Intelligent Methods in Engineering Sciences 5(2): 46-54, 2026

IMIENS

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

https://www.imiens.org

International
Open Access

August, 2026

e-ISSN 2979-9236

Research Article

https://doi.org/10.58190/imiens.2026.174

Performance Analysis of Machine Learning Methods for Classifying Types of Dry Beans

Hasan Selim Sindir [a,*], ***Yavuz Selim Taspinar*** [b]

[a] *AYD Automotive Industry, Konya, Türkiye*
[b] *Technology Faculty, Mechatronics Engineering, Selcuk University, Konya, Türkiye*

ARTICLE INFO

Article history:
Received 27 December 2025
Accepted 28 August 2026

Keywords:
Artificial neural networks,
Classification,
Dry beans,
Machine learning,
Optimization

ABSTRACT

Dry beans are a staple food of strategic importance worldwide and in our country due to their high protein content and nutritional value. In agricultural production, preserving seed quality and ensuring product standardization play a critical role in food safety and commercial sustainability. In this context, the accurate and rapid differentiation of morphologically similar bean varieties is essential for maintaining product purity. The main objective of this study is to enable the automatic and highly accurate classification of seven different dry bean varieties registered in Turkey using modern artificial intelligence technologies. Unlike traditional methods, the goal is to develop a decision support system that minimizes human-induced classification errors arising from physical similarities. The study utilized the Dry Bean Dataset obtained from open-source platforms. The dataset includes 16 different numerical morphological and shape features such as Area, Perimeter, Principal Axis Length, and Eccentricity, along with 7 different target classes: Seker, Barbunya, Bombay, Bush, Dermason, Horoz, and Sira. Decision Tree, Random Forest, Logistic Regression, and Artificial Neural Networks algorithms were used to solve the classification problem and perform performance analysis. As a result of the experimental analyses conducted, the highest classification success was achieved with the Artificial Neural Networks model, with a general accuracy rate of 93.4% and an AUC value of 0.995. In contrast, the lowest performance was observed in the Decision Tree algorithm, with an accuracy rate of 90.6%. This developed system has the potential to automate quality control processes in the food industry, thereby reducing manual sorting costs and increasing processing speed. Furthermore, it can provide industrial benefits as a reliable digital tool in seed certification processes and in verifying the purity of commercial products.

1. INTRODUCTION

Dry beans are the general name given to the mature and dry seeds of the Phaseolus vulgaris species, belonging to the legume (Fabaceae) family. Consumed worldwide as a staple food for thousands of years, dry beans hold an important place in the cuisine of many cultures. In Turkish cuisine, in particular, the dish “dry beans” stands out as a national delicacy. In terms of nutritional value, dry beans are considered an excellent source of plant-based protein. They also contain high amounts of soluble and insoluble fiber, which is critical for digestive health. This fiber helps balance blood sugar and provides a feeling of fullness for a long time. They are also rich in important minerals and vitamins such as folate, iron, potassium, and magnesium. This research focuses on seven different types of dry beans grown and officially registered in Turkey. The main objective of the study is to use modern technologies to distinguish these seven types. Machine learning algorithms and image processing techniques were preferred for this classification process. The seven registered varieties examined in the article include types such as Barbunya, Dermason, Seker, Sira, Horoz, Bombay, and Cali. These varieties differ from each other in terms of physical characteristics such as the shape, size, texture, and color of their seeds. The article has developed a highly accurate classification model by digitally measuring these physical differences. This type of classification is of great importance for ensuring quality control in agriculture, standardization in the food industry, and maintaining product purity in trade. There are many studies in the literature on this subject. Among these studies, those relevant to the current research are listed in order.

In a study, Słowiński used the “Dry Bean Dataset” published in the UCI machine learning repository in 2021. Using Support Vector Machines (SVM), Decision Tree, Random Forest, and Artificial Neural methods, they

* **Corresponding Author:** hsnselim.sindir@gmail.com

achieved an accuracy rate ranging from 88.35% to 93.61%. The literature indicates that a higher classification accuracy was achieved after reducing the features used in the dataset [1].

In a study, Taspinar et al. created a dataset consisting of 33,064 images in 2022 and trained Convolutional Neural Networks (CNN). By comparing transfer learning (InceptionV3, VGG16, VGG19) with hybrid (SVM/LR) and end-to-end methods, they determined that the highest success rate of 84.48% was achieved with the InceptionV3 model. They proposed to the literature that dry bean types can be classified quickly and accurately using image processing [2].

In a study, Mendigoria et al. examined the expanded morphological characteristics and variety classification of dry beans and applied the CIELab color channel thresholding method. KNN7 achieved the highest classification success with 93.69% accuracy, 93.64% precision, 93.66% specificity, and 93.69% F1 score. They presented their machine learning models as innovative in seed variety classification and dry bean seed phenotyping [3].

In a study, Krishnan et al. used the "Dry Bean Dataset" published in UCI Machine Learning (ML) in 2023. They applied Box-Cox transformation (BCT) to improve the dataset. To determine the best method, they used 22 different methods and performed 10-fold cross-validation. The CatBoost ML algorithm achieved the highest accuracy rate of 93.8% [4].

