Malaysian Journal of Analytical Sciences (MJAS)
Published by Malaysian Analytical Sciences Society

ANALYSIS OF HYPERSPECTRAL REFLECTANCE FOR DISEASE CLASSIFICATION OF SOYBEAN FROGEYE LEAF SPOT USING KNIME ANALYTICS

(Analisis Refleksi Hiperspektra untuk Klasifikasi Penyakit Bintik Daun Frogeye Kacang Soya Menggunakan Analisis KNIME)

Yuhao Ang, and Helmi Zulhaidi Mohd Shafri

Department of Civil Engineering and Geospatial Information Science Research Centre (GISRC), Faculty of Engineering,
Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia

**Corresponding author: helmi@upm.edu.my*

Received: 10 October 2022; Accepted: 13 April 2023; Published: 23 June 2023

Abstract

The feasibility of classifying soybean frogeye leaf spot (FLS) has been investigated with the advance of hyperspectral technology. Hyperspectral reflectance data of healthy and FLS disease soybeans were used. The first step was to smooth out the data by using a filtering technique namely Savitzky-Golay to eliminate the noise of the spectrum. In order to select the most significant wavelengths, genetic algorithm (GA) was used as a forward feature selection technique. This analysis involved the implementation of machine learning (ML) algorithms, including decision trees, random forests, and stacking, to classify soybean FLS severity levels. Preprocessing ML steps including converting class numbers to strings, identifying and removing missing values, partitioning and normalizing data were implemented prior to the development of the model. Overall accuracy and the receiver operating characteristic curve measure were used to assess the performance of this analysis. All of these steps were carried out through KNIME analytical platform. Based on the results of the analysis, GA-stacking and random forest algorithms achieved the best overall accuracy of 85.9% and 84.3%, respectively. In terms of reproducibility, data flow control, data exploration, analysis and visualization, KNIME Analytics Platform provided great convenience in connecting tools graphically and ensuring the same results on different operating systems. The rapid implementation of workflow in KNIME Analytics Platform provided the opportunity to process hyperspectral reflectance data to classify crop diseases.

Keywords: hyperspectral reflectance, forward feature selection, genetic algorithms, machine learning, KNIME analytics platform

Abstrak

Kebolehlaksanaan mengklasifikasikan bintik daun frogeye kacang soya (FLS) telah disiasat dengan kemajuan teknologi hiperspektral. Data reflektan hiperspektral kacang soya yang sihat dan penyakit FLS telah digunakan. Langkah pertama ialah melicinkan data dengan menggunakan teknik penapisan iaitu Savitzky-Golay untuk menghapuskan hingar spektrum. Untuk memilih panjang gelombang yang paling ketara, algoritma genetik (GA) digunakan sebagai teknik pemilihan ciri ke hadapan. Analisis ini melibatkan pelaksanaan algoritma pembelajaran mesin (ML), termasuk pohon keputusan, hutan rawak, dan pemadatan, untuk mengklasifikasikan tahap keparahan FLS kacang kedelai. Langkah-langkah pra pemprosesan ML termasuk menukar nombor

kelas kepada rentetan, mengenal pasti dan mengeluarkan nilai yang hilang, mempartisi dan menormalisasi data telah dilaksanakan sebelum pembangunan model. Ketepatan keseluruhan dan ukuran lengkung ciri operasi penerima digunakan untuk menilai prestasi analisis ini. Kesemua langkah ini telah dijalankan melalui platform analisis KNIME. Berdasarkan keputusan analisis, algoritma GA-stacking dan random forest masing-masing mencapai ketepatan keseluruhan terbaik iaitu 85.9% dan 84.3%. Dari segi kebolehulangan, kawalan aliran data, penerokaan data, analisis dan visualisasi, Platform Analisis KNIME memberikan kemudahan besar dalam menyambungkan alatan secara grafik dan memastikan hasil yang sama pada sistem pengendalian yang berbeza. Pelaksanaan pantas aliran kerja dalam Platform Analisis KNIME memberi peluang untuk memproses data reflektan hiperspektral dalam klasifikasi penyakit tanaman.

Kata kunci: reflektan hiperspektral, pemilihan ciri ke hadapan, algoritma genetik, pembelajaran mesin, platform analisis KNIME

Introduction

As a legume, soybean is an essential source of proteins and fatty acids, the major feed sources of proteins, and the second primary producer of editable oils. The total production for soybean around the world was 367.76 million tons in 2021 and it is estimated that there will be a decrease of 17.04 million tons or 4.63% in the soybean production across the globe [1]. Apparently, the use as a primary source in pig feed increases the consumption of soybean worldwide. However, several diseases have severely harmed the quality and the seed of soybean. For example, frogeye leaf spot (FLS) is a soybean foliar disease caused by the fungus *Cercospora sojina* Hara (CSH) leads to yield losses, seed damage, and economic losses. FLS infections may result in yield losses of up to 60% [2, 3]. FLS is a polycyclic illness which its infection, symptom development, and reproduction may occur several times within the span of a single season [4]. It is crucial to assess and to detect the expanse of disease so that its economic losses can be estimated and proper management can be implemented.

