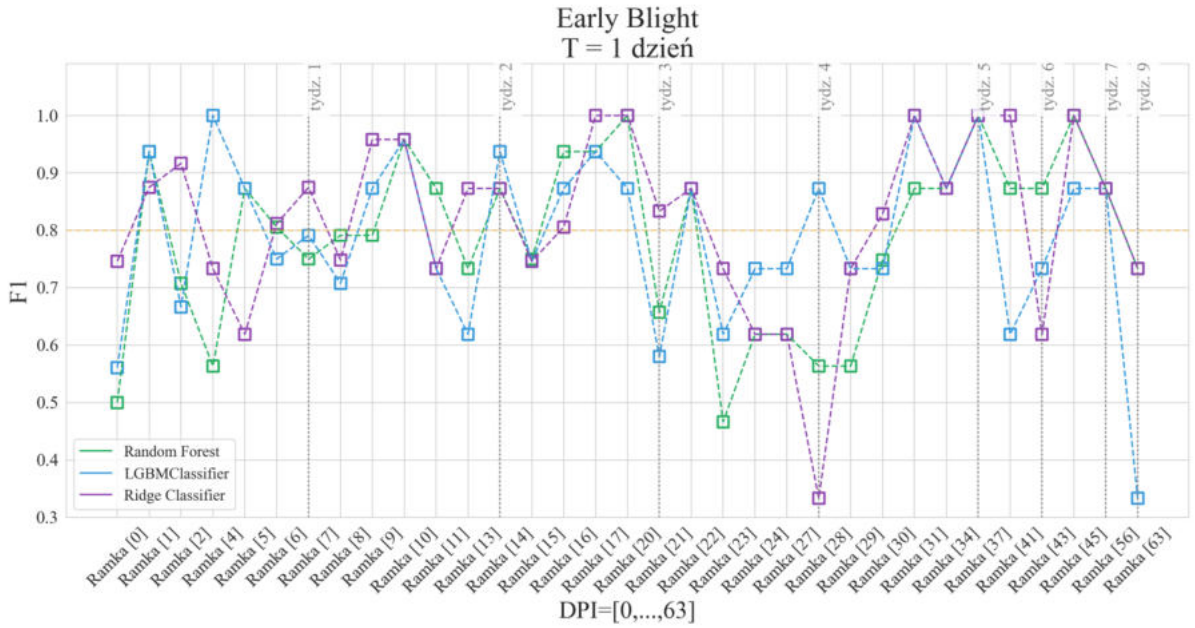
Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.81 (Random Forest, LGBM) do 0.89 (Ridge) przy odchyleniu standardowym na poziomie 0.06-0.072.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (T=1), a pozostałymi (T=3, T=5) wynosiła:

- Dla T=1 vs. T=3: 0.0402
- Dla T=1 vs. T=5: 0.0501

10.1.4. Early Blight (EB)

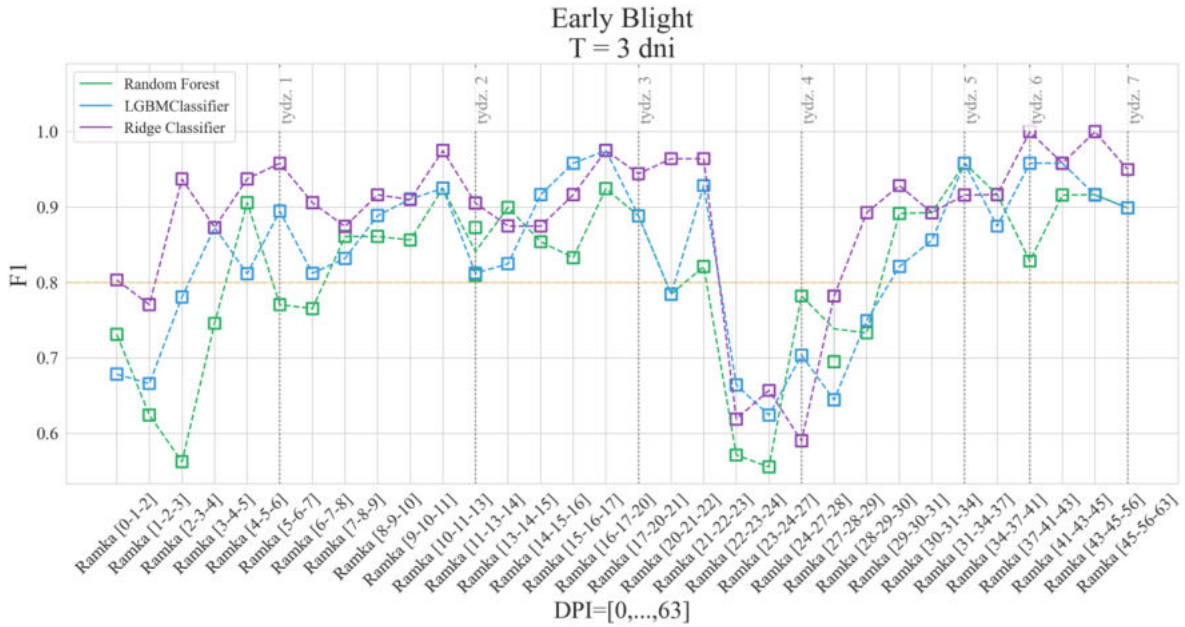
Pomiary z podziałem na poszczególne dni pomiarowe (T =1)



rys. 63 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę EB, T=1

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.79 (LGBM) do 0.83 (Ridge) przy odchyleniu standardowym na poziomie 0.149-0.151.

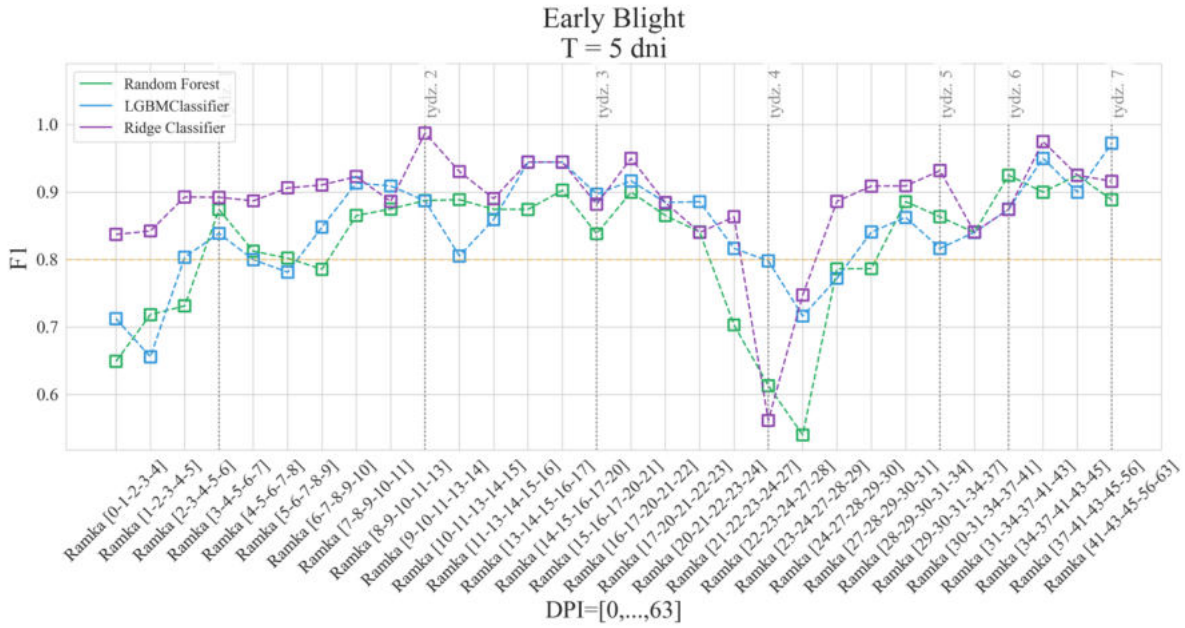
Pomiary z 3 kolejnych dni (T =3)



rys. 64 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę EB, T=3

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.84 (Random Forest, LGBM) do 0.91 (Ridge) przy odchyleniu standardowym na poziomie 0.101-0.109.

Pomiary z 5 kolejnych dni (T =5)



rys. 65 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę EB, T=5

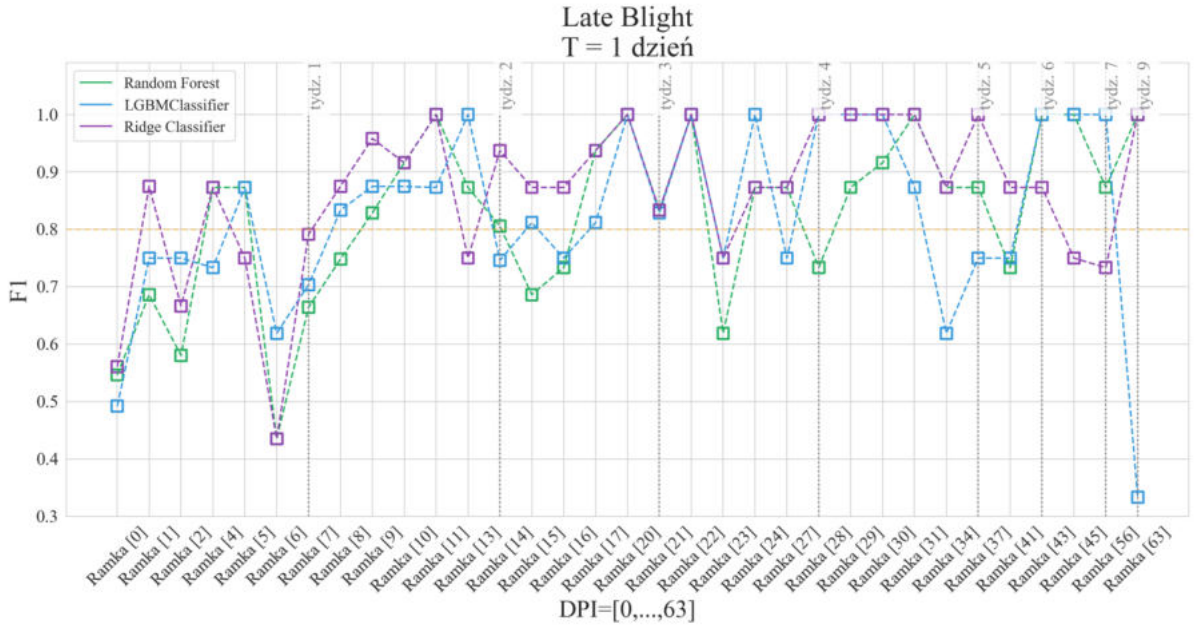
Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.85 (LGBM) do 0.89 (Ridge) przy odchyleniu standardowym na poziomie 0.075-0.096.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (T=1), a pozostałymi (T=3, T=5) wynosiła:

- Dla T=1 vs. T=3: 0.0539
- Dla T=1 vs. T=5: 0.0506

10.1.5. Late Blight (LB)

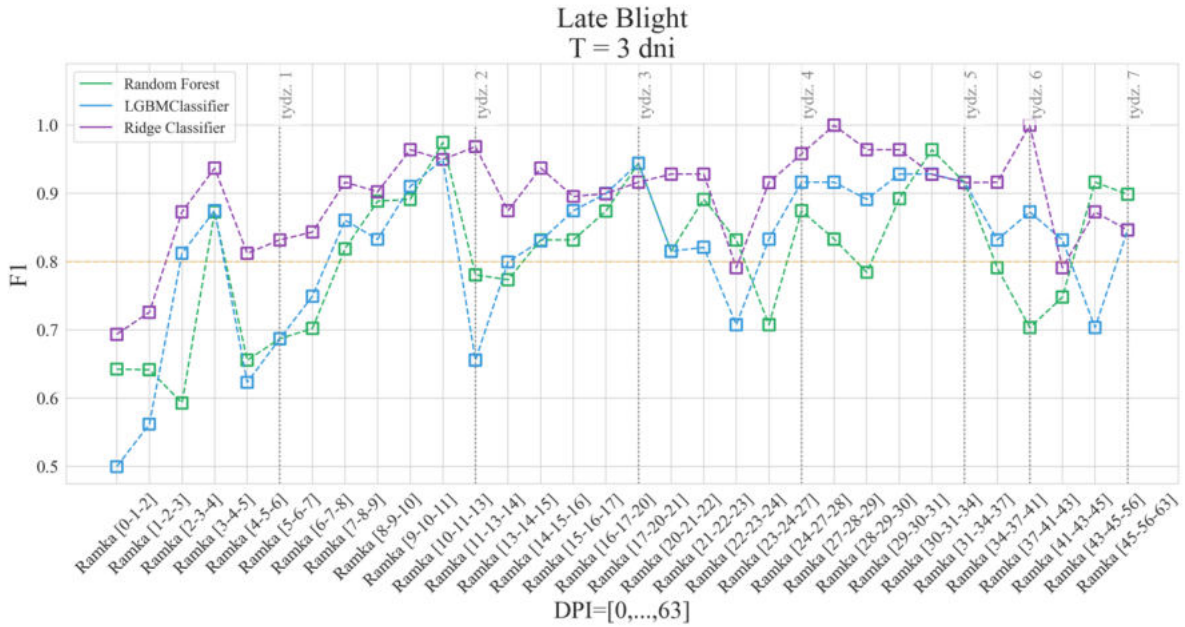
Pomiary z podziałem na poszczególne dni pomiarowe (T =1)



rys. 66 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę LB, T=1

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.83 (LGBM) do 0.87 (Random Forest, Ridge) przy odchyleniu standardowym na poziomie 0.135-0.161.

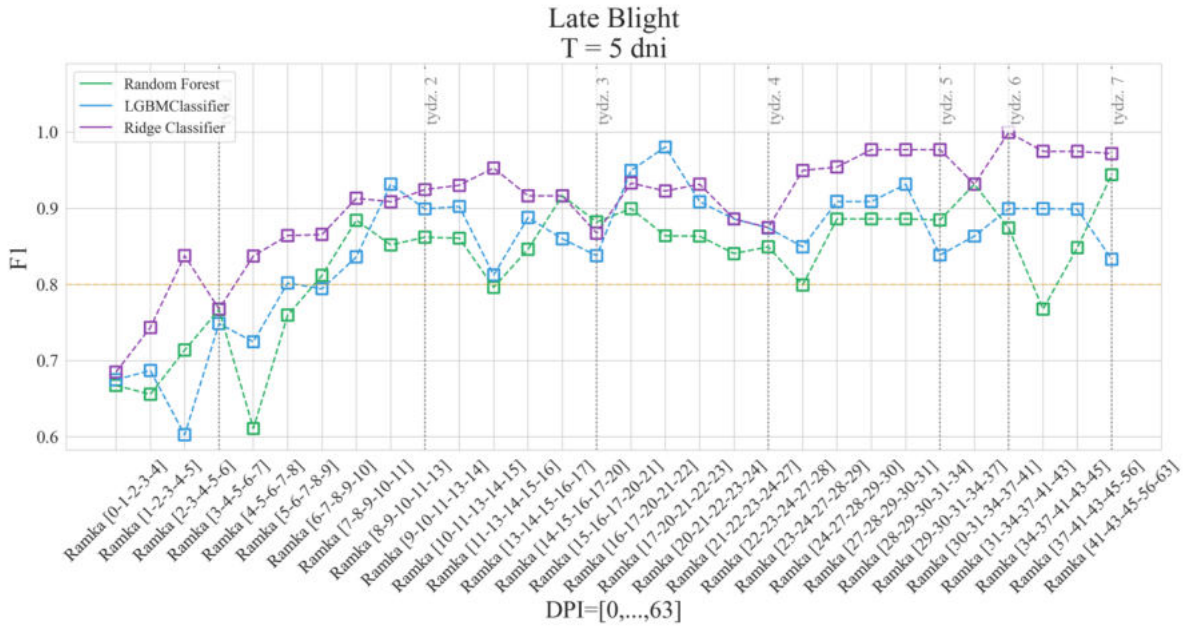
Pomiary z 3 kolejnych dni (T =3)



rys. 67 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę LB, T=3

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.83 (Random Forest, LGBM) do 0.92 (Ridge) przy odchyleniu standardowym na poziomie 0.074-0.115.

Pomiary 5 kolejnych dni (T =5)



rys. 68 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę LB, T=5

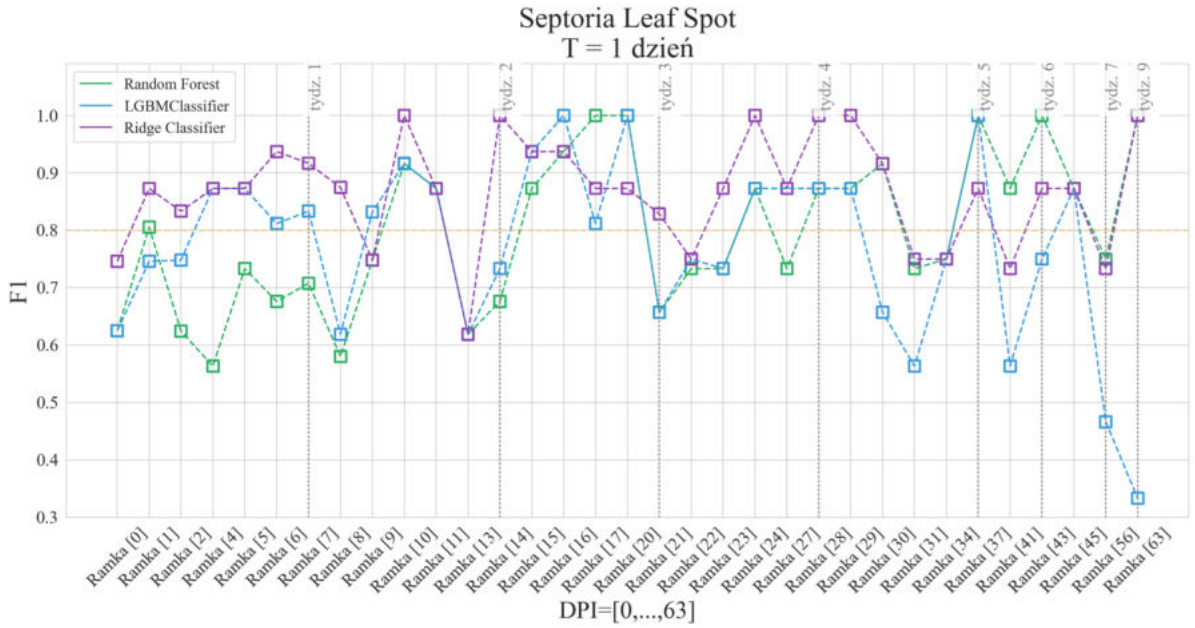
Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.86 (Random Forest) do 0.92 (Ridge) przy odchyleniu standardowym na poziomie 0.074-0.087.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (T=1), a pozostałymi (T=3, T=5) wynosiła:

- Dla *T*=1 vs. *T*=3: 0.0001
- Dla *T*=1 vs. *T*=5: 0.0058

10.1.6. Septoria Leaf Spot (SL)

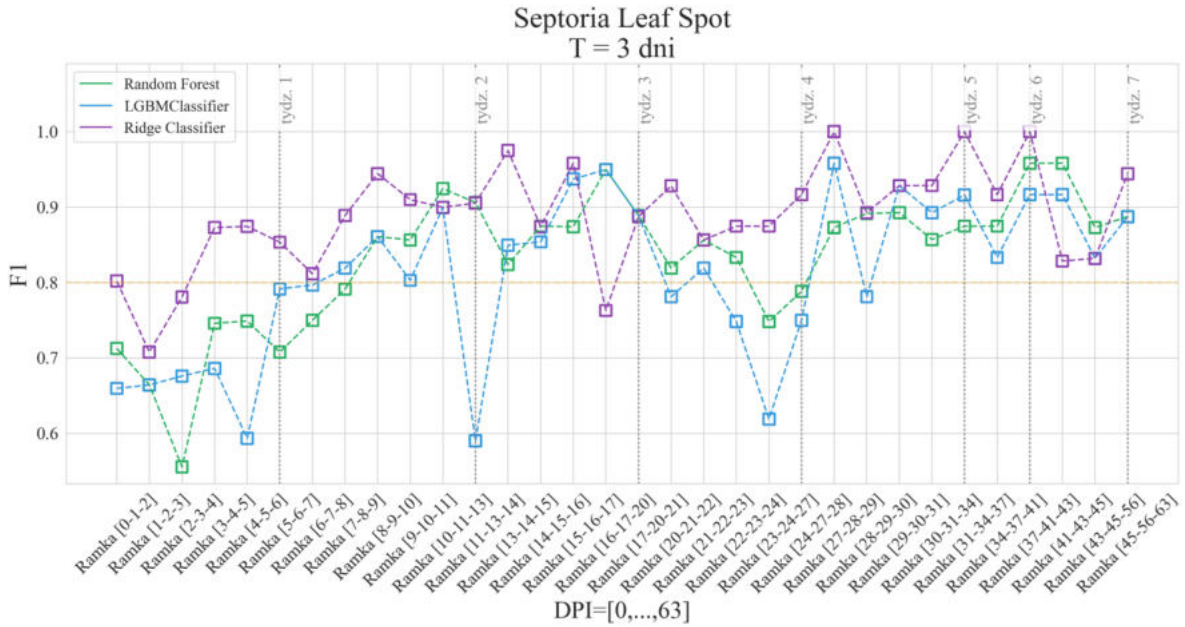
Pomiary z podziałem na poszczególne dni pomiarowe (T =1)



rys. 69 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę SL, T=1

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.81 (LGBM) do 0.91 (Ridge) przy odchyleniu standardowym na poziomie 0.101-0.109.

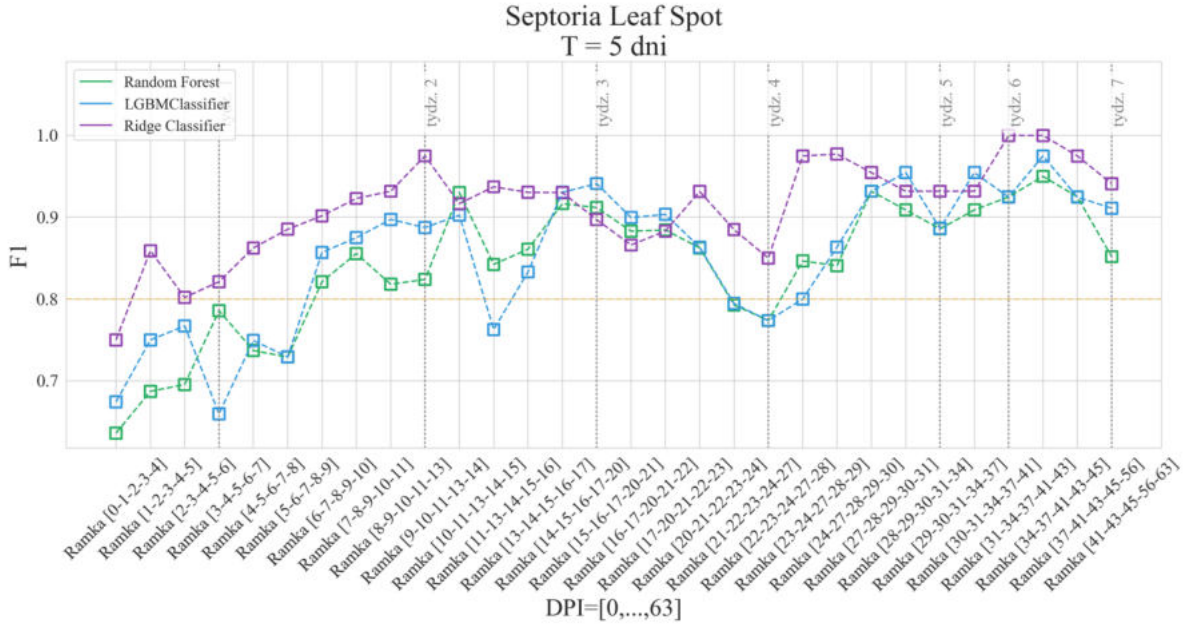
Pomiary z 3 kolejnych dni (T =3)



rys. 70 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę SL, T=3

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.83 (LGBM) do 0.9 (Ridge) przy odchyleniu standardowym na poziomie 0.069-0.114.

