Opole University of Technology
Faculty of Electrical Engineering, Automatic Control, and Informatics
MSc. Eng. Krzysztof Smykała

Application of Machine Learning Methods and Hyperspectral Measurements in Diagnosis of *Solanum lycopersicum* Cultivation

doctoral thesis in the field of Automatics, Electronics, Electrical Engineering and Space Technologies under the supervision of PhD, Dsc, Eng. Michał Tomaszewski and with support of PhD Eng. Bogdan Ruszczak

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**'.