VERİ BİLİMİ DERGİSİ
www.dergipark.gov.tr/veri

Classification of Some Biochemical Properties with J48 Classification Tree Algorithms in Hyperspectral Data

Sinan DEMİR[1]*, Levent BAŞAYİĞİT[2],

[1,2]*Isparta University of Applied Sciences. Faculty of Agriculture, Department of Soil Science and Plant Nutrition*

Makale Bilgisi

Başvuru:
21/09/2022
Kabul:
28/11/2022

Abstract

Hyperspectral sensing methods have been used in agriculture for many years to determine physiological events against abiotic and biotic stress factors of plants. Studies on the determination of wavelengths associated with biochemical properties using hyperspectral data of plants in the visible (400-700 nm) region and near-infrared (700-1300 nm) region under sustainable agricultural production models are limited. In this research, Isparta Oil rose with the geographical indication product from organic farming medicine and aromatic plants from sustainable production models were used considering that the effect of cultural practices from abiotic and biotic stress factors is minimal. Hyperspectral measurements and feature selection of chlorophyll a (mg g^{-1}), chlorophyll b (mg g^{-1}), chlorophyll a+b (mg g^{-1}), total flavonoid content (mg catechin g^{-1}), total phenolic content (mg GAE g^{-1}), and total antioxidant capacity (mg TEAC g^{-1}) biochemical properties are made using the J48 classification tree algorithm in organically grown Isparta Oil Rose leaves. In the classification algorithm, different hyperspectral data (bands) are used as independent variables for each biochemical feature, while the amounts of each biochemical feature of the dependent variable are lower and higher than the mean, and the binary response variable (binary) is taken into the model. In the selection of independent variables, the correlation-based CfsSubsetEval algorithm, which does not cause multicollinearity, was used. The areas under the classification accuracies, sensitivity, specificity, and receiver operating characteristic (ROC) curves were determined, which are the classification performances of chlorophyll a, chlorophyll b, chlorophyll a+b, total flavonoid content, total phenolic substance content, and total antioxidant capacity. The highest classification performance selected the hyperspectral features for chlorophyll a+b content was determined as “76.25%”, “0.553”, “0.802”, and “0.631” respectively. As a result of the study, it was concluded that the chlorophyll a+b (mg g^{-1}) content was determined with the J48 classification tree algorithm with the highest accuracy in the classification of the biochemical contents of made the organic farming Isparta Oil rose leaves with visible and near-infrared hyperspectral data.

Keywords: *Hyperspectral Sensing, Biochemical Feature, Organic Farming, Cfssubseteval, J48 Decision Tree Algorithm, Classification*

* İletişim e-posta: demirsinan.07@gmail.com

Hiperspektral Verilerde J48 Sınıflandırma Ağacı Algoritmaları ile Bazı Biyokimyasal Özelliklerin Sınıflandırılması

