IJAMEC

INTERNATIONAL JOURNAL OF APPLIED METHODS IN ELECTRONICS AND COMPUTERS

www.ijamec.org

International Open Access

Volume 13
Issue 03
September, 2025

Research Article

Performance Comparison of SVM Kernel Functions for Date Fruit Classification

Huseyin BULDUK [a,*], ***Kadir SABANCI***

[a] *Electrical and Electronics Engineering, Karamanoglu Mehmetbey University, Karaman, Türkiye*

ARTICLE INFO

Article history:
Received 28 May 2025
Accepted 22 July 2025

Keywords:
Classification,
Date Fruit,
Support Vector Machines (SVM),
Feature Selection,
Minimum Redundancy Maximum Relevance (MRMR)

ABSTRACT

In this study, the Support Vector Machines (SVM) algorithm was employed to classify different types of date fruits. The performances of various SVM kernel functions - namely Linear, Quadratic, Cubic, Medium Gaussian, and Coarse Gaussian- were compared during the classification process. The analyses were conducted using the Date Fruit Dataset, which was published on the Kaggle platform and comprises 34 numerical features. The Minimum Redundancy Maximum Relevance (MRMR) feature selection algorithm was utilized to identify the 13 most effective features for classification. Subsequently, classification was performed using both the complete feature set (34 features) and the selected subset (13 features). The findings revealed that the highest classification accuracy was achieved with the Linear kernel SVM model in both cases. When all features were used, the Linear SVM model reached an accuracy of 91.79%, whereas the accuracy increased to 92.07% when the 13 features selected by MRMR were employed. These results indicate that feature selection plays a significant role in improving classification performance.

1. Introduction

In recent years, agriculture has emerged as a key driver of global economic development [1]. Date fruit is a valuable agricultural product, with approximately 8.46 million tons produced annually worldwide, and is primarily cultivated in regions such as Southwest Asia, North Africa, and the Middle East [2]. With more than forty species and over four hundred variations, it exhibits significant diversity in terms of taste, shape, and color. This diversity necessitates automated classification for quality control and market segmentation [3]. However, manual classification is inefficient due to its time-consuming nature, high cost, and susceptibility to human bias [2]. Different machine learning approaches (especially Support Vector Machines (SVM) and image processing techniques) have been proposed as effective solutions for date fruit classification [4,5].

Extensive research on date fruit classification has confirmed the effectiveness of both machine learning and deep learning algorithms. As an example, Albarrak et al. [2] reported a 99% classification accuracy using the MobileNetV2 architecture on a dataset containing eight distinct types of date fruits. Similarly, Muhammad [6] reported an accuracy of 98.5% through the use of feature extraction and SVM. In the study of Muhammad, Support Vector Machines (SVM) were employed as the classifier for the categorization of date fruit varieties. Demonstrating robust performance in high-dimensional and non-linearly separable feature spaces, SVM utilized selected texture (LBP and WLD) and shape-size features as input to determine the optimal decision boundaries between classes. In their study, Koklu et al. [7] utilized the Otsu thresholding technique to extract features from a dataset of 898 date images, achieving an accuracy of 92.8% through the application of logistic regression and artificial neural networks. Altaheri et al. [8] utilized transfer learning with CNN to classify five date types with 99.01% accuracy. Abi

* Corresponding author. E-mail address: *huseyinbulduk@kmu.edu.tr*
DOI: 10.58190/ijamec.2025.129