The present method for determining the severity of foliar disease is mostly dependent on visual assessment. Experts physically evaluate the size of lesion area, the leaf color patterns, the distribution and the shape of leaves, the number of stems and branches, and also the density of soybean [5]. However, the visual assessment method is bias and can be altered by the experimental observations of experts. Another assessment method is molecular procedures such as enzyme-linked immunosorbent assay (ELISA) and polymerase chain reaction (PCR), which are very sensitive and capable of detecting disease. The drawback of this approach is time-consuming and destructive [6]. Thus, non-destructive highly efficient methods such as remote sensing tool remains warranted to assess the severity of soybean diseases.

Hyperspectral reflectance has attracted much focus in identifying and estimating the severity level of soybean disease. The hyperspectral technique consists of a narrow bandwidth and more than ten different channels that are capable of identifying subtle changes in an object of interest. Hyperspectral technique is frequently used to determine biophysical properties of certain crops, such as chlorophyll content, moisture content, trace element content, crop cell structure and nutrient contents [7-9]. Several studies have demonstrated the effectiveness and the potential of hyperspectral reflectance in identifying disease severity, such as rice false smut disease [10], citrus disease [11], oil palm *Ganoderma* disease [12], and soybean stem rot disease [13].

Rapid development of an automation system for classifying disease of the infected crops is an emerging area in precision agriculture. Machine learning (ML) method has been adopted to develop classification models using hyperspectral reflectance data for crop disease detection such as support vector machine, random forest (RF) and artificial neural network [14]. [15] has successfully developed random forest algorithms to classify wheat and rye leaves rust based on pure spectra. The result showed that the RF model based on the selected spectral wavelength achieved promising accuracy with the accuracy of 96.6%. In addition, hyperspectral data contain extra, redundant and highly correlated information across a large number of wavelengths, which can increase the complexity of analysis. Therefore, data dimensionality reduction, which cut down and optimize the whole wavelength, is

required. Several methods have been used effectively to reduce the dimensionality of hyperspectral data, including generic algorithms [16], principal component analysis (PCA) [17], and successive projection algorithm 18], and others [19-21]. To detect Huanglongbing disease in citrus, [11] used PCA to reduce the number of spectral features, and the selected PC scores were used in the classification algorithms such as quadratic discriminant analysis (QDA) and k-nearest neighbor (KNN). The results showed that the selected features achieved an overall average classification accuracy of 86% and 93% using KNN-based and QDA-based algorithms, respectively. In another similar study, [22] applied feature selection technique based on successive projection algorithm (SPA) to detect vineyard disease. The result demonstrated that SPA technique performed better than common VIs with the overall classification accuracy of about more than 96% for both DA and SVM. Nevertheless, implementing data dimensionality reduction in analyzing hyperspectral reflectance data can sometimes be computationally demanding.

The KNIME analytics platform is an open-source software that aims to solve these problems by creating a platform that can easily be extended with new tool integrations, which features a strongly typed data system, and permits creators to record in detail each step of the workflow [23]. One of the KNIME's benefits is its multimode of nodes for performing a sequence of ML process with its offerings of various advanced ML algorithms. In this study, the hyperspectral analysis was conducted to classify the FLS disease on codeless platform. This is the first study to use such platform to perform the hyperspectral processing including the feature selection and the classification approaches in KNIME analytics platform. Therefore, the main objectives of this study are to estimate FLS disease class using hyperspectral reflectance, to evaluate the performance of proposed feature selection by comparing with full dataset and to determine the feasibility of classifying FLS through the proposed ML method.

Materials and Methods

Data acquisition

Leaf hyperspectral reflectance data were collected using a FieldSpec® HandHeld 2 spectrometer (Analytical Spectral Devices, Inc., Colorado, USA). The hyperspectral region ranged from 325 to 1075 nm, with a resolution of 3 nm. The number of hyperspectral channels is 512. Each diseased or healthy leaf was considered as one sample. A total of 440 samples, comprising 340 diseased and 100 healthy leaves, were collected and evaluated. Four severity classes such 0%–1% (class 1), 1%–3% (class 2), 3%–6% (class 3), 6%–20% (class 4), and lastly the healthy leaves were regarded as class 0. The data was obtained from [24], which has been appropriately cited.

Preprocessing

The wavelengths were cut to remove the excitation wavelength and keep the useful wavelength that represent the reflectance of the disease. Only useful wavelength interval from 450 nm to 1000nm was remained. Then, the preprocessing technique such as Savitzky-Golay filter was used. The purpose of Savitzky-Golay filter is to smooth the original data while also eliminating noise from the spectrum [25]. The length of the filtering window and the order of the polynomial interpolation were set to 11 and 3, respectively. In addition, the first and second derivatives orders were generated over the course of the investigation. The preprocessing workflow was implemented using the installed KNIME interactive R statistics integration to use the R package and to execute them in KNIME analytic platform, as shown in Figure 1.

