Proceedings Paper DOI: 10.58190/icisna.2024.91

PROCEEDINGS OF

INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS AND NEW APPLICATIONS

https://proceedings.icisna.org/

2nd International Conference on Intelligent Systems and New Applications (ICISNA'24), Liverpool, April 26-28, 2024.

Machine Learning-Based Classification of Mulberry Leaf Diseases

Elham Tahsin YASIN[1], Ramazan KURSUN[2], Murat KOKLU[3]

[1] Graduate School of Natural and Applied Sciences, Computer Engineering, Selcuk University, TÜRKİYE
ilham.tahsen@gmail.com, ORCID: 0000-0003-3246-6000

[2]Guneysinir Vocational School, Selcuk University, Guneysinir, TÜRKİYE
rkursun@selcuk.edu.tr, ORCID: 0000-0002-6729-1055

[3]Department of Computer Engineering, Selcuk University, TÜRKİYE
mkoklu@selcuk.edu.tr, ORCID: 0000-0002-2737-2360

***Abstract*— This research examines the potential of machine learning methods in the classification of Mulberry leaf diseases. By applying SqueezeNet's deep feature extraction, the study aimed to identify disease patterns efficiently. The dataset used in the study consisted of ten distinct classes of Mulberry leaf diseases, which was divided into an 80% training set and a 20% testing set. The Support Vector Machine (SVM) supervised machine learning algorithm was used to classify the diseases, and the classification model achieved an accuracy of 77.5%. The results of the study demonstrate the effectiveness of machine learning approaches in aiding the detection and management of Mulberry leaf diseases, which can contribute to advancements in agricultural disease monitoring and mitigation strategies.**

***Keywords*— Computer Vision, Image classification, Machine Learning, Mulberry leaf image, Plant Disease, SVM.**

I. INTRODUCTION

The occurrence of conditions in Mulberry leaves presents a notable challenge in agriculture, resulting in considerable losses for producers on a worldwide basis. Conventional approaches to detect these disorders are frequently complicated and may exhibit limitations in terms of precision. Consequently, there is an increasing interest in investigating alternate methods to improve disease categorization and treatment. Applying machine learning algorithms offers a viable approach to properly tackle this problem [1, 2]. Support Vector Machine (SVM) techniques have demonstrated significant promise for addressing classification problems in this scenario. This work seeks to utilise Support Vector Machines (SVM) to classify mulberry leaf diseases with enhanced accuracy and efficiency in comparison to traditional techniques. The application of SVM-based categorization shows potential for enhancing disease management tactics and reducing economic losses in agriculture [3]. Automated systems that use leaf images for plant recognition have gained attention in fields such as food engineering, industry, botany, and medicine. This work expands upon previous research in automated plant recognition by especially concentrating on the categorization of mulberry leaf disease using Support Vector Machines (SVM). Integrating a recognition system. The strength and accuracy of the classification process are enhanced by extracting significant characteristics from the leaf images, including length, width, area, perimeter, and other similar features. This collaborative task seeks to make a valuable contribution to both agricultural science and technical innovation by providing a practical solution for managing mulberry leaf disease.

In order to assess the current state of research concerning image classification through machine learning algorithms and establish a theoretical foundation for the study, we conducted a comprehensive review of existing literature on plant disease and their classification using machine learning or deep learning models [4-9].

Plant leaf diseases have garnered significant attention due to their considerable effects on agriculture and daily life [10, 11]. The conventional methods of disease identification are time-consuming and prone to errors, leading to the development of innovative techniques based on deep learning. These techniques include the use of deep convolutional neural networks (CNNs) for disease identification and classification, which have achieved impressive accuracy rates [12-14]. Another study proposes an ensemble CNN method that utilizes multiple CNN models to enhance plant leaf recognition performance, achieving high accuracies when employing the Xception model along with data augmentation techniques [15, 16]. Additionally, plant disease recognition systems optimized for smartphones have been developed, showcasing their effectiveness in identifying leaf diseases when incorporating offline training and data augmentation methods. Overall, these studies highlight the potential of deep learning techniques, such as CNNs [17-19], in transforming plant disease identification and classification and providing accurate and efficient solutions for mitigating the impact of plant diseases on agriculture and human livelihoods [20, 21].