Sen et al [9] used SVM to classify four types of date palms with 73.8% accuracy. In their study, Support Vector Machines (SVM) emerged as the most accurate classification method among various machine learning algorithms used for the classification of date palm fruits. Experiments conducted on a dataset consisting of a limited number of images collected under various conditions showed that SVM outperformed other methods such as Decision Tree, Random Forest, and Neural Networks in terms of stability and accuracy. Alsirhani et al. [10] achieved 95.21% test accuracy through transfer learning with DenseNet. Alhadhrami et al. [11] obtained 98.33% accuracy using a pre-trained CNN, while Nasiri et al. [12] achieved 96.98% accuracy with a deep CNN model. Faisal et al. (2020) attained 99.4% accuracy by combining computer vision and deep learning approaches. SVM-based classification methods have found wide applicability in the agricultural sector. For example, Sonmez et al. achieved 99.84% success in wheat classification with SVM [13]. In this study, which focused on the classification of wheat varieties based on color features, Support Vector Machines (SVM) was employed as one of six machine learning algorithms evaluated for performance. To assess the model's flexibility, the researchers experimented with three distinct kernel functions—Linear, Polynomial, and RBF—and identified the optimal c (penalty) hyperparameter for each via an automated loop. The results demonstrated that SVM was among the top-performing models, achieving an accuracy of over 99%. In the research conducted by Adige et al. [14], the classification of apple varieties was performed by using Support Vector Machines (SVM) together with the Bag of Visual Words (BoVW) method. In this method, instead of directly processing the raw image pixels, the feature vectors extracted from the images with the BoVW method were classified by SVM. The procedure involves converting images into numerical vectors based on histograms of visual words, which are then fed as input to the SVM. The experimental results indicated that the BoVW-SVM approach, particularly with a polynomial kernel function, achieved a test accuracy of 95.9%, outperforming the ResNet-50 deep learning model that was evaluated for comparison in the same study. In a similar vein, Arshaghi et al. [15] applied an integrated approach that combined deep learning techniques with Support Vector Machines (SVM) to identify potato diseases. Gencturk et al. [16] achieved 100% accuracy in hazelnut classification using an InceptionV3+ResNet50 fusion model. These studies indicate that SVM delivers high performance in the classification of agricultural products and that its effectiveness can be enhanced through feature extraction. Özaltın's study presented an effective method for the automatic classification of date varieties and demonstrated that artificial neural networks outperformed other algorithms with an accuracy of 93.85% [17]. These findings underline the strong performance of SVM and deep learning methods in date classification and highlight the critical importance of feature extraction.

In this study, the classification of date fruit varieties was performed based on the features selected through the Minimum Redundancy Maximum Relevance (MRMR) algorithm. The process involved the implementation of various kernel functions within the Support Vector Machines (SVM) framework.

2. MATERIAL AND METHODS

2.1. Dataset

The Date Fruit Dataset [18] consists of seven distinct date fruit types, namely Berhi, Deglet, Dokol, Iraqi, Rotana, Safavi, and Sogay. It includes 898 individual samples, each described by 34 numerical attributes. These attributes capture a range of physical characteristics such as dimensions, shape, surface texture, and color distribution. Such attributes enable classification algorithms to learn the distinctive patterns among different date varieties.

2.2. Method

The analyses for this study were conducted using the Classification Learner application within the MATLAB (R2024a) software environment. All computations were executed on a workstation equipped with an Intel Core i7-13700 processor and 32 GB of RAM, running the Windows 10 operating system. During the classification phase, several SVM kernel types were utilized, including Linear, Quadratic, Cubic, Medium Gaussian, and Coarse Gaussian. The dataset was split into two subsets: 80% for training and 20% for testing. Model performance was assessed through multiple evaluation metrics, such as accuracy, precision, recall, F1 score, Matthews Correlation Coefficient (MCC), confusion matrices, and the ROC curve of the top-performing model.

Support Vector Machines (SVM) represent a type of supervised learning algorithm grounded in the foundations of statistical learning theory and the principles of convex optimization [19]. Originally developed in the 1990s by Vapnik and his colleagues for binary classification problems, SVMs have been successfully applied in various domains such as bioinformatics, text classification, and computer vision [20]. The hard-margin SVM was initially designed for linearly separable data; however, it proved inadequate when dealing with noisy or outlier data. To overcome this limitation, Cortes and Vapnik [21] proposed a soft-margin approach that enables nonlinear classification. SVM transforms data into higher-dimensional feature spaces using kernel functions, allowing nonlinearly separable data to become linearly separable. Commonly used kernel functions include linear, quadratic, cubic, and Gaussian-based kernels (e.g., Medium and Coarse Gaussian), which provide flexibility in handling complex data structures and improve the model's generalization performance.