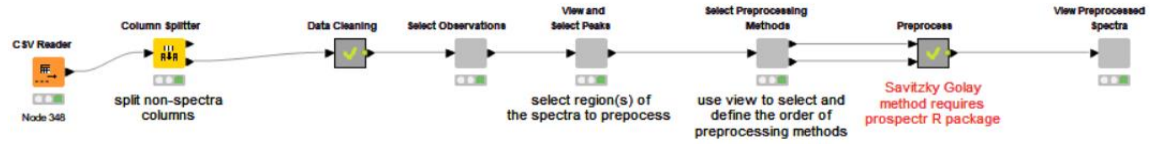

Figure 1. The workflow of preprocessing

Processing

Forward feature selection method such as genetic algorithm was applied to select the useful wavelengths and to eliminate the redundant wavelengths. The genetic algorithm is a chaotic approach for optimizing functions that is based on evolution and genetics. In genetic algorithm, an optimal set is determined based on evolution. The first step in selecting the features is to generate a population based on the subsets of possible features. The subsets from this population are then evaluated using a predictive model to obtain the weights of each individual on the populations. By that, N individuals are chosen, where N is the size of the population, given that previously calculated weights, and two new individuals are generated from each pair of individuals. In this way, each generation have the same number of individuals as the previous one which can bring it forwards when the algorithms run in many iterations. The next step is reproduction, which entails this numbers will be used to randomly split the parents' genes into pieces of varying sizes. The first new person will have the first piece from parent A and the second chunk from parent B. The second new person will be made out of the first and the second halves of parent B. Lastly, mutation mitigates the risk of the search falling into a local optimum and getting stuck. At each generation, in addition to the crossover, a random mutation is added. The reproduction and the mutation processes are then repeated once the new population has been produced, with the weights of each individual being recalculated. A workflow illustrating the forward feature selection process is shown in Figure 2. Using the KNIME workflow, raw data was recorded using the 'CSV Reader' node, and then combined with the 'Number to String' node to represent the disease severity classes into a string, and finally a metanode that comprised of several nodes was created to execute genetic algorithms via a sequence of forward feature selections.

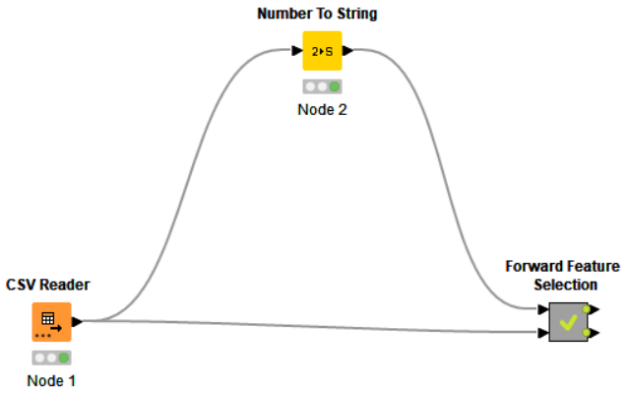

Figure 2. The workflow of feature selection

Modelling methods

Several ML algorithms were used to classify the severity level of FLS disease in soybean. Decision tree, random forest and stacking were used as the ML algorithms in this study. Each of the ML algorithms were implemented in the KNIME workflow. In general, raw data was recorded using a CSV Reader node, then the preprocessing ML steps such as converting the class numbers to strings, and identifying and removing missing values were carried out, followed by data partitioning (X-partitioner and X-aggregator) and normalization (normalizer and apply normalizer). Next, ML algorithms namely random forest, decision tree, and stacking were implemented and evaluated using a scorer and receiver operating characteristic (ROC) curve. The mechanisms of each ML algorithms are stated as followed.

Random forest: Random forests combine multiple trees and assemble decision trees through 'bagging' to create a final model. In this method, bootstrapping is used to select 'k' samples from the original dataset at random. The output is finalized by aggregating the predictions of each decision trees. These samples are used to build decision trees and create output depending on its predictions. The model parameters are adjusted accordingly through trial and error. The split criterion is assigned as information gain. The tree depth and the minimum node size were tuned accordingly and set to 5 and 2, respectively. The workflow of random forest is shown in Figure 3.

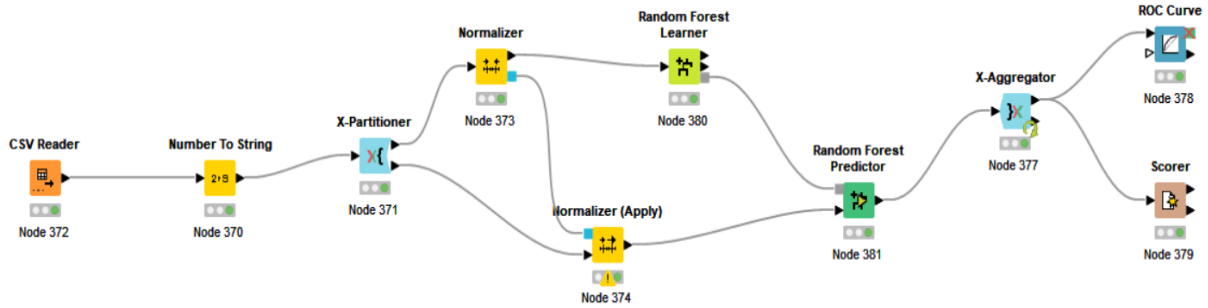

Figure 3. The workflow of random forest algorithm

Stacking: In stacking, heterogeneous base learners consisting different ML models are applied and trained in parallel. Then, heterogeneous base-learners are combined in aggregating the output of each single base-leaners by training a meta-learner to create a prediction based on different base learner's predictions. The purpose of the meta-learner is to determine how to combine the input predictions made by the base-learners with the output made by the training dataset to achieve better predictions [26]. Stacking algorithm was implemented by utilizing KNIME Weka Data Mining Integration, which provide Knime plugins that include the Weka data mining framework's functionality [27]. In this study, Naïve Bayes, J48 and Kstar algorithms were selected as base learners, and logistic regression was used as meta-learner. The workflow of stacking algorithm is shown in Figure 4.

