CLASSIFICATION OF DIFFERENT FOREST TYPES WITH MACHINE LEARNING ALGORITHMS

Kadir Sabanci[1], M.Fahri Ünlerşen[2], Kemal Polat[3]
[1]Karamanoglu Mehmetbey University, Turkey
[2]Necmettin Erbakan University, Turkey
[3]Abant Izzet Baysal University, Turkey
kadirsabanci@kmu.edu.tr; unlersen@yandex.com; kpolat@ibu.edu.tr

Abstract
In this study, forest type mapping data set taken from UCI (University of California, Irvine) machine learning repository database has been classified using different machine learning algorithms including Multilayer Perceptron, k-NN, J48, Naïve Bayes, Bayes Net and KStar. In this dataset, there are 27 spectral values showing the type of three different forests (Sugi, Hinoki, mixed broadleaf). As the performance measure criteria, the classification accuracy has been used to evaluate the classifier algorithms and then to select the best method. The best classification rates have been obtained 90.43% with MLP, and 89.1013% with k-NN classifier (for k=5). As can be seen from the obtained results, the machine learning algorithms including MLP and k-NN classifier have obtained very promising results in the classification of forest type with 27 spectral features.
Key words: Forest types, Multilayer perceptron, k-NN classifier, Data Mining.

Introduction

Today, as the number of measuring devices increases, so does the number and types of data. As a result of these advancements, it is required that so much information is stored in databases, and that this stored information is needed to be analyzed by intelligent and automated processes which convert the data into useful information and knowledge (Dener, Dörterler, & Orman, 2009). Consequently, data mining has become an important research area.

Data mining is a computational process that reveals patterns in data sets by using such methods like artificial intelligence, machine learning, statistics etc (Chen, Han, & Yu, 1996). The methods used in data mining are investigated in two groups as predictive and descriptive. In predictive methods, a model is created by using a dataset whose results are known. For example in a bank, the properties of customers who pay their credits back can be revealed, and a model can be created by using previous data sets about funding of them. Afterwards this model can be used on new customers for determining the possibility of paying their credits back. In descriptive methods, a relationship can be searched between two data sets. For example, the shopping haAabits of two different cultures may be investigated for similarity (Özekes, 2003).

Data mining methods can be divided into three groups due to their function.

- Classification and Regression
- Clustering
- Association Rules

In this study data mining methods are used to classify the data set. In classification, training examples are used to learn a model that can classify the data samples into known classes. The classification process involves following these steps: creating a training data set, identifying class attributes and classes, identifying useful attributes for classification, relevance analysis, learning a model using training examples in the training set and using the model to classify the unknown data (Sharma & Jain, 2013). The causes of selecting of machine learning algorithms in data mining are that we can identify the tree types automatically based on the spectral features of trees and we can get very high identification success by means of machine learning algorithms.

There are many studies in the literature in which data mining classification algorithms are used. The main areas are medical, food and agriculture. Jamuna *et al*. (2010) used different classification algorithms and compared these methods in order to ascertain the productivity of cotton seed in the oncoming stages of development. Although Decision Tree Classifier and Multilayer Perceptron Methods produce results at the same level of accuracy, it was observed that the Decision Tree Classifier method produces results in a much shorter time. Sabanci and Aydin (2014) used image processing techniques to detect and spray weeds on rows in sugar beet fields. The images captured with the CCD camera on the spraying robot were processed using image processing algorithms in Matlab software. Weeds in the row were detected by using Multilayer Perceptron Algorithm with the data which are obtained from the images and spraying liquid was applied on them. Kiani *et al*. (2010) pointed out at the fact that the use of chemicals for weed control in wheat fields caused environmental pollution. Due to this pollution, they stated that alternative methods such as image processing could be used for detecting weeds. Accordingly, Multilayer Perceptron Algorithm was used in the study in order to classify and analyze the