The MRMR algorithm is employed as a supervised feature selection method and aims to enhance classification accuracy. The principle of Maximum Relevance selects features that have high mutual information with the target class, while the principle of Minimum Redundancy reduces overlapping information among features to avoid redundancy. By jointly optimizing these objectives, the goal is to achieve accurate and efficient classification with a reduced number of features [22].

The flowchart presented in Figure 1 illustrates the steps followed for the classification of date fruit types. In the first phase, the complete feature set comprising 34 attributes was used directly in classification with various SVM kernel functions. In the second stage, feature selection is applied to the same dataset. The Minimum Redundancy Maximum Relevance (MRMR) algorithm is used to identify the 13 most relevant features for classification. Classification is then repeated using the same SVM kernel functions but with the selected subset of features.

In both approaches, the primary objective is to classify the date fruit varieties (Berhi, Deglet, Dokol, Iraqi, Rotana, Safavi, and Sogay) as accurately as possible, and the effect of feature selection on classification performance is examined comparatively.

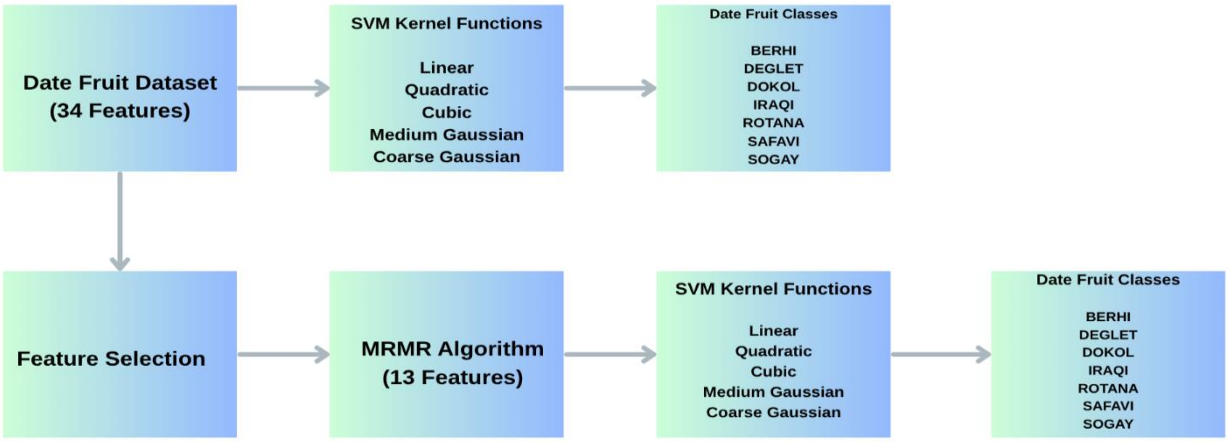

Figure 1. Flowchart of the Study

3. RESULTS AND DISCUSSION

3.1. Classification Performance Obtained Using 34 Features

In the initial phase of the study, classification was performed using all 34 features of the date fruit dataset with various SVM kernel functions. The models were trained using the MATLAB Classification Learner interface, and their performances were visualized through confusion matrices. The obtained confusion matrices clearly illustrate the classification accuracy of each model as well as their ability to distinguish between the classes. In Figure 2 (a–e), the confusion matrices for the Linear, Quadratic, Cubic, Medium Gaussian, and Coarse Gaussian kernel functions are presented, respectively. The highest classification accuracy was achieved with the SVM model employing the Linear kernel.