Özet

Hiperspektral algılama yöntemleri tarımda uzun yıllar bitkilerin abiyotik ve biyotik stres faktörlerine karşı fizyolojik olayların belirlenmesinde kullanılmaktadır. Sürdürülebilir tarımsal üretim modelleri altındaki bitkilerin görünür (400-700 nm) bölge ve yakın kızılötesi (700-1300 nm) bölgedeki hiperspektral verileri kullanılarak biyokimyasal özellikler ile ilişkili dalga boylarının belirlenmesi üzerine araştırmalar sınırlı sayıdadır. Bu araştırmada, sürdürülebilir üretim modellerinden organik tarım yapılan tıbbı ve aromatik bitkilerden coğrafi işaretli Isparta Yağ gülü abiyotik ve biyotik stres faktörlerinden kültürel uygulamalardan kaynaklı etkinin en az olması göz önüne alınarak kullanılmıştır. Organik tarım yapılan Isparta Yağ gülü yapraklarında hiperspektral ölçümler ve klorofil a (mg g^{-1}), klorofil b (mg g^{-1}), klorofil a+b (mg g^{-1}), toplam flavonoid madde içeriği (mg catechin g^{-1}), toplam fenolik madde içeriği (mg GAE g-1) ve toplam antioksidan kapasitesi (mg TEAC g^{-1}) biyokimyasal özelliklerin J48 sınıflandırma ağacı algoritması kullanılarak öznitelik seçimi yapılmıştır. Sınıflandırma algoritmasında her bir biyokimyasal özellik için bağımsız değişken olarak farklı hiperspektral veriler (bantlar) kullanırken, bağımlı değişken her bir biyokimyasal özelliklerinin miktarlarının ortalamaya göre düşük ve yüksek olması ikili yanıt değişkeni (binary) şeklinde modele alınmıştır. Bağımsız değişken seçiminde ise çoklu doğrusal bağlantıya (multicollinearity) neden olmayan korelasyon tabanlı CfsSubsetEval algoritması kullanılmıştır. Klorofil a, klorofil b, klorofil a+b, toplam flavonoid madde içeriği, toplam fenolik madde içeriği ve toplam antioksidan kapasitesi özelliklerinin sınıflandırma performansları olan sınıflama doğrulukları, duyarlılık, özgüllük, alıcı işletim karakteristiği (ROC) eğrileri altında kalan alanlar belirlenmiştir. Hiperspektral özelliklerden seçilen klorofil a+b içeriği için en yüksek sınıflandırma performansı sırasıyla "76.25", "0.553", "0.802" ve "0.631" olarak belirlenmiştir. Çalışma sonucunda organik tarım yapılan Isparta Yağ gülü yapraklarının biyokimyasal içeriklerinin görünür ve yakın kızılötesi hiperspektral veriler ile sınıflamasında klorofil a+b (mg g^{-1}) içeriğinin J48 sınıflandırma ağacı algoritması ile en yüksek doğrulukta belirlendiği sonucuna varılmıştır.

Anahtar Kelimeler: *Hiperspektral Algılama, Biyokimyasal Özellik, Organik Tarım, Cfssubseteval, J48 Karar Ağacı Algoritması, Sınıflama*

1 Introduction

Hyperspectral sensors, which can detect in a wide wavelength range of the electromagnetic spectrum, have increased their academic and agricultural use with the developments in remote sensing technologies in the last 20 years. The increasing use of experimental and algorithmic developments due to the development of remote sensing technologies allows the successful use of data with high spectral resolution obtained with hyperspectral sensors in agricultural applications [1]. Research continues on the integration of technological developments in terms of optimum management of scarce resources in agriculture. It has increased the use of remote sensing sensors for Agriculture 4.0, which increases profitability with the integration of technology in agriculture, includes sustainable environmentally friendly practices, and aims to provide optimum efficiency with scarce resources. It is an important tool in the early determination of abiotic and biotic stress factors at the field scale with the remote sensing data, and technology developments that have been used in agricultural applications for many years [2]-[3]. Hyperspectral sensing techniques make it possible to detect small changes in spectral channels narrower than 10 nanometers, with the development of faster computational techniques for narrow absorption properties resulting from biochemical and biophysical properties of vegetation. The absorption properties of plants are formed as a result of electron transitions in chlorophyll, bending, and stretching of chemical bonds in plant sap and other chemical compounds [4]. The characteristic absorption properties of organic compounds at wavelengths in the visible and near-infrared region in the range of 325-1075 nanometers, it has allowed the traceability of biochemical changes in living vegetation under farmland conditions. Early estimation of leaf

biochemical content has been significantly improved by narrowband sensor (hyperspectral) applications compared to broadband (multispectral) sensor applications [5],[6],[7],[8]. Hyperspectral remote sensing techniques have been successfully used in agricultural plants, especially at the leaf level, to obtain parameters such as chlorophyll, nitrogen, cellulose, and lignin [8],[9],[10],[11].

The extent and threshold of physiological responses to abiotic and biotic stressors in agricultural production depend on the duration and magnitude of environmental stress. Also, the extent and threshold of physiological responses to abiotic and biotic stressors depend on the synthesis of anti-stress substances and the plant's adaptive potential and morpho-functional processes, particularly metabolic adjustments [18]. It is believed that secondary metabolites synthesized by the cell structures of plants under conditions of abiotic and biotic stress protect from oxidative effects [9]. These are named secondary metabolites that include certain vitamins, carotenoids, terpenoids, essential oils, and phenolic compounds. These are important for normal plant growth and defense against infection and injury [11]. Secondary metabolites have been detected in many plant sources. Examples of these plants are field crops, horticultural crops, and medicinal and aromatic plants [12],[13],[14]. Medicinal and aromatic plants are plants with high secondary metabolite content to reduce the effect of abiotic and biotic stress factors [15], [16]. Isparta oil rose (*Rosa Damascena* Mill.) is the most important rose species producing high-value essential oil. Organic farming of *R. Damascena* is carried out to be used in perfume, medicine, and the food industry in Isparta conditions. Considering these reasons, it is important to determine the effect of abiotic and biotic stress factors on food security and safe food supply in agriculture. Following the change in the secondary metabolite content of *R. Damascena* during the vegetation period can be done by analytical means [17],[18],[19].