Pomiary z 5 kolejnych dni (T =5)



rys. 71 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę SL, T=5

Mediana wskaźnika F1 dla poszczególnych klasyfikatorów wahała się od 0.85 (Random Forest) do 0.93 (Ridge) przy odchyleniu standardowym na poziomie 0.058-0.086.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (T=1), a pozostałymi (T=3, T=5) wynosiła:

- Dla T=1 vs. T=3: 0.0331
- Dla T=1 vs. T=5: 0.0488

10.1.7. Wnioski dla wyników w ujęciu chorób

Wnioski:

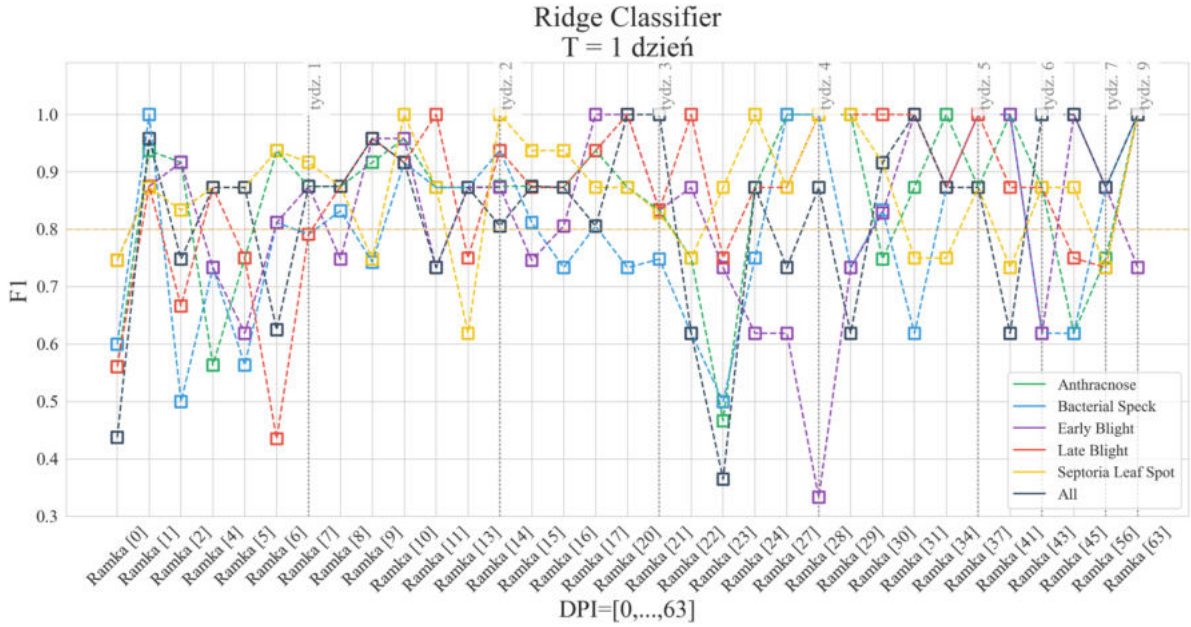
1. Odchylenie standardowe uzyskanych wyników dla ujęcia jednodniowego dawało gorsze wyniki wskaźnika F1 przy jednoczesnej większej dystrybucji wartości (odchylenie standardowe),
2. Dla dnia DPI=23 uzyskano znacznie pogorszone wyniki dla choroby Early Blight. Może być to spowodowane błędnie wykonanymi pomiarami spektralnymi, jednak wymagałoby to pogłębionych badań,
3. Największą poprawę wyników pomiędzy podejściem jednodniowym (T=1), a pozostałymi (T=3, T=5) odnotowano dla choroby Anthracnose, tj. T=3: poprawa o 0.054 w porównaniu do T=1; T=5: poprawa o 0.0598 w porównaniu do T=1.

10.2. Wyniki w ujęciu rodzaju algorytmu klasyfikacyjnego

Zestawienie wyników w ujęciu klasyfikatorów pozwala na porównanie skuteczności działania poszczególnych klasyfikatorów dla konkretnych chorób.

10.2.1. Ridge Classifier

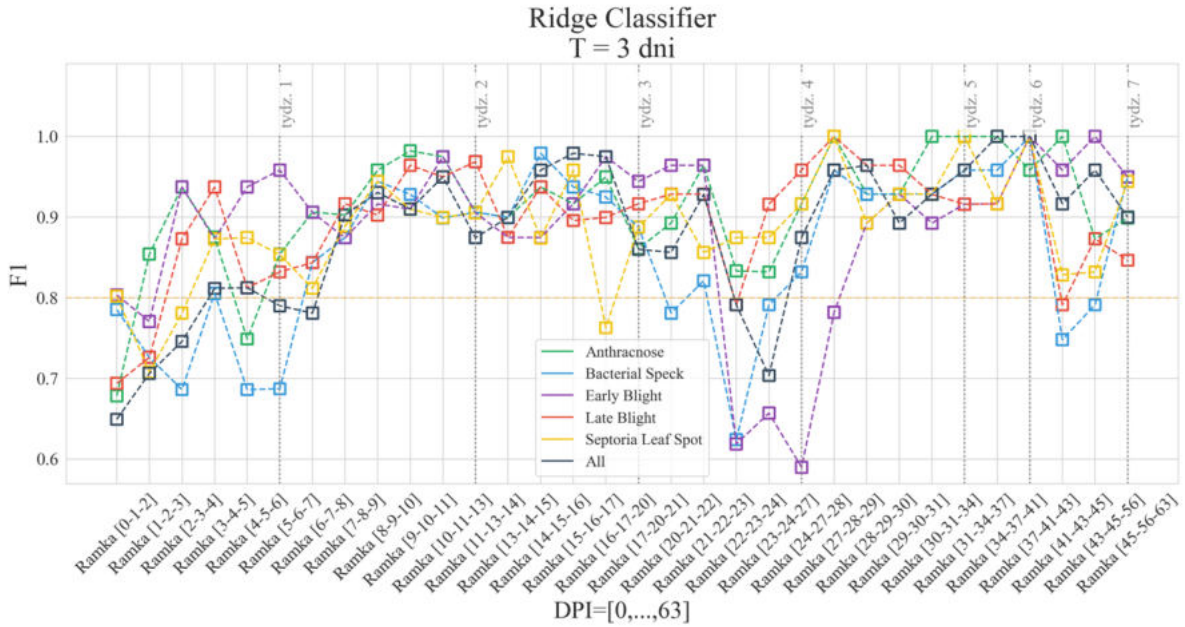
Pomiary z podziałem na poszczególne dni pomiarowe (T =1)



rys. 72 Wyniki treningu klasyfikatora Ridge Classifier dla poszczególnych przedziałów czasowych (DPI), T=1

Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.81 (BS) do 0.87 (AN, LB, SL, ALL) przy odchyleniu standardowym na poziomie 0.096-0.16. Jak można zauważyć na przedstawionym wykresie, różnice dla konkretnych chorób w tym podejściu były wysokie, rzędu nawet ponad 0.65 (EB, DPI=28 i DPI=31)

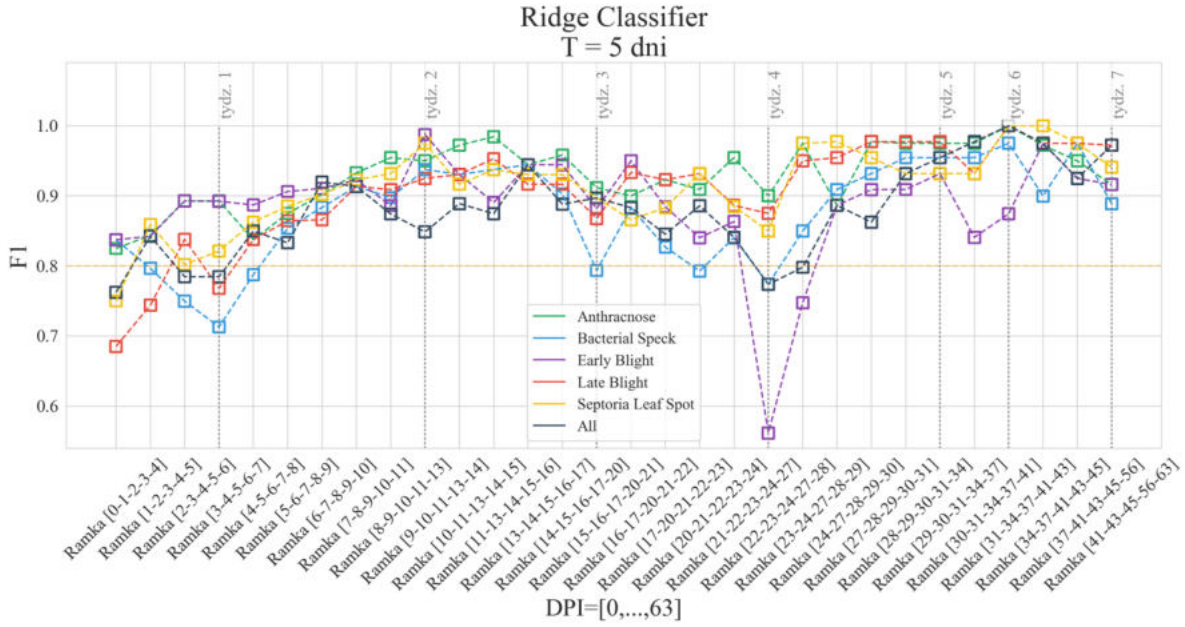
Pomiary 3 kolejnych dni (T =3)



rys. 73 Wyniki treningu klasyfikatora Ridge Classifier dla poszczególnych przedziałów czasowych (DPI), T=3

Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.9 (BS, ALL) do 0.92 (AN, LB) przy odchyleniu standardowym na poziomie 0.069-0.101. Zauważalny jest bardzo duży spadek wyników dla BS i EB dla DPI=[22-23-24] oraz DPI=[24-27-28].

Pomiary 5 kolejnych dni (T =5)



rys. 74 Wyniki treningu klasyfikatora Ridge Classifier dla poszczególnych przedziałów czasowych (DPI), T=5

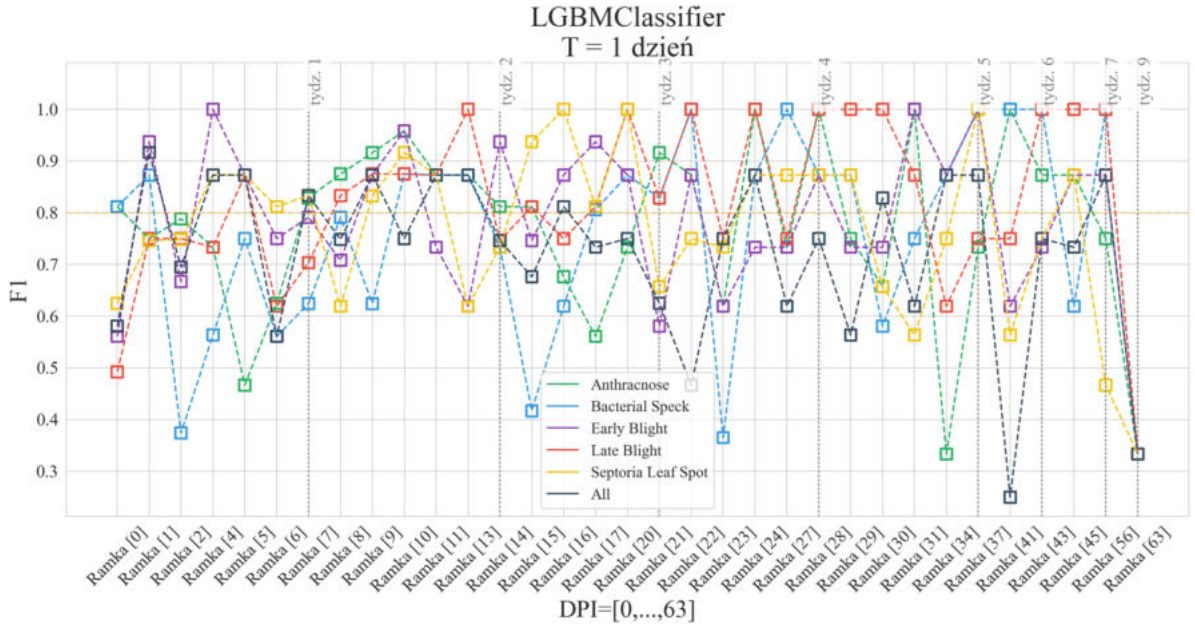
Wyniki dla klasyfikatora brzegowego (Ridge Classifier), przy *T*=5 dla ramki danych [22-23-24-27-28] dla choroby Early Blight spadły do poziomu poniżej 0.5616.
Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.88 (ALL) do 0.94 (AN) przy odchyleniu standardowym na poziomie 0.047-0.077.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (T=1), a pozostałymi (T=3, T=5) wynosiła:

- Dla T=1 vs. T=3: 0.0528
- Dla T=1 vs. T=5: 0.0547

10.2.2. LGBM Classifier

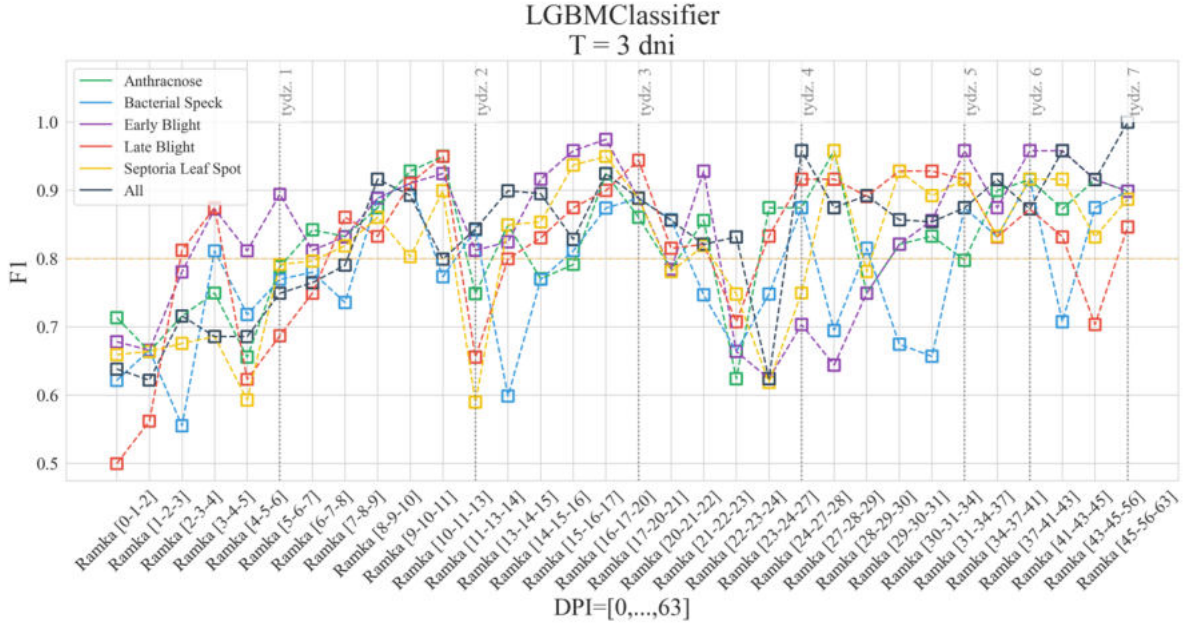
Pomiary z podziałem na poszczególne dni pomiarowe (T =1)



rys. 75 Wyniki treningu klasyfikatora LGBM Classifier dla poszczególnych przedziałów czasowych (DPI), T=1

Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.75 (ALL) do 0.83 (LB) przy odchyleniu standardowym na poziomie 0.151-0.199.

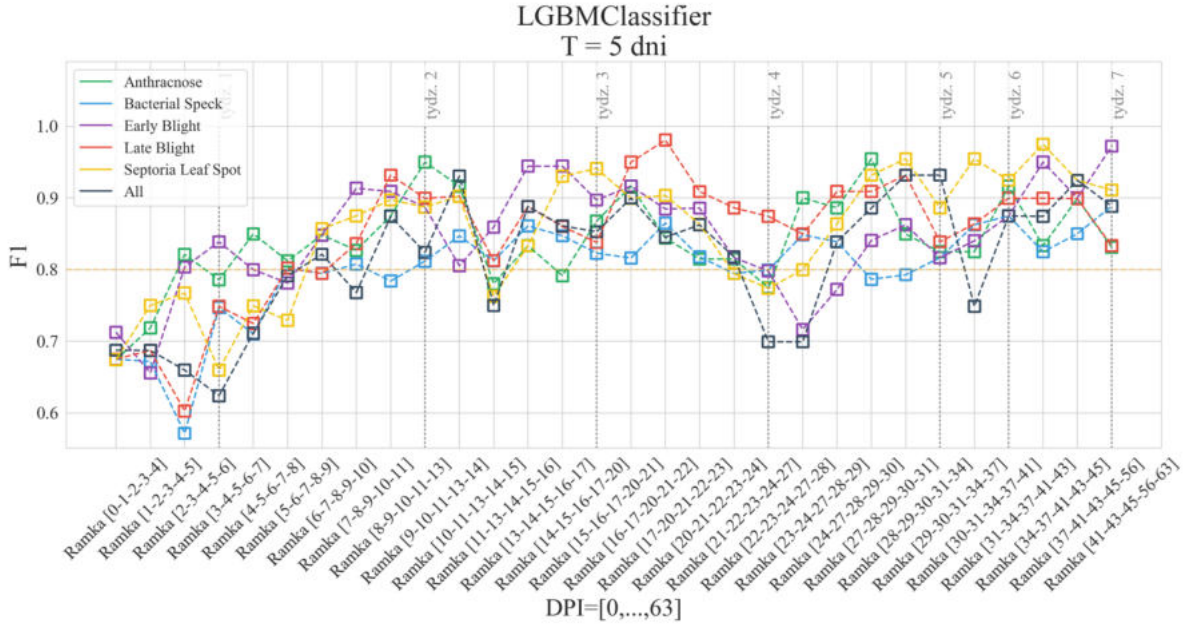
Pomiary 3 kolejnych dni (T =3)



rys. 76 Wyniki treningu klasyfikatora LGBM Classifier dla poszczególnych przedziałów czasowych (DPI), T=3

Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.78 (BS) do 0.86 (ALL) przy odchyleniu standardowym na poziomie 0.089-0.115.

Pomiary 5 kolejnych dni (T =5)



rys. 77 Wyniki treningu klasyfikatora LGBM Classifier dla poszczególnych przedziałów czasowych (DPI), T=5

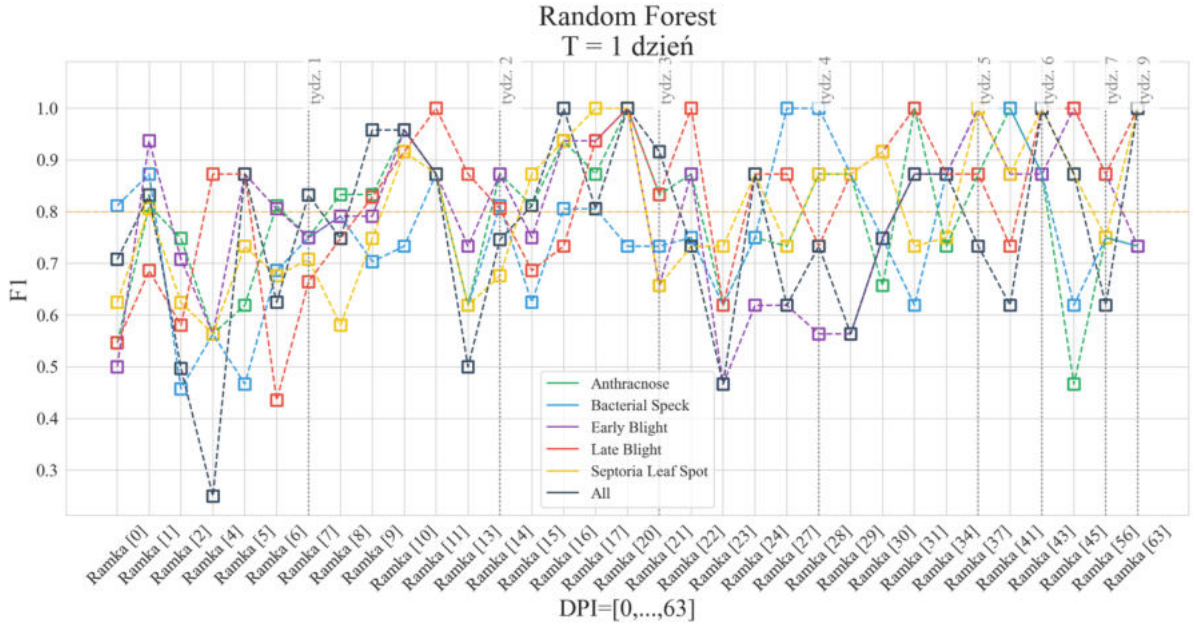
Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.81 (BS) do 0.88 (SL) przy odchyleniu standardowym na poziomie 0.062-0.09.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (T=1), a pozostałymi (T=3, T=5) wynosiła:

- Dla T=1 vs. T=3: 0.0489
- Dla T=1 vs. T=5: 0.0554

10.2.1. Random Forest Classifier

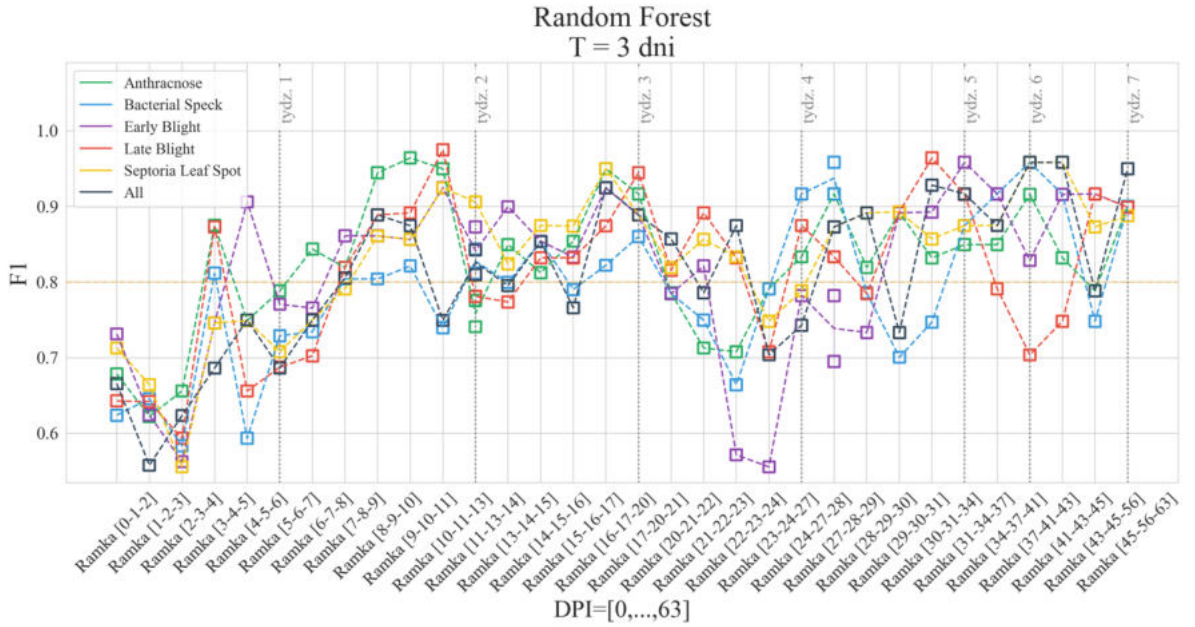
Pomiary z podziałem na poszczególne dni pomiarowe (T =1)



rys. 78 Wyniki treningu klasyfikatora Random Forest Classifier dla poszczególnych przedziałów czasowych (DPI), T=1

Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.75 (BS, SL) do 0.87 (LB) przy odchyleniu standardowym na poziomie 0.135-0.179.