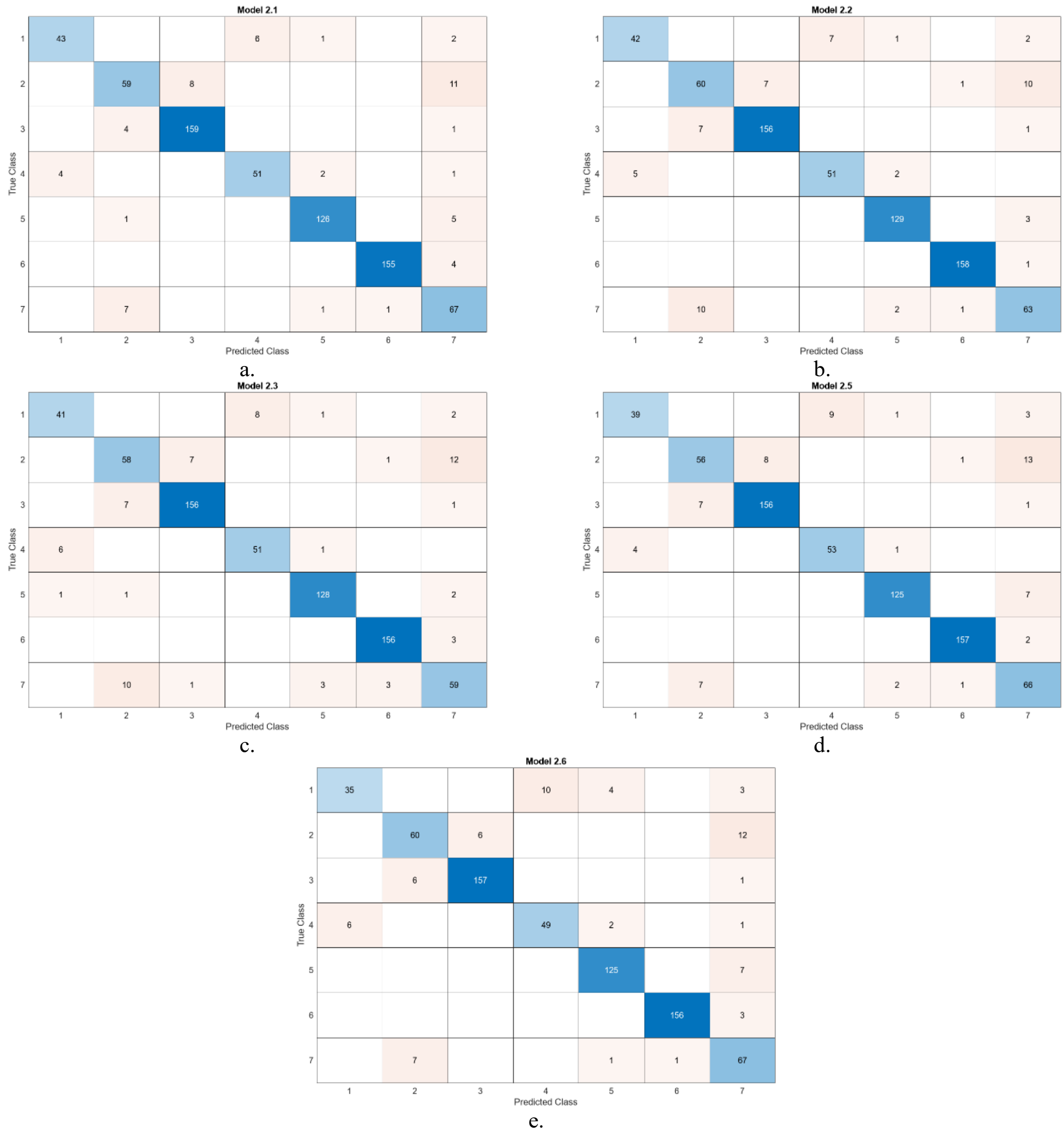

Figure 2. Confusion matrices obtained using SVM models with different kernel functions based on 34 features: (a) Linear, (b) Quadratic, (c) Cubic, (d) Medium Gaussian, (e) Coarse Gaussian