Studies were conducted to determine its biochemical contents such as chlorophyll a, b, a+b, total substance contents of phenolic and flavonoid, and total antioxidant capacity while *R. Damascena* was in vegetation seasons under abiotic and biotic stress factors. [17],[18],[19]. In previous studies conducted by analytical means, the determination of the characteristics of the change in leaf content by hyperspectral measurements has not been adequately investigated. It is necessary to investigate the developments in machine learning algorithms and the use of plant datasets with high spectral characteristics in organic farming systems. This study was conducted using the hyperspectral data of the leaves in the range of 325-1075 nanometers in the vegetation period of the chlorophyll a, b, a+b, total substance contents of phenolic and flavonoid, and total antioxidant capacity of the *R. Damascena* plant grown under organic farming systems in Isparta conditions. With the J48 algorithm, the classification of characteristic bands was investigated. The results of this study will provide an understanding of the detectability of secondary metabolites that cannot be measured in farmland conditions for smart agriculture and digital agriculture, and the usability of the relationship of hyperspectral data with the biochemical contents of users.

2 Materials

In this study, *R. Damascena* leaves collected from 10 plots from Isparta oil rose gardens, which were grown organically under the same ecological conditions in Ardıçlı village of Keçiborlu district of Isparta province, according to the randomized plot design, were used to create a hyperspectral data set. *R. Damascena* leaves were collected in 8 periods in the 10-days apart during the 2019-2020 vegetation season. Hyperspectral measurements from leaf samples taken from each period were using the ASD Fieldspec RS3 [20] spectroradiometer that the technical characteristics of are given in Table 1.

Plant probe uses a halogen bulb as a 12 Volt direct current light source. The plant probe allows the R. *Damascena* leaf between it and the ASD leaf clip to quickly record the energy coming from the light source in the 325-1075 nanometer range of the electromagnetic spectrum by the spectroradiometer device with a fiber optic cable. The hyperspectral measurement application from *R. Damascena* leaves using the Plant probe is given in Figure 1.

Table 1. Hyperspectral data parameters.

Name	Parameters
Electromagnetic spectrum range	325-1075 nm
Number of bands	750
Spectral resolution	3.5 nm
Spectral width	1.5 nm
Scanning time	17 ms
Wavelength Accuracy	±1 nm
Plant probe spot size	10 mm
Plant probe power requirements	12-18 VDC, 6.5 W

Figure 1. The Hyperspectral measurement from *R. Damascena leaf*

3 Method Applied

In this study, a method has been proposed to determine the change in leaf biochemical content of abiotic and biotic stress factors occurring during the vegetation season of Isparta oil roses with organic farming, using hyperspectral data. First, Chlorophyll (a, b, and a+b) content was determined according to the method defined by Zhang and Huang [21]. The total phenolic content was determined according to the Folin-Ciocalteau method defined by Singleton and Rossi [22]. Total flavonoid content was made according to the method specified by Zhishen et al. [23]. Total antioxidant capacity was determined according to the DPPH (1,1-diphenyl-2-picrylhydrazyl) reported in a method defined by Kumaran and Karunakaran [24]. Chlorophyll a, b, a+b, total substance contents of phenolic and flavonoid, and total antioxidant capacity contents of *R. Damascena* leaves were analyzed under laboratory conditions. Then, descriptive statistics of determining chlorophyll a, b, a+b, total substance contents of phenolic and flavonoid, and total antioxidant capacity contents were calculated. And then, correlations of biochemical contents were calculated. Biochemical analyzes were performed for 240 samples according to the methods mentioned above. According to the center of the data set of the determined biochemical contents, categorical classes were created as high and low as two classes. From the hyperspectral data set from features in the range of 325-1075 nanometers were determined 501 bands in the range of 400-900 nanometers. The reflection values of the bands, which are highly correlated with biochemical contents and do not show multicorrelation with each other, were calculated with the feature selection process CfsSubsetEval algorithm, which is widely used in weka software. After the feature selection was completed, the J48 algorithm, one of the machine learning algorithms, was used to classify the biochemical contents of *R. Damascena* leaves with hyperspectral data. The proposed method is given in Figure 2. The proposed method is shown in Weka (Figure 3).