properties of energy, entropy, contrast, homogeneity and inertia. Babalik *et al*. (2010) used artificial neural networks and image processing techniques to determine the vitreousness of hard wheat, they used artificial neural networks to classify the vitreous and non-vitreous kinds of Type-1252 wheat. In this study, the classification success rates of self-regulated mapping (SRM) and multilayer perceptron (MP) were examined. Sabanci *et al*. (2012) classified potatoes in terms of their size with the help of image processing techniques and artificial neural network. Before the classification process, potatoes with surface defects and deformities were detected using Otsu method and morphological processes, and they were excluded from the classification. Then potatoes were classified based on their sizes. For this process, the images of small, medium and big sized potatoes were captured and the system was trained using multilayer artificial neural networks. Using image processing techniques and artificial neural networks, the classification success rates of potatoes were analyzed. Karthikeyan *et al*. (2015) obtained the dataset values (from the UCI database) of hepatitis disease that occurs on the liver and applied J48, Naïve Bayes, Multilayer Perceptron, random forest classification algorithms using these datasets. As a result of the study, the highest percentage rate in the classification of hepatitis patients based on sick cells was obtained using the Naïve Bayes algorithm. Polat and Gunes (2009) offered a genuine hybrid classification system that is based on the C4.5 decision tree classifier and a one-against-all approach to classify the multi-class problems including image segmentation, dermatology and lymphography datasets obtained from the UCI MRL database. Sabanci *et al*. (2015) used the EEG eye state dataset obtained from the UCI machine learning repository database. 14 continuous EEG measurements constitute the basics of the dataset. The duration of the measurement was 117 seconds (each measurement had 14980 samples). They used Multilayer Perceptron Neural Network Models and k-Nearest Neighbor Algorithm to calculate the classification success rate. The classification success rates were measured for varying number of neurons in the hidden layer of the Multilayer Perceptron Neural Networks model. The highest classification success rate was achieved when the number of neurons in the hidden layer was 7. And the success rate was 56.45%. The classification success rates were measured using k-Nearest neighbor algorithm for varying neighborhood values. The highest classification success rate was achieved using kNN algorithm. In k-Nearest neighbor model, the success rate regarding 3 nearest neighbors was measured as 84.05%. Yu *et al*. (2015) classified the forest trees in Zijin Mountains National Forest Park in Nanjing, China. The data is of the year 2011. Three types of band combinations were compared based on the accuracy of the classification. Using the obtained optimal band combination, decision tree classifier, neural networks and support vector machine classification methods were compared. The best result was obtained using 8-bant combination for decision tree classification, and the success rate was 87.10%. It was determined that artificial neural networks produced the worst results with the success rate of 73.85%. Aguiar *et al*. (2010) identified forages in the state of Mato Grosso do Sul in Brazil and their varying decomposition levels. In order to obtain the plant cover index and fractional images, MODIS duration series were used. Using ripple technique at various decomposition levels, the input parameter required for WEKA J48 classifier was obtained. This way, forages were successfully selected from Cerrado. The segregation between different forages caused lower performance; the best results were obtained for forages that include common plants. Yang *et al*. (2014) classified trees in Boreal forests in Canada. Using LiDAR, RapidEye and the combination of these two data along with the support vector machine classification method, the success rates were compared. The data they used composed of six components. These are digital elevation model, slope, red-edge NDVI, red-edge, canopy height and near infrared bands of RapidEye data. The best result was obtained using the combination of LiDAR and RapidEye data.

In this paper, the forest type mapping including three tree types have been automatically classified based on machine learning algorithms including k-NN, Multilayer Perceptron, J48, Bayes Net, Naïve Bayes and K-Star classification methods using spectral features belonging to these tree types.