Classification was performed using the full dataset comprising 34 features with five distinct SVM kernel functions, and the corresponding results are summarized in Table 1. The performance assessment was based on several metrics, including accuracy (%), precision, recall, F1 score, and the Matthews Correlation Coefficient (MCC). The Linear SVM model attained the highest classification accuracy at 91.79%. Furthermore, it achieved the top F1 score (0.8938) and MCC value (0.8813), establishing it as the most effective model in terms of class balance and overall predictive performance.

The Quadratic SVM model attained a classification accuracy of 91.66%, which was comparable to the Linear SVM. Nevertheless, its performance in terms of precision, recall, F1 score, and MCC was slightly inferior, suggesting that despite its high accuracy, its ability to consistently distinguish between classes was less robust. Meanwhile, the Cubic, Medium Gaussian, and Coarse Gaussian SVM models exhibited similar accuracy levels (between 90.26% and 90.68%) and produced nearly equivalent results across the other evaluation metrics. The lower MCC values observed in these models, compared to the Linear and Quadratic SVMs, suggest that their ability to distinguish between classes was relatively weaker.

Although the Medium Gaussian SVM offered a better balance in terms of accuracy (90.68%) and F1 score (0.8776) compared to the Cubic and Coarse Gaussian kernels, it still performed inferiorly to the Linear SVM.

Overall, when all metrics are considered, the Linear SVM model delivered the most consistent and successful results in the classification of date fruit varieties.

Table 1. Classification Performance Metrics of Different SVM Kernel Functions (34 Features)

Kernel Function	Accuracy (%)	Precision	Recall	F1 Score	MCC
Linear SVM	91.79	0.8988	0.8889	0.8938	0.8813
Quadratic SVM	91.66	0.8924	0.8861	0.8892	0.8758
Cubic SVM	90.26	0.8739	0.8695	0.8717	0.8559
Medium Gaussian SVM	90.68	0.8838	0.8715	0.8776	0.8639
Coarse Gaussian SVM	90.26	0.8747	0.8591	0.8668	0.8528

3.2. Classification Performance Obtained Using 13 Features

Using the MRMR algorithm, 13 features were selected from the original set of 34 as the most effective for classification. These selected features were then used to reclassify the data with the best-performing model, the Linear SVM. The performance metrics for the classification conducted using the 13 features—including accuracy, precision, recall, F1 score, and Matthews Correlation Coefficient (MCC)—are presented in Table 2.

Table 2. Performance Metrics of the Linear SVM Model (13 Features)

Metric	Value
Accuracy	0.9207
Precision	0.9018
Recall	0.8938
F1 Score	0.8978
MCC	0.8856

The classification performance metrics obtained using the Linear SVM model with both the full set of 34 features and the selected subset of 13 features are presented in Table 3. It was observed that the classification with 13 features yielded higher accuracy (92.07%), precision (0.9018), recall (0.8938), F1 score (0.8978), and MCC (0.8856) compared to the full feature set. These findings demonstrate that the classification performance of the model was improved by employing a smaller but more informative set of features selected through the MRMR algorithm.