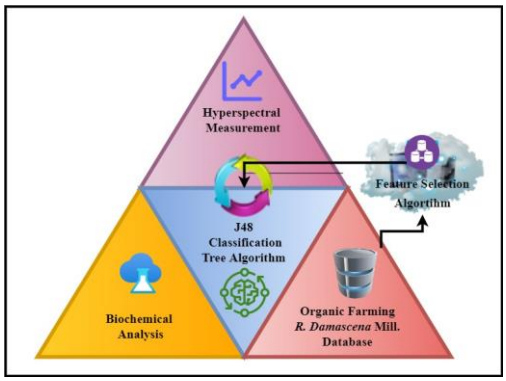

Figure 2. Flowchart of the method applied

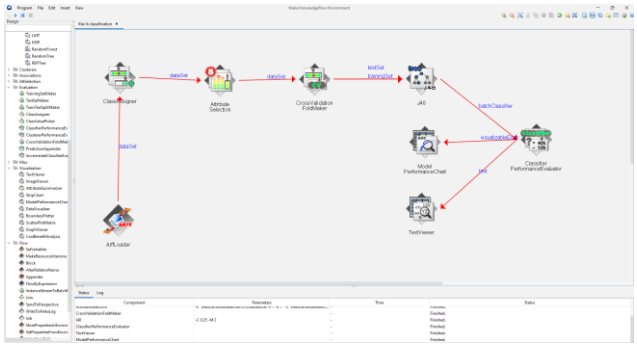

Figure 3. Weka knowledge flow

As seen Fig. 2, the Organic Farming *R. Damascena* Mill. Database is collected by hyperspectral measurement and biochemical analysis result transferred to the Weka software including machine learning algorithms and methods in the computer, and the proposed method has been applied feature selection and J48 classification tree algorithm. Correlation-based CfsSubsetEval algorithm was used in the selection of hyperspectral features to be used in the classification of biochemical contents. CfsSubsetEval uses a search algorithm. In the algorithm, the attributes that have the best relationship with the class label are determined. In this case, the determined hyperspectral feature group contains a high correlation with the class label. However, other hyperspectral features are found to be less important [25]. In this study, hyperspectral features were determined for each biochemical content class label by using the BestFirst search algorithm in the CfsSubSetEval algorithm. The following is a description of the evaluation criteria calculated using dataset values. The confusion matrix was obtained by selected 10-fold cross-validation and computing of the evaluation results. Performance criteria of J48 machine learning models developed to predict biochemical contents from the hyperspectral dataset are given equation below;

$$Accuracy\ (\%) = [\frac{TP+TN}{TP+TN+FP+FN}] \times 100 \quad (1)$$

$$Sensitivity\ (\%) = [\frac{TP}{TP+FN}] \times 100 \quad (2)$$

$$Specificity\ (\%) = [\frac{TN}{TP+TN}] \times 100 \quad (3)$$

ROC analysis is another important analysis used in the performance measurement of the classification process performed on the data set. Simply defined, An ROC curve can be defined as a graph of sensitivity and 1-specificity calculated from the confusion matrix value from a classification result [26]. The overall accuracy of the classification is determined using the area under the ROC curve (AUC). The AUC values are calculated from 0 to 1. In general, an AUC of 0.7 to 0.8 is acceptable.