Materials and Methods

Dataset

In this study, Advanced Spaceborne Thermal Emission and Reflection Raidometer (ASTER) satellite images (15m resolution) in a forestland of approximately 13 x 12 km in Ibaraki Prefecture, Japan were used. In this area, there were mainly Cryptomeria japonica (Sugi) trees, Chamaecyparis obtuse (Hinoki) trees, mixed broadleaf, angiosperm natural trees and also a few non-forest structures (such as buildings, roads and cultivated areas) (Johnson, Tateishi, & Xie, 2012). The orthorectificated ASTER images were obtained at three different dates in order to determine the coniferous and broadleaf tree types. Each pixel identifies a distance of 15m. The images were obtained at green (0.52 – 0.60 μm), red (0.63 – 0.69 μm) and near infrared (NIR) (0.76 – 0.86 μm) bands (as a total of nine bands) (Johnson, Tateishi, & Xie, 2012). The data obtained from the UCI Forest

type mapping data set are composed of two parts as training and test. There is a total of 524 data. The 38% of the data is for training and 62% is for test. Each data consists of 27 attributes. Each data is classified as Sugi, Hinoki, mixed broadleaf and others (UCI, 2016). Using the data obtained for each channel by processing the orthorectificated ASTER images using the inverse distance weighting (IDW) method, a map for Sugi and Hinoki type trees was created. The nearest 15 neighbor pixels were used for IDW process. Using the training data, the average spectral values of Sugi and Hinoki types at each band were obtained. For Sugi, pred_minus_obs_S was obtained by subtracting the values obtained with IDW from average values. For Hinoki, pred_minus_obs_H was obtained by subtracting the values obtained with IDW from average values. Therefore, a 27 attribute set composed of 9 original values, 9 pred_minus_obs values and 9 pred_minus_obs_H values was obtained (Johnson, Tateishi, & Xie, 2012).

Software-WEKA

Developed by Waikato University in New Zealand, WEKA is an open-source data mining software with a functional graphical interface which incorporates machine learning algorithms (Witten, Frank, & Hall, 2011.). WEKA includes various data pre-processing, classification, regression, clustering, association rules, and visualization tools. The algorithms can be applied on the data cluster either directly or by calling via Java code (Patterson *et al*., 2008; Hall *et al*., 2009). They are also suitable for developing new machine learning algorithms.

Machine learning algorithms

K-Nearest Neighbor Algorithm: The k-NN is a supervised learning algorithm that solves classification problems. Classification is the examination of the attributes of an image and the designation of this image to a predefined class. The important point is the determination of the features of each category in advance (Wang, Neskovic, & Cooper, 2007). According to the kNN algorithm used in the classification, based on the attributes drawn from the classification stage, the distance of the new individual that is wanted to be classified to all previous individuals is considered and the nearest k class is used. As a result of this process, test data belongs to the k-nearest neighbor category that has more members in a certain class. The most important optimization problems in the kNN method are the identification of the number of neighbors and the method of distance calculation algorithm. In the study, the identification of the optimum k number is performed with experiments, and the Euclidean Distance Calculations method is used as a distance calculation method.

Euclidean calculation method (Zhou, Li, & Xia, 2009):

$$d(x_i, x_j) = \left(\sum_{s=1}^{p} (x_{is} + x_{js})^2 \right)^2$$

x_i and x_j are two different points, and we need distance calculation process in between.

Multilayer Perceptron: It is a feed forward type artificial neural network model which maps input sets onto appropriate output sets. A multilayer perceptron (MLP) is composed of multiple layers of nodes where each layer is connected to the next. Each node is a processing element or a neuron that has a nonlinear activation function except the input nodes. It uses a supervised learning technique named back propagation and it is used for training the network. The alteration of the standard linear perceptron, MLP is capable of distinguishing data which are not linearly separable (Hall *et al*., 2009).

J48: It is a widely used machine learning algorithm that is based on J.R. Quilan C4.5 algorithm. Data that will be examined will belong to the categorical type, so continuous data will not be examined at this step. However, the algorithm will leave room for adaptation in a way to include this capability (Hall *et al*., 2009; Arora, 2012).