In a study, Khan et al. used a dataset containing 16 features, 12 dimensions, and 4 different types of grains found in the UCI ML repository to compare Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Machine (SVM), and Multilayer Perception (MLP) methods. Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Machine (SVM), and Multilayer Perception (MLP) methods and compared them with each other. The XGB method achieved an accuracy (ACC) of 93.0% and 95.4% in imbalanced and balanced classes, respectively. They contributed to the literature by investigating a reliable classifier with the lowest noise effects and creating an effective algorithm for dry bean classification [5].

In a study, Dogan et al. created a dataset consisting of 14 classes for dry bean classification and investigated the effects of extreme learning machine (ELM) and GoogLeNet transfer learning on this data. Their proposed SSA-ELM model successfully classified 14 different dry bean types with a success rate of 91.43%. They explained that the use of image data in dry bean classification, the extraction of deep features, and classification with optimized ELM achieved a higher success rate compared to other ML methods [6].

In a study, Macuácua et al. (2023) developed a computer vision system for the automatic classification of different seed varieties based on machine learning models combined with data mining techniques using the Dry Beans Classifications dataset. The KNN classifier demonstrated better performance with approximately 95% accuracy and 96% precision. They aimed to demonstrate that the combined use of data mining and machine learning classification methods can effectively and efficiently improve classification accuracy [7].

In a study, Mehta et al. (2023) analyzed and compared different Support Vector Machine (SVM) classification algorithms—linear, polynomial, and radial basis function (RBF)—with other popular classification algorithms. Using the RBF SVM kernel algorithm, they achieved the highest Accuracy of 93.34%, Precision of 92.61%, Recall of 92.35%, and F1 Score of 91.40%. They emphasized the importance of considering different SVM algorithms for complex and non-linear structured datasets in the literature [8].

In a study, Nayak et al. used Extreme Gradient Boosting ensembles employing the Synthetic Minority Over-Sampling Methodology (SMOTE) along with Decision Tree (DT), Support Vector Machine (SVM), and Support Vector Regression (SVR) models to analyze the effectiveness of the SMOTE method in improving the accuracy of the - April 10, 2023, developed and compared machine learning-based classification algorithms using Extreme Gradient Boosting ensembles with the Synthetic Minority Over-Sampling Methodology (SMOTE), Decision Tree (DT), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Adaptive Boosting classifier, and Bagging Classifier. The Extreme Gradient Boosting ensembles using SMOTE achieved the highest accuracy of 97.32% [9].

In a study, Shobana et al. suggested that the features used in bean classification do not contribute equally to accurate classification and applied estimators for classification using the Pipelined Linear Discriminant Analysis methodology. They suggested that Support Vector Machine, Random Forest, and Multilayer Perceptron had a higher accuracy rate of 93%. They also explained that this methodology would effectively solve the overfitting problem and improve the overall performance of the models [10].

In a study, Rimi et al. from the Bangladesh Agricultural Research Institute (BARI) collected data from local markets and the internet, performing image preprocessing, resizing, and enhancement. To categorize the data, they used machine learning models such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees, and deep learning models such as VGG-16, Inception-v3, and ResNet-50 convolutional neural networks. Inception-v3 delivered the best results in terms of performance metrics compared to other models, achieving an accuracy rate of 98.6% [11].

In a study, Vaidya et al. used the dry bean data found in

the UC Irvine machine learning repository. Using Principal Component Analysis (PCA), they isolated the most informative features of the data set and performed measurements using the Artificial Neuron Network (ANN) confusion matrix, accuracy, precision, Recall, and F1-Score. It was determined that the PCA-based ANN model had the highest classification accuracy (94%) and the highest F1-score (95%) for dry bean seeds [12].

In a study, Koklu and Ozkan subjected bean images obtained using a Computer Vision System (CVS) to segmentation and feature extraction stages in 2020. Multilayer perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), and Decision Tree (DT) classification models were created using 10-fold cross-validation, and their performance metrics were compared. The SVM classification model achieved the highest accuracy results. The literature suggests that using machine learning methods after segmenting images and extracting features using CVS will increase the accuracy rate [13].

In a study, Słowinski analyzed a dataset containing over 13,000 examples of geometric properties of dry beans in 2024 using Machine Learning (ML) and Deep Learning (DL) techniques. He visualized the dataset and analyzed it using the Shallow Learning technique and a Simple Artificial Neural Network. In the case of the Multilayer Perceptron, ReLu, ELU, and Sigmoid activation functions were tested. Random Forest achieved an average accuracy of up to 92.61%, claiming to be the best model for the dry bean classification task [14].

In a study, Coronel-Reyes et al. used INIAP-420, INIAP-422, and INIAP-425 data to create classification models using Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and k-Nearest Neighbors (k-NN) methods with 10-fold cross-validation. The SVM classification model achieved the highest accuracy, correctly classifying the INIAP-420, INIAP-422, and INIAP-425 bean types at rates of 97.72%, 98.33%, and 97.46%, respectively [15].

According to research in the literature, the contributions of this study can be listed as follows:

Comprehensive Algorithm Analysis: This study comparatively examined four fundamental machine learning algorithms (DT, RF, LR, NN) from different disciplines on the same dataset. Thus, the performance differences between traditional methods and modern artificial neural networks were revealed through experimental data.

Success of Artificial Neural Networks: Experimental results have proven that the Artificial Neural Networks model is statistically superior to other algorithms with an accuracy rate of 93.4%. It has been determined that this method is the most suitable solution, especially for complex and non-linear data structures.