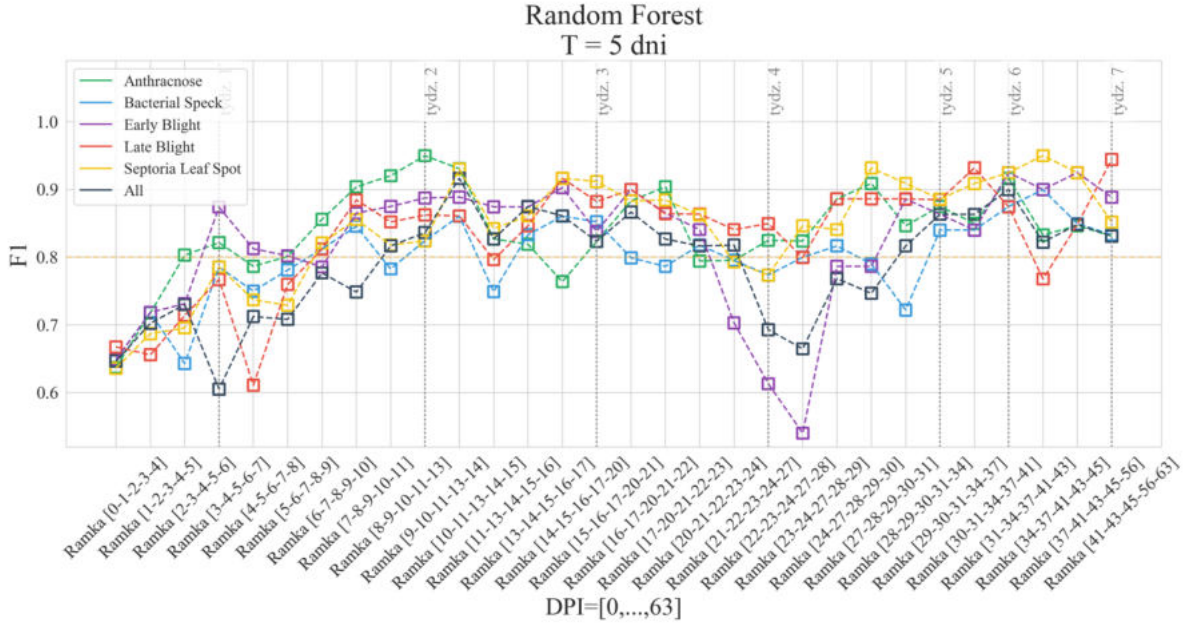
Pomiary 3 kolejnych dni (T =3)



rys. 79 Wyniki treningu klasyfikatora Random Forest Classifier dla poszczególnych przedziałów czasowych (DPI), T=3

Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.8 (BS) do 0.87 (SL) przy odchyleniu standardowym na poziomie 0.087-0.109.

Pomiary 5 kolejnych dni (T =5)



rys. 80 Wyniki treningu klasyfikatora Random Forest Classifier dla poszczególnych przedziałów czasowych (DPI), T=5

Mediana wskaźnika F1 dla poszczególnych chorób wahała się od 0.81 (BS) do 0.86 (EB, LB) przy odchyleniu standardowym na poziomie 0.06-0.96.

Średnia poprawa wyniku dla poszczególnych podejść pomiędzy podejściem jednodniowym (*T*=1), a pozostałymi (*T*=3, *T*=5) wynosiła:

- Dla *T*=1 vs. *T*=3: 0.0265,
- Dla *T*=1 vs. *T*=5: 0.0221.

10.2.2. Wnioski dla wyników w ujęciu klasyfikatorów

Wnioski:

1. Największą poprawę wyników pomiędzy podejściem jednodniowym (*T*=1), a pozostałymi (*T*=3, *T*=5) odnotowano dla klasyfikatorów:
    a. Ridge Classifier dla *T*=3: poprawa o 0.0554 w porównaniu do *T*=1,
    b. LGBM Classifier dla *T*=5: poprawa o 0.0554 w porównaniu do *T*=1.

10.3. Zestawienie wyników dla różnych wartości parametru *T*

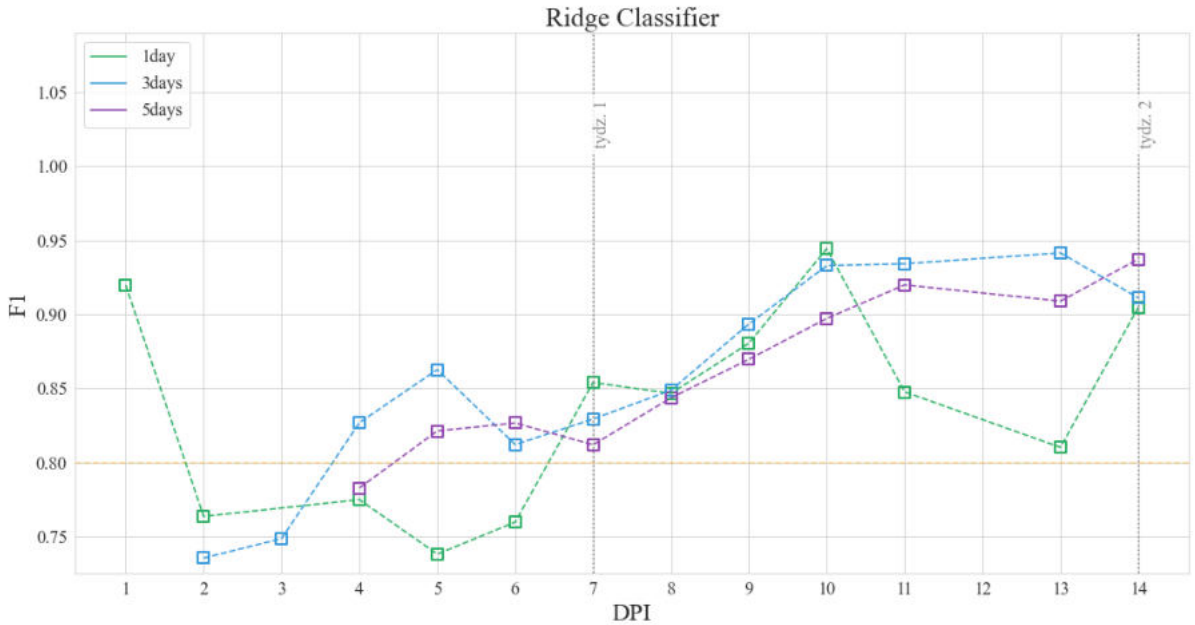
Powyższe podrozdziały prezentują wyniki w ujęciu chorobowym oraz klasyfikacyjnym. W poniższym podrozdziale postanowiono zestawić różne podejścia przetwarzania wstępnego zbioru.

Uśrednione wyniki dla *T* równego kolejno 1, 3 oraz 5 obrazują, że trening klasyfikatorów na danych tylko z 1 dnia jest bardzo czuły na błędy, co zaobserwować można chociażby dla DPI=23 zawierający najprawdopodobniej błędne pomiary dla EB. Z punktu widzenia

diagnostyki uprawy najważniejszy jest okres 2 pierwszych tygodni od zakażenia, dlatego w poniższej analizie skupiono się na tym newralgicznym okresie.
Najlepszy uśredniony wynik uzyskano dla klasyfikatora brzegowy (Ridge Classifier), dlatego przedstawiono wyniki dla niego:

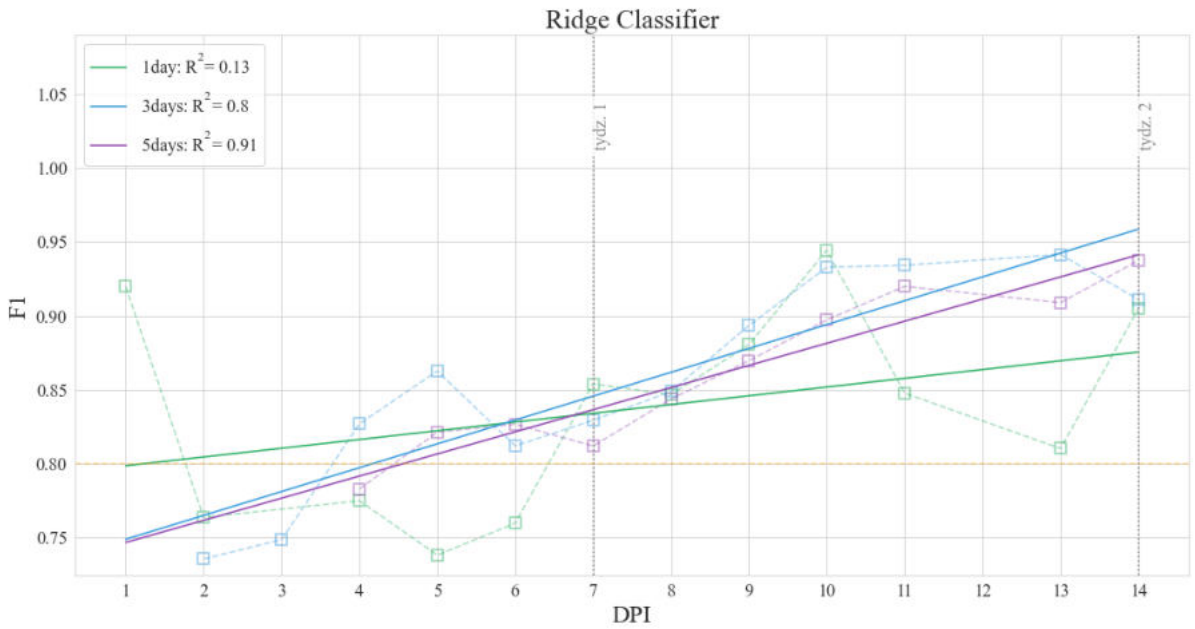
a) Dla *T*=1 uzyskano średni wynik F1=0.836,
b) Dla *T*=3 uzyskano średni wynik F1=0.889,
c) Dla *T*=5 uzyskano średni wynik F1=0.898.

Dodatkowo na przedstawionych wykresach wyznaczono funkcję liniową oraz wielomianową drugiego i trzeciego stopnia.

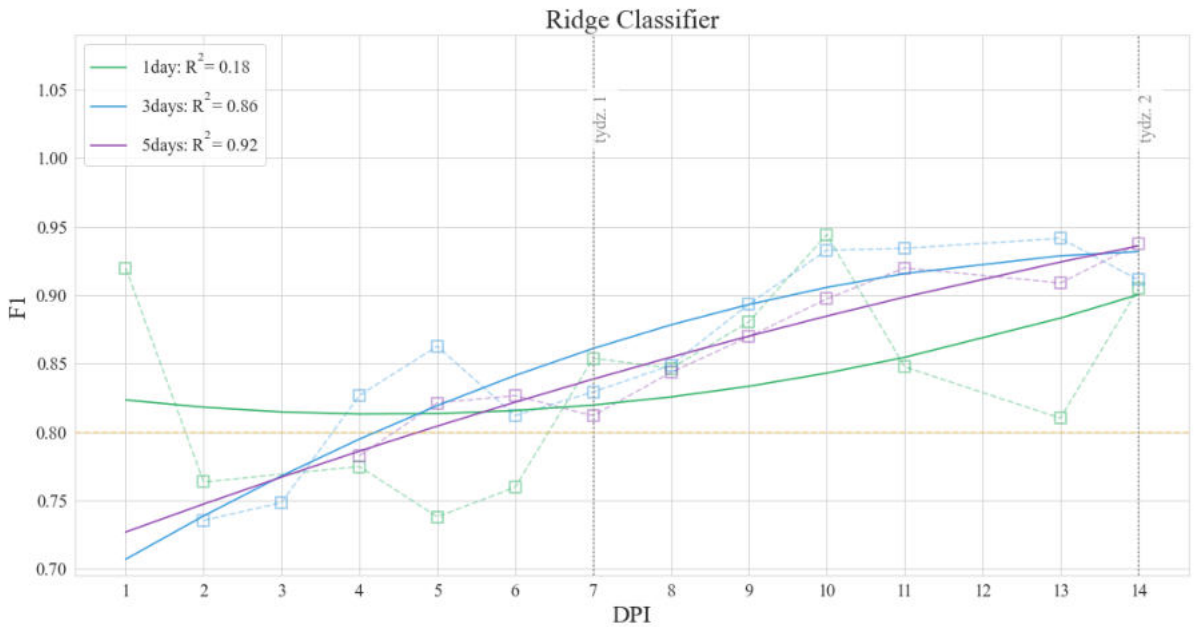


rys. 81 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla najlepszego klasyfikatora (Ridge Classifier)

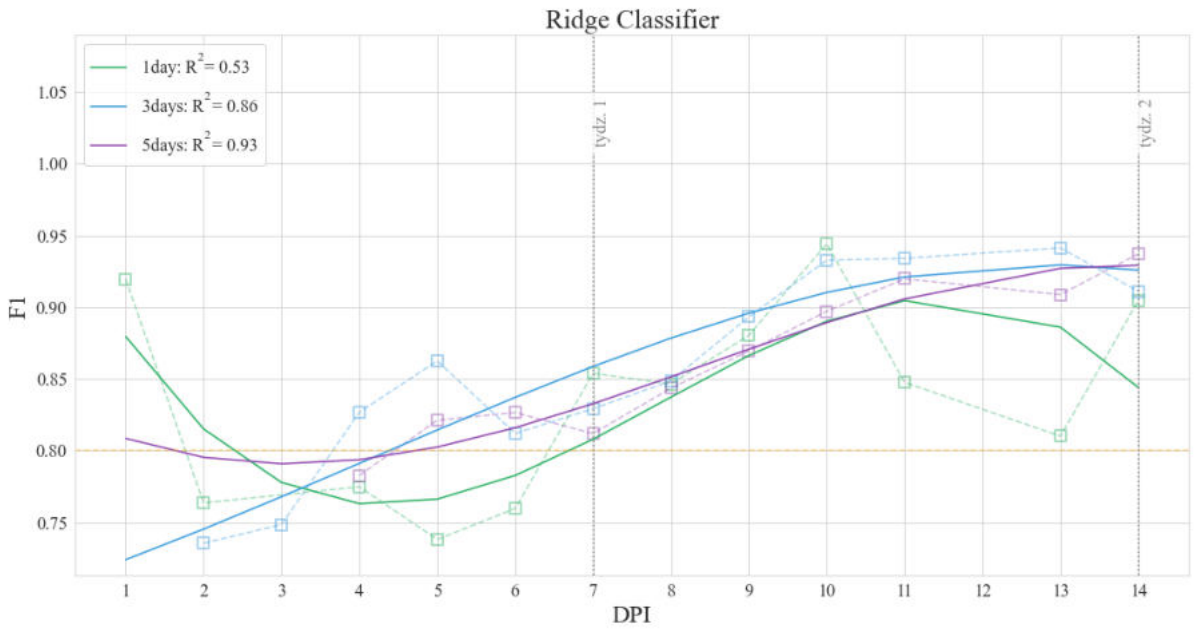
Dla przedstawionych wyników wyznaczono funkcje trendu na podstawie funkcji liniowej oraz wielomianu I-ego i II-ego stopnia.



rys. 82 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla najlepszego klasyfikatora (Ridge Classifier) z wyrysowanymi liniami trendu wielomianu pierwszego stopnia



rys. 83 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla najlepszego klasyfikatora (Ridge Classifier) z wyrysowanymi liniami trendu wielomianu drugiego stopnia

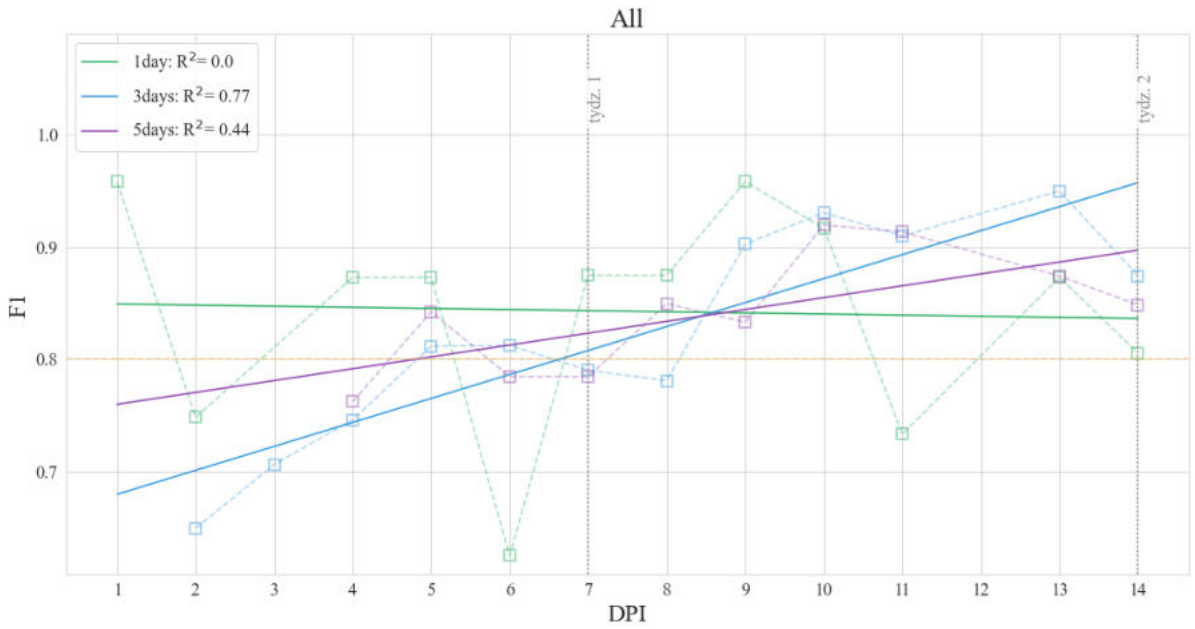


rys. 84 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla najlepszego klasyfikatora (Ridge Classifier) z wyrysowanymi liniami trendu wielomianu trzeciego stopnia

Dla uzyskanych wyników można zauważyć trend rosnący, co wynikać może z faktu pojawiania się kolejnych objawów chorobowych i jednocześnie wpływać na jakość dopasowania modelu. Zaobserwować można również, że model dla agregacji jednodniowej (*T*=1) charakteryzuje się wyższym odchyleniem standardowym, co może świadczyć o podatności na potencjalne błędy pomiarowe.

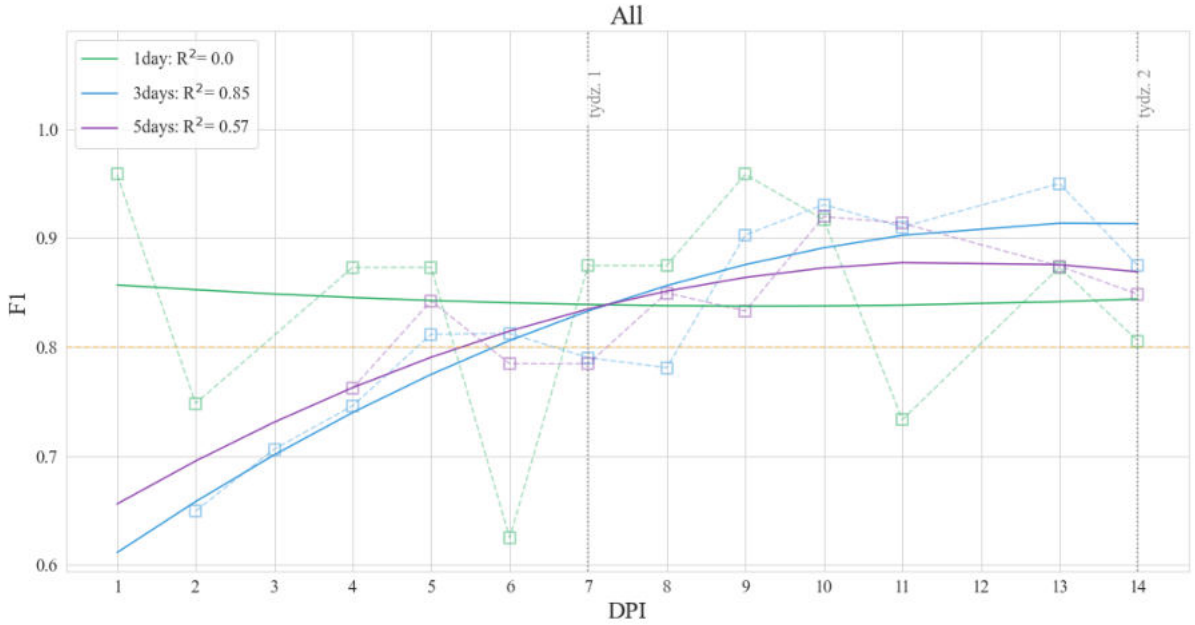
10.3.1. Podział na poszczególne choroby

Dla otrzymanych wyników określono funkcje trendu w celu porównaniu różnych podejść do zagadnienia diagnostyki na podstawie pomiarów hiperspektralnych. W tym celu wyznaczono modele trendu poszczególnych podejść z rozbiciem na poszczególne choroby dla wskaźnika F1, a następnie dla każdego z modelu trendu określono współczynnik determinacji (R^2) oraz miejsce przecięcia z punktem 0.8 osi Y (F1). Sprawdzano kilka funkcji trendu. Najlepsze wyniki dopasowania spośród badanych funkcji wykazały: funkcja liniowa oraz funkcja wielomianu drugiego stopnia.
Niektóre z funkcji trendu osiągały wyniki powyżej F1=0,8 lub przecięcie z funkcją y=0,8 wykraczało poza rozpatrywany zakres (1-14 dnia pomiarowego).