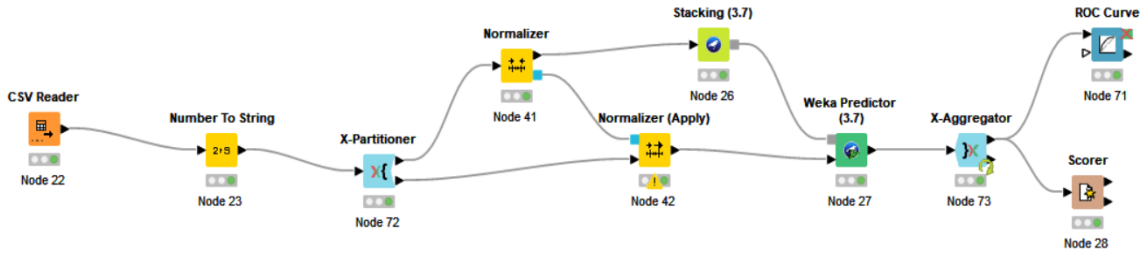

Figure 4. The workflow of stacking algorithm

Decision trees: A decision tree is a tree-representation classifier in which each leaf node corresponds to a class label and each inside node represents a feature. Branches represent conjunctions of features that lead to those class labels. Pruning is applied as it is the important part in decision trees, used to decrease the size of the decision tree. Several parameters were set such as quality measures (Gini index), minimum number records per node (8) and pruning method (Minimum Description Length). The workflow of decision tree is shown in Figure 5.

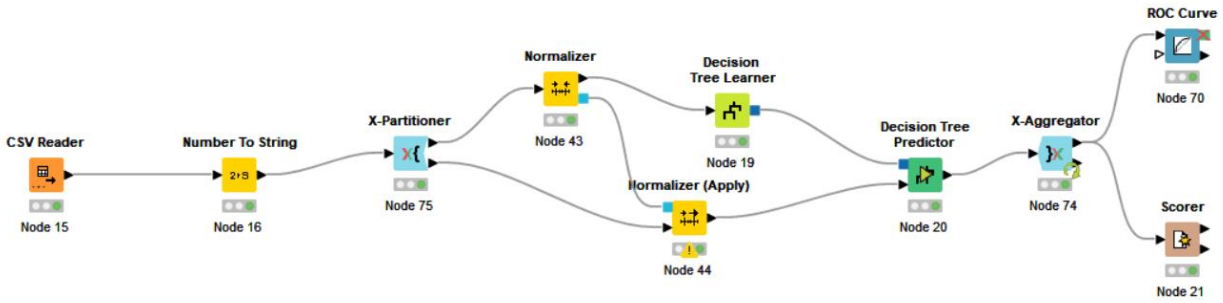

Figure 5. The workflow of decision tree algorithm

Performance evaluation

Cross-validation was used as a resampling strategy in the evaluation process to assess the ML models. The number of groups into which a particular data sample to be divided is denoted by K-fold, which is used to separate each observation from the original dataset into an equal probability of occurring in the training and the test sets. In this investigation, K was set to 10, training the (K-1) folds and testing the model with the K^{th} fold remaining. The procedure is repeated K times until the final average score is determined. Cross-validation is applied for the evaluation metrics in the accuracy assessment. In order to analyze the efficacy of the ML algorithms, accuracy assessment is important in the classification of FLS diseases. The accuracy of each ML algorithms was evaluated using ground truth data for the five severity classes of soybean FLS. The confusion matrix approach was used to assess the performance of a classification model by comparing the actual classes to the predictions of the ML model. The overall accuracy was computed by summing the number of correctly classified classes and dividing it by the total number of classes. This overall accuracy is the outcome of ML algorithms classifying the severity levels of soybean FLS using leaf hyperspectral data. The ROC measure is also used to quantify the prediction accuracy of a predictive model. It is based on the trade-off between the true positive rate and the false positive rate when probability threshold is used.

Results and Discussion

Spectral reflectance for raw, first and second derivatives of FLS disease

Spectral reflectance profile taken from sample spectra is shown graphically for each severity stages of soybean FLS (class 0, 1, 2, 3 and 4) in Figure 6. Basically, the spectral profile showed a low reflectance in the visible region with small peak in the green region indicating a sudden increase beginning at 690 nm reaching a peak in the NIR region. These are regular for green plant reflectance spectra. A comparison of profile showed larger difference in the reflectance between each severity classes of soybean FLS at some wavelengths compared to other wavelengths.

In the visible spectrum (607– 690 nm), soybean FLS with different severity stages were very similar and many overlapped each other, while in the infrared region (702 – 966 nm), the spectra of soybean FLS with different severity classes separated from each other. Healthy soybean had the highest NIR reflectance compared to others in 774 nm till 806 nm (Figure 6). The second highest were soybeans with class 1 and 2. Soybean FLS with class 4 had the lowest reflectance.