Bayes Net: Bayes Net is a probabilistic graphical model and a statistical model representing a group of random variables in addition to their conditional dependencies through a directed acyclic graph. For instance, a Bayesian network can represent the probabilistic relations between diseases and symptoms. When the symptoms are given, the network can calculate the probabilities of the existence of various diseases (Hall *et al*., 2009).

Naive Bayes: In a learning problem, Naïve Bayes classifiers have a high degree of scalability and they entail a number of parameters that are linear with the number of variables (predictors/features). The maximum-likelihood training could be performed by examining a closed-form expression that takes linear time instead of by expensive iterative approximation unlike many other types of classifiers (Hall *et al*., 2009).

KStar: K* or K-Star is a classifier based on instance. A test instance class depends on the training instances that are similar to it, and it is determined by various similarity functions. The point it is different from other instance-based learners is that it uses a distance function that is based on entropy (Cleary & Trigg, 1995).

Results and Discussion

In the study, WEKA software was used in order to classify 3 different forest types (Sugi, Hinoki, mixed broadleaf). Using the kNN algorithm, the classification success rates of different forest types were obtained for different k-neighbor values. Also, root mean square error (RMSE) and mean absolute error (MAE) values were obtained. The classification success rates obtained with kNN algorithm, and MAE and RMSE values can be seen in Table 1. The diagram demonstrating the changes in MAE and RMSE error values based on the number of neighbors in the classification performed with the kNN algorithm is shown in Figure 1.

The data in the same dataset were processed using the multilayer perceptron model, and the classification success rates of different forest types were obtained. The classification success rates of different number of neurons in the hidden layer, and MAE and RMSE error values were obtained. In the multilayer perceptron model, the training was performed by taking the learning rate value as 0.3, momentum value as 0.2 and iteration number as 500. The classification success rates, and MAE and RMSE values obtained using the multilayer perceptron model can be seen in Table 2. The diagram demonstrating the changes in MAE and RMSE error values based on the number of neighbors in the hidden layer is demonstrated in Figure 2.

Then the same data was processed using J48, Naïve Bayes, Bayes Net, KStar machine learning algorithms and classification success rates and MAE and RMSE error values of different tree types in the forest were obtained. The success and error rates obtained using 6 different classification algorithms (Multilayer Perceptron, kNN, J48, Naïve Bayes, Bayes Net, KStar) can be seen in Table 3. The diagram demonstrating the error values obtained based on different machine learning algorithms can be seen in Figure 3.

Table 1

The Success Rate and Error Values Obtained by using kNN Classifier

Neighborliness Number (k)	Classification accuracy (%)	MAE	RMSE
1	83.3652	0.0856	0.2872
2	83.3652	0.0839	0.2412
3	88.1453	0.0834	0.2271
4	87.9541	0.0836	0.2198
5	**89.1013**	**0.0877**	**0.2169**
6	88.7189	0.0903	0.2181
7	88.7189	0.0908	0.2158
8	88.1453	0.0922	0.2139
9	88.9101	0.0941	0.2145
10	88.3365	0.0954	0.2141

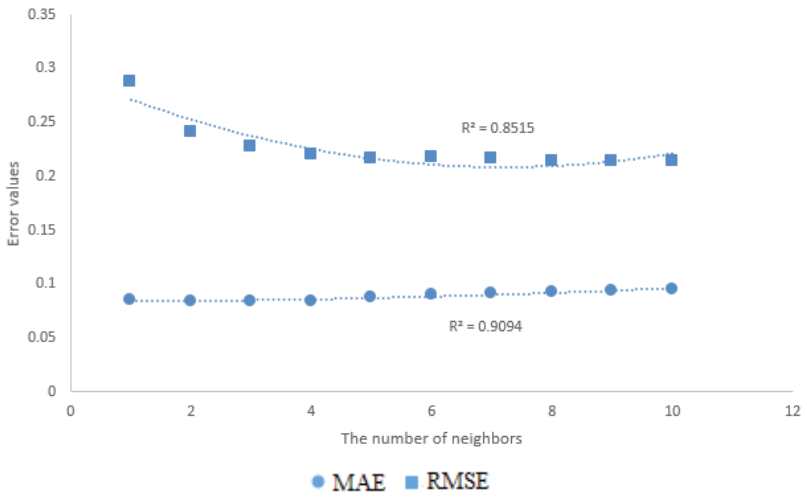

Figure 1. Variation of error rate based on the number of neighborhood.