Error Analysis for Difficult Classes: Detailed confusion matrix analyses were performed for the “Sira” and “Dermason” species, which have very similar physical characteristics and are often confused. As a result of these analyses, the behavior of the model that minimizes the classification difficulty created by morphological similarity was explained.

High Discrimination Power: The developed models achieved high AUC values reaching 0.99, particularly for Logistic Regression and ANN. This finding demonstrates that the proposed system performs the distinction between classes with near-perfect accuracy.

Industrial Applicability: The study proposes a high-accuracy decision support system that can be used to automate quality control processes in the food industry. This proposed structure has the potential to standardize processes while reducing the costs associated with manual classification.

The remainder of the article is structured as follows. The second section presents the dataset used in the study, the methods employed, and the performance metrics. The third section discusses the experimental results obtained in the study. The fourth section presents the study's conclusions, contributions, limitations, and potential applications.

2. MATERIAL AND METHODS

2.1. Dry Bean Dataset

The Dry Bean Dataset, obtained from the Kaggle platform and originally introduced through the UCI Machine Learning Repository, was selected as the benchmark dataset because of its high diversity, balanced multiclass structure, and widespread acceptance in agricultural machine learning research [16]. The dataset consists of 13,611 instances representing seven dry bean cultivars (Barbunya, Bombay, Cali, Dermason, Horoz, Seker, and Sira), each described by 16 numerical morphological descriptors extracted through computer vision analysis. These descriptors quantify the geometric characteristics of individual beans, including size, shape, compactness, convexity, elongation, and boundary properties, thereby providing a comprehensive numerical representation of seed morphology. Unlike image-based datasets, the feature extraction stage has already been completed, allowing the study to focus entirely on evaluating the discriminative capability of machine learning algorithms. One of the major challenges of this dataset is the substantial morphological similarity between certain cultivars, particularly Dermason and Sira, whose overlapping feature distributions make accurate classification difficult. This characteristic makes the dataset particularly suitable for assessing the robustness, generalization ability, and decision boundary learning capability of different classification algorithms. Owing to its realistic complexity and extensive use in previous

studies, the Dry Bean Dataset enables reliable benchmarking and facilitates direct comparison of the proposed models with existing state-of-the-art approaches. Table 1 shows sample data according to sample classes of the dataset.

Table 1. Sample data from the dataset

A	P	L	I	K	Ec	C	Ed	Ex	S	R	CO	SF1	SF2	SF3	SF4	Class
28395	610.291	208.178	173.888	1.197	0.549	28715	190.141	0.763	0.988	0.958	0.913	0.007	0.003	0.834	0.998	Seker
28734	638.018	200.524	182.734	1.097	0.411	29172	191.272	0.783	0.984	0.887	0.953	0.006	0.003	0.909	0.998	Seker
41487	815.9	299.046	177.081	1.688	0.805	42483	229.832	0.068	0.976	0.783	0.768	0.007	0.001	0.590	0.997	Barbunya
64510	989.333	348.661	236.351	1.475	0.735	65406	286.594	0.785	0.986	0.828	0.821	0.005	0.001	0.675	0.996	Barbunya
114004	1279.35	451.361	323.747	1.394	0.696	115298	380.991	0.748	0.988	0.875	0.844	0.003	0.001	0.712	0.993	Bombay
188247	1691.96	664.188	368.013	1.804	0.832	194252	489.574	0.727	0.969	0.826	0.737	0.003	0.000	0.543	0.980	Bombay
45504	793.417	295.469	196.311	1.505	0.747	45972	240.702	0.737	0.989	0.908	0.814	0.006	0.001	0.663	0.998	Cali
64930	990.342	377.208	220.674	1.709	0.811	65973	287.526	0.745	0.984	0.831	0.762	0.005	0.001	0.581	0.993	Cali
33006	710.496	283.020	149.623	1.891	0.848	33354	204.998	0.635	0.989	0.821	0.724	0.008	0.001	0.524	0.992	Horoz
45968	838.333	339.333	174.233	1.947	0.858	46698	241.926	0.705	0.984	0.821	0.712	0.007	0.001	0.508	0.989	Horoz
31519	676.641	255.073	157.802	1.616	0.785	32065	200.327	0.758	0.982	0.865	0.785	0.008	0.001	0.616	0.997	Sira
44364	791.197	304.876	185.827	1.640	0.792	44811	237.667	0.679	0.990	0.890	0.779	0.006	0.001	0.607	0.998	Sira
22777	563.861	204.878	141.801	1.444	0.721	23071	170.295	0.772	0.987	0.900	0.831	0.008	0.002	0.690	0.998	Dermason
42159	772.237	295.142	182.204	1.619	0.786	42600	213.686	0.788	0.989	0.888	0.784	0.007	0.001	0.616	0.998	Dermason

2.2. Decision Tree

Decision Trees are a supervised learning algorithm frequently preferred in data mining and machine learning applications, applicable to both classification and regression problems. Essentially, it is a hierarchical model structure that uses the features in the data set to transform a complex decision-making process into simple and interpretable 'If-Then' rule sets. Its key advantages include not requiring statistical assumptions (such as normal distribution of data) and producing models that are easily understandable by humans [17].

In this study, various optimization strategies were applied to the Decision Tree model to maximize classification accuracy and minimize the risk of overfitting. Due to their algorithmic structure, decision trees tend to model noise in the training data and produce high variance. To prevent the performance loss this could cause in the test data, the model's complexity was limited to increase its generalization ability.