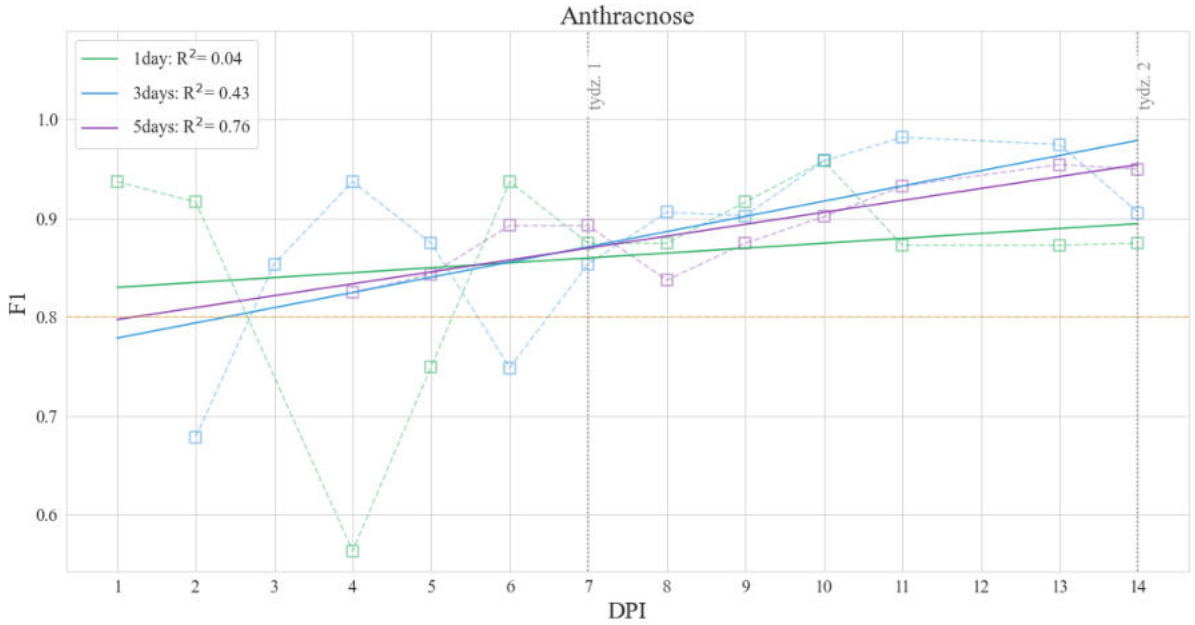
rys. 85 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku wszystkich chorób z wyrysowanymi liniami trendu funkcji liniowej

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,72, a poziom 0,8 został osiągnięty dla t=6,2.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,3, a poziom 0,8 został osiągnięty dla t=3,4.



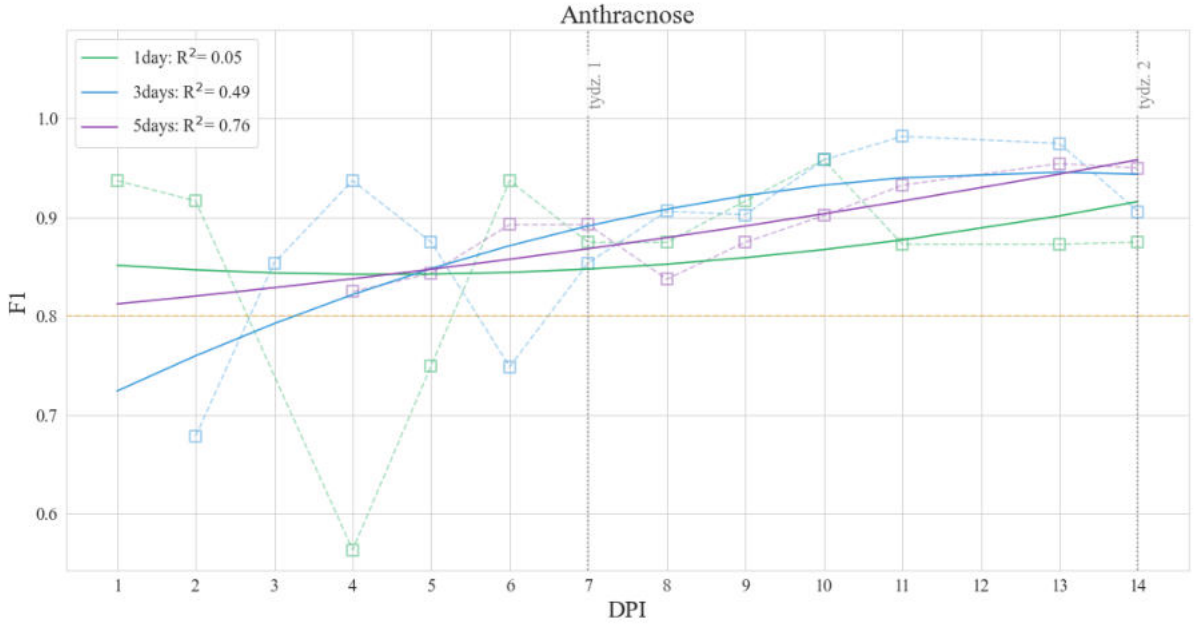
rys. 86 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku wszystkich chorób z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,78, a poziom 0,8 został osiągnięty dla t=5,8.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,4, a poziom 0,8 został osiągnięty dla t=5,4.



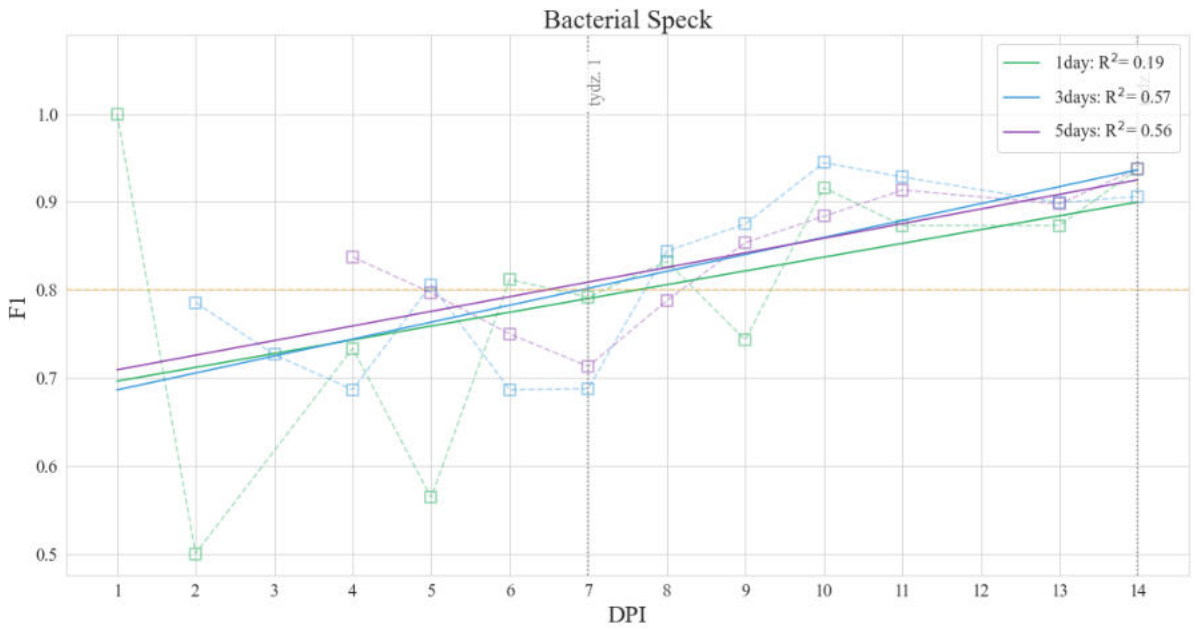
rys. 87 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Anthracnose z wyrysowanymi liniami trendu funkcji liniowej

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,28, a poziom 0,8 został osiągnięty dla każdego t w badanym zakresie.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,69, a poziom 0,8 został osiągnięty dla każdego t w badanym zakresie.



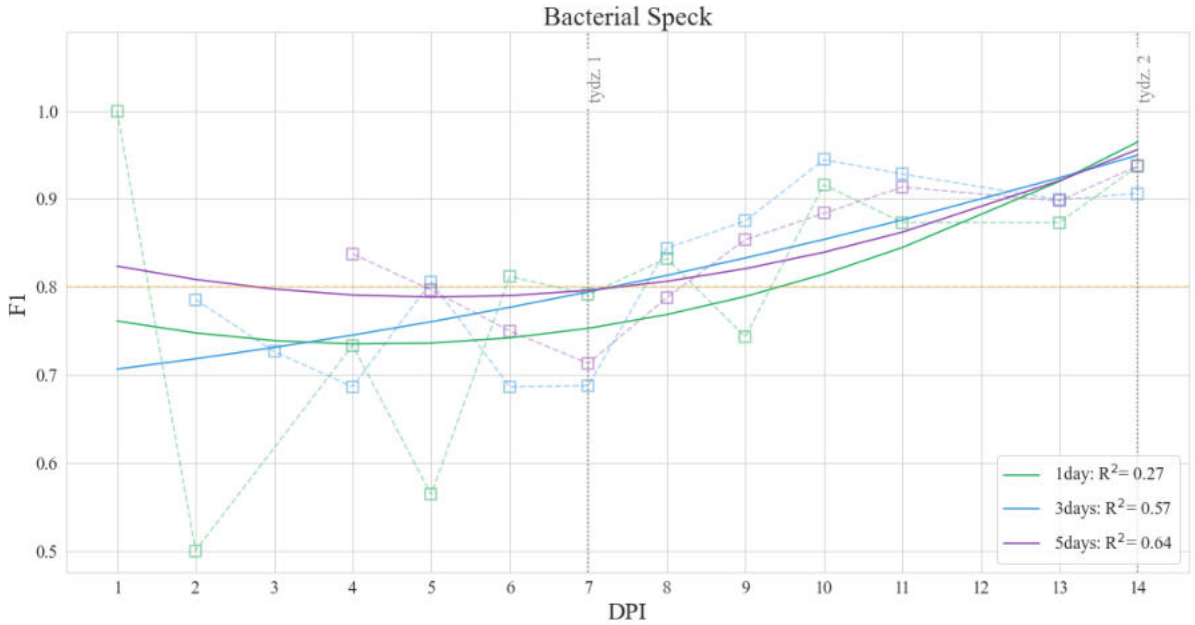
rys. 88 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Anthracnose z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,28, a poziom 0,8 został osiągnięty dla każdego t w badanym zakresie.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,71, a poziom 0,8 został osiągnięty dla każdego t w badanym zakresie.



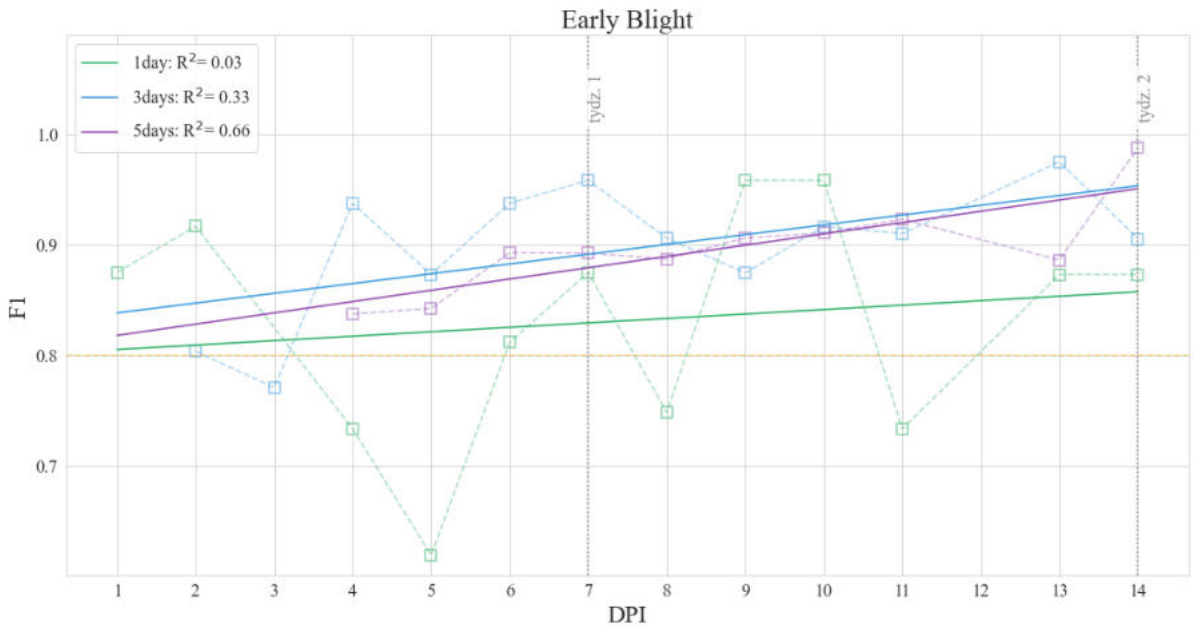
rys. 89 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Bacterial Speck z wyrysowanymi liniami trendu funkcji liniowej

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,65, a poziom 0,8 został osiągnięty dla t=7,4.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,75, a poziom 0,8 został osiągnięty dla t=7,6.



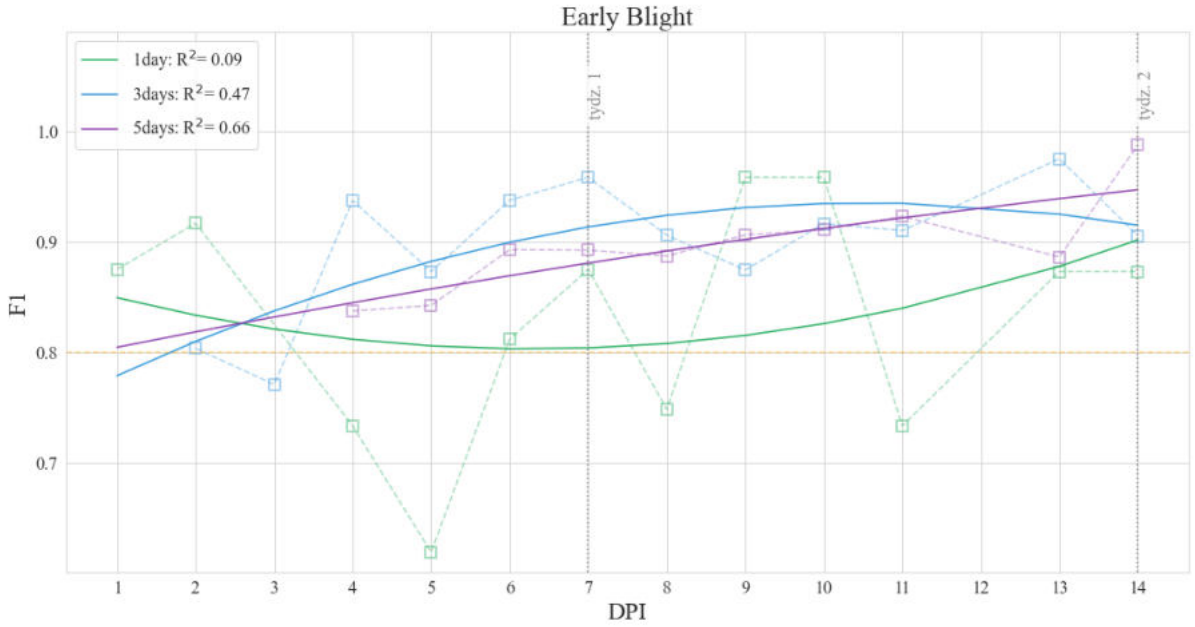
rys. 90 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Bacterial Speck z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,66, a poziom 0,8 został osiągnięty dla t=7.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,75, a poziom 0,8 został osiągnięty dla t=7,6.



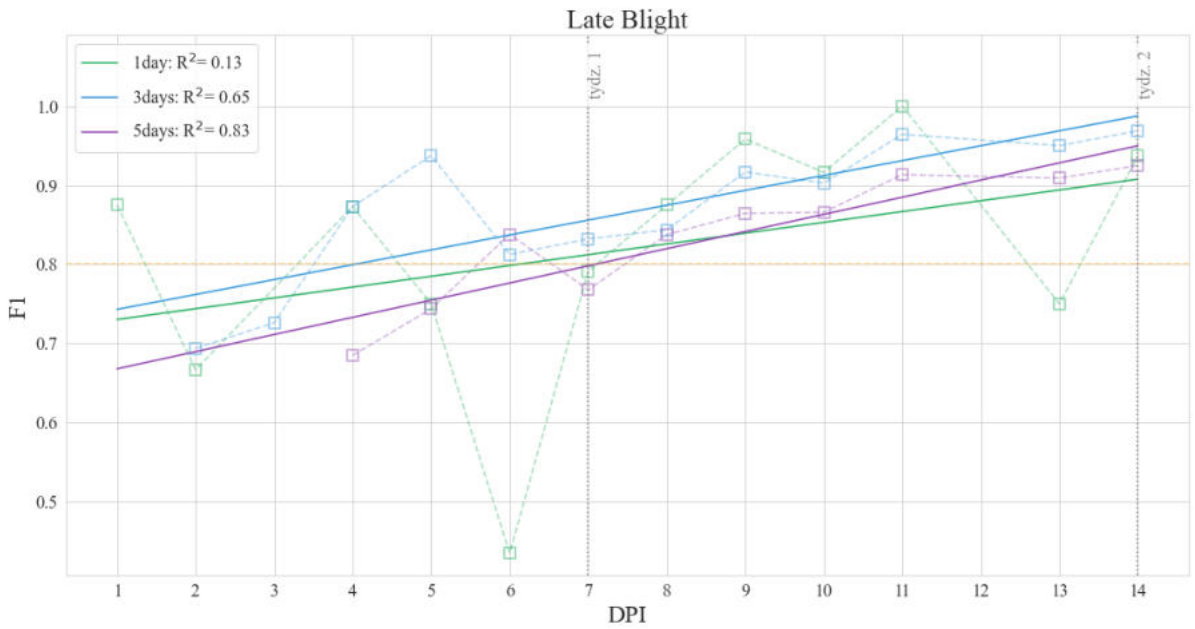
rys. 91 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Early Blight z wyrysowanymi liniami trendu funkcji liniowej

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,2, a poziom 0,8 został osiągnięty dla każdego t w badanym zakresie.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,56, a poziom 0,8 został osiągnięty dla każdego t w badanym zakresie.



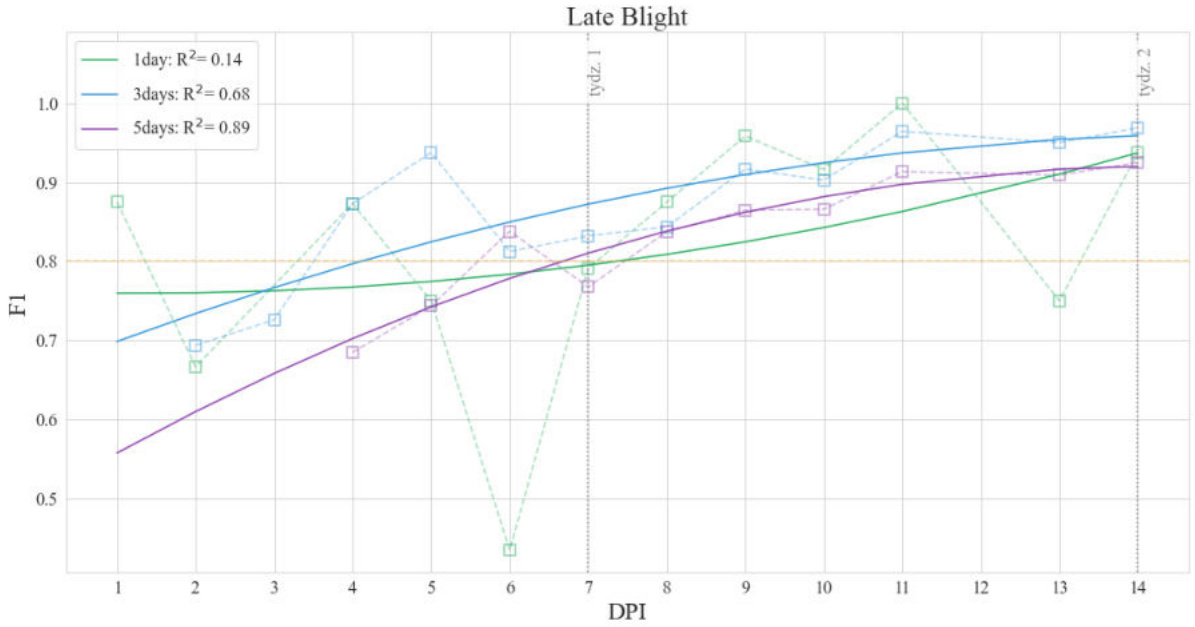
rys. 92 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Early Blight z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,31, a poziom 0,8 został osiągnięty dla t=1,4.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,56, a poziom 0,8 został osiągnięty dla każdego t w badanym zakresie.