Table 2

Success Rate Obtained By Using Multilayer Perceptron Classifier Error Values

The number of neurons in the hidden layer	Classification accuracy (%)	MAE	RMSE
5	88.9101	0.0611	0.2157
10	90.0574	0.0585	0.2068
15	88.7189	0.0621	0.2127
20	90.0574	0.0563	0.2007
25	89.4837	0.059	0.2091
30	89.6750	0.0569	0.2094
40	90.0574	0.0563	0.2054
50	89.6750	0.0566	0.2062
60	89.6750	0.0549	0.2056
70	89.8662	0.0543	0.2027
80	88.9101	0.0601	0.2146
90	**90.4398**	**0.0572**	**0.2078**
100	89.8662	0.0557	0.204

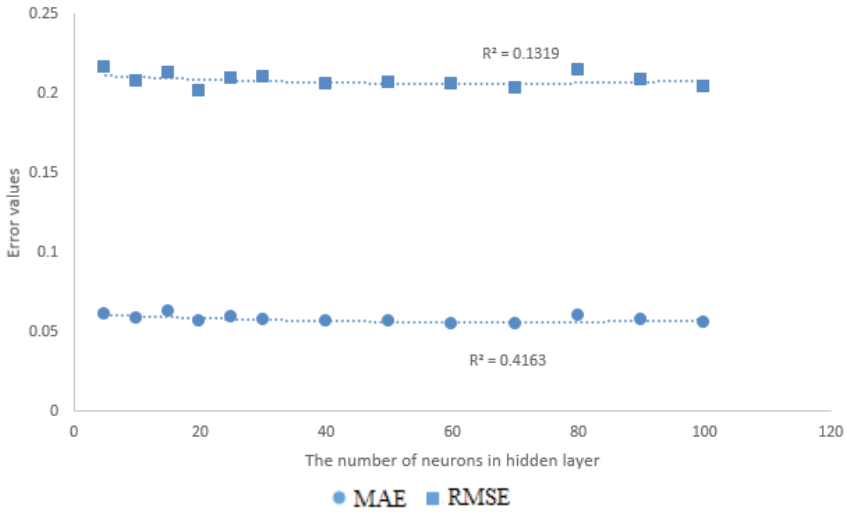

Figure 2. Variation of error rate based on the number of neurons in hidden layer.

Table 3

Success Rate Obtained By Using Various Machine Learning Algorithms

Machine learning algorithms	Classification accuracy (%)	MAE	RMSE
Multilayer Perceptron	**90.4398**	**0.0572**	**0.2078**
kNN	89.1013	0.0877	0.2169
J48	86.0421	0.0810	0.2543
Naive Bayes	85.6597	0.0708	0.2562
Bayes Net	85.4685	0.0729	0.2593
KStar	81.4532	0.0933	0.2933