The first derivative of spectral reflectance is shown in Figure 7(a). Overall, the first derivative of spectral reflectance of FLS soybean disease exhibited a tiny reflectance peak about 515 nm and a substantial first derivative reflectance peak in the red-edge area (679 – 714 nm). Healthy soybean with class 0 indicates the highest peak at 717.87nm in the red edge region, which is more prone to shifting to the right as compared to FLS soybean disease. The second derivative of spectral reflectance of FLS soybean disease is shown in Figure 7(b). Second derivative reflectance spectra with FLS soybean disease showed distinct peaks at 685 nm and 732 nm.

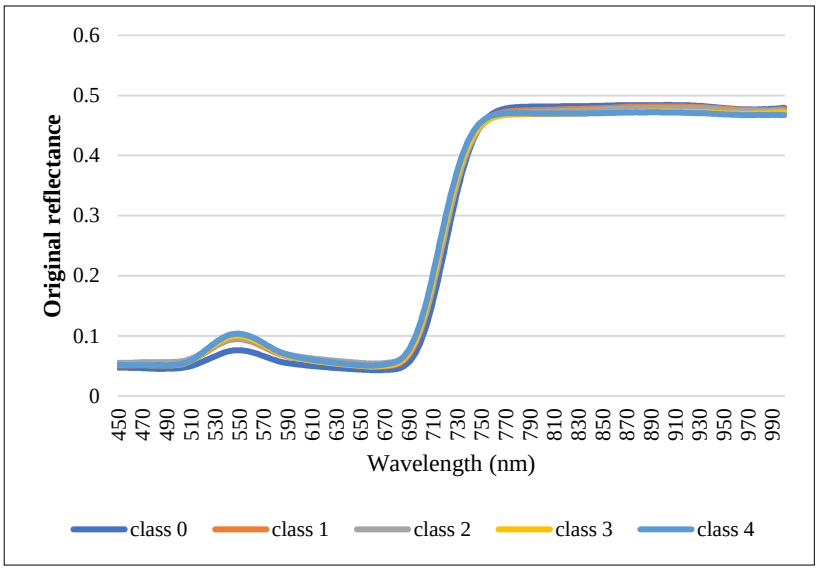

Figure 6. Original spectral reflectance of FLS disease

The first derivative equation is crucial in plant disease detection with space of the lesion density function at a fixed position [28]. The spectrum will become complicated and complex as the number of disease symptoms increases. Different stages of disease severity attributed to the collected spectrum variation when plants in the particular area facing disease spreads. Therefore, the first derivative can remove both additive and multiplicative effects in spectrum [29]. Lu et al. [30] showed that the first derivative between 500 – 1000 nm was better in discriminating leaf curl on tomatoes than the original reflectance spectrum. In general, the first and the second derivatives indicated the potential for estimating crop disease because of their insensitivity to the soil background [31, 32].

Classification of soybean FLS disease using ML models

Table 1 shows the overall accuracy of ML algorithms used in the classification of soybean FLS disease. Random forest and stacking achieved the best overall accuracy using the selected first and second derivatives of spectral reflectance. Random forests achieved an overall accuracy of 84.3% and 77.7% for both selected first and second derivatives of spectral reflectance, respectively. The stacking algorithms have improved to 85.9% and 79.1% from 70% and 56.1% on both selected first and second derivatives of spectral reflectance, respectively.