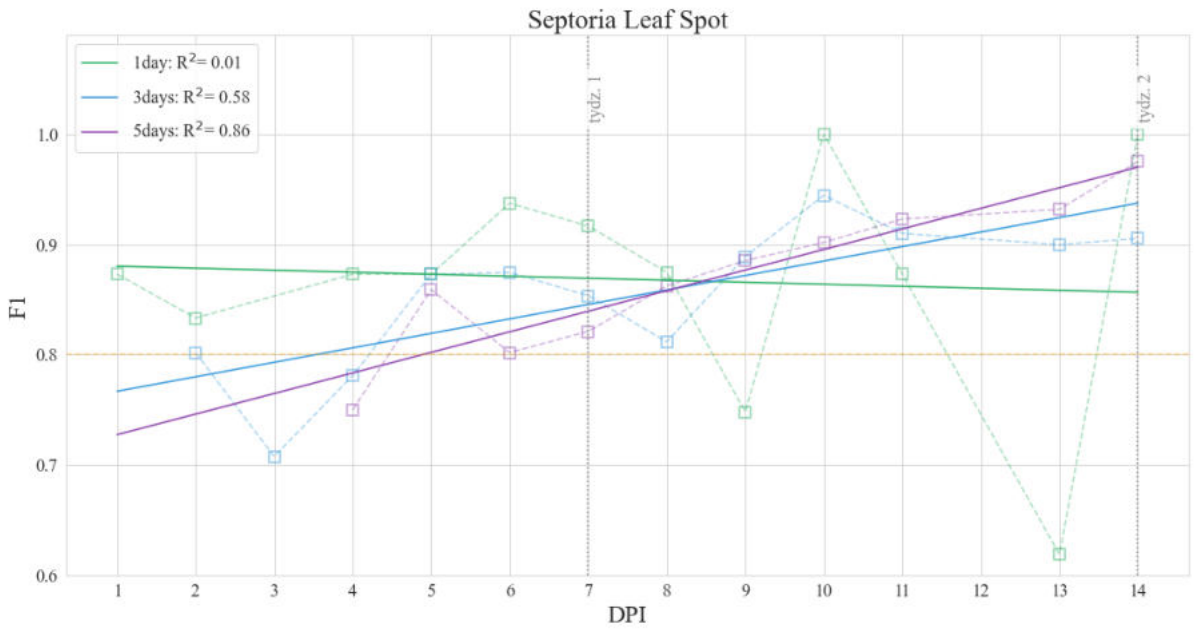
rys. 93 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Late Blight z wyrysowanymi liniami trendu funkcji liniowej

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,55, a poziom 0,8 został osiągnięty dla t=2,8.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,8, a poziom 0,8 został osiągnięty dla t=6,4.



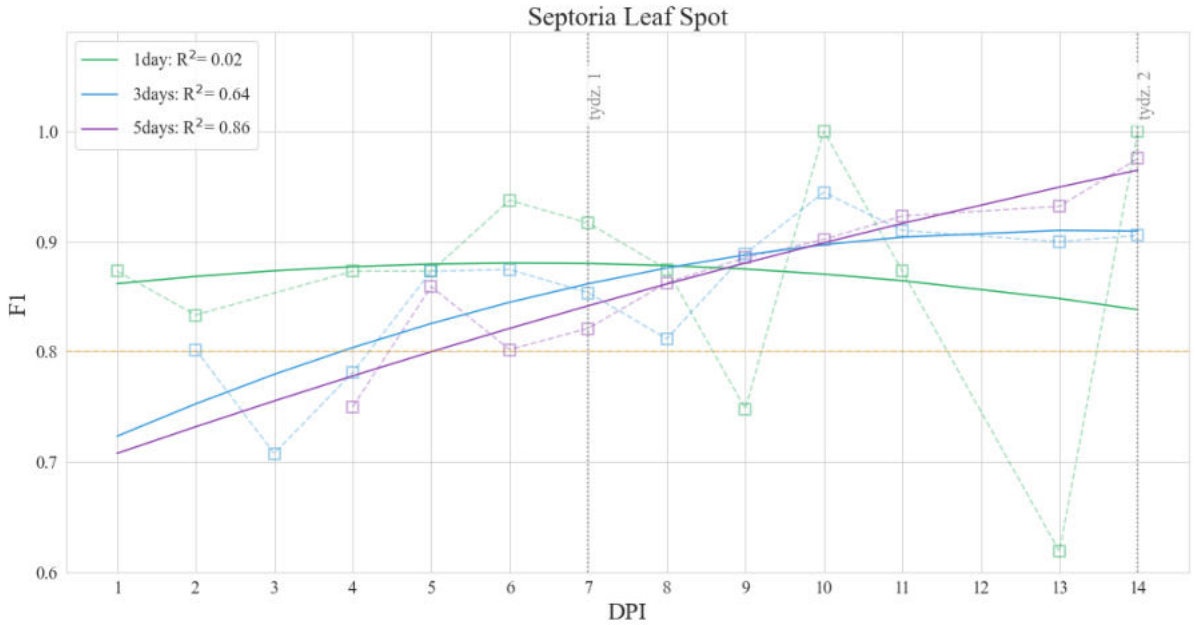
rys. 94 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Late Blight z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,56, a poziom 0,8 został osiągnięty dla t=3,2.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,83, a poziom 0,8 został osiągnięty dla t=6,5.



rys. 95 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Septoria Leaf Spot z wyrysowanymi liniami trendu funkcji liniowej

Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,56, a poziom 0,8 został osiągnięty dla t=4.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,83, a poziom 0,8 został osiągnięty dla t=4.

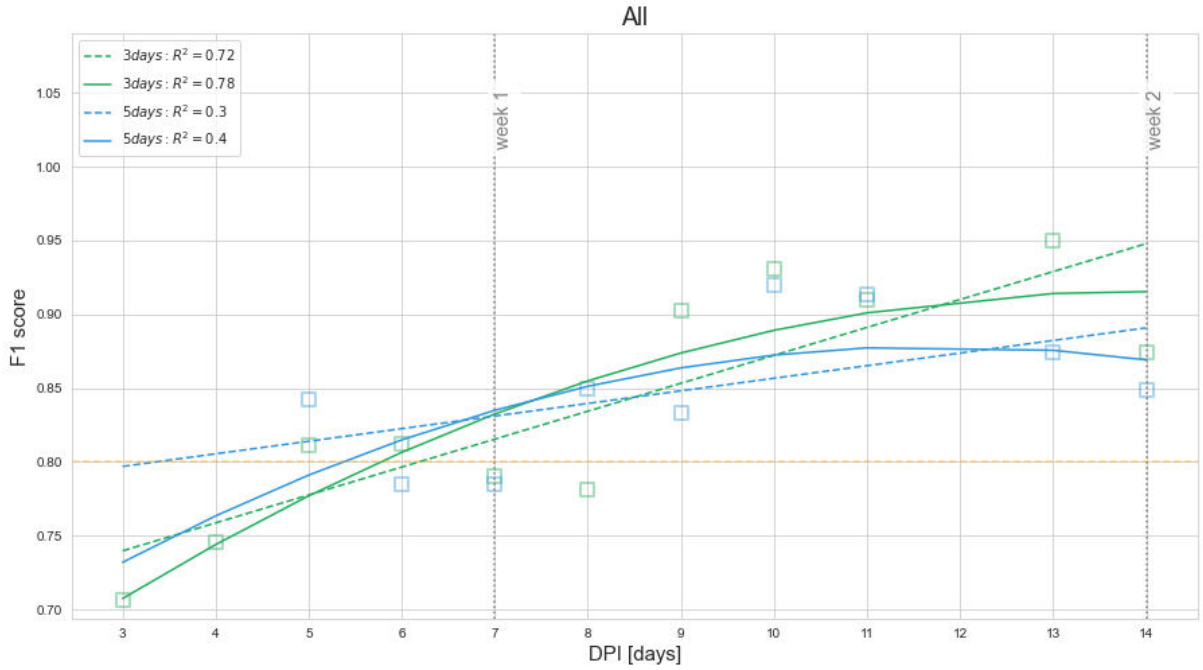


rys. 96 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Septoria Leaf Spot z wyrysowanymi liniami funkcji wielomianowej II-ego stopnia

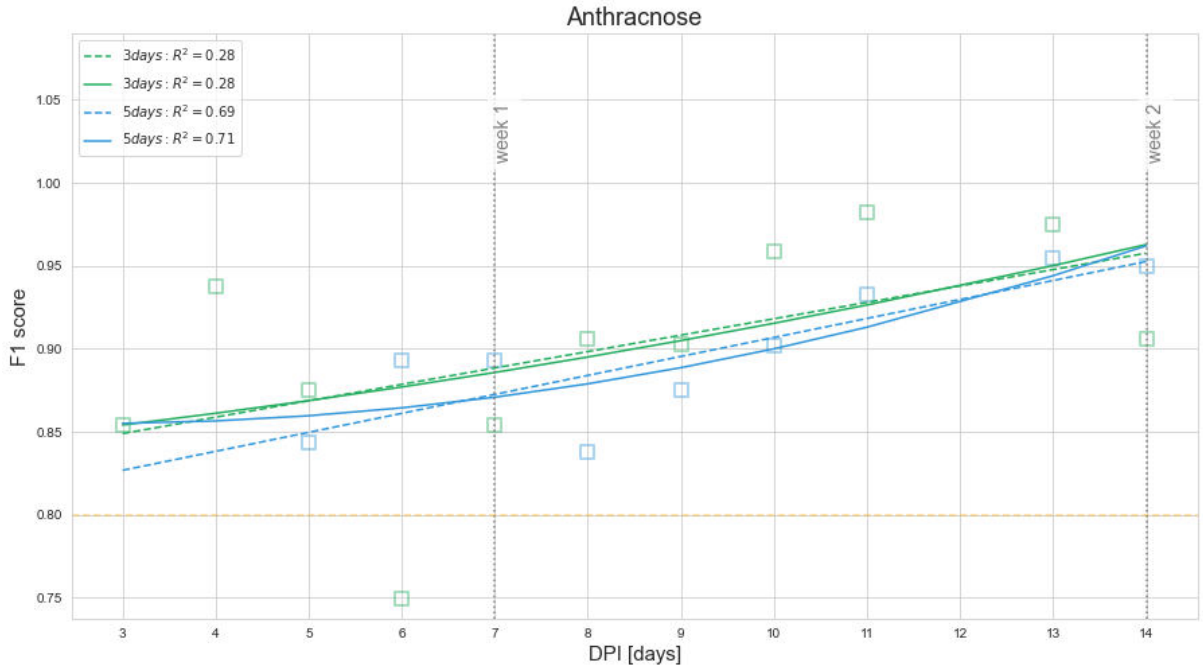
Dla ujęcia 3-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,72, a poziom 0,8 został osiągnięty dla t=4,6.
Dla ujęcia 5-dniowego współczynnik determinacji modelu dopasowania trendu wyniósł 0,85, a poziom 0,8 został osiągnięty dla każdego t w badanym zakresie.

10.3.2. Połączenie wykresów

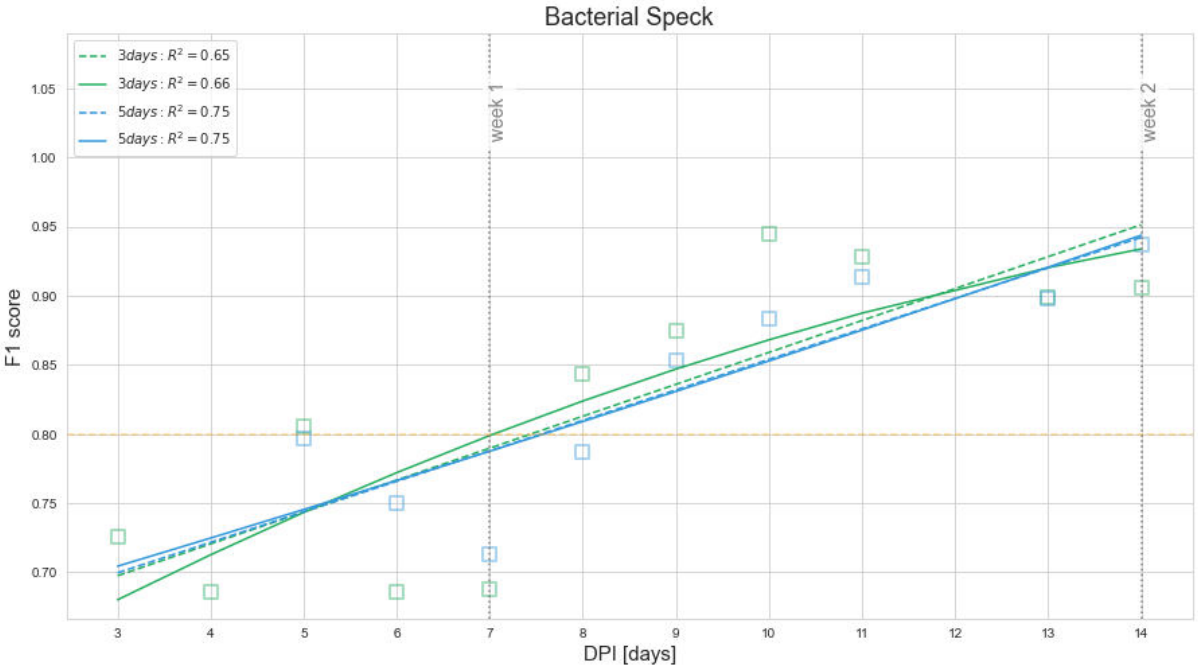
Dla lepszego porównania wyników zdecydowano się na naniesienie funkcji liniowej oraz wielomianowej II-ego stopnia na jeden wykres. Pojedynczy wykres omawia podejście do jednego przypadku badawczego (przedstawionych w rozdziale 10.1). Wyniki pomiarowe przedstawiono w formie punktów, bez ich łączenia za pomocą linii. Wyniki zawężono do klasyfikatora brzegowego (Ridge Classifier) oraz ujęć 3- i 5-dniowych.



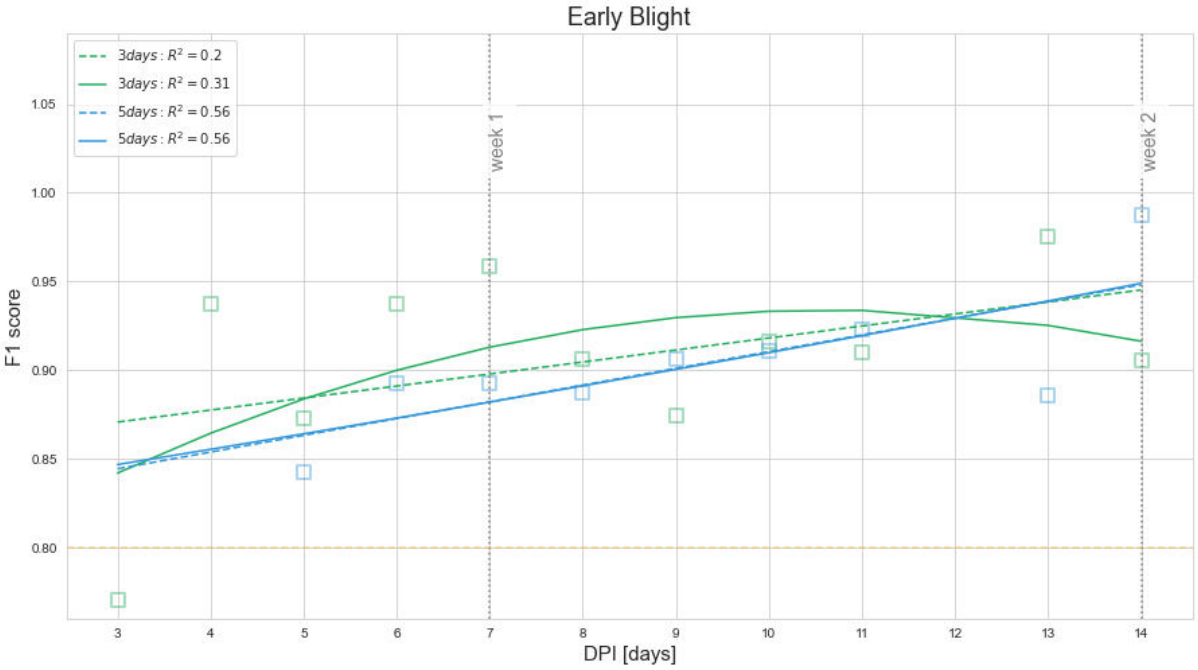
rys. 97 Trend rosnący efektywności klasyfikatora dla wszystkich chorób, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5



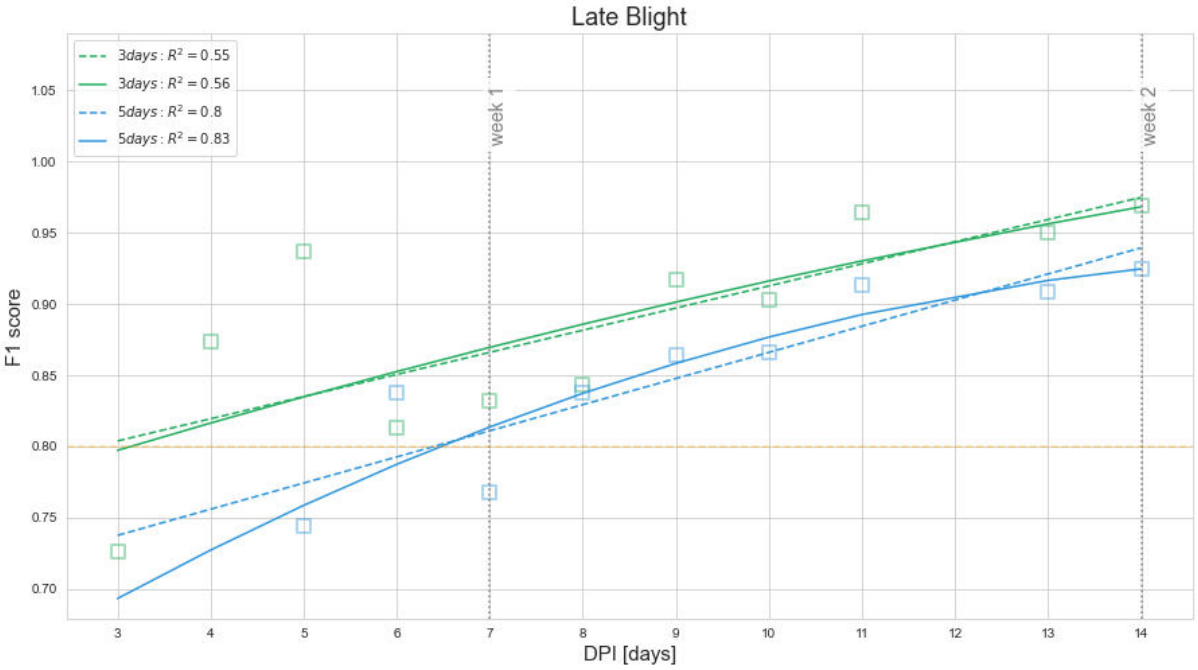
rys. 98 Trend rosnący efektywności klasyfikatora dla Anthracnose, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5



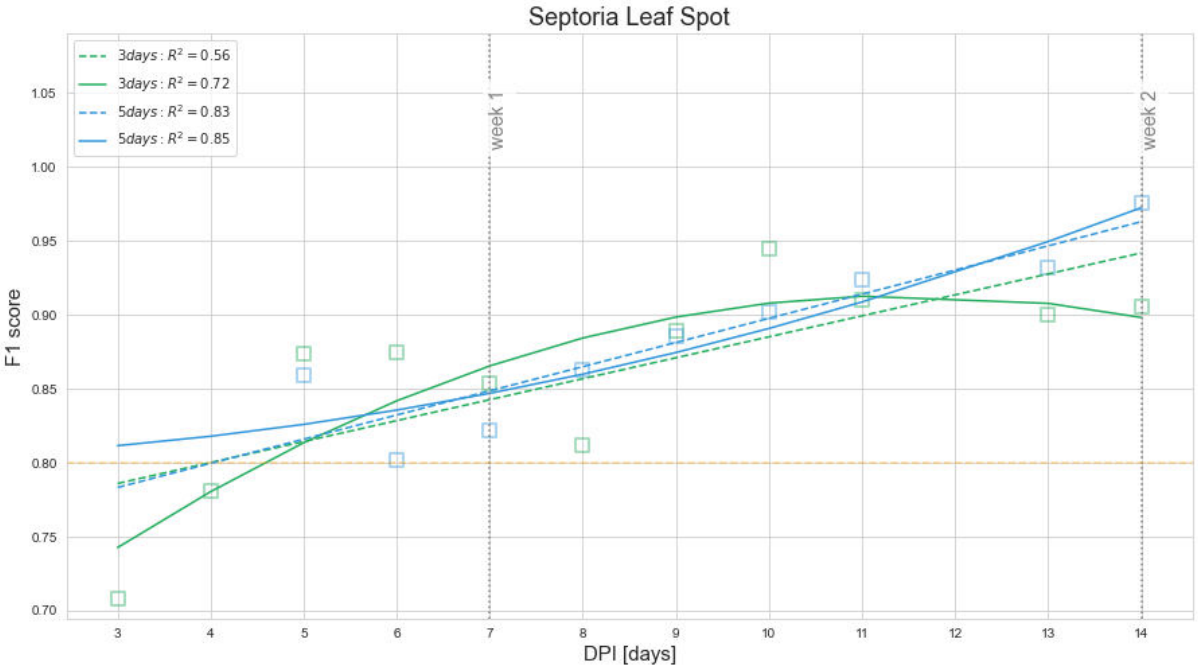
rys. 99 Trend rosnący efektywności klasyfikatora dla Bacterial Speck, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5



rys. 100 Trend rosnący efektywności klasyfikatora dla Early Blight, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5



rys. 101 Trend rosnący efektywności klasyfikatora dla Late Blight, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5



rys. 102 Trend rosnący efektywności klasyfikatora dla Septoria Leaf Spot, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5

W poniższej tabeli przedstawiono syntetyczne zestawienie punktu przecięcia z osią Y na poziomie 0.8 współczynnika F1, co determinuje najkrótszy możliwy czas, w jakim można przeprowadzić wystarczająco pewną diagnostykę chorób za pomocą zaproponowanych narzędzi oraz metod.

Wyznaczono również współczynnik determinacji linii trendu do uzyskanych wyników, a całość podzielono na sposób agregacji (*T*), choroby oraz funkcję dopasowania. Ze względu na

wysokie odchylenie standardowe wyników dla *T*=1 dzień wyniki te zostały pominięte w zestawieniu.

tab. 25 Podsumowanie wyników analizy diagnostyki infekcji w czasie z podziałem na badane choroby oraz wyznaczenie najwcześniejszego momentu możliwej diagnostyki podany w nawiasie

Sposób agregacji (*T*)	Choroba	Funkcja dopasowania	R^2	Punkt przecięcia z y=0,8 dla t=<1,14>
3days	Wszystkie analizowane	Liniowa	0,72	6,2 (7)
5days	Wszystkie analizowane	Liniowa	0,30	3,4 (5)
3days	Wszystkie analizowane	Wielomian II stopnia	0,78	5,8 (6)
5days	Wszystkie analizowane	Wielomian II stopnia	0,40	5,4 (6)
3days	Anthracnose	Liniowa	0,28	- (3)
5days	Anthracnose	Liniowa	0,69	0,7 (5)
3days	Anthracnose	Wielomian II stopnia	0,28	- (3)
5days	Anthracnose	Wielomian II stopnia	0,71	- (5)
3days	Bacterial Speck	Liniowa	0,65	7,4 (8)
5days	Bacterial Speck	Liniowa	0,75	7,6 (8)
3days	Bacterial Speck	Wielomian II stopnia	0,66	7,0 (7)
5days	Bacterial Speck	Wielomian II stopnia	0,75	7,6 (8)
3days	Early Blight	Liniowa	0,20	- (3)
5days	Early Blight	Liniowa	0,56	- (5)
3days	Early Blight	Wielomian II stopnia	0,31	1,4 (3)
5days	Early Blight	Wielomian II stopnia	0,56	- (5)
3days	Late Blight	Liniowa	0,55	2,8 (3)
5days	Late Blight	Liniowa	0,80	6,4 (7)
3days	Late Blight	Wielomian II stopnia	0,56	3,2 (4)
5days	Late Blight	Wielomian II stopnia	0,83	6,5 (7)
3days	Septoria Leaf Spot	Liniowa	0,56	4,0 (7)
5days	Septoria Leaf Spot	Liniowa	0,83	4,0 (5)
3days	Septoria Leaf Spot	Wielomian II stopnia	0,72	4,6 (5)
5days	Septoria Leaf Spot	Wielomian II stopnia	0,85	- (5)

Uzyskane wyniki wskazują na możliwość diagnostyki badanych chorób w okresie krótszym niż 7 dni od zakażenia, co potwierdza tezę trzecią niniejszej pracy doktorskiej („Hiperspektralne pomiary reflektancji pozwalają na klasyfikację jednostek chorobowych *Solanum lycopersicum* w czasie szybszym niż metoda oceny wizualnej stosowana w praktyce"). Najlepsze wyniki dla wariantu kontrola – wszystkie rodzaje badanych chorób, uzyskano dla okresu agregacji *T*=5 dni. Dla modelu liniowego uzyskano wynik 5 dni, dla wielomianu II stopnia 6 dni (wynik 5,4 zaokrąglony w górę). W przypadku analizy indywidualnych chorób najlepsze wyniki uzyskano dla Early Blight, Late Blight oraz Anthracnose: 3 dni dla *T*=3. Z kolei dla Septoria Leaf Spot uzyskano wynik 4 dni dla *T*=3 dni, natomiast dla Bacterial Speck uzyskano najgorszy wynik 8 dni również dla *T*=3.