Table 3. Comparison of Linear SVM Performance Based on the Number of Features

Kernel Function	Accuracy (%)	Precision	Recall	F1 Score	MCC
Linear (34 Features)	91.79	0.8988	0.8889	0.8938	0.8813
Linear (13 Features)	92.07	0.9018	0.8938	0.8978	0.8856

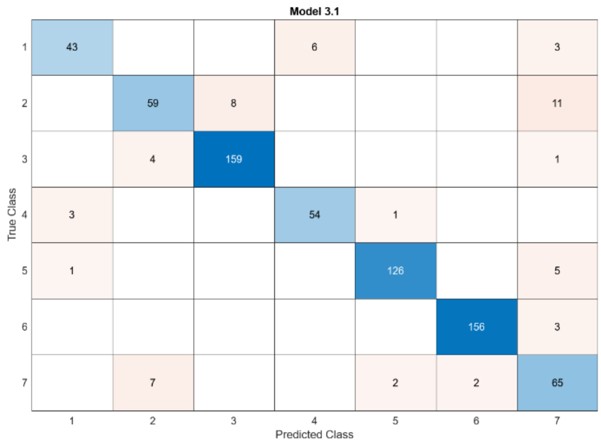

Figure 3. Confusion matrix of the Linear SVM kernel function obtained using 13 features

The ROC curve obtained with the Linear SVM kernel function using 34 features is given in Figure 4. The ROC curve demonstrates the trade-off between the true positive rate and the false positive rate, serving as a visual tool to evaluate the classification capability of the model. Figure 5 displays the ROC curve generated using the Linear SVM kernel with the 13 selected features.

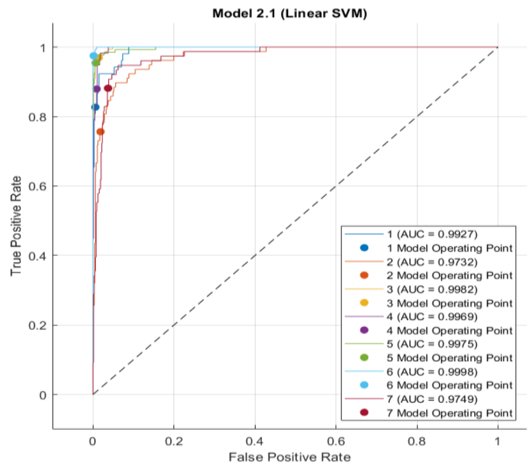

Figure 4. ROC Curve of the Linear SVM Model (34 Features)

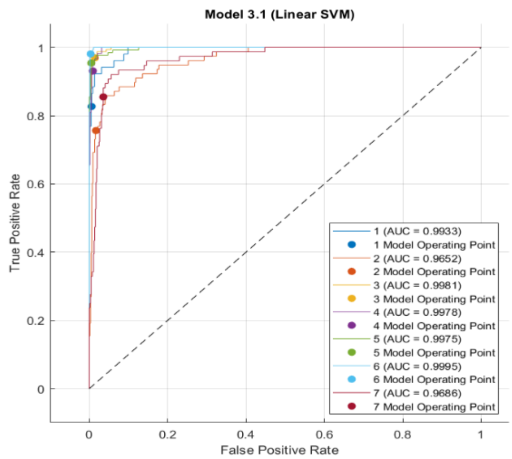

Figure 5. ROC Curve of the Linear SVM Model (13 Features)

3.3. Discussion

In this study, various SVM kernel functions were evaluated for date fruit classification, and the Linear SVM demonstrated the highest performance among all models. Reducing the number of features from 34 to 13 increased the classification accuracy from 91.79% to 92.07%, and the MCC from 0.8813 to 0.8856. This indicates that feature reduction contributed to the improvement of the model. While the Quadratic SVM followed with an accuracy of 91.66%, the other kernel functions exhibited lower performance. In the classification with 13 features, only the Linear SVM was analyzed, and the ROC curve confirmed its high discriminative power. Consistent with the results in the literature, this approach was found to be both competitive in terms of classification accuracy and efficient in terms of computational cost.

4. CONCLUSIONS

This study comparatively examined the performance of various SVM kernel functions in the classification of date fruits using the Date Fruit Dataset. The Linear SVM emerged as the most effective method, achieving the highest performance with 13 features—yielding an accuracy of 92.07% and an MCC of 0.8856. In the analysis conducted with 34 features, the Linear SVM also outperformed the other kernel functions, achieving 91.79% accuracy and an MCC of 0.8813, compared to 91.66% for Quadratic, 90.68% for Medium Gaussian, and 90.26% for both Cubic and Coarse Gaussian kernels. It was observed that feature reduction enhanced the generalization ability of the model by eliminating redundant information and reduced computational cost.