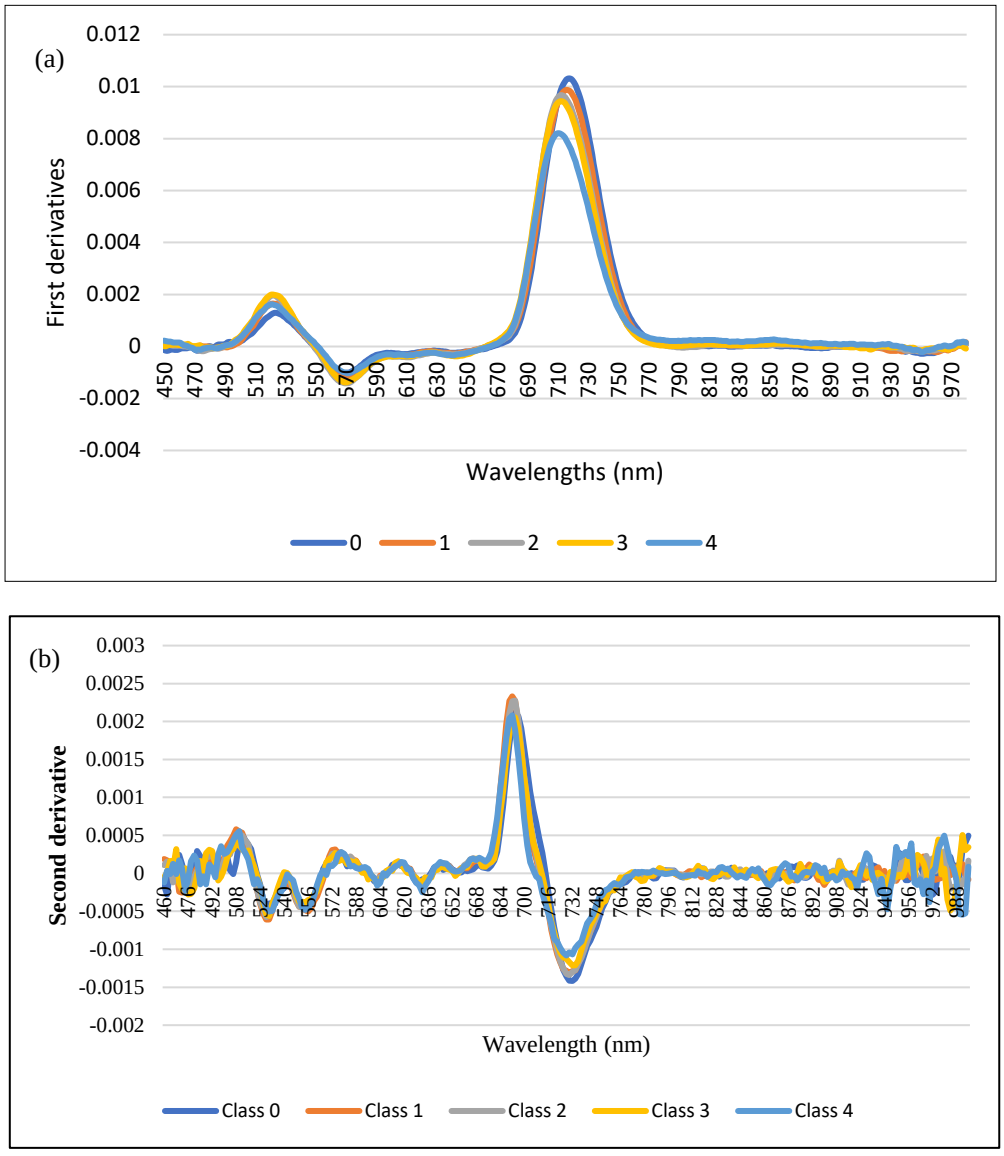

Figure 7. (a) First derivative of spectral reflectance of FLS disease and (b) Second derivative of spectral reflectance of FLS disease

Table 1. Overall accuracy of ML models

Data	Raw	First derivative (full)	Second derivative (full)	First derivative+ Savitzky-Golay+feature selection (genetic algorithms)	Second derivative+ Savitzky-Golay+feature selection (genetic algorithms)
Decision tree	55.4	64.7	55.9	65.0	62.7
Random forest	68.4	75.9	69.5	84.3	77.7
Stacking	69.7	70.0	56.1	85.9	79.1

Decision tree had the lowest and the poorest accuracy performance even after the feature selection algorithm was implemented. For example, only small increases with 0.2% for implementing the feature selection using the first derivative data. Noticeably, there was not much difference between the accuracy of decision tree algorithms using raw data and second derivative.

Based on the above results, incorporating genetic algorithms (GA) as feature selections has improved the overall accuracy for most of the ML algorithms since GA can effectively screen the useful characteristic bands for soybean FLS disease detection, thereby reducing the number of variables in the model, and thus producing high classification accuracy. Our result agrees with the studies of Chen et al. [33], who proved that genetic algorithms and following support vector machine is the best performance model with overall accuracy of 90.7% for leaves and 92.6% for stems.

Adding to performance analysis, Figure 8 illustrates the ROC curve for stacking algorithm, which is the best among the other algorithms. This analysis contributes an insight into how well the model classify the different severity classes (classes 0, 1, 2, 3 and 4) of soybean FLS. Overall, the curve reached a high value for the metric area under curve, in this case, 95% area of the graph was under the curve, showing the model is good in separating the soybean FLS classes.