Optimization techniques were used to prevent the overfitting problem frequently encountered in Decision Trees and to increase the model's generalization capacity. Thus, the model was cleaned of noise in the training set, and the consistency of predictions on external data was strengthened [18].

2.3. Random Forest

Random Forest is an ensemble learning-based algorithm that combines the predictions generated by multiple decision trees to produce a more stable, accurate, and generalizable result [19]. Developed by Breiman (2001), this method balances the tendency of a single decision tree to overfit the data by creating multiple and diverse trees. Its fundamental logic is based on the principle of 'wisdom of the crowd'; that is, many weak learners (decision trees) come together to form a strong learner.

The Random Forest model configuration is designed to optimize both computational cost and prediction accuracy. Within the scope of this study, the Random Forest algorithm, a robust ensemble learning method in terms of classification performance and model stability, was utilized. This method minimizes the high variance problem observed in individual trees by combining the predictions of numerous decision trees created using the Bagging technique [20].

2.4. Logistic Regression

Logistic Regression is a statistical method used to model the relationship between input variables and outcomes when the dependent variable has a categorical (classification) structure [21]. Despite the term 'regression' in its name, this method is fundamentally a classification algorithm. Unlike linear regression, it produces outputs not as values in an infinite range, but as probability values between 0 and 1. Thanks to this feature, it is accepted as a standard solution tool for problems where the outcome is divided into binary classes such as 'Yes/No' or 'Sick/Healthy', ranging from medical diagnoses to financial risk analysis. Logistic regression uses the Sigmoid function to classify data into categories. The optimization process aims to find the coefficients that minimize the error (Log-Loss) between the probabilities predicted by the model and the actual class labels. However, in its raw form, the model may overfit if it uses the features as they are. Therefore, the "C" parameter (the inverse of regularization strength) and Penalty (L1/L2) types are optimized to improve the model's generalization ability. Although the logistic regression model was preferred due to its simplicity and interpretability, it carries the risk of overfitting in multidimensional data sets. In this study, regularization techniques were applied to control the model's coefficients and minimize generalization error. During the model optimization process, the L1 (Lasso) and L2 (Ridge) penalty methods, which penalize the magnitude of the coefficients, were compared, and fine-tuning was performed on the inverse regularization

parameter (C parameter) that controls the model complexity. This ensured that the model was not affected by noise in the training data and aimed to draw the decision boundary between classes in the most optimal way [22, 23].

2.5. Artificial Neural Network

The Artificial Neural Network model has a layered architecture that receives, processes, and produces information. This architecture essentially consists of three main components: the Input Layer, where data enters the system; one or more Hidden Layers, where features are processed and learning occurs; and the Output Layer, where the system produces its final decision (class prediction). Each neuron in the network multiplies the inputs by specific weights and passes them through an Activation Function to the next layer. The learning process occurs by calculating the error between the predicted value and the actual value and using this error to update the weights in the network via the Backpropagation algorithm [24]. Structural optimization techniques were applied to improve the performance of the established Artificial Neural Network architecture and eliminate the overfitting problem. The number of hidden layers and the number of neurons in each layer were determined by balancing the complexity of the model with the processing cost. To prevent overfitting to the training data, Dropout layers were added to the network, and randomly selected neurons were deactivated during training to enhance the network's generalization ability [25]. In addition, the Early Stopping technique was used to record the optimal model weights by stopping training at the point where the error in the validation set began to increase [26].

2.6. Confusion Matrix

The Confusion Matrix is a fundamental table used to visualize the results of a classification problem and analyze the model's performance in detail. The rows of the matrix represent the actual class labels, while the columns represent the class labels predicted by the model. This structure clearly shows which classes the model predicted correctly and which classes it confused with each other.

A classification results in four basic cases: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

- True Positive (TP): The model correctly predicts the positive class.
- True Negative (TN): The model correctly predicting the negative class.
- False Positive (FP): The model incorrectly predicting a negative example as positive (Type 1 Error).
- False Negative (FN): The model incorrectly predicting a positive example as negative (Type 2 Error).

In this study, complexity matrix analyses were used to determine the extent to which bean varieties with similar morphological characteristics (e.g., Sira and Dermason) could be distinguished by the model [27].

2.7. Performance Metrics

Various statistical metrics derived from the confusion matrix have been used to measure and compare the success of classification models. The accuracy metric alone may not be sufficient to measure the success of a model, especially in imbalanced datasets; therefore, Precision, Recall, F1-Score, and ROC-AUC values were also considered [28]. The metrics used in the study are defined below:

• Accuracy (CA): This refers to the percentage of correct predictions (TP+TN) within the total data set.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

• Precision: Shows how many of the values predicted to be positive are actually positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

• Sensitivity (Recall): Measures how many true positives are correctly predicted by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{Tp} + \text{FN}}$$

• F1-Score: The harmonic mean of Precision and Sensitivity values. It is a critical metric for evaluating model performance on imbalanced datasets.

$$\text{F1} = 2\ \text{X}\ \frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}}$$

• ROC-AUC: The area under the Receiver Operating Characteristic (ROC) curve (AUC) is the most robust metric indicating the model's ability to distinguish between classes (between 0 and 1).

2.8. K-Fold Cross Validation

To measure the model's generalization ability and minimize the risk of overfitting to the training data, the K-Fold Cross Validation method was applied. In this method, the data set is randomly divided into 10 equal parts (folds). In each iteration, one of the parts is set aside as test data, while the remaining 10-1 parts are used as training data. After this process is repeated 10 times, the average of the success rates obtained is taken to determine the final performance of the model [29].