The findings of the study indicate that machine learning-based approaches offer practical and effective solutions for date fruit classification in the agricultural sector. The high accuracy achieved by the Linear SVM with 13 features provides a solid foundation for developing automated classification systems in quality control and market segmentation processes. The ROC curve analysis further confirmed the model's high discriminative ability and demonstrated that the classification performance was more balanced when using the selected 13 features.

Future studies may aim to go beyond the current linear SVM-based approach to improve both classification accuracy and industrial applicability. In addition to traditional methods such as MRMR, more sophisticated feature selection strategies—such as genetic algorithms, particle swarm optimization (PSO), and deep learning-based automatic feature extraction (e.g., autoencoders)—can be employed to enhance the model's accuracy and generalization capability. Hybridizing SVM with deep learning models, particularly those utilizing Convolutional Neural Networks (CNNs) for robust feature extraction followed by SVM classification, may further boost classification performance. Moreover, cutting-edge architecture such as EfficientNet and Vision Transformer (ViT), when integrated with transfer learning techniques, can yield high performance even on limited datasets. Strategies such as data augmentation, GAN-based synthetic image generation, and the incorporation of multispectral imagery can further enrich the dataset.

Declaration of Ethical Standards

The authors confirm that this study adheres to all ethical standards, including proper authorship attribution, accurate citation, appropriate data reporting and the publication of original research.

Credit Authorship Contribution Statement

The conceptualization of the research and the data collection process were carried out by Hüseyin Bulduk. The evaluation and analysis of the results were performed by Kadir Sabancı. The original draft of the manuscript was written by Hüseyin Bulduk, while the review and editing were undertaken by Kadir Sabancı.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Funding / Acknowledgements

No funding or research grants were received during the preparation of this study.

Availability of Data and Material

The authors confirm that the data supporting the findings of this study are available within the manuscript.

References

[1] Das and Rupa (2023). *ResNet for Leaf-based Disease Classification in Strawberry Plant.* International Journal of Applied Methods in Electronics and Computers 11(03): 151-157, 2023. https://doi.org/10.58190/ijamec.2023.42

[2] Albarrak K, Gulzar Y, Hamid Y, Mehmood A, Soomro AB. *A Deep Learning-Based Model for Date Fruit Classification*. Sustainability. 2022; 14(10):6339. https://doi.org/10.3390/su14106339

[3] A. Haidar, Haiwei Dong and N. Mavridis, *"Image-based date fruit classification," 2012 IV* International Congress on Ultra-Modern Telecommunications and Control Systems, St. Petersburg, Russia, 2012, pp. 357-363, doi: 10.1109/ICUMT.2012.6459693

[4] Osisanwo, F., Akinsola, J., Awodele, O., Hinmikaiye, J., Olakanmi, O., & Akinjobi, J. (2017). *Supervised machine learning algorithms: Classification and comparison*. International Journal of Computer Trends and Technology, 48(3), 128–138. https://doi.org/10.14445/22312803/IJCTT-V48P126

[5] Lionel J Garcia, Mathilde Timmermans, Francisco J Pozuelos, Elsa Ducrot, Michaël Gillon, Laetitia Delrez, Robert D Wells, Emmanuël Jehin, *prose: a python framework for modular astronomical images processing,* Monthly Notices of the Royal Astronomical Society, Volume 509, Issue 4, February 2022, Pages 4817–4828, https://doi.org/10.1093/mnras/stab3113

[6] Muhammad, G. (2015). *Date fruits classification using texture descriptors and shape-size features*. Engineering Applications of Artificial Intelligence, 37, 361–367. https://doi.org/10.1016/j.engappai.2014.10.001