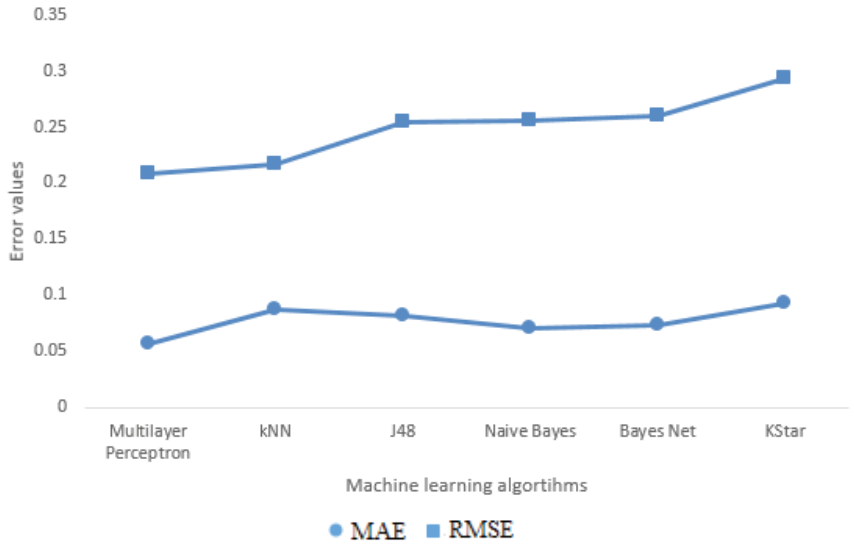

Figure 3. Variation of error rate based on the machine learning algorithms.

Conclusions

In this study, three different forest types in Japan (Sugi, Hinoki, mixed broadleaf) were classified using machine learning algorithms (Multilayer Perceptron, kNN, J48, Naïve Bayes, Bayes Net, KStar). The classification success and error values of machine learning algorithms were calculated. It was observed that the success rate was higher for the classifications performed using the Multilayer Perceptron Algorithm. The highest classification success rate was achieved when the number of neurons in the hidden layer was 80 and the success rate was 90.4398%. The MAE error value was 0.0572 and the RMSE error value was 0.2078 for the number of neurons in the hidden layer. For the classification success rates obtained using K-Nearest Neighbor Algorithm, the highest classification success rate was achieved for 5 neighborhood values, and it was 89.1013%. For this neighborhood value, the MAE error value was 0.0877 and the RMSE error value was 0.2169. The success rates obtained using J48, Naïve Bayes, Bayes Net and KStar classification algorithms were found as 86.0421%, 85.6597%, 85.4685% and 81.4532% respectively.

References

1. Aguiar, D.A., Adami, M., Fernando Silva, W., Rudorff, B.F.T., Mello, M.P., & Da Silva, J.D.S.V. (2010). MODIS time series to assess pasture land. In Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International (pp. 2123-2126). IEEE.
2. Arora, R. (2012). Comparative analysis of classification algorithms on different datasets using WEKA. *International Journal of Computer Applications*, 54(13), pp. 21-25.
3. Babalık, A., & Botsalı, F.M. (2010). Yapay sinir ağı ve görüntü işleme teknikleri kullanarak durum buğdayının camsılığının belirlenmesi (Determination of durum wheat vitreousness using artificial neural networks and image processing techniques). Selcuk Teknik Online Dergisi, 9(1): 163-174 (in Turkish).
4. Chao, Y., Mingyang, L., & Mifang Z. (2015). Classification of Dominant Tree Species in an Urban Forest Park Using the Remote Sensing Image of WorldView-2. 2015 8th International Congress on Image and Signal Processing (CISP 2015) 742-747. IEEE.
5. Chen, M.S., Han, J., & Yu, P.S. (1996). Data mining: an overview from a database perspective. Knowledge and data Engineering, IEEE Transactions on, 8(6), 866-883.
6. Cleary, J.G., & Trigg, L.E. (1995). K*: An instance-based learner using an entropic distance measure. In Proceedings of the 12th International Conference on Machine learning Vol. 5, pp. 108-114.
7. Dener, M., Dörterler, M., & Orman, A. (2009). Açık kaynak kodlu veri madenciliği programları: Weka'da örnek uygulama (A sample application of Weka which is an open source data mining software). Akademik Bilişim, 11 – 13 Şubat 2009 (pp. 787-796). Harran Üniversitesi, Şanlıurfa, Türkiye (in Turkish).
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009). The WEKA Data Mining Software: An Update. ACM SIGKDD Explorations Newsletter, 11(1), 10-18.