In this study, K was set to 10, and 10-fold cross-validation was applied. This balanced the bias and variance issues in the dataset, thereby increasing the model's reliability on real-world data. The average of the 10 results obtained was taken.

3. RESULTS AND DISCUSSION

In this study, the classification performance of Decision

Tree, Random Forest, Logistic Regression, and Neural Network algorithms was analyzed using the Dry Bean dataset. The performance of the models was evaluated based on the following metrics: Classification Accuracy (CA), F1-Score, Precision, Recall, and Area Under the Curve (AUC).

3.1. Decision Tree

The classification tests performed using the Decision Tree algorithm employed in this study yielded a model accuracy (CA) rate of 0.906 (90.6%), an F1-score of 0.906, and an Area Under the Curve (AUC) value of 0.945. Although the AUC score evaluated in the ROC analysis indicates that the model has a high ability to distinguish between classes, the Decision Tree algorithm showed the lowest performance among the other methods (Random Forest, Logistic Regression, and Artificial Neural Network) compared in this study. When examining the model's ROC curve, it was observed that the balance between sensitivity and specificity was within acceptable limits, but it offered weaker generalization ability compared to other models, especially in regions with high class transitions. Figure 1 shows the ROC of the DT model.

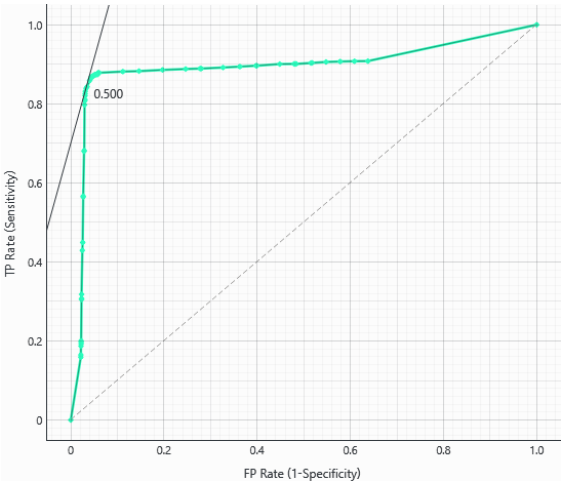

Figure 1. Decision Tree (Sira) ROC analysis

These findings reveal that the single Decision Tree structure has limitations in distinguishing classes (such as Order-Dermason) that exhibit very close distributions in the data space and have complex, non-linear relationships. The rule-based decision boundaries generated by the algorithm proved insufficient for such transitional classes, which in turn caused the overall accuracy rate to lag behind other methods. Figure 2 shows the confusion matrix of the DT model.

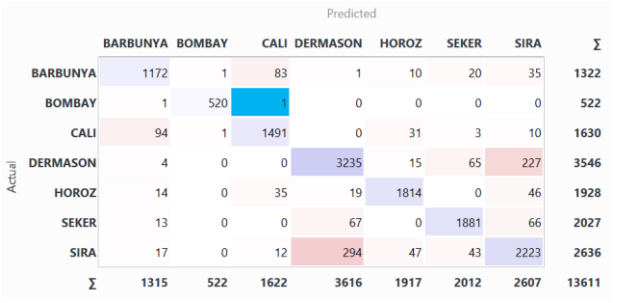

Figure 2. Decision Tree confusion matrix

3.2. Random Forest

As a result of classification experiments conducted using the Random Forest algorithm based on ensemble learning, the model's overall accuracy (CA) rate was determined to be 92.1% (0.921) and its F1-score was 0.921. In particular, the AUC (Area Under the Curve) value of 0.989 obtained in the ROC analysis indicates that the model has a very high capacity for inter-class discrimination and approaches the perfect classification threshold (1.0). These results, which show an increase in performance compared to the Single Decision Tree method, prove that the algorithm reduces variance by using a multi-tree structure and provides a more robust prediction process.

However, the transition between the “Sira” and “Dermason” types, which constitute the most challenging part of the dataset and exhibit high physical similarity, continued to be the main source of error in the Random Forest model. According to the analysis data, 281 “Sira” beans were mistakenly labeled as “Dermason,” while 215 ‘Dermason’ beans were mistakenly labeled as “Sira”. Nevertheless, the Random Forest algorithm managed to balance the disadvantages of individual tree structures with its high AUC value and overall accuracy rate, demonstrating that it is a reliable classifier for complex data sets. Figure 3 shows the ROC of the RF model. Figure 4 shows the confusion matrix of the RF model.