4 Results

In this study, the descriptive statistics of the biochemical analysis results in the workflow given in Figure 2 were calculated in the R studio program [27]. Descriptive statistics of chlorophyll a (Chl a), chlorophyll b (Chl b), chlorophyll a+b (Chl a+b), total flavonoid content (TFLC), total phenolic content (TPEC), and total antioxidant capacity (TAC) in *R. Damascena* leaves are given in Table 2. The arithmetic mean (mean) value was calculated by dividing the values of the biochemical contents by the number of variables. The lowest mean value of leaf TPEC content is 8.61. The highest mean value of leaf TAC content is 86.67 (Table 2). The standard deviation is the value used to summarize the data values. It shows the distribution of the data according to the mean. If the data set is close to the mean, the standard deviation gets smaller. The standard deviation values calculated according to the results of the biochemical analysis are given in Table 2. The lowest standard deviation value of leaf Chl a content is 0.65. The highest standard deviation value of leaf Chl b content is 5.50 (Table 2). The standard deviation is a measure of the prevalence of the distribution. However, it is impossible to say much about the standard deviation and distribution. Because the value found is an absolute value. Is this value big or small? It is necessary to look at the coefficient of variation to decide whether. The coefficient of variation indicates how many percent the standard deviation changes with respect to the mean. The coefficient of variation values calculated according to the results of the biochemical analysis is given in Table 2. The lowest change coefficient of variation value of leaf Chl a content is 2.43%. The highest change coefficient of variation value of leaf TPEC content is 37.73% (Table 2). The difference between the largest and smallest value in an attribute of the data set is called the distribution width (Range). The range calculated for the biochemical data set is given in table 2. The lowest value distribution width of leaf Chl a content is 3.48. The highest value distribution width of leaf TAC content is 71.78 (Table 2).

Biochemical analysis was done on leaves of *R. Damascena* that were made hyperspectral measurements. Using the hyperspectral data set created for the *R. Damascena* 2019-2020 vegetation season, 2 classes were determined according to the arithmetic mean value of the biochemical data in

order to make the biochemical content faster with machine learning algorithms. Hyperspectral features were selected with the CfsSubset algorithm, which is widely used in machine learning and data mining studies. The feature selection process was determined for each biochemical feature in the weka software. Biochemical contents and features were determined from 502 hyperspectral data sets. The CfsSubsetEval algorithm was used to determine the features with the highest correlation of biochemical contents from the hyperspectral data set. The determined hyperspectral features are given in Table 3.

Table 2. Descriptive statistics of biochemical analysis results

Biochemical Analysis	Sample (N)	Minimum	Mean	Maximum	Standard Deviation	Coefficient of Variation	Range
Chl a (mg g^{-1})	240	26.05	26.63	29.53	0.65	2.43	3.48
Chl b (mg g^{-1})	240	20.05	46.63	50.27	5.50	11.79	30.22
Chl a+b (mg g^{-1})	240	49.32	73.23	76.91	4.89	6.67	27.59
TFLC (mg catechin g^{-1})	240	5.66	13.66	27.07	4.33	31.68	21.41
TPEC (mg GAE g^{-1})	240	4.08	8.61	17.10	3.25	37.73	13.02
TAC (mg TEAC g^{-1})	240	19.09	86.67	90.87	5.15	5.94	71.78

Table 3. Feature selection from hyperspectral dataset

Biochemical Analysis	Algorithm	Hyperspectral Feature (400-900 nm)	Selected Attributes
Chl a	**CfsSubsetEval (CFS Subset Evaluator)**	423, 529, 561, 662, 663, 687, 707, 710, 770, 892	10
Chl b		404, 423, 528, 598, 610, 703, 710, 715, 766, 867, 898	11
Chl a+b		404, 423, 528, 561, 616, 694, 703, 716, 766, 865, 898	11
TFLC		405, 497, 502, 513, 553, 562, 563, 569, 667, 701, 712, 716, 718, 724, 732, 733, 781, 783, 875	19
TPEC		415, 482, 553, 556, 557, 611, 674, 678, 679, 715, 717, 719, 724, 737, 811, 871, 889	17
TAC		419, 548, 549, 586, 620, 711, 724	7

Figure 4. Confusion matrix results obtained with the applied method

The feature selection process was determined from 502 features between 400-900 nm. The different number of attributes were determined for biochemical features. However, properties in the 550-700 nm range in the red and green regions of the electromagnetic spectrum were determined to be highly correlated with all biochemical features. Wavelengths were determined for organically grown oil roses in monitoring abiotic and biotic stress levels in farmland conditions with hyperspectral sensing (Table 3).