10.3.3. Wnioski dla zestawienia wyników dla różnej wartości parametru *T*

1. W ciągu pierwszych 2 tyg. wyniki znacznie się poprawiają (0.6 >~0.85+) dla zastosowanego podejścia,
2. Odchylenie standardowe w tym okresie (pierwsze 2 tyg.) nie jest duże, a zwiększa się po 4 tygodniu,
3. Dla DPI=23 występuje wyraźne pogorszenie wyników, co może świadczyć o istotnym błędzie pomiarowym, co przedstawiono w krótkim nagraniu z aplikacji służącej do przeglądania danych pomiarowych (https://bit.ly/DPI_23),
4. Analiza w ujęciu 5-dniowym skutkowała lepszymi wynikami modelu zarówno liniowego, jak i kwadratowego.
5. Analiza w ujęciu 3-dniowym jest podejściem, które pozwala na otrzymywanie relatywnie wysokich wyników (wskaźnik F1 > 0.8 oraz R^2 > 0.8) przy wykorzystaniu mniejszej ilości danych, a co jest z tym związane, jest podejściem pozwalającym na sprawniejsze reagowanie. Podejście oparte na pomiarach jedynie z 1 dania cechują się wysoką zmiennością wyników, natomiast podejście oparte na pomiarach 5-dniowych wydłużają czas reakcji na zagrożenie o 2 dni przy podobnych wynikach w stosunku do analizy pomiarów w ujęciu 3-dniowym.

10.4. Podsumowanie rozdziału

W niniejszym rozdziale przedstawiono podejście do badanego zagadnienia z podziałem zbioru uczącego w zależności od DPI, czyli liczby dni, które minęły od inokulacji. Zaproponowano podział na 3 różne wartości parametru *T* (1, 3 oraz 5). Badania przeprowadzono w sposób analogiczny do metodyki przedstawionej w 9, tzn. wyodrębniono 6 wariantów eksperymentalnych, tzn. jeden wariant uwzględniający wszystkie choroby oraz 5 wariantów uwzględniający kolejno AN, BS, EB, LB, SL. Wykonano analizę wyników w ujęciu klasyfikatorów, przypadków chorobowych, a także zastosowania różnych wartości parametru *T*. Uzyskane wyniki wskazują na możliwość diagnostyki badanych chorób w okresie krótszym niż 7 dni od zakażenia, co potwierdza trzecią tezę niniejszej pracy doktorskiej.

11. Podsumowanie i wnioski

Metody bezinwazyjnej diagnostyki odgrywają istotną rolę w wielu branżach, takich jak medycyna czy energetyka. Powiększająca się populacja ludzkości, wymuszająca zwiększenie efektywności produkcji żywności przy jednoczesnym zachowaniu zasad zrównoważonego rolnictwa, wytwarza potrzebę tworzenia i wprowadzania nowych metod diagnostycznych również w branży rolno-spożywczej. Z tego powodu badania nad takimi metodami cieszą się szerokim zainteresowaniem wśród naukowców oraz praktyków.

W badaniach tych znajdują swoje miejsce również rozwiązania diagnostyczne oparte o metody przetwarzania sygnałów hiperspektralnych z wykorzystaniem metod uczenia maszynowego. Dokładnie ta tematyka została poruszona w niniejszej pracy, której celem było opracowanie metod klasyfikacji obiektów chorych i zdrowych na podstawie danych hiperspektralnych, a także opracowanie metod wstępnego przetwarzania tego typu danych na potrzeby wspomnianej klasyfikacji. Opracowano metody oraz narzędzia, które zweryfikowano na podstawie posiadanego zbioru danych.

Podstawy teoretyczne rozważanego problemu, w tym zarysowanie potrzeb naukowych i ekonomicznych zagadnienia, rozeznanie dotychczas wykorzystywanych metod konwencjonalnych, jak i nieinwazyjnych, w tym metod wykorzysujących dane hiperspektralne i metody uczenia maszynowego zostały opisane w rozdziałach 3-4 i wskazują na zasadność podjętego kierunku badań.

Koncepcja opracowanej metody została w sposób syntetyczny podsumowana w rozdziale 5. Na rysunku rys. 12 przedstawiono również koncepcję urządzenia pomiarowego wykorzystującego opisaną metodę i zbudowane klasyfikatory w działaniu polowym, zaproponowano urządzenie ręczne oraz urządzenie zamontowane na platformie latającej. Urządzenie zintegrowane z bezzałogową platformą latającą mogłoby mieć następujące zalety: większa automatyzacja rozwiązania, możliwość szybszego wykonania pomiaru na większych obszarach, jak również większa skalowalność rozwiązania oraz możliwość wykonania pomiaru bez konieczności inspekcji lokalnej np. w środku pola. Praktyczna implementacja takiego rozwiązania wiąże się jednak z wieloma ograniczeniami i wyzwaniami. Wśród nich można wymienić m.in. relatywnie duże wymiary i waga aktualnych instrumentów hiperspektralnych, co skutecznie ogranicza czas przelotu i zasięg, dynamicznie zmieniające się warunki atmosferyczne, relatywnie krótki czas lotu w przypadku zastosowania platform wielowirnikowych czy narażenie na uszkodzenie urządzenia pomiarowego podczas lądowania przy wykorzystaniu platform samolotowych w przypadku montażu urządzenia pod platformą.

Metody i materiały w przeważającej formie zostały przedstawione w rozdziałach 6-8. W rozdziale 8 przedstawiono metodę postępowania z danymi w celu uzyskania najlepszych wyników dla wybranych klasyfikatorów (liniowy klasyfikator wektorów nośnych, regresja logistyczna, klasyfikator grzbietowy, klasyfikator lasu losowego, klasyfikator lekkiej maszyny wzmocnienia gradientowego oraz klasyfikator ekstremalnego wzmocnienia gradientowego), a także przedstawiono metodę łączenia danych z uwzględnieniem wymiaru czasu (DPI, liczba dni po inokulacji).

Pozostałe rozdziały (9 i 10) doprecyzowują opis metod tworzenia modeli klasyfikacji binarnej, ale przede wszystkim opisują uzyskane wyniki i ich analizę. Rozdział 9 skupia się na tworzeniu

klasyfikatorów w ujęciu całościowym (tzn. dla całego zbioru, bez uwzględnienia wymiaru czasu – DPI), z uwzględnieniem wieloetapowego postępowania z danymi zawierającym procesy normalizacji danych, eliminacji cech oraz dostrajania hiperparametrów modeli. Rozdział 10 skupia się na metodzie klasyfikacji uwzględniającej wymiar czasu w newralgicznym dla rozwoju choroby okresie (DPI <= 14).

W wyniku realizacji niniejszej pracy wykonano programistyczną implementację metod wstępnego przetwarzania danych hiperspektralnych na potrzeby klasyfikacji binarnej oraz procesów treningowych rzeczonych klasyfikatorów w przetwarzaniu wieloetapowym dla podejścia całościowego i podejścia uwzględniającego wymiar czasowy, na przykładzie uprawy pomidora (*Solanum lycopersicum*).

Najważniejsze osiągnięcia przedstawione w niniejszym doktoracie to:

- Opracowanie metody pozwalających na wykrywanie chorób na podstawie danych hiperspektralnych,
- Opracowanie metody wstępnego przetwarzania danych hiperspektralnych na potrzeby diagnostyczne roślin charakteryzującej się poprawą wyników klasyfikacji na poziomie,
- Opracowanie metody tworzenia podzbiorów danych hiperspektralnych w celu uwzględnienia wymiaru czasu (DPI) w procesie treningu klasyfikatorów,
- Zaprojektowanie i implementacja wieloetapowego procesu treningu klasyfikatorów, w tym z uwzględnieniem wymiaru czasowego posiadanych danych,
- Wyznaczenie klasyfikatorów najlepiej sprawdzających się w procesie klasyfikacji binarnej,
- Wyznaczenie zakresów spektralnych w przestrzeni VNIR oraz SWIR istotnych w klasyfikacji chorób roślin.

Zrealizowane badania wykonane w celu weryfikacji postawionych tez pomocniczych zaowocowały wnioskami i wskazówkami w zakresie metod diagnostyki roślin korzystających z pomiarów hiperspektralnych i metod uczenia maszynowego:

1. Istotne zakresy spektralne wykazane w ramach przeprowadzonych prac są tożsame z wcześniejszymi wyniki innych autorów w przestrzeni VNIR, jednak dodatkowo jako istotne wskazano zakresy z przestrzeni SWIR (1800-1890 nm oraz 2269 nm),
2. Spośród grona 40 klasyfikatorów wyłoniono najlepsze dla omawianego zagadnienia: liniowy klasyfikator wektorów nośnych, regresja logistyczna, klasyfikator grzbietowy, klasyfikator lasu losowego, klasyfikator lekkiej maszyny wzmocnienia gradientowego oraz klasyfikator ekstremalnego wzmocnienia gradientowego,
3. Z wybranych klasyfikatorów najlepszy średni wynik uzyskały klasyfikatory regresji logistycznej (0,878 w ETAPIE 6),
4. Z wybranych klasyfikatorów najgorszy średni wynik uzyskały klasyfikatory lasów losowych (0,708 w ETAPIE 1),
5. Poszczególne etapy postępowania z danymi (ETAPY 1-6) powodują osiąganie coraz lepszych wyników, nie licząc redukcji zakresów brzegowych wskazanych przez dystrybutora sprzętu (ETAP 3),
6. Wykonanie normalizacji minimum-maksimum (ETAP 2) poprawia parametry modeli klasyfikacyjnych w porównaniu do danych surowych (ETAP 1),
7. Usunięcie brzegowych wartości zakresów spektralnych (ETAP 3) wskazanych przez dystrybutora wykorzystanego urządzenia pomiarowego jako najbardziej zaszumione pogarsza lub nie poprawia wyników klasyfikacyjnych w porównaniu do normalizacji minimum-maksimum (ETAP 2),

8. Rekurencyjna eliminacja cech (ETAP 4) poprawia wyniki w stosunku do normalizacji minimum-maksimum (ETAP 2),
9. Przeprowadzenie pełnego procesu rekurencyjnej eliminacji cech (ETAP 4) nie gwarantuje najlepszych rezultatów klasyfikacyjnych (ETAP 5),
10. Dostrajanie hiperparametrów pozwala na osiągnięcie najlepszych rezultatów (ETAP 6) w porównaniu z pozostałymi etapami – średnia poprawa wskaźnika F1 pomiędzy etapami 1 i 6 wyniosła 9,7%,
11. Analiza danych uwzględniająca wymiar czasowy (DPI) przy zastosowaniu parametru szerokości ramki agregacji danych (*T*) równej 3 dni dała lepsze rezultaty od zastosowania szerokości *T*=1 ze względu na poziom uzyskanych wyników oraz lepsze od *T*=5 ze względu na czas potrzebny do przeprowadzenia badania przy porównywalnych wynikach (F1 wyniósł 0,836, 0,889 i 0,898 dla *T* równego odpowiednio 1, 3 i 5),
12. Zastosowanie parametru *T*=1 było związane z uzyskiwaniem wyższego odchylenia standardowego dla otrzymanych wyników niż w przypadku *T*=3 lub *T*=5,
13. Największą poprawę wyników pomiędzy wartościami podejściem jednodniowym (*T*=1), a pozostałymi (*T*=3, *T*=5) odnotowano dla choroby AN, tj. *T*=3: poprawa o 0.054 w porównaniu do *T*=1; *T*=5: poprawa o 0.0598 w porównaniu do *T*=1,
14. Największą poprawę wyników pomiędzy podejściem jednodniowym (*T*=1), a pozostałymi (*T*=3, *T*=5) odnotowano dla klasyfikatorów:
    a. Ridge Classifier dla *T*=3: poprawa o 0.0554 w porównaniu do *T*=1,
    b. LGBM Classifier dla *T*=5: poprawa o 0.0554 w porównaniu do *T*=1,
15. Mediana uzyskiwanych wyników (wskaźnika F1) wskazuje, że we wszystkich przypadkach (nie licząc choroby BS dla *T*=1) udało się uzyskać zakładany próg F1≥0.8, tj.
    a. dla podziału wg chorób mediana mieściła się w zakresie:
        i. dla *T*=1: 0.791-0.873,
        ii. dla *T*=3: 0.812-0.873,
        iii. dla *T*=5: 0.825-0.886,
    b. dla podziału wg klasyfikatorów mediana mieściła się w zakresie:
        i. dla *T*=1: 0.806-0.873,
        ii. dla *T*=3: 0.832-0.906,
        iii. dla *T*=5: 0.835-0.909.

Wyniki opisane w **rozdziale 9.2** udowadniają I tezę pomocniczą *„Możliwe jest określenie algorytmów klasyfikacyjnych, które z większą efektywnością pozwalają na rozpoznanie następujących chorób pomidora zwyczajnego: alternarioza, antraknoza, bakteryjna cętkowatość pomidora, septorioza, zaraza ziemniaka w pomidorze z zastosowaniem ograniczonego zbioru danych uczących”*.

Wyniki opisane w **rozdziale 9.3 i 9.4** udowadniają II tezę pomocniczą *„Możliwe jest określenie specyficznych długości zakresu widma odpowiadających określonym jednostkom chorobowym Solanum lycopersicum”*.

Wyniki opisane w **rozdziale 10.3** udowadniają III tezę pomocniczą *„Hiperspektralne pomiary reflektancji pozwalają na klasyfikację jednostek chorobowych Solanum lycopersicum w czasie szybszym niż metoda oceny wizualnej stosowana w praktyce”*.

Przedstawione powyżej rezultaty badań udowodniają tezę główną pracy: *„Istnieje możliwość wykorzystania danych hiperspektralnych reflektancji z zakresu fal o długości 350-2500nm i algorytmów uczenia maszynowego w procesie automatycznej diagnostyki Solanum lycopersicum, takich jak liście roślin, pod kątem klasyfikacji wybranych jednostek chorobowych”.*

Przeprowadzone badania potwierdziły przydatność danych hiperspektralnych w procesie diagnostycznym i wykrywania niektórych chorób uprawy *Solanum lycopersicum*, a także w procesie wczesnego wykrywania chorób tej uprawy. Zaproponowana w pracy metoda przetwarzania danych hiperspektralnych (ETAPY 1-6) pozwoliła na uzyskanie poprawy wyników klasyfikacji w prawie każdym przypadku. Wskazano również ważne zakresy spektralne, które były szczególnie ważne w procesie decyzyjnym i mogą być rozważane przy opracowywaniu nowych metod i przyrządów pomiarowych dedykowanych diagnostyce chorób uprawy *Solanum lycopersicum*. Przeprowadzone badanie literaturowe i wykonane w ramach pracy badania wykazują, że zebranie kompleksowego zestawu danych pomiarowych, obejmujące różne przypadki chorobowe w różnych fazach rozwoju, najlepiej wykonywanych dzienną częstotliwością, może pozwolić na stworzenie możliwie niezawodnych i szybkich metod diagnostyki chorób roślin.Niektórym autorom udało się uzyskać lepsze wartości metryk dopasowania modelu, jednak wyniki te były uzyskiwane dla widocznych symptomów chorobowych, części roślin niemożliwych do sprawdzenia za pomocą technik teledetekcyjnych (np. łodyg, które zazwyczaj są przykryte liśćmi) lub otrzymane wyniki nie były stałe (np. 90% dokładności klasyfikacyjnej dla DPI=4, ale już zaledwie 33,33% dla DPI=5).
Pomiary wykonywane z dzienną częstotliwością są bardziej wiarygodne niż pomiary wykonywane sporadycznie. Zwiększenie częstotliwości z pomiarów codziennych częstszych pomiarów może poprawić uzyskane wyniki.
Wykorzystanie najnowszych osiągnięć w dziedzinie uczenia maszynowego może pomóc w poprawieniu wyników klasyfikacji, jednak wymagałoby to pozyskania znacznej ilości danych.
Pomiary wykorzystane w niniejszej pracy były wykonywane w środowisku laboratoryjnym w sposób manualny. Proces ten był uciążliwy, czaso- i pracochłonny, a co z tym związane, również podlegał błędom ludzkim. Rozwój nieinwazyjnych metod teledetekcyjnych wymaga automatyzacji i standaryzacji procesu wykonywania tego typu pomiarów.
Zastosowanie opisanych metod może pomóc w skutecznym i szybkim diagnozowaniu chorób roślin, co może przyczynić się do poprawy wydajności i jakości produkcji roślinnej oraz ochrony środowiska.

12. Kierunek dalszych prac

Metody przedstawione w niniejszej pracy doktorskiej mogą zostać zastosowane do innych upraw i ich chorób, co wymaga opracowania odpowiednich, dedykowanych zbiorów danych zbudowanych na podstawie pomiarów wykonanych w warunkach laboratoryjnych i/lub terenowych.

Zaprezentowana metoda diagnostyczna może być przetestowana podczas przelotów inspekcyjnych wykonywanych, np. za pomocą bezzałogowych i załogowych statków powietrznych, czy satelitów, jednak wymagałoby to pogłębionych badań w warunkach rzeczywistych. Wykonanie takich badań jest jednak zadaniem nietrywialnym z uwagi na organiczny charakter rozprzestrzeniania się chorób i brak możliwości wykluczenia wystąpienia chorób nieuwzględnionych w procesie badawczym, jednak jest to kolejny krok na drodze do automatyzacji i skalowalności procesu diagnostycznego z wykorzystaniem zaproponowanych metod.

Zasadnym jest powtórzenie badania z wykorzystaniem różnych modeli sztucznych sieci neuronowych, w tym uczenia głębokiego. Może to pozwolić na poprawę osiąganych wyników zwłaszcza w bardzo złożonych przypadkach. Wymagałoby to jednak pozyskanie dużo większego zestawu danych treningowych. Z dotychczasowych doświadczeń autora wynika, że w zestawach danych złożonych z tysięcy czy dziesiątek tysięcy pomiarów hiperspektralnych lepiej sprawdzają się metody klasyczne uczenia maszynowego, a zbiory danych złożonych z większej liczby pomiarów hiperspektralnych nie są ogólnodostępne.

Niewątpliwe dalszym kierunkiem badań powinno być odtworzenie zaproponowanych metod w badaniach poświęconych innym uprawą i ich chorobom. Istotnym może okazać się również zwiększenie częstotliwości pomiarowej do wykonywania pomiarów w odstępach 12-godzinnych. Dodatkowo w dalszych rozważaniach można wykorzystać inne metody ekstrakcji i transformacji cech modelu uczenia maszynowego.

13. Bibliografia

[1] J. Dwojak, "Opracowanie Efektywnej Diagnostyki Eksploatacyjnej Zespołów Maszynowych W Energetyce Na Przykładzie PGE Elektrowni Opole S.A.," 2012.

[2] G. Lemaire, M. H. Jeuffroy, and F. Gastal, "Diagnosis tool for plant and crop N status in vegetative stage. Theory and practices for crop N management," *Eur. J. Agron.*, vol. 28, no. 4, pp. 614–624, 2008, doi: 10.1016/j.eja.2008.01.005.

[3] FAO, "The State of the World's Land and Water Resources for Food and Agriculture," 2011. Accessed: Dec. 04, 2019. [Online]. Available: http://www.fao.org/fileadmin/templates/solaw/files/executive_summary/SOLAW_EX_SUMM_WEB_EN.pdf.

[4] FAOSTAT, "Land Use." http://www.fao.org/faostat/en/#data/RL/visualize (accessed Aug. 09, 2020).

[5] AgTechStudio, "Raport Stanu Pola - dostarczanie bieżących informacji o rozwoju upraw." https://raportstanupola.pl/ (accessed Nov. 29, 2020).

[6] SatAgro Sp. z o.o., "SatAgro." https://www.satagro.pl/ (accessed Nov. 29, 2020).

[7] X. Jin, L. Jie, S. Wang, H. Qi, and S. Li, "Classifying Wheat Hyperspectral Pixels of Healthy Heads and Fusarium Head Blight Disease Using a Deep Neural Network in the Wild Field," *Remote Sens.*, vol. 10, no. 3, p. 395, Mar. 2018, doi: 10.3390/rs10030395.

[8] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Deep Convolutional Neural Networks for the Classification of Snapshot Mosaic Hyperspectral Imagery," 2017, Accessed: May 28, 2018. [Online]. Available: https://pdfs.semanticscholar.org/bc53/89bec5adaaa0e8ac4caa0fc939dfcd964579.pdf.

[9] F. Cao, Z. Yang, J. Ren, M. Jiang, and W.-K. Ling, "Does Normalization Methods Play a Role for Hyperspectral Image Classification?," 2017, Accessed: May 23, 2018. [Online]. Available: https://arxiv.org/pdf/1710.02939.pdf.

[10] Q. Gao, S. Lim, and X. Jia, "Hyperspectral image classification using convolutional neural networks and multiple feature learning," *Remote Sens.*, vol. 10, no. 2, 2018, doi: 10.3390/rs10020299.

[11] S. Keller, F. M. Riese, J. Stötzer, P. M. Maier, and S. Hinz, "Is It Possible To Retrieve Soil-Moisture Content From Measured Vnir Hyperspectral Data?," 2018, Accessed: May 14, 2018. [Online]. Available: https://arxiv.org/pdf/1804.09046.pdf.

[12] H. G. Jones and R. A. Vaughan, *Remote sensing of vegetation: principles, techniques and applications.* Oxford University Press, 2010.