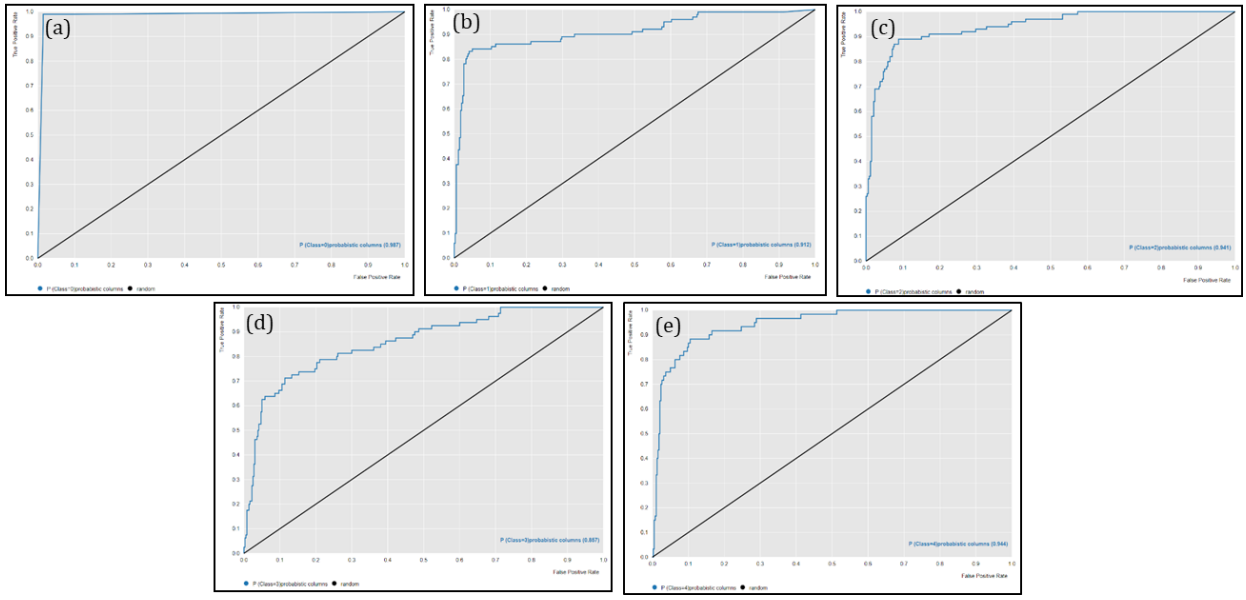

Figure 8. Stacking model ROC curve for class 0-4: (a) Class 0, (b) Class 1, (c) Class 2, (d) Class 3 and (e) Class 4

Conclusions

Based on the results, it was concluded that stacking algorithm achieved the highest accuracy after genetic algorithm implementation. Future studies will assess different feature selection techniques to analyze hyperspectral reflectance in the classification of crop disease. Different preprocessing techniques will also be adopted to ensure the proper removal of noise in spectrum. Besides, several ML algorithms will be proposed and applied on the analysis of hyperspectral reflectance. KNIME integrations provide additional functionality such as processing of complex hyperspectral preprocessing (e.g. Savitzky-Golay), as well as the use of advanced ML algorithms. Therefore, the functionality of KNIME as a codeless tool for analyzing hyperspectral reflectance should be further investigated in order to confirm its practicality and usefulness. The results obtained make it possible to

provide fast implementation of workflow, including model development for evaluating the degree of crop disease using hyperspectral reflectance data. Future studies will utilize other advanced machine learning algorithms to improve the accuracy of the model.

Acknowledgement

We would like to acknowledge the Malaysian Ministry of Higher Education (MOHE) for supporting us through the Malaysian Research University Network (MRUN) Long-Term Research Grant Scheme (LRGS). It is supported under Grant no: 203.PKOMP.6770007 with UPM vote no: 6300268-10801. A credit is also given to [24] for the permission to use their data in this study.

References

1. USDA Foreign Agricultural Service (2022). World Agricultural Production. Access from https://www.fas.usda.gov/data/world-agricultural-production.
2. Dashiell, K. E. and Akem, C. N. (1991). Yield losses in soybeans from frogeye leaf spot caused by *Cercospora sojina*. *Crop Protection*, 10(6): 465-468.
3. Mwase, W. F. and Kapooria, R. G. (2001). Incidence and severity of frogeye leaf spot and associated yield losses in soybeans in agroecological zone II of Zambia. *Mycopathologia*, 149(2): 73-78.
4. Mian, M. A. R., Missaoui, A. M., Walker, D. R., Phillips, D. V. and Boerma, H. R. (2008). Frogeye leaf spot of soybean: A review and proposed race designations for isolates of *Cercospora sojina* Hara. *Crop Science*, 48(1): 14-24.
5. George, A. P., Broadley, R. H., Nissen, R. J. and Ward, G. (2002). Effects of new rest-breaking chemicals on flowering, shoot production and yield of subtropical tree crops. *Acta Horticulturae*, 575(4): 835-840.
6. Bukhari, K. A. I. M., Nithya, K., Valluvaparidasan, V., Paranidharan, V. and Velazhahan, R. (2012). Detection of *Colletotrichum falcatum* causing red rot of sugarcane by enzyme-linked immunosorbent assay. *Archives of Phytopathology and Plant Protection*, 45(7): 823-830.
7. Clevers, J. G. P. W., Kooistra, L. and Schaepman, M. E. (2010). Estimating canopy water content using hyperspectral remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 12(2): 119-125.
8. Peng, Y., Zhang, M., Xu, Z., Yang, T., Su, Y., Zhou, T., Wang, H., Wang, Y. and Lin, Y., 2020. Estimation of leaf nutrition status in degraded vegetation based on field survey and hyperspectral data. *Scientific Reports*, 10(1): 1-12.
9. Sonobe, R., Yamashita, H., Nofrizal, A. Y., Seki, H., Morita, A. and Ikka, T. (2021). Use of spectral reflectance from a compact spectrometer to assess chlorophyll content in *Zizania latifolia*. *Geocarto International*, 2021: 1-13.
10. Liu, Z. Y., Shi, J. J., Zhang, L. W. and Huang, J. F. (2010). Discrimination of rice panicles by hyperspectral reflectance data based on principal component analysis and support vector classification. *Journal of Zhejiang University: Science B*, 11(1): 71-78.
11. Sankaran, S. and Ehsani, R. (2010). Spectral evaluations for developing optical methods of Huanglongbing (HLB) detection in citrus orchards. *Biosystems*, 2021: 121-123.
12. Izzuddin, M. A., Seman Idris, A., Nisfariza, M. N., Nordiana, A. A., Shafri, H. Z. M. and Ezzati, B. (2017). The development of spectral indices for early detection of Ganoderma disease in oil palm seedlings. *International Journal of Remote Sensing*, 38(23): 6505-6527.
13. Vigier, B. J., Pattey, E. and Strachan, I. B. (2004). Narrowband vegetation indexes and detection of disease damage in soybeans. *IEEE Geoscience and Remote Sensing Letters*, 1(4): 255-259.
14. Zhang, N., Yang, G., Pan, Y., Yang, X., Chen, L. and Zhao, C. (2020). A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. *Remote Sensing*, 12(19): 1-34.
15. Wójtowicz, A., Piekarczyk, J., Czernecki, B. and Ratajkiewicz, H. (2021). A random forest model for the classification of wheat and rye leaf rust symptoms based on pure spectra at leaf scale. *Journal of Photochemistry and Photobiology B: Biology*, 223: 112278.