The determined features were used for the classification of biochemical features with the J48 algorithm. Confusion matrix results obtained in the applied method are shown in Figure 4. Less than equal to the mean content of chlorophyll a, chlorophyll b, chlorophyll a+b, total flavonoid content, total phenolic content, and total antioxidant capacity in *R. Damascena* leaves were given class I. The higher than mean was given class II (Figure 4). As can be seen from the Confusion matrix, the best result for biochemical content *R. Damascena* leaves is calculated Chl a+b with J48 algorithm.

As can be seen from the Confusion matrix, the best result for biochemical content *R. Damascena* leaves is calculated Chl a+b with the J48 algorithm. Performance tests for the best classification results for biochemical contents were calculated according to in the confusion matrix. The performance test was applied to the statistical parameters calculated in the confusion matrix and given in equations 1-3. The classification accuracies, sensitivity, specificity, and receiver operating characteristic (ROC) curves, which are the classification performances of chlorophyll a, chlorophyll b, chlorophyll a+b, total flavonoid content, total phenolic substance content, and total antioxidant capacity were determined. The classification performance of biochemical contents with the J48 algorithm is given in Table 4.

Table 4. J48 algorithm performance test

Biochemical Contents	Sensitivity %	Specificity %	ROC Area (AUC)	Accuracy %
Chl a	35.3	78.2	0.558	72.1
Chl b	44.0	81.1	0.643	73.3
Chl a+b	55.3	80.2	0.631	76.25
TFLC	71.4	65.3	0.638	69.6
TPEC	77.1	62.1	0.723	71.7
TAC	65.7	68.2	0.625	67.5

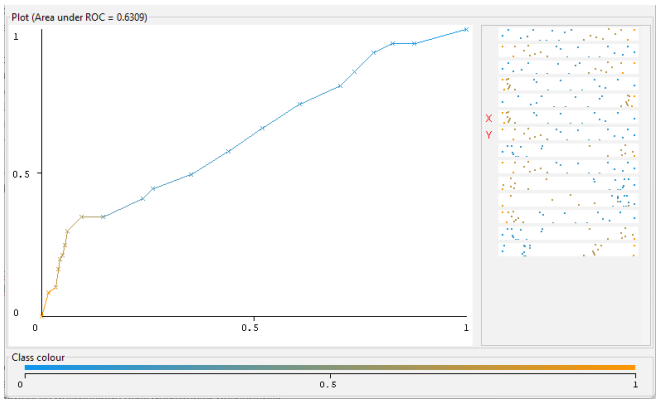

Figure 5. ROC curve for chlorophyll a+b classification

As seen in Table 4, the best result was calculated with the J48 algorithm with an accuracy of 76.26% for Chl a+b. The results calculated in Table 4 were obtained with 10-Fold Cross Validation from the hyperspectral data set. The purpose of ROC (Receiver Operating Characteristic) Curves is to examine the performance of a binary classifier, by creating a graph of the True Positives ve False Positives for every classification threshold. The ROC curve emerges as the ratio of sensitivity to precision in binary classification systems where the discrimination threshold differs. The ROC graph for chlorophyll a+b is given in figure 5, as the J48 algorithm classification accuracy is clearly indicated in the confusion matrix. In addition to analytical methods, a method for the use of hyperspectral data was applied in the determination of biochemical contents. Organic Farming *Rosa Damascena* Mill. Database was created by using hyperspectral data for the determination of biochemical contents with the J48 algorithm, one of the machine learning algorithms. Hyperspectral bands that can be used for monitoring physiological events under abiotic and biotic stress factors were determined for *Rosa Damascena* Mill. in Isparta ecological conditions.

It has been stated that feature selection in the determination of biochemical contents using hyperspectral data increases the overall accuracy of classification by 75% and 100% [28]. Zhang et al. [29] have also reported that the GA-SVM classifier performed the best, which produced overall accuracy, according to an investigation of the performance of different machine learning classifiers in the spectral library application, including K-nearest neighbour (KNN), Random Forest (RF), and a genetic algorithm coupled with a

support vector machine (GA-SVM). Dedeoğlu [30] has also reported that Gaussian mixture discriminant analysis (GMDA) findings showed that the model developed using hyperspectral reflectance data can discriminate different N nutritional statuses in plants with an accuracy of ≥70% and can be applied under field conditions. It supports that hyperspectral data and the J48 algorithm can be used in the classification of chlorophyll a+b content of the Damascena leaf. The results of the study are in line with the results of the literature [28],[29],[30].