[13] S. O. Ihuoma and C. A. Madramootoo, "Recent advances in crop water stress detection," 2017, doi: 10.1016/j.compag.2017.07.026.

[14] M. P. Finn *et al.*, "Remote Sensing of Soil Moisture Using Airborne Hyperspectral Data," *GIScience Remote Sens.*, vol. 48, no. 4, pp. 522–540, Oct. 2011, doi: 10.2747/1548-1603.48.4.522.

[15] "Polska Klasyfikacja Wyrobów i Usług (PKWiU)," *Rozporządzeniem Rady Ministrów z dnia 4 września 2015 r. (DZ. U. z 2015 r. poz. 1676)*. https://stat.gov.pl/Klasyfikacje/doc/pkwiu_15/pdf/Schemat_PKWiU2015.pdf (accessed Nov. 04, 2020).

[16] M. R. Foolad, "Genome Mapping and Molecular Breeding of Tomato," *Int. J. Plant Genomics*, vol. 2007, pp. 1–52, Aug. 2007, doi: 10.1155/2007/64358.

[17] B. Nawrocka, J. Robak, C. Ślusarski, and W. Macias, *Choroby i szkodniki pomidora w polu i pod osłonami*. Kraków: Wydawnictwo Plantpress Sp. z o.o., 2001.

[18] V. V. Datar and C. D. Mayee, "Conidial dispersal of Alternaria solani in tomato," *Indian Phytopathol*, vol. 35, pp. 68–70, 1982.

[19] J. F. J. Grigolli *et al.*, “Characterization of tomato accessions for resistance to early blight,” *Crop Breed. Appl. Biotechnol.*, vol. 11, no. 2, pp. 174–180, 2011, doi: 10.1590/S1984-70332011000200010.
[20] P. Sharma and S. Sharma, “Paradigm Shift in Plant Disease Diagnostics: A Journey from Conventional Diagnostics to Nano-diagnostics,” in *Part of the Fungal Biology book series (FUNGBIO)*, 2016, pp. 237–264.
[21] K. Golhani, S. K. Balasundram, G. Vadamalai, and B. Pradhan, “A review of neural networks in plant disease detection using hyperspectral data,” *Inf. Process. Agric.*, vol. 5, no. 3, pp. 354–371, Sep. 2018, doi: 10.1016/j.inpa.2018.05.002.
[22] A. Sakudo, Y. Suganuma, T. Kobayashi, T. Onodera, and K. Ikuta, “Near-infrared spectroscopy: Promising diagnostic tool for viral infections,” *Biochem. Biophys. Res. Commun.*, vol. 341, no. 2, pp. 279–284, Mar. 2006, doi: 10.1016/j.bbrc.2005.12.153.
[23] Minghua Zhang and Zhihao Qin, “Spectral analysis of tomato late blight infections for remote sensing of tomato disease stress in California,” in *IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS ’04. Proceedings. 2004*, 2004, vol. 6, no. February, pp. 4091–4094, doi: 10.1109/IGARSS.2004.1370031.
[24] C. H. Bock, G. H. Poole, P. E. Parker, and T. R. Gottwald, “Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging,” *CRC. Crit. Rev. Plant Sci.*, vol. 29, no. 2, pp. 59–107, 2010, doi: 10.1080/07352681003617285.
[25] S. Kryczyński and Z. Weber, *Fitopatologia Tom 2 Choroby roślin uprawnych*. 2011.
[26] Google, “Analiza popularności fraz machine learning, deep learning, neural networks w Google Trends.” https://trends.google.com/trends/explore?date=2010-11-07 2020-08-11&q=machine learning,deep learning,neural networks (accessed Mar. 09, 2023).
[27] J. Lu, R. Ehsani, Y. Shi, A. I. de Castro, and S. Wang, “Detection of multi-tomato leaf diseases (late blight, target and bacterial spots) in different stages by using a spectral-based sensor,” *Sci. Rep.*, vol. 8, no. 1, p. 2793, Dec. 2018, doi: 10.1038/s41598-018-21191-6.
[28] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,” *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/3289801.
[29] C. Xie, C. Yang, and Y. He, “Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities,” *Comput. Electron. Agric.*, vol. 135, pp. 154–162, Apr. 2017, doi: 10.1016/j.compag.2016.12.015.
[30] P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka, and E. Hernandez, “Plant Disease Detection Using Hyperspectral Imaging,” in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Nov. 2017, vol. 2017-Decem, pp. 1–8, doi: 10.1109/DICTA.2017.8227476.
[31] K. Smykała, B. Ruszczak, and K. Dziubański, “Application of Ensemble Learning to Detect Alternaria Solani Infection on Tomatoes Cultivated Under Foil Tunnels,” in *Intelligent Environments 2020*, 2020, pp. 145–150, doi: 10.3233/AISE200033.
[32] B. Ruszczak, K. Smykała, and K. Dziubański, “The detection of Alternaria solani infection on tomatoes using ensemble learning,” *J. Ambient Intell. Smart Environ.*, vol. 12, no. 5, pp. 407–418, Sep. 2020, doi: 10.3233/AIS-200573.
[33] R. Van De Vijver *et al.*, “In-field detection of Alternaria solani in potato crops using hyperspectral imaging,” *Comput. Electron. Agric.*, vol. 168, no. November 2019, Jan. 2020, doi: 10.1016/j.compag.2019.105106.
[34] M. S. Tasrif Anubhove, N. Ashrafi, A. M. Saleque, M. Akter, and S. U. Saif, “Machine Learning Algorithm based Disease Detection in Tomato with Automated

Image Telemetry for Vertical Farming," in *2020 International Conference on Computational Performance Evaluation (ComPE)*, Jul. 2020, pp. 250–254, doi: 10.1109/ComPE49325.2020.9200129.

[35] J. Fahrentrapp, F. Ria, M. Geilhausen, and B. Panassiti, "Detection of Gray Mold Leaf Infections Prior to Visual Symptom Appearance Using a Five-Band Multispectral Sensor," *Front. Plant Sci.*, vol. 10, no. May, pp. 1–14, May 2019, doi: 10.3389/fpls.2019.00628.

[36] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput. Electron. Agric.*, vol. 145, no. February, pp. 311–318, Feb. 2018, doi: 10.1016/j.compag.2018.01.009.

[37] T. Sylvain and L. G. Cecile, "Disease Identification: A Review of Vibrational Spectroscopy Applications," *Compr. Anal. Chem.*, vol. 80, no. January, pp. 195–225, 2018, doi: 10.1016/bs.coac.2018.03.005.

[38] R. L. Whetton, K. L. Hassall, T. W. Waine, and A. M. Mouazen, "Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 1: Laboratory study," *Biosyst. Eng.*, vol. 166, pp. 101–115, 2018, doi: 10.1016/j.biosystemseng.2017.11.008.

[39] K. Nagasubramanian, S. Jones, S. Sarkar, A. K. Singh, A. Singh, and B. Ganapathysubramanian, "Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems," *Plant Methods*, vol. 14, no. 1, pp. 1–20, 2018, doi: 10.1186/s13007-018-0349-9.

[40] S. Thomas *et al.*, "Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective," *J. Plant Dis. Prot.*, vol. 125, no. 1, pp. 5–20, 2018, doi: 10.1007/s41348-017-0124-6.

[41] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep Learning for Tomato Diseases: Classification and Symptoms Visualization," *Appl. Artif. Intell.*, vol. 31, no. 4, pp. 299–315, Apr. 2017, doi: 10.1080/08839514.2017.1315516.

[42] H. Zhu, H. Cen, C. Zhang, and Y. He, "Early detection and classification of tobacco leaves inoculated with tobacco mosaic virus based on hyperspectral imaging technique," *2016 Am. Soc. Agric. Biol. Eng. Annu. Int. Meet. ASABE 2016*, pp. 1–7, 2016, doi: 10.13031/aim.20162460422.

[43] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Front. Plant Sci.*, vol. 7, 2016, doi: 10.3389/fpls.2016.01419.

[44] A. K. Mahlein, "Plant disease detection by imaging sensors – Parallels and specific demands for precision agriculture and plant phenotyping," *Plant Dis.*, vol. 100, no. 2, pp. 241–254, 2016, doi: 10.1094/PDIS-03-15-0340-FE.

[45] M. Desai, A. Kumar Jain, N. K. Jain, and K. Jethwa, "Detection and Classification of Fruit Disease : A Review," *Int. Res. J. Eng. Technol.*, 2016, Accessed: Feb. 17, 2020. [Online]. Available: www.irjet.net.

[46] D. Ashourloo, A. A. Matkan, A. Huete, H. Aghighi, and M. R. Mobasheri, "Developing an Index for Detection and Identification of Disease Stages," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 851–855, 2016, doi: 10.1109/LGRS.2016.2550529.

[47] S. Adhikari, E. KC, L. Balkumari, B. Shrestha, and B. Baiju, "Tomato Plant Diseases Detection System Using Image Processing," 2018, [Online]. Available: https://www.researchgate.net/profile/Saban_Kc2/publication/327930730_TOMATO_PLANT_DISEASES_DETECTION_SYSTEM_USING_IMAGE_PROCESSING/links/5bada15645851574f7ebdd24/TOMATO-PLANT-DISEASES-DETECTION-

SYSTEM-USING-IMAGE-PROCESSING.pdf.
[48] C. Xie, Y. Shao, X. Li, and Y. He, "Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging," *Sci. Rep.*, vol. 5, no. 1, p. 16564, Dec. 2015, doi: 10.1038/srep16564.
[49] F. Martinelli *et al.*, "Advanced methods of plant disease detection. A review," *Agron. Sustain. Dev.*, vol. 35, no. 1, pp. 1–25, 2015, doi: 10.1007/s13593-014-0246-1.
[50] Y. M. Awad and A. A. Abdullah, "Early Detection of Powdery Mildew Disease in Wheat (Triticum aestivum L.) Using Thermal Imaging Technique," *Intell. Syst.*, vol. 322, no. November, 2015, doi: 10.1007/978-3-319-11313-5.
[51] P. Baranowski, M. Jedryczka, W. Mazurek, D. Babula-Skowronska, A. Siedliska, and J. Kaczmarek, "Hyperspectral and thermal imaging of oilseed rape (Brassica napus) response to fungal species of the genus Alternaria," *PLoS One*, 2015, doi: 10.1371/journal.pone.0122913.
[52] S. N. Ghaiwat and P. Arora, "Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review," *Int. J. Recent Adv. Eng. Technol. ISSN (Online*, no. 2, pp. 2347–2812, 2014.
[53] R. Ishimwe, K. Abutaleb, and F. Ahmed, "Applications of Thermal Imaging in Agriculture — A Review," no. September, 2014, doi: 10.4236/ars.2014.33011.
[54] Y.-H. F. Yeh, W.-C. Chung, J.-Y. Liao, C.-L. Chung, Y.-F. Kuo, and T.-T. Lin, "A Comparison of Machine Learning Methods on Hyperspectral Plant Disease Assessments," *IFAC Proc. Vol.*, vol. 46, no. 4, pp. 361–365, Jan. 2013, doi: 10.3182/20130327-3-JP-3017.00081.
[55] A. K. Mahlein *et al.*, "Development of spectral indices for detecting and identifying plant diseases," *Remote Sens. Environ.*, vol. 128, pp. 21–30, 2013, doi: 10.1016/j.rse.2012.09.019.
[56] A. K. Mahlein, U. Steiner, C. Hillnh??tter, H. W. Dehne, and E. C. Oerke, "Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases," *Plant Methods*, vol. 8, no. 1, 2012, doi: 10.1186/1746-4811-8-3.
[57] E. Bauriegel, A. Giebel, M. Geyer, U. Schmidt, and W. B. Herppich, "Early detection of Fusarium infection in wheat using hyper-spectral imaging," *Comput. Electron. Agric.*, vol. 75, no. 2, pp. 304–312, Feb. 2011, doi: 10.1016/J.COMPAG.2010.12.006.
[58] R. Van De Vijver *et al.*, "Towards automated field detection of Alternaria solani using hyperspectral imaging," no. July, 2018, doi: 10.13140/RG.2.2.13011.43049.
[59] S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, "A review of advanced techniques for detecting plant diseases," *Comput. Electron. Agric.*, vol. 72, no. 1, pp. 1–13, 2010, doi: 10.1016/j.compag.2010.02.007.
[60] T. Rumpf, A. K. Mahlein, U. Steiner, E. C. Oerke, H. W. Dehne, and L. Plümer, "Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance," *Comput. Electron. Agric.*, vol. 74, no. 1, pp. 91–99, 2010, doi: 10.1016/j.compag.2010.06.009.
[61] X. Wang, M. Zhang, J. Zhu, and S. Geng, "Spectral prediction of Phytophthora infestans infection on tomatoes using artificial neural network (ANN)," *Int. J. Remote Sens.*, vol. 29, no. 6, pp. 1693–1706, 2008, doi: 10.1080/01431160701281007.
[62] D. Moshou, C. Bravo, J. West, S. Wahlen, A. McCartney, and H. Ramon, "Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks," *Comput. Electron. Agric.*, vol. 44, no. 3, pp. 173–188, Sep. 2004, doi: 10.1016/j.compag.2004.04.003.
[63] P. Thenkabail, J. Lyon, and A. Huete, *Hyperspectral Remote Sensing of Vegetation, Second Edition*. 2018.
[64] H. Zhu, B. Chu, C. Zhang, F. Liu, L. Jiang, and Y. He, "Hyperspectral Imaging for

Presymptomatic Detection of Tobacco Disease with Successive Projections Algorithm and Machine-learning Classifiers," *Sci. Rep.*, vol. 7, no. 1, p. 4125, Dec. 2017, doi: 10.1038/s41598-017-04501-2.
[65] M. Islam, Anh Dinh, K. Wahid, and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Apr. 2017, no. November 2018, pp. 1–4, doi: 10.1109/CCECE.2017.7946594.
[66] A. Mahlein *et al.*, "Plant disease detection by hyperspectral imaging: from the lab to the field," *Adv. Anim. Precis. Agric.*, 2017, doi: 10.1017/S2040470017001248.
[67] A. Lowe, N. Harrison, and A. P. French, "Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress," *Plant Methods*, vol. 13, no. 1, pp. 1–12, 2017, doi: 10.1186/s13007-017-0233-z.
[68] C. D. Jones, J. B. Jones, and W. S. Lee, "Diagnosis of bacterial spot of tomato using spectral signatures," *Comput. Electron. Agric.*, vol. 74, no. 2, pp. 329–335, Nov. 2010, doi: 10.1016/j.compag.2010.09.008.
[69] Q. Wang, F. Qi, M. Sun, J. Qu, and J. Xue, "Identification of Tomato Disease Types and Detection of Infected Areas Based on Deep Convolutional Neural Networks and Object Detection Techniques.," *Comput. Intell. Neurosci.*, Dec. 2019, doi: 10.1155/2019/9142753.
[70] D. Bienkowski, M. J. Aitkenhead, A. K. Lees, C. Gallagher, and R. Neilson, "Detection and differentiation between potato (Solanum tuberosum) diseases using calibration models trained with non-imaging spectrometry data," *Comput. Electron. Agric.*, vol. 167, no. June, p. 105056, 2019, doi: 10.1016/j.compag.2019.105056.
[71] I. Dhau, E. Adam, O. Mutanga, and K. K. Ayisi, "Detecting the severity of maize streak virus infestations in maize crop using in situ hyperspectral data," *Trans. R. Soc. South Africa*, vol. 73, no. 1, pp. 8–15, 2018, doi: 10.1080/0035919X.2017.1370034.
[72] A. Siedliska, P. Baranowski, and M. Zubik, "Visible and near infrared hyperspectral imaging to early detection of fungal infection in strawberry fruits," 2014, Accessed: Jun. 12, 2018. [Online]. Available: https://www.hsi2016.com/wp-content/uploads/2014/10/P13.pdf.
[73] Parrot, "Oferta Parrot Sequoia." https://www.parrot.com/en/shop/accessories-spare-parts/other-drones/sequoia (accessed Mar. 09, 2023).
[74] DJI, "Oficjalna strona DJI P4 Multispectral." https://www.dji.com/pl/p4-multispectral (accessed Mar. 09, 2023).
[75] ESA, "Oficjalna strona misji Sentinel-2." https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed Mar. 09, 2023).
[76] NASA, "Oficjalna strona misji Landsat 8." https://landsat.gsfc.nasa.gov/satellites/landsat-8/ (accessed Mar. 09, 2023).
[77] Digital Globe, "Oficjalna strona misji WorldView3." http://worldview3.digitalglobe.com/ (accessed Mar. 09, 2023).
[78] Headwall Photonics, "Oficjalna specyfikacja kamery Co-Aligned VNIR-SWIR." https://www.headwallphotonics.com/hubfs/Co-Aligned VNIR-SWIR Sensor 04May2022.pdf (accessed Mar. 09, 2023).
[79] ASI, "Oficjalna strona misji PRISMA." https://www.asi.it/en/earth-science/prisma/ (accessed Mar. 09, 2023).
[80] DLR, "Oficjalna strona misji EnMap." https://www.enmap.org/ (accessed Mar. 09, 2023).
[81] KP Labs, "Oficjalna strona misji Intuition-1." https://kplabs.space/intuition-1/ (accessed Mar. 09, 2023).
[82] K. M. Gold *et al.*, "Hyperspectral measurements enable pre-symptomatic detection and

differentiation of contrasting physiological effects of late blight and early blight in potato," *Remote Sens.*, vol. 12, no. 2, 2020, doi: 10.3390/rs12020286.
[83] Y. Xie, D. Plett, and H. Liu, "The Promise of Hyperspectral Imaging for the Early Detection of Crown Rot in Wheat," *AgriEngineering*, vol. 3, no. 4, pp. 924–941, 2021, doi: 10.3390/agriengineering3040058.
[84] D. Ashourloo, M. R. Mobasheri, and A. Huete, "Evaluating the effect of different wheat rust disease symptoms on vegetation indices using hyperspectral measurements," *Remote Sens.*, vol. 6, no. 6, pp. 5107–5123, 2014, doi: 10.3390/rs6065107.
[85] Y. Peng *et al.*, "Early detection of plant virus infection using multispectral imaging and spatial–spectral machine learning," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, 2022, doi: 10.1038/s41598-022-06372-8.
[86] C. Nguyen, V. Sagan, M. Maimaitiyiming, M. Maimaitijiang, S. Bhadra, and M. T. Kwasniewski, "Early detection of plant viral disease using hyperspectral imaging and deep learning," *Sensors (Switzerland)*, vol. 21, no. 3, pp. 1–23, 2021, doi: 10.3390/s21030742.
[87] Y. Cen, Y. Huang, S. Hu, L. Zhang, and J. Zhang, "Early Detection of Bacterial Wilt in Tomato with Portable Hyperspectral Spectrometer," *Remote Sens.*, vol. 14, no. 12, p. 2882, Jun. 2022, doi: 10.3390/rs14122882.
[88] M. Kuska *et al.*, "Hyperspectral phenotyping on the microscopic scale: Towards automated characterization of plant-pathogen interactions," *Plant Methods*, vol. 11, no. 1, 2015, doi: 10.1186/s13007-015-0073-7.
[89] Ushio, "Oficjalna specyfikacja lamp Eurostar Reflekto MR16 użytych w badaniu." https://www.bulbspro.com/media/pdf/Ushio_EuroReflektoMR16.pdf (accessed Mar. 09, 2023).
[90] G. T. Georgiev and J. J. Butler, "Long-term calibration monitoring of Spectralon diffusers BRDF in the air-ultraviolet," *Appl. Opt.*, vol. 46, no. 32, p. 7892, Nov. 2007, doi: 10.1364/AO.46.007892.
[91] "Oficjalna dokumentacja pakietu XGBoost." https://xgboost.readthedocs.io/en/stable/ (accessed Mar. 09, 2023).
[92] Microsoft Corporation, "Oficjalna dokumentacja pakietu LightGBM." https://lightgbm.readthedocs.io/en/latest/index.html (accessed Mar. 09, 2023).
[93] "Oficjalna dokumentacja pakietu Scikit-Learn." https://scikit-learn.org/stable/ (accessed Mar. 09, 2023).
[94] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, 2001, doi: 10.14569/ijacsa.2016.070603.
[95] G. Ke *et al.*, "LightGBM : A Highly Efficient Gradient Boosting Decision Tree," no. Nips, pp. 1–9, 2017.
[96] H. Wickham, "Tidy Data," *J. Stat. Softw.*, vol. 59, no. 10, pp. 1–23, 2014, [Online]. Available: http://www.jstatsoft.org/.
[97] M. Tomaszewski, J. Nalepa, E. Moliszewska, B. Ruszczak, and K. Smykała, "Early detection of Solanum lycopersicum diseases from temporally - aggregated hyperspectral measurements using machine learning," *Sci. Rep.*, pp. 1–13, 2023, doi: 10.1038/s41598-023-34079-x.
[98] J. Brownlee, "How to Calculate Feautre Importance With Python," *Machine Learning Mastery*, 2020. https://machinelearningmastery.com/calculate-feature-importance-with-python/.
[99] G. F. Hughes, "On the Mean Accuracy of Statistical Pattern Recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, 1968, doi: 10.1109/TIT.1968.1054102.
[100] W. Ma, C. Gong, Y. Hu, P. Meng, and F. Xu, "The Hughes phenomenon in

hyperspectral classification based on the ground spectrum of grasslands in the region around Qinghai Lake," *Int. Symp. Photoelectron. Detect. Imaging 2013 Imaging Spectrom. Technol. Appl.*, vol. 8910, p. 89101G, 2013, doi: 10.1117/12.2034457.

[101] L. P. Coelho and W. Richert, *Building Machine Learning Systems with Python*. Packt Publishing, 2015.

[102] S. Guido and A. Müller, *Introduction to Machine Learning with Python: A Guide for Data Scientists*. 2017.

[103] S. Allwright, "What is a good F1 score and how do I interpret it?," 2022. https://stephenallwright.com/good-f1-score/ (accessed Aug. 06, 2022).

Streszczenie

Celem pracy było stworzenie metody pozwalającej na wstępną, automatyczną diagnostykę obiektu za pomocą pomiarów reflektancji hiperspektralnej oraz algorytmów uczenia maszynowego na przykładzie diagnostyki wybranych chorób w uprawach roślin. W pracy doktorskiej autor posłużył się zbiorem pomiarów reflektancji hiperpektralnej z zakresu 350-2500nm wykonanych za pomocą urządzenia ASD FieldSpec 4 Hi-Res w środowisku kontrolowanym oraz wybranymi metodami klasycznymi uczenia maszynowego, takimi jak liniowy klasyfikator wektorów nośnych, regresja logistyczna, klasyfikator grzbietowy, klasyfikator lasu losowego, klasyfikator lekkiej maszyny wzmocnienia gradientowego oraz klasyfikator ekstremalnego wzmocnienia gradientowego.