16. Chen, X., Dong, Z., Liu, J., Wang, H., Zhang, Y., Chen, T., Du, Y., Shao, L. and Xie, J. (2020). Hyperspectral characteristics and quantitative analysis of leaf chlorophyll by reflectance spectroscopy based on a genetic algorithm in combination with partial least squares regression. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 243: 118786.
17. Liu, Z. Y., Wu, H. F. and Huang, J. F. (2010). Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. *Computers and Electronics in Agriculture*, 72(2): 99-106.
18. Zhu, H., Chu, B., Zhang, C., Liu, F., Jiang, L. and He, Y. (2017). Hyperspectral imaging for presymptomatic detection of tobacco disease with successive projections algorithm and machine-learning classifiers. *Scientific Reports*, 7(1): 1-12.
19. Wei, X., Johnson, M. A., Langston, D. B., Mehl, H. L. and Li, S. (2021). Identifying optimal wavelengths as disease signatures using hyperspectral sensor and machine learning. *Remote Sensing*, 13(14): 1314283.
20. Tian, L., Xue, B., Wang, Z., Li, D., Yao, X., Cao, Q., Zhu, Y., Cao, W. and Cheng, T. (2021). Spectroscopic detection of rice leaf blast infection from asymptomatic to mild stages with integrated machine learning and feature selection. *Remote Sensing of Environment*, 257(8): 112350.
21. Zhang, J., Cheng, T., Guo, W., Xu, X., Qiao, H., Xie, Y. and Ma, X. (2021). Leaf area index estimation model for UAV image hyperspectral data based on wavelength variable selection and machine learning methods. *Plant Methods*, 17(1): 1-14.
22. Al-Saddik, H., Simon, J. C. and Cointault, F. (2019). Assessment of the optimal spectral bands for designing a sensor for vineyard disease detection: The case of 'Flavescence dorée.' *Precision Agriculture*, 20(2): 398-422.
23. Fillbrunn, A., Dietz, C., Pfeuffer, J., Rahn, R., Landrum, G. A. and Berthold, M. R. (2017). KNIME for reproducible cross-domain analysis of life science data. *Journal of Biotechnology*, 261(July): 149-156.
24. Liu, S., Yu, H., Sui, Y., Zhou, H., Zhang, J., Kong, L., Dang, J. and Zhang, L. (2021). Classification of soybean frogeye leaf spot disease using leaf hyperspectral reflectance. *PLoS ONE*, 16(9): 1-23.
25. Ruffin, C. and King, R. L. (1999). The analysis of hyperspectral data using Savitzky-Golay filtering-theoretical basis (part 1). In *IGARSS'99 Proceedings: Remote Sensing of the System Earth—A Challenge for the 21st Century: IEEE 1999 International Geoscience and Remote Sensing Symposium, 28 June–2 July 1999, Congress Centrum Hamburg*: pp. 756-758.
26. Džeroski, S. and Ženko, B. (2004). Is combining classifiers with stacking better than selecting the best one? *Machine Learning*, 54(3): 255-273.
27. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1): 10-18.
28. Gregory, P. H. (1968). Interpreting plant disease dispersal gradients. *Annual review of Phytopathology*, 6(1): 189-212.
29. Tsai, F. and Philpot, W. (1998). Derivative analysis of hyperspectral data. *Remote Sensing of Environment*, 66(1): 41-51.
30. Lu, J., Zhou, M., Gao, Y. and Jiang, H. (2018). Using hyperspectral imaging to discriminate yellow leaf curl disease in tomato leaves. *Precision Agriculture*, 19(3): 379-394.
31. Li, Y., Demetriades-Shah, T. H., Kanemasu, E. T., Shultis, J. K. and Kirkham, M. B. (1993). Use of second derivatives of canopy reflectance for monitoring prairie vegetation over different soil backgrounds. *Remote Sensing of Environment*, 44(1): 81-87.
32. Muhammed, H. H. (2005). Hyperspectral crop reflectance data for characterising and estimating fungal disease severity in wheat. *Biosystems Engineering*, 91(1): 9-20.
33. Cen, Y., Huang, Y., Hu, S., Zhang, L. and Zhang, J. (2022). Early detection of bacterial wilt in tomato with portable. *Remote Sensing*, 14(12): 2882.