5 Conclusion

In this study, a method application has been made for the processing of big data collected by land and remote sensing platforms as a result of the increasing use of technology in agriculture with Agriculture 4.0 and smart agriculture. In this study, for Rosa Damascena, which was grown organically in Isparta ecological conditions, due to similar climatic conditions and cultural practices, hyperspectral data and average biochemical contents within the vegetation period were classified. Rosa Damascena is the agricultural product with the highest added value among organic products that are in high demand for the cosmetics, food, and pharmaceutical industries. For this reason, it is suggested that hyperspectral data can be used successfully to determine the abiotic and biotic stress factors of products grown under organic farming.

In this study, the CfsSubsetEval algorithm was used for the feature selection of hyperspectral data. Eleven features were identified that were highly correlated with Chl a+b content. As a result of the classification made with the J48 algorithm, Chl a+b was determined with an accuracy of 76.25%. As a result of the study, the methodology that can be used in determining the physiological events in plants with hyperspectral data was determined for *Rosa Damascena* Mill..

Acknowledgment

This study is prepared from the data of the project numbered TOVAG 120 O 166 supported by TUBITAK, Turkey. Part of this work was presented orally at the V. International Conference on Data Science and Applications 2022(ICONDATA'22).

References

[1] Mahlein, A. K., Oerke, E. C., Steiner, U., & Dehne, H. W., "Recent advances in sensing plant diseases for precision crop protection". European Journal of Plant Pathology, 133(1), 197-209, 2012.

[2] Chlingaryan, A., Sukkarieh, S., & Whelan, B., "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review". Computers and electronics in agriculture, 151, 61-69, 2018.

[3] Demir, S., Başayiğit, L., "The effect of restricted irrigation applications on vegetation index based on UAV multispectral sensing". Yuzuncu Yıl University Journal of Agricultural Sciences, 31(3), 629-643, 2021.

[4] Teke, M., Deveci, H. S., Haliloğlu, O., Gürbüz, S. Z., Sakarya, U., "A short survey of hyperspectral remote sensing applications in agriculture". In 2013 6th international conference on recent advances in space technologies (RAST) (pp. 171-176). IEEE, 2013.

[5] Sahoo, R. N., Ray, S. S., & Manjunath, K. R., "Hyperspectral remote sensing of agriculture". Current science, 848-859, 2015.

[6] Ang, K. L. M., Seng, J. K. P., "Big data and machine learning with hyperspectral information in agriculture". IEEE Access, 9, 36699-36718, 2021.

[7] Singh, P.; Pandey, P.C.; Petropoulos, G.P.; Pavlides, A.; Srivastava, P.K.; Koutsias, N.; Deng, K.A.K.; Bao, Y., "Hyperspectral remote sensing in precision agriculture: Present status, challenges, and future trends". In Hyperspectral Remote Sensing; Elsevier: Amsterdam, The Netherlands, pp. 121–146, 2020.

[8] Khan, A., Vibhute, A. D., Mali, S., Patil, C. H., "A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications". Ecological Informatics, 101678, 2022.

[9] Cheng, T., Zhu, Y., Li, D., Yao, X., Zhou, K., "Hyperspectral Remote Sensing of Leaf Nitrogen Concentration in Cereal Crops". In Hyperspectral Indices and Image Classifications for Agriculture and Vegetation (pp. 163-182). CRC Press, 2018.

[10] Candiani, G., Tagliabue, G., Panigada, C., Verrelst, J., Picchi, V., Rivera Caicedo, J. P., Boschetti, M., "Evaluation of Hybrid Models to Estimate Chlorophyll and Nitrogen Content of Maize Crops in the Framework of the Future CHIME Mission". Remote Sensing, 14(8), 1792, 2022.

[11] Krishna, G., Sahoo, R. N., Singh, P., Bajpai, V., Patra, H., Kumar, S., ... Sahoo, P. M., "Comparison of various modelling approaches for water deficit stress monitoring in rice crop through hyperspectral remote sensing". Agricultural water management, 213, 231-244, 2019.

[12] Jan, R., Asaf, S., Numan, M., Kim, K. M., "Plant secondary metabolite biosynthesis and

transcriptional regulation in response to biotic and abiotic stress conditions". Agronomy, 11(5), 968, 2021.