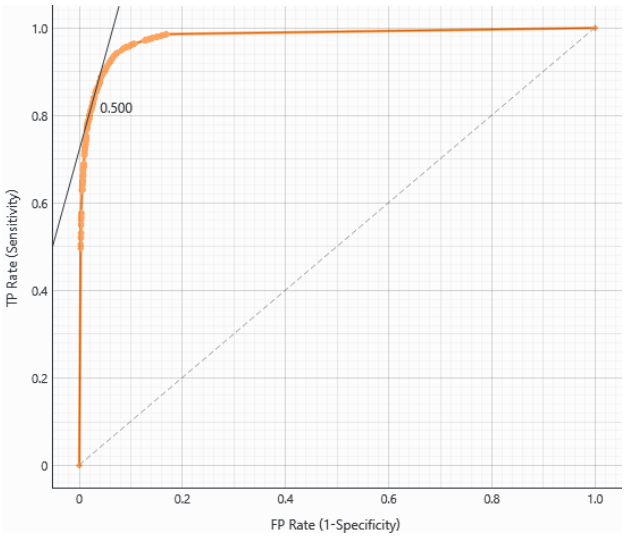

Figure 3. Random Forest (Sira) ROC analysis

Predicted

Actual	BARBUNYA	BOMBAY	CALI	DERMASON	HOROZ	SEKER	SIRA	Σ
BARBUNYA	1195	1	77	0	11	14	24	1322
BOMBAY	1	521	0	0	0	0	0	522
CALI	67	0	1520	0	27	2	14	1630
DERMASON	0	0	0	3288	8	51	199	3546
HOROZ	7	0	27	16	1831	0	47	1928
SEKER	6	0	1	43	0	1917	60	2027
SIRA	11	0	6	273	45	33	2268	2636
Σ	1287	522	1631	3620	1922	2017	2612	13611

Figure 4. Random Forest confusion matrix

3.3. Logistic Regression

As a result of the classification analyses performed using the Logistic Regression model, the model's overall accuracy (CA) rate was measured as 91.3% (0.913) and its F1-score as 0.914. The most notable performance of this method is the AUC (Area Under the Curve) value of 0.992 obtained within the scope of the ROC analysis. This score, which is extremely close to the perfect classification threshold (1.0), shows that the model is quite successful in predicting the probability values of the classes and has a high discriminatory power. Although Logistic Regression lags behind Artificial Neural Networks in terms of accuracy rate, its high AUC value proves that its potential to distinguish between classes in the dataset is statistically reliable. Figure 5 shows the ROC of the DT model.

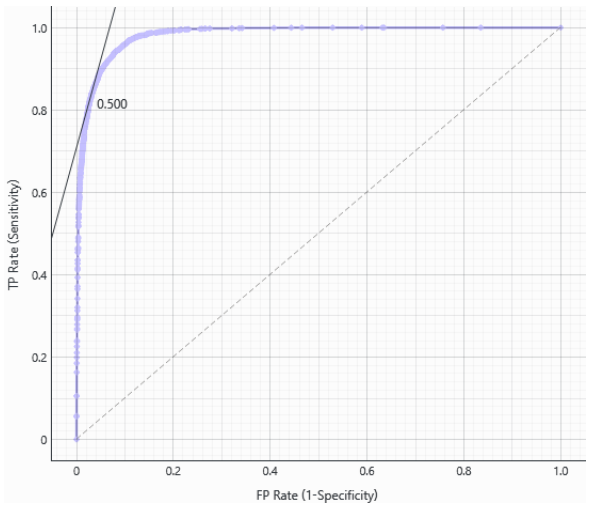

Figure 5. Logistic Regression (Sira) ROC Analysis

However, the “Sira” and “Dermason” species, which have very similar morphological characteristics and overlap in the data space, have also been the primary source of error for this model. According to the analysis results, 252 samples belonging to the “Sira” class were incorrectly predicted as “Dermason,” while 248 samples belonging to the ‘Dermason’ class were incorrectly predicted as “Sira”. This situation demonstrates that the linear decision boundaries of Logistic Regression are limited in fully modeling the complex and non-linear transitions between these two species, yet still exhibit competitive performance in overall classification success. Figure 6 shows the confusion matrix of the LR model.

Predicted

Actual	BARBUNYA	BOMBAY	CALI	DERMASON	HOROZ	SEKER	SIRA	Σ
BARBUNYA	1157	0	101	0	7	12	45	1322
BOMBAY	1	521	0	0	0	0	0	522
CALI	59	0	1522	0	30	2	17	1630
DERMASON	1	0	0	3234	2	61	248	3546
HOROZ	6	0	35	23	1825	0	39	1928
SEKER	16	0	0	36	1	1909	65	2027
SIRA	3	0	8	252	59	49	2265	2636
Σ	1243	521	1666	3545	1924	2033	2679	13611

Figure 6. Logistic Regression confusion matrix

3.4. Neural Network

In the classification experiments conducted within the scope of this study, the highest success was achieved with the Neural Network model, with a general accuracy rate of 93.4% (0.934) and an F1-score of 0.934. In particular, when examining the ROC analysis that validates the model's performance, the AUC (Area Under the Curve) value reaching 0.995 statistically proves that the algorithm's capacity to distinguish between classes is very close to the perfect level (1.0). The distinct proximity of the ROC curves to the upper left corner indicates that the model establishes the balance between sensitivity and specificity in the most optimal way and offers a more stable prediction structure compared to other methods. Figure 7 shows the ROC of the ANN model.