Dla poruszanej tematyki wykonane zostało badania aktualnego stanu wiedzy, którego wyniki wskazującą na zasadność przyjętego kierunku działań i możliwości wykorzystania danych do osiągnięcia założonego celu, a jednocześnie wskazują na istotne braki w dotychczas wykonywanych badaniach, takie jak niedobór wyników dla pomiarów wykonywanych z dzienną częstotliwością czy brak badań obejmujących metodę przetwarzania danych hiperspektrlanych zaproponowaną przez autora. Wyznaczono 6 wariantów eksperymentalnych uwzględniających różne przypadki chorobowe zarówno w badaniu dla zbioru całościowego, jak i fragmentarycznego, uwzględniającego wymiar czasu.

W pracy zaprezentowana została metoda przetwarzania i klasyfikacji danych hiperspektralnych w celu automatyzacji procesu wczesnej diagnostyki obiektów roślinnych z użyciem metod nieinwazyjnych. Zaproponowana metoda pozwoliła na poprawę średniej skuteczności diagnostycznej z poziomu 0.749 do 0.846 (poprawa o 9,7%) mierzonej wskaźnikiem F1.

Zaprezentowano metodę uwzględniającą wymiar czasu wykonywanych pomiarów. Wyznaczono parametr T agregacji czasowej danych, który pozwala na zwiększenie pewności decyzyjnej przy jednoczesnym skróceniu czasu analizy. Autor konkluduje, że dla T równemu 3 dniom pomiarowych uzyskuje się wyniki lepsze niż dla agregacji 1-dniowej, a porównywalne do agregacji 5-dniowej.

Wyznaczono zakresy spektralne w przestrzeni światła widzialnego i bliskiej podczerwieni (VNIR) oraz podczerwieni fal krótkich (SWIR) istotne w procesie diagnostycznym. Autor zauważa, że o ile zakresy VNIR pokrywają się z dotychczasowymi wynikami innych autorów, to zakresy wskazane w przestrzeni SWIR, tj. 1800-1890 nm oraz 2269 nm stanowią nowość naukową.

Podsumowanie

Wyniki uzyskane w trakcie realizacji prac zawartych w pracy doktorskiej udowadniają tezy pomocnicze oraz tezę główną pracy „***Istnieje możliwość wykorzystania danych hiperspektralnych reflektancji z zakresu fal o długości 350-2500nm i algorytmów uczenia maszynowego w procesie automatycznej diagnostyki Solanum lycopersicum, takich jak liście roślin, pod kątem klasyfikacji wybranych jednostek chorobowych”***.

Abstract

The objective of the thesis was to develop a method enabling preliminary, automated object diagnosis using hyperspectral reflectance measurements and machine learning algorithms on the example of the diagnosis of selected diseases in plant crops. In the doctoral thesis, the author utilized a dataset of hyperspectral reflectance measurements in the range of 350-2500nm collected with the spectroradiometer ASD FieldSpec 4 Hi-Res in controlled environment, along with selected classical machine learning methods such as linear support vector classifier, logistic regression, ridge classifier, random forest classifier, light gradient boosting machine classifier and extreme gradient boosting classifier.

Research on the current state of knowledge was conducted, the results of which indicate the validity of the adopted approach and the feasibility of hyperspectral data usage to obtain the set goal. At the same time, these results highlight significant gaps in the research conducted heretofore, such as a shortage of datasets consisting of the measurements taken at daily frequencies or a lack of studies involving the hyperspectral data processing method proposed by the author. Six experimental variants were determined, including various disease cases, both in the comprehensive dataset study and a fragmented one, considering the temporal dimension.

The thesis presents a method of processing and classifying temporally-aggregated hyperspectral data to automate the process of early detection of plant diseases using non-invasive methods. The proposed method resulted in an enhancement of the average diagnosis efficiency from 0.749 to 0.846 (an improvement of 9.7%) as measured by the F1 score.

A method considering the temporal dimension of measurements was presented. A parameter T of temporal data aggregation was determined, which enhances decision-making confidence while simultaneously reducing the time of analysis. The author concludes that for T equals 3 measurement days, the obtained results are better than for 1-day aggregation and comparable to 5-day aggregation.

Spectral wavelengths in the visible and near-infrared (VNIR) and short-wave infrared (SWIR) spectral ranges, crucial for the diagnosis process, were identified. The author notes that while the wavelengths located in VNIR range align with prior findings from other authors, the wavelengths located in SWIR range, i.e., 1800-1890 nm and 2269nm, represent a scientific novelty.

The results obtained during the implementation of the work included in the doctoral thesis confirm the auxiliary theses as well as the main thesis of the work, '**It is possible to use hyperspectral reflectance data in the range of wavelengths from 350 to 2500nm and machine learning algorithms in the process of automatic diagnostics of *Solanum lycopersicum* for classification of selected diseases**'.

Spis rysunków

rys. 1 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019 13
rys. 2 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019 14
rys. 3 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019 14
rys. 4 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019 15
rys. 5 Zdjęcia referencyjne chorych roślin wykonane 10.10.2019 16
rys. 6 Zainteresowanie technologiami a) uczenia maszynowego, b) uczenia głębokiego oraz c) sieci neuronowych w ujęciu czasowym za lata 2010-2020 [26] 16
rys. 7 Przykładowe pomiary spektralne wykonane za pomocą urządzenia pomiarowego ASD FieldSpec-4 Hi-Res 19
rys. 8 Kamera hiperspektralna Zebra X1 przymocowana do bezzałogowego statku latającego DJI Matrice 600 [autor zdjęcia: Krzysztof Płatek] 20
rys. 9 Kamera hiperspektralna Zebra X1 21
rys. 10 Pasma wyróżnione przez autorów publikacjii 28
rys. 11 Główne etapy realizacji proponowanej metody diagnostyki wybranych chorób w uprawie pomidorów zwyczajnych 32
rys. 12 Koncepcja przyrządu pomiarowego, który umożliwi precyzyjną diagnostykę chorób roślin przy wykorzystaniu technologii pomiarów hiperspektralnych i uczenia maszynowego 33
rys. 13 Wykres Gantta przedstawiający na osi czasu dni siewu, wysadzania i zarażania roślin oraz wykonywania pomiarów; każdy pojedynczy kwadrat reprezentuje 1 dzień pomiarowy 36
rys. 14 Spektroradiometr ASD FieldSpec Hi-Res wykonujący pomiary hiperspektralne w zakresie 350-2500nm 40
rys. 15 Krzywa spektralna panelu kalibracyjnego przed przeprowadzeniem optymalizacji ... 41
rys. 16 Krzywa spektralna reflektancji panelu kalibracyjnego po przeprowadzeniu optymalizacji i kalibracji wzornika bieli 41
rys. 17 Stanowisko pomiarowe w trakcie wykonywania procesu kalibracji instrumentu pomiarowego ASD FieldSpec 4 HiRes wraz z oświetleniem halogenowym 42
rys. 18 Ortofotomapa wskaźnika wegetacyjnego (NDVI) pola doświadczalnego w miejscowości Ernestynów z ledwie zauważalnymi lub niewidocznymi ogniskami chorobowymi (26.07.2019) 43
rys. 19 Ortofotomapa wskaźnika wegetacyjnego (NDVI) pola doświadczalnego w miejscowości Ernestynów z widocznymi ogniskami chorobowymi (02.08.2019) 43
rys. 20 Ortofotomapa wskaźnika wegetacyjnego (NDVI) pola doświadczalnego w miejscowości Ernestynów z widocznymi, obszernymi zmianami chorobowymi sprzyjających dla patogenu warunkach atmosferycznych i po zastosowaniu nieadekwatnych środków ochrony roślin (05.08.2019) 44
rys. 21 Schemat obrazujący proces zbierania danych pomiarowych w tunelach foliowych ... 45
rys. 22 Rozkład pomiarów hiperspektralnych dla poszczególnych odmian i chorób; kolorowe tło stanowi jest wizualizacją zakresu min-max wszystkich wykonanych pomiarów; linią czerwoną oznaczono medianę wszystkich pomiarów 46
rys. 23 Proces wstępnego przetwarzania danych i zmiana ilości danych podczas wykonywania tego procesu 55
rys. 24 Uproszczony schemat przetwarzania danych do postaci czystego zbioru danych 56
rys. 25 Liczba pomiarów w zależności od dni, które minęły od inokulacji w środowisku kontrolowanym z podziałem na analizowane choroby 61
rys. 26 Schemat tworzenia podzbiorów danych do analizy przedziałów czasowych, DPI (days post inoculum) liczba dni, które minęły od inokulacji, źródło: [97] 62

rys. 27 Zestawienie zdjęć referencyjnych badanych roślin w czasie rozwoju chorób w pierwszych dwóch tygodniach od inokulacji, czerwonym obwodem zaznaczono widoczne zmiany chorobowe, numeracja kolumn oznacza liczbę dni, które minęły od inokulacji, wiersze AN, BS, CS, EB, LB i SL oznaczają przypadki chorobowe opisane w rozdziale 9 „Wykrywanie chorób z wykorzystaniem wybranych algorytmów uczenia maszynowego", źródło: [97]....63
rys. 28 Liczba wykonanych pomiarów w zależności od użytego patogenu....64
rys. 29 Liczba wykonanych pomiarów w zależności od odmiany rośliny....65
rys. 30 Fragment wykresu rys. 22; Rozkład pomiarów hiperspektralnych dla poszczególnych odmian i chorób; kolorowe tło stanowi jest wizualizacją zakresu min-max wszystkich wykonanych pomiarów; linią czerwoną oznaczono medianę wszystkich pomiarów....65
rys. 31 Liczba wykonanych pomiarów w zależności od fitotronu....66
rys. 32 Histogram wskaźnika wegetacyjnego wykonanych pomiarów hiperspektralnych....67
rys. 33 Wykres KDE dla danych NDVI zbioru, czyli rozkładu prawdopodobieństwa wystąpienia NDVI dla losowej wartości w przedziale 0-1 na podstawie wartości NDVI otrzymanej ze zbioru....67
rys. 34 Rozkład wartości wskaźnika wegetacyjnego w poszczególnych dniach od inokulacji z uwzględnieniem liczby pomiarów w poszczególnych obszarach....68
rys. 35 Mediana widma pomiarów dla wysokości 5 cm....71
rys. 36 Mediana widma pomiarów dla wysokości 30 cm....72
rys. 37 Mediana widma pomiarów dla wysokości 60 cm....72
rys. 38 Widmo wszystkich pomiarów dla wysokości 5 cm. Zielonym kolorem oznaczono pomiary kontrolny, czerwonym - obiekty zakażone....73
rys. 39 Widmo wszystkich pomiarów dla wysokości 5 cm poddanych normalizacji. Zielonym kolorem oznaczono pomiary kontrolny, czewonym - obiekty zakażone....74
rys. 40 Schemat planowanych etapów eksperymentu....75
rys. 41 Kolejne etapy eksperymentu badawczego. Kolorem czerwonym oznaczono rekomendowaną procedurę przetwarzania danych....76
rys. 42 Wyniki treningu klasyfikatorów na różnych etapach metody dla pełnego zbioru danych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym....78
rys. 43 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Antracnose" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym....79
rys. 44 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Bacterial Speck" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym....80
rys. 45 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Early Blight" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym....81
rys. 46 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Late Blight" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym....82
rys. 47 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Septoria Leaf Spot" od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym....83
rys. 48 Średnia poprawa wyników dla poszczególnych klasyfikatorów....84
rys. 49 Proces rekurencyjnej eliminacji cech i wyznaczania optymalnego etapu eliminacji..87
rys. 50 Udział poszczególnych pasm spektralnych w procesie decyzyjnym....88

rys. 51 Istotne przedziały spektralne w procesie klasyfikacji. Kreskowanie białą linią wyznacza przedziały pasm istotne również w przypadku poszczególnych chorób. Kreskowanie jest niezależne od zbioru. 89
rys. 52 Istotne przedziały spektralne w procesie klasyfikacji. Kreskowanie białą linią wyznacza wspólne przedziały spektralne dla wszystkich chorób. Kreskowanie jest niezależne od zbioru. 90
rys. 53 Zestawienie porównawcze z odniesieniem do pasm spektralnych zaobserwowanych w literaturze (góra – najistotniejsze pasma spektralne wg literatury, dolny – najistotniejsze pasma spektralne wg literatury i przeprowadzonych badań) 92
rys. 54 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI) w scenariuszu a. pomiary kontrolne kontra pomiary obiektów zakażonych (AN-SL), T=1 96
rys. 55 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu a. pomiary kontrolne kontra pomiary obiektów zakażonych (AN-SL), T=3 97
rys. 56 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu a. pomiary kontrolne kontra pomiary obiektów zakażonych (AN-SL), T=5 97
rys. 57 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę AN, T=1 98
rys. 58 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę AN, T=3 99
rys. 59 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę AN, T=5 99
rys. 60 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę BS, T=1 100
rys. 61 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę BS, T=3 101
rys. 62 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę BS, T=5 101
rys. 63 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę EB, T=1 102
rys. 64 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę EB, T=3 103
rys. 65 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę EB, T=5 103
rys. 66 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę LB, T=1 104
rys. 67 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę LB, T=3 105

rys. 68 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę LB, T=5 105
rys. 69 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę SL, T=1 106
rys. 70 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę SL, T=3 107
rys. 71 Wyniki treningu klasyfikatorów dla poszczególnych przedziałów czasowych (DPI)w scenariuszu b. pomiary kontrolne kontra pomiary obiektów zakażonych patogenem powodującym chorobę SL, T=5 107
rys. 72 Wyniki treningu klasyfikatora Ridge Classifier dla poszczególnych przedziałów czasowych (DPI), T=1 109
rys. 73 Wyniki treningu klasyfikatora Ridge Classifier dla poszczególnych przedziałów czasowych (DPI), T=3 110
rys. 74 Wyniki treningu klasyfikatora Ridge Classifier dla poszczególnych przedziałów czasowych (DPI), T=5 110
rys. 75 Wyniki treningu klasyfikatora LGBM Classifier dla poszczególnych przedziałów czasowych (DPI), T=1 111
rys. 76 Wyniki treningu klasyfikatora LGBM Classifier dla poszczególnych przedziałów czasowych (DPI), T=3 112
rys. 77 Wyniki treningu klasyfikatora LGBM Classifier dla poszczególnych przedziałów czasowych (DPI), T=5 112
rys. 78 Wyniki treningu klasyfikatora Random Forest Classifier dla poszczególnych przedziałów czasowych (DPI), T=1 113
rys. 79 Wyniki treningu klasyfikatora Random Forest Classifier dla poszczególnych przedziałów czasowych (DPI), T=3 113
rys. 80 Wyniki treningu klasyfikatora Random Forest Classifier dla poszczególnych przedziałów czasowych (DPI), T=5 114
rys. 81 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla najlepszego klasyfikatora (Ridge Classifier) 115
rys. 82 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla najlepszego klasyfikatora (Ridge Classifier) z wyrysowanymi liniami trendu wielomianu pierwszego stopnia 116
rys. 83 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla najlepszego klasyfikatora (Ridge Classifier) z wyrysowanymi liniami trendu wielomianu drugiego stopnia 116
rys. 84 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla najlepszego klasyfikatora (Ridge Classifier) z wyrysowanymi liniami trendu wielomianu trzeciego stopnia 117
rys. 85 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku wszystkich chorób z wyrysowanymi liniami trendu funkcji liniowej 118
rys. 86 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku wszystkich chorób z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia 118
rys. 87 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Anthracnose z wyrysowanymi liniami trendu funkcji liniowej 119
rys. 88 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Anthracnose z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia 119

rys. 89 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Bacterial Speck z wyrysowanymi liniami trendu funkcji liniowej 120
rys. 90 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Bacterial Speck z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia 120
rys. 91 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Early Blight z wyrysowanymi liniami trendu funkcji liniowej 121
rys. 92 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Early Blight z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia 121
rys. 93 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Late Blight z wyrysowanymi liniami trendu funkcji liniowej 122
rys. 94 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Late Blight z wyrysowanymi liniami trendu funkcji wielomianowej II-ego stopnia 122
rys. 95 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Septoria Leaf Spot z wyrysowanymi liniami trendu funkcji liniowej 123
rys. 96 Zestawienie uśrednionych wyników dla T=[1, 3, 5], DPI < 15, dla przypadku Septoria Leaf Spot z wyrysowanymi liniami funkcji wielomianowej II-ego stopnia 123
rys. 97 Trend rosnący efektywności klasyfikatora dla wszystkich chorób, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5 124
rys. 98 Trend rosnący efektywności klasyfikatora dla Anthracnose, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5 124
rys. 99 Trend rosnący efektywności klasyfikatora dla Bacterial Speck, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5 125
rys. 100 Trend rosnący efektywności klasyfikatora dla Early Blight, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5 125
rys. 101 Trend rosnący efektywności klasyfikatora dla Late Blight, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5 126
rys. 102 Trend rosnący efektywności klasyfikatora dla Septoria Leaf Spot, przerywaną linią oznaczono funkcję liniową, a ciągłą funkcję wielomianową II-ego stopnia, kolorem zielonym oznaczono wyniki dla T=3, kolorem niebieskim – T=5 126

Spis tabel

tab. 1 Zestawienie publikacji dot. wykorzystania różnych źródeł danych w detekcji chorób.17
tab. 2 Zestawienie publikacji nt. wykrywania choroba na podstawie danych hiperspektralnych w zależności od gatunku odmiany 17
tab. 3 Zestawienie wyników badań nad wykrywanie chorób roślin za pomocą danych hiperspektralnych 22
tab. 4 Syntetyczne zestawienie charakterystyki przeprowadzonych pomiarów i badań 26
tab. 5 Syntetyczne zestawienie publikacji dotyczących analizy pomiarów hiperspektralnych w celu wczesnego wykrywania chorób 30
tab. 6 Parametry urządzenia pomiarowego wykorzystanego w badaniu 38
tab. 7 Obszar skanowania próbki w zależności od wysokości, na której umieszczona była soczewka urządzenia skanującego przy zastosowaniu domyślnego osprzętu (brak dodatkowych soczewek korygujących FOV) 40
tab. 8 Wybrane wiersze z tabeli mapującej służącej do wstępnego przetwarzania danych (pozostawiono oryginalne nazwy kolumn odpowiadające kolumną znajdującym się w przetwarzanym zbiorze danych – opis znajduje się na końcu podrozdziału) 57
tab. 9 Przykładowe rekordy zawarte w zbiorze danych wykorzystanych w prowadzonym badaniu. Kolumny "..." reprezentują brakujące kolumny opisane poniżej tabeli, m.in. istotne daty zabiegów agrotechnicznych (siew, rozsada, inokulacja), wartość wskaźnika wegetacyjnego dla wykonywanego pomiaru, ilość dni, które minęły od inokulacji do dnia wykonywania pomiaru oraz wartości pomiarowych dla zakresu spektralnego 351-2499[nm] (pozostawiono oryginalne nazwy kolumn odpowiadające kolumną znajdującym się w przetwarzanym zbiorze danych – opis znajduje się na końcu podrozdziału) 58
tab. 10 Tabela mapująca zawierająca informacje o badanym obiekcie w zależności od numeru referencyjnego (field_no, pozostawiono oryginalne nazwy kolumn odpowiadające kolumną znajdującym się w przetwarzanym zbiorze danych – opis znajduje się na końcu podrozdziału) 59
tab. 11 Tabela mapująca dla dat siania, rozsady oraz inokulacji obiektów badawczych (pozostawiono oryginalne nazwy kolumn odpowiadające kolumną znajdującym się w przetwarzanym zbiorze danych – opis znajduje się na końcu podrozdziału) 59
tab. 12 Miary tendencji centralnej wskaźnika NDVI dla poszczególnych grup badawczych, podzielonych wg badanych chorób 67
tab. 13 Zestawienie wyników najlepszych klasyfikatorów (nieprzetrenowanych) w zależności od wysokości wykonywanych pomiarów oraz poziomu zbalansowania zbioru 69
tab. 14 Zestawienie wyników najlepszych klasyfikatorów (przetrenowanych) w zależności od wysokości wykonywanych pomiarów oraz poziomu zbalansowania zbioru 70
tab. 15 Wyniki treningu klasyfikatorów na różnych etapach metody dla pełnego zbioru danych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym 77
tab. 16 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Antracnose” od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym 78
tab. 17 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Bacterial Speck” od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym 79
tab. 18 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Early Blight” od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym 80

tab. 19 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Late Blight” od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym 81
tab. 20 Wyniki treningu klasyfikatorów na różnych etapach metody odróżniającej chorobę „Septoria Leaf Spot” od przypadków zdrowych z zastosowaniem stratyfikacji zbioru uczącego oraz walidacji krzyżowej na zbiorze treningowym..... 82
tab. 21 Wyniki treningu najlepszych klasyfikatorów na różnych etapach metody dla poszczególnych chorób 83
tab. 22 Wyniki multiklasyfikacji dla nieprzetworzonego zbioru danych z wykorzystaniem metody dotrenowania hiperparametrów 93
tab. 23 Wyniki multiklasyfikacji dla znormalizowanego zbioru danych z ograniczonym zakresem spektralnym (475-2400nm) z wykorzystaniem metody dotrenowania hiperparametrów 93
tab. 24 Wyniki multiklasyfikacji dla znormalizowanego zbioru danych z pełnym zakresem spektralnym (350-2500nm) z wykorzystaniem metody dotrenowania hiperparametrów 93
tab. 25 Podsumowanie wyników analizy diagnostyki infekcji w czasie z podziałem na badane choroby oraz wyznaczenie najwcześniejszego momentu możliwej diagnostyki podany w nawiasie..... 127

14. DODATEK A

14.1. Załączniki

1. Załącznik 1 - Upoważnienie do wykorzystania danych firmowych
2. Dziennik eksperymentów przeprowadzonych w ramach badań: https://qzsolutions-my.sharepoint.com/:w:/g/personal/ksmykala_qzsolutions_pl/EaS9iweIM-1Nv1Sp9cCOLHwBILSQoB0d55FjgC7iAgZOQA?e=sGg4KX

14.2. Załączniki w formie skoroszytów

1. Pełna lista wyników klasyfikacji binarnej: https://qzsolutions-my.sharepoint.com/:x:/g/personal/ksmykala_qzsolutions_pl/EaN2mtw6wSNGvrJlwWxvbygBBYgiXwCVLCq5__8izOR2MA?e=6fCQrA

2. Tabela przestawna wyników klasyfikacji binarnej: https://qzsolutions-my.sharepoint.com/:x:/g/personal/ksmykala_qzsolutions_pl/Ee9uwHpbBIFAovgg1kyu4UkBtwDYxENpDaeCgfo0AGm5FA?e=BpY1IM

QZSolutions

Opole, dn. 09.02.2021

Oświadczenie

Panu **Krzysztofowi Smykała** udziela się upoważnienia na nieodpłatne wykorzystanie danych będących własnością firmy QZ Solutions sp. z o.o. w celach realizacji prac naukowo-badawczych, w tym przede wszystkim danych pozyskanych w ramach projektu o sygnaturze POIR.01.01.01-00-1317/17.

Zbigniew Kawalec
Prezes Zarządu
QZ Solutions sp. z o.o.