9. Jamuna, K.S., Karpagavalli, S., Revathi, P., Gokilavani, S., & Madhiya E. (2010). Classification of seed cotton yield based on the growth stages of cotton crop using machine learning techniques. International Conference on Advances in Computer Engineering, 20 – 21 June 2010 (pp. 312-315). Bangalore, Karnataka, India. IEEE. DOI: 10.1109/ACE.2010.71.
10. Johnson, B., Tateishi, R., & Xie, Z. (2012). Using geographically weighted variables for image classification. Remote Sensing Letters, 3(6), 491-499.
11. Karthikeyan, T., & Thangaraju, P. (2015). Best First and Greedy Search based CFS-Naive Bayes Classification Algorithms for Hepatitis Diagnosis.Biosciences and Biotechnology Research Asia, 12(1), 983-990.
12. Kiani, S., Azimifar, Z., & Kamgar, S. (2010). Wavelet-based crop detection and classification. In Electrical Engineering (ICEE), 2010 18th Iranian Conference on. 11 – 13 May 2010 (pp. 587-591). Isfahan. IEEE. DOI: 10.1109/IRANIANCEE.2010.5507003.
13. Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları (Data mining methods and application areas). Istanbul Ticaret Universitesi Dergisi, vol 3, 65-82. (in Turkish).
14. Patterson, D., Liu, F., Turner, D., Concepcion, A., & Lynch, R. (2008). Performance Comparison of the Data Reduction System. Proceedings of the SPIE Symposium on Defense and Security, Mart, Orlando, FL, pp. 27-34.
15. Polat, K., & Güneş, S. (2009). A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. Expert Systems with Applications, 36(2), 1587-1592.
16. Sabanci, K., Aydin, C., & Ünlerşen, M.F. (2012). Görüntü İşleme ve Yapay Sinir Ağları Yardımıyla Patates Sınıflandırma Parametrelerinin Belirlenmesi (Determination of the classification parameters of potatoes by the help of image processing and artificial neural network). Igdir Universitesi Fen Bilimleri Enstitusu Dergisi, 2(10), 59-66. (in Turkish).
17. Sabancı, K., & Aydın, C. (2014). Görüntü İşleme Tabanlı Hassas İlaçlama Robotu (An image processing based precision spraying robot). Tarim Bilimleri Dergisi, 20(4), pp. 406-414. DOI: 10.15832/tbd.33629 (in Turkish).
18. Sabancı, K., & Koklu, M. (2015). The Classification of Eye State by Using kNN and MLP Classification Models According to the EEG Signals. *International Journal of Intelligent Systems and Applications in Engineering*, 3(4), 127-130.
19. Sharma, T.C., & Jain, M. (2013). WEKA approach for comparative study of classification algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(4), 1925-1931.
20. Johnson, B. (2015). UCI Machine Learning Repository: Forest type mapping Data Set, Retrieved March 1, 2016, from https://archive.ics.uci.edu/ml/datasets/Forest+type+mapping.
21. Wang, J., Neskovic, P., & Cooper, L.N. (2007). Improving nearest neighbor rule with a simple adaptive distance measure, Pattern Recognition Letters, 28(2):207-213.
22. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009). The WEKA Data Mining Software: An Update, SIGKDD Explorations, Volume 11, Issue 1, pp. 10-18.
23. Witten, I.H., Frank, E., & Hall, M.A. (2011). Data mining: practical machine learning tools and techniques. Elsevier, London.
24. Yang, X., Rochdi, N., Zhang, J., Banting, J., Rolfson, D., King, C., & Purdy, B. (2014). Mapping tree species in a boreal forest area using RapidEye and LiDAR data. In Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International (pp. 69-71). IEEE.
25. Zhou, Y., Li, Y., & Xia, S. (2009). An improved KNN text classification algorithm based on clustering, *Journal of computers*, 4(3):230-237.