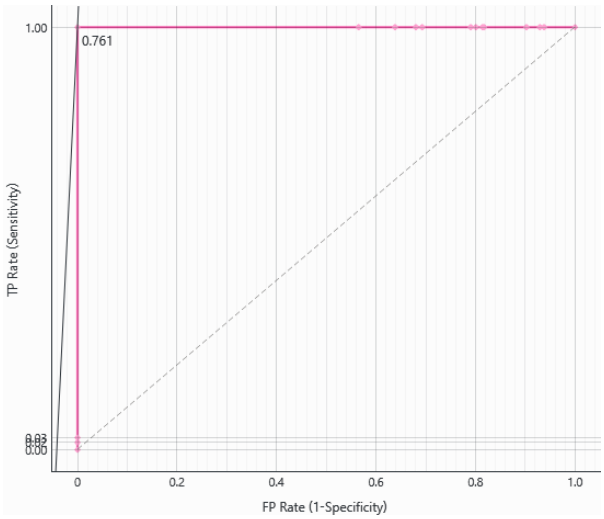

Figure 7. Neural Network (BOMBAY) ROC analysis

The model's superior performance can be attributed to the ability of Artificial Neural Networks to model the complex and non-linear relationships that exist between the features in the dataset. Even in types such as “Sira” and “Dermason”, which are morphologically and geometrically highly similar and difficult to distinguish using methods such as Decision Trees and Logistic Regression, the Neural Network architecture has succeeded in minimizing errors. The model has demonstrated that it is the method with the strongest generalization ability among the algorithms used in the study, accurately identifying not only classes with distinct

features such as "Bombay," but also complex classes where features overlap. Figure 8 shows the confusion matrix of the ANN model.

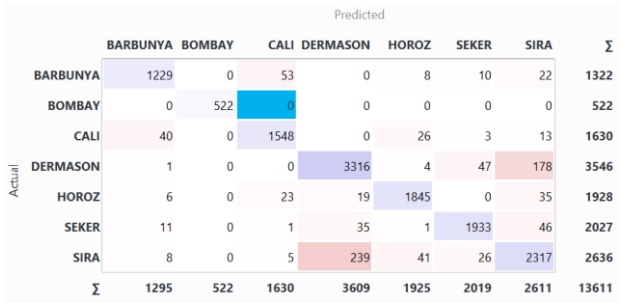

Figure 8. Neural Network confusion matrix

The experimental findings obtained within the scope of the study are presented in Table 2. When comparing the performance of classification algorithms, the Neural Network model demonstrated the highest success. The model demonstrated a statistically superior generalization ability compared to other methods, with a 93.4% (0.934) overall accuracy (CA) rate and an AUC value of 0.995.

In contrast, the Decision Tree algorithm showed the lowest performance among the tested models, with an accuracy of 90.6% (0.906) and an AUC value of 0.945. Although the Random Forest and Logistic Regression models produced competitive results with accuracy rates of 92.1% and 91.3%, respectively, they lagged behind the Artificial Neural Network model. A similar hierarchy was observed when examining the F1-scores; the Artificial Neural Network (0.934) proved to be the most stable model for this dataset in terms of balanced data distribution and class separation success. Table 2 shows the performance metrics of all models.

Table 2. Performance metrics of all models

Model	AUC	CA	F1	Prec	Recall	Spec
Decision Tree	0.945	0.906	0.906	0.906	0.906	0.978
Random Forest	0.989	0.921	0.921	0.921	0.921	0.981
Logistic Regression	0.992	0.913	0.914	0.914	0.913	0.980
Neural Network	0.995	0.934	0.934	0.934	0.934	0.984

4. CONCLUSIONS

The ROC (Receiver Operating Characteristic) analyses performed were examined to validate the discrimination capabilities of the classification models. When comparing the AUC (Area Under the Curve) values of the models, it was observed that all methods performed above the 0.90 threshold. In particular, the Neural Network and Logistic Regression models achieved near-perfect AUC values of 0.995 and 0.992, respectively, demonstrating superior consistency in separating classes. In contrast, the Decision Tree model, with an AUC value of 0.945, remained within acceptable limits but showed a lower level of discriminative power compared to other methods. The proximity of the ROC curves to the upper left corner statistically supports the high sensitivity rates of the models.

When class-based performances were examined, the sensitivity of the models to morphological characteristics showed significant differences. The "Bombay" bean type, whose physical structure clearly distinguishes it from others, was successfully identified by all models; in particular, the Neural Network model demonstrated flawless performance by correctly classifying all 522 examples belonging to this class (100%). However, the "Sira" and "Dermason" types, whose morphological and geometric characteristics are very similar, stood out as the groups where the algorithms made the most errors. For example, in the Decision Tree model, a significant portion of the samples belonging to the "Sira" class were mislabeled as "Dermason".

More complex architectures such as Random Forest (0.989 AUC) and Artificial Neural Networks have been more successful than singular Decision Trees in decomposing such overlapping classes (Rank-Dermason). The findings show that the Artificial Neural Network model, with an overall accuracy rate of 93.4% and a balanced error distribution, is the most suitable method for modeling complex and nonlinear relationships on this dataset.

Declaration of Ethical Standards

The authors declare that this manuscript is original and has not been published or submitted elsewhere. The study used a publicly available dataset and did not involve human participants or animals. All ethical principles regarding authorship, citation, data use, and research integrity were followed.

CRediT Authorship Contribution Statement

Hasan Selim Sindir: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Investigation, Writing – Original Draft, Visualization.

Yavuz Selim Taspinar: Conceptualization, Methodology, Validation, Formal Analysis, Writing – Review & Editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding / Acknowledgements

The authors received no specific funding for this research.

Data Availability Statement

he dataset used in this study is publicly available on Kaggle:https://www.kaggle.com/datasets/whenamancode

s/dry-beans-dataset.

References

[1] G. Słowiński, "Dry beans classification using machine learning," in Proc. 29th Int. Workshop Concurrency, Specification and Programming (CS&P 2021), Berlin, Germany, 2021, CEUR Workshop Proc., vol. 2951, pp. 166–173.