[7] Koklu, M., Kursun, R., Taspinar, Y. S., & Cinar, I. (2021b). *Classification of date fruits into genetic varieties using image analysis.* Mathematical Problems in Engineering, 2021, 1–13. https://doi.org/10.1155/2021/4793293

[8] H. Altaheri, M. Alsulaiman and G. Muhammad, "Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep *Learning,"* in IEEE Access, vol. 7, pp. 117115-117133, 2019, doi: 10.1109/ACCESS.2019.2936536

[9] A. A. Abi Sen, N. M. Bahbouh, A. B. Alkhodre, A. M. Aldhawi, F. A. Aldham and M. I. Aljabri, *"A Classification Algorithm for Date Fruits,"* 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 235-239, doi: 10.23919/INDIACom49435.2020.9083706

[10] Alsirhani, Amjad, Muhammad Hameed Siddiqi, Ayman Mohamed Mostafa, Mohamed Ezz, and Alshimaa Abdelraof Mahmoud. 2023. *"A Novel Classification Model of Date Fruit Dataset Using Deep Transfer Learning"* Electronics 12, no. 3: 665. https://doi.org/10.3390/electronics12030665

[11] N. Alhadhrami, A. Abobakr, A. Alhammadi and M. Shatnawi, *"Multiple Classification of Date Fruit using Transfer Learning,"* 2023 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 2023, pp. 1-5, doi: 10.1109/ASET56582.2023.10180537

[12] Nasiri, A., Taheri-Garavand, A., & Zhang, Y.-D. (2019). *Image-based deep learning automated sorting of date fruit.* Postharvest Biology and Technology, 153, 133–141. https://doi.org/10.1016/j.postharvbio.2019.04.003

[13] M.E. Sonmez, Kadir Sabanci and Nevzat Aydin. *Classification of Wheat Rootstock and Their Hybrids According to Color Features by Machine Learning Algorithms.* International Journal of Applied Mathematics Electronics and Computers 10(02): 039-048, 2022. https://doi.org/10.18100/ijamec.1098276

[14] Adige, S., Kurban, R., Durmuş, A. *et al. Classification of apple images using support vector machines and deep residual networks*. Neural Comput & Applic **35**, 12073–12087 (2023). https://doi.org/10.1007/s00521-023-08340-3

[15] [15] Arshaghi, A., Ashourian, M. & Ghabeli, L. *Potato diseases detection and classification using deep learning methods*. Multimed Tools Appl 82, 5725–5742 (2023). https://doi.org/10.1007/s11042-022-13390-1

[16] Gencturk, B., Arsoy, S., Taspinar, Y.S. *et* al. *Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models*. Eur Food Res Technol 250, 97–110 (2024). https://doi.org/10.1007/s00217-023-04369-9

[17] Ozaltın, O. (2024). *Date fruit classification by using image features based on machine learning algorithms.* Research in Agricultural Sciences, 55(1),26-35. https://doi.org/10.5152/AUAF.2024.23171

[18] Koklu, M. (2021). *Date Fruit Dataset. Kaggle.* Retrieved from https://www.kaggle.com/datasets/muratkokludataset/date-fruit-datasets

[19] Mammone, A., Turchi, M., & Cristianini, N. (2009). *Support vector machines.* WIREs Computational Statistics, 1(3), 283–289. https://doi.org/10.1002/wics.49

[20] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers.* Proceedings of the fifth annual workshop on Computational Learning Theory*, 144-152.* doi: 10.1145/130385.130401

[21] Cortes, C., & Vapnik, V. (1995). *Support-vector networks.* Machine Learning, 20(3), 273–297. doi: 10.1007/%2FBF00994018

[22] Peng, H., Long, F., & Ding, C. (2005). *Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 1226–1238. doi: 10.1109/TPAMI.2005.159