[13] Yilmaz, R., Yildirim, A., Çelik, C., Karakurt, Y., "Determination of nut characteristics and biochemical components of some pecan nut cultivars". Yuzuncu Yıl University Journal of Agricultural Sciences, 31(4), 906-914, 2021.

[14] Yıldırım, F., Esen, M., Binici, S., Çelik, C., Yıldırım, A., Karakurt, Y., "Antioxidant Enzymes activities of walnut nursery trees to drought stress progression". International Journal of Agriculture Forestry and Life Sciences, 5(2), 217-225, 2021.

[15] Medini, F., Fellah, H., Ksouri, R., Abdelly, C., "Total phenolic, flavonoid and tannin contents and antioxidant and antimicrobial activities of organic extracts of shoots of the plant *Limonium delicatulum*". Journal of Taibah University for science, 8(3), 216-224, 2014.

[16] Mazid, M., Khan, T. A., Mohammad, F., "Role of secondary metabolites in defense mechanisms of plants". Biology and medicine, 3(2), 232-249,2011.

[17] Khanam, U. K. S., Oba, S., Yanase, E., Murakami, Y., "Phenolic acids, flavonoids and total antioxidant capacity of selected leafy vegetables". Journal of Functional Foods, 4(4), 979-987, 2012.

[18] Baydar, N. G., Baydar, H., "Phenolic compounds, antiradical activity and antioxidant capacity of oil-bearing rose (*Rosa damascena* Mill.) extracts". Industrial Crops and Products, 41, 375-380, 2013.

[19] Hassan, F., Al-Yasi, H., Ali, E., Alamer, K., Hessini, K., Attia, H., El-Shazly, S., "Mitigation of salt-stress effects by moringa leaf extract or salicylic acid through motivating antioxidant machinery in damask rose". Canadian Journal of Plant Science, 101(2), 157-165, 2020.

[20] Janse, P. V., Kayte, J. N., Agrawal, R. V., Deshmukh, R. R., "Standard spectral reflectance measurements for ASD FieldSpec Spectroradiometer". In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 729-733). IEEE, 2018.

[21] Zhang, Z., Huang, R., "Analysis of malondialdehyde, chlorophyll proline, soluble sugar, and glutathione content in Arabidopsis seedling". Bio-protocol, 3(14), e817-e817, 2013.

[22] Singleton, V. L., Rossi, J. A., "Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents". American journal of Enology and Viticulture, 16(3), 144-158, 1965.

[23] Zhishen, J., Mengcheng, T., Jianming, W., "The determination of flavonoid contents in mulberry and their scavenging effects on superoxide radicals". Food chemistry, 64(4), 555-559, 1999.

[24] Kumaran, A., & Karunakaran, R. J., "Activity-guided isolation and identification of free radical-scavenging components from an aqueous extract of Coleus aromaticus". Food Chemistry, 100(1), 356-361, 2007.

[25] Taghizadeh-Mehrjardi, R., Ayoubi, S., Namazi, Z., Malone, B. P., Zolfaghari, A. A., Sadrabadi, F. R., "Prediction of soil surface salinity in the arid region of central Iran using auxiliary variables and genetic programming". Arid Land Research and Management, 30(1), 49-64, 2016.

[26] Mandrekar, J. N., "Receiver operating characteristic curve in diagnostic test assessment. Journal of Thoracic Oncology", 5(9), 1315-1316, 2010.

[27] R Core Team, R., "R: A language and environment for statistical computing". 2018.

[28] Kandpal, K. C., Kumar, S., Venkat, G. S., Meena, R., Pal, P. K., & Kumar, A., "Onsite age discrimination of an endangered medicinal and aromatic plant species Valeriana jatamansi using field hyperspectral remote sensing and machine learning techniques". International Journal of Remote Sensing, 42(10), 3777-3796, 2021.

[29] Zhang, J., He, Y., Yuan, L., Liu, P., Zhou, X., & Huang, Y., "Machine learning-based spectral library for crop classification and status monitoring". Agronomy, 9(9), 496, 2019.

[30] Dedeoglu, M., "Estimation of critical nitrogen contents in peach orchards using visible-near infrared spectral mixture analysis". Journal of Near Infrared Spectroscopy, 28(5), 315-327, 2020.