[2] Y. S. Taspinar, M. Dogan, I. Cinar, R. Kursun, I. A. Ozkan, and M. Koklu, "Computer vision classification of dry beans (Phaseolus vulgaris L.) based on deep transfer learning techniques," European Food Research and Technology, vol. 248, no. 11, pp. 2707–2725, 2022, doi: 10.1007/s00217-022-04080-1.

[3] C. H. Mendigoria et al., "Seed architectural phenes prediction and variety classification of dry beans (Phaseolus vulgaris) using machine learning algorithms," in Proc. 2021 IEEE 9th Region 10 Humanitarian Technology Conf. (R10-HTC), Bangalore, India, 2021, pp. 1–6, doi: 10.1109/R10-HTC53172.2021.9641554.

[4] S. Krishnan, S. K. Aruna, K. Kanagarathinam, and E. Venugopal, "Identification of dry bean varieties based on multiple attributes using CatBoost machine learning algorithm," Scientific Programming, vol. 2023, Art. no. 2556066, 2023, doi: 10.1155/2023/2556066.

[5] M. S. Khan et al., "Comparison of multiclass classification techniques using dry bean dataset," International Journal of Cognitive Computing in Engineering, vol. 4, pp. 6–20, 2023, doi: 10.1016/j.ijcce.2023.01.002.

[6] M. Dogan, Y. S. Taspinar, I. Cinar, R. Kursun, I. A. Ozkan, and M. Koklu, "Dry bean cultivars classification using deep CNN features and salp swarm algorithm based extreme learning machine," Computers and Electronics in Agriculture, vol. 204, Art. no. 107575, 2023, doi: 10.1016/j.compag.2022.107575.

[7] J. C. Macuácua, J. A. S. Centeno, and C. Amisse, "Data mining approach for dry bean seeds classification," Smart Agricultural Technology, vol. 5, Art. no. 100240, 2023, doi: 10.1016/j.atech.2023.100240.

[8] A. Mehta, P. Sengupta, D. Garg, H. Singh, and Y. Shacham-Diamand, "Benchmarking the effectiveness of classification algorithms and SVM kernels for dry beans," arXiv preprint arXiv:2307.07863, 2023, doi: 10.48550/arXiv.2307.07863.

[9] J. Nayak, P. B. Dash, and B. Naik, "An advance boosting approach for multiclass dry bean classification," Journal of Engineering Science and Technology Review, vol. 16, no. 2, pp. 107–115, 2023, doi: 10.25103/jestr.162.14.

[10] G. Shobana, S. N. Bushra, K. U. Maheswari, and N. Subramanian, "Multivariate classification of dry beans using pipelined dimensionality reduction technique," in Proc. 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1–6, doi: 10.1109/ICSES55317.2022.9914079.

[11] I. F. Rimi, M. T. Habib, S. Supriya, M. A. A. Khan, and S. A. Hossain, "Traditional machine learning and deep learning modeling for legume species recognition," SN Computer Science, vol. 3, no. 6, Art. no. 430, 2022, doi: 10.1007/s42979-022-01268-w.

[12] H. Vaidya, K. V. Prasad, S. Renuka, and K. Kumar Swamy, "Multiclass classification of dry beans using artificial neural network," Journal of International Academy of Physical Sciences, vol. 27, no. 2, pp. 109–124, 2023.

[13] M. Koklu and I. A. Ozkan, "Multiclass classification of dry beans using computer vision and machine learning techniques," Computers and Electronics in Agriculture, vol. 174, Art. no. 105507, 2020, doi: 10.1016/j.compag.2020.105507.

[14] G. Słowiński, "Python machine learning. Dry beans classification case," Zeszyty Naukowe WWSI, vol. 18, no. 30, pp. 7–26, 2024, doi: 10.26348/znwwsi.30.7.

[15] J. Coronel-Reyes, C. Delgado-Vera, J. Chavez-Urbina, and A. Sinche-Guzmán, "Multiclass classification of dry bean grains using machine learning techniques," in Technologies and Innovation: CITI 2024, R. Valencia-García et al., Eds. Cham, Switzerland: Springer, 2025, vol. 2276, pp. 16–27, doi: 10.1007/978-3-031-75702-0_2.

[16] "Dry Beans Classifications," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/whenamancodes/dry-beans-dataset. [Accessed: Aug. 28, 2026].

[17] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Boca Raton, FL, USA: Chapman & Hall/CRC, 2017.

[18] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.

[19] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[20] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1007/BF00058655.

[21] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, 2013, doi: 10.1002/9781118548387.

[22] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[23] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55–67, 1970, doi: 10.1080/00401706.1970.10488634.

[24] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.

[25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, vol. 15, pp. 1929–1958, 2014.

[26] L. Prechelt, "Early stopping—But when?" in Neural Networks: Tricks of the Trade, G. B. Orr and K.-R. Müller, Eds., Lecture Notes in Computer Science, vol. 1524. Berlin, Germany: Springer, 1998, pp. 55–69, doi: 10.1007/3-540-49430-8_3.

[27] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," Remote Sensing of Environment, vol. 62, no. 1, pp. 77–89, 1997, doi: 10.1016/S0034-4257(97)00083-7.

[28] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[29] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Proc. 14th International Joint Conference on Artificial Intelligence (IJCAI), 1995, pp. 1137–1145.