A study demonstrated the effectiveness of the selected methodologies, as the mint leaves were accurately classified using InceptionV3-extracted features at 94.8%, and VGG19-extracted features resulted in an improved accuracy of 96.8 [22]. Similar to the investigation of the Pudina Leaf freshness investigation identified three distinct categories of: dried, fresh, and rotten, another study proposed the classification of black gram plant leaf diseases [23]. By applying advanced techniques such as Darknet-53, ResNet-101, GoogLeNet, and EfficientNet-B0, authors aimed to contribute valuable insights for effective disease management.

From another study an automated plant recognition system that utilizes leaf photos was created. This system designed to assist food engineers, industrialists, botanists, and physicians. The system was implemented using a k-Nearest Neighbor classifier. Initially, it obtained an accuracy of 83.5%, which improved to 87.3% after incorporating color histogram data [24].

This study aims to employ machine learning techniques, specifically deep feature extraction [25] with SqueezeNet, for the classification of Mulberry leaf diseases. Using a dataset comprising ten classes of Mulberry leaf diseases, the research evaluates the efficacy of the Support Vector Machine (SVM) algorithm in disease classification.

II. MATERIALS AND METHODS

The primary purpose of this work was to use the SVM algorithm to categorize diseases affecting mulberry leaves and evaluate its effectiveness compared to traditional approaches. To accomplish this, we used a cross-sectional research design. The study utilized image data obtained from the publicly accessible Mendeley data repository [26]. Before being classified, the dataset was pre-processed by extracting features [27] using SqueezeNet. Afterwards, the Support Vector Machine (SVM) algorithm was used for classification. To guarantee a thorough evaluation, the dataset was divided into two sets: a training set and a testing set, with a split ratio of 80% for training and 20% for testing. The flow diagram of the study was shown in Figure 1.

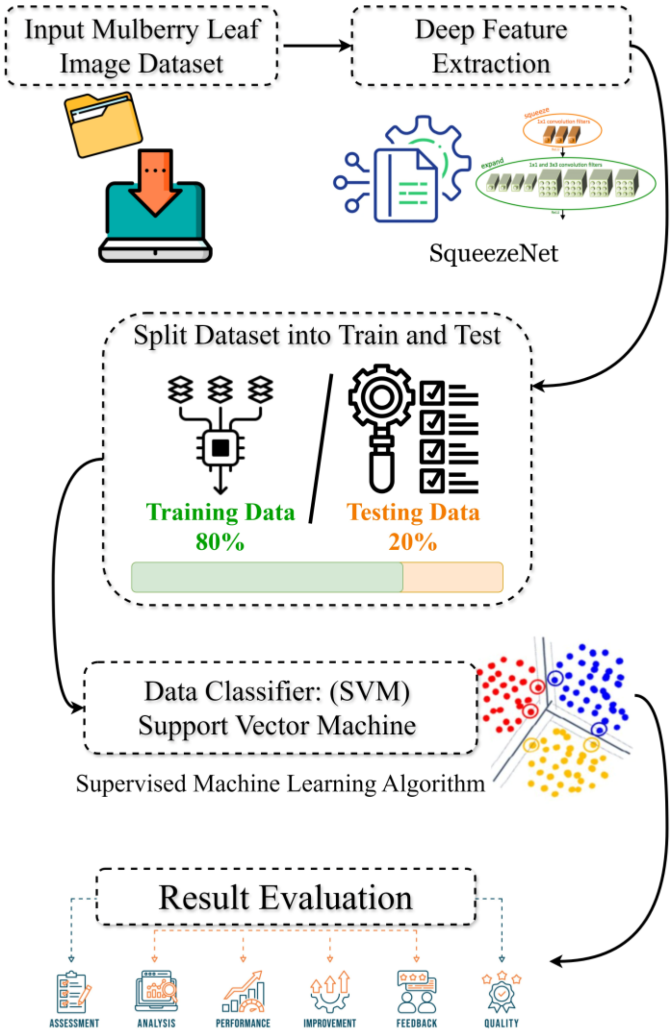


Fig 1. Mulberry leaf disease study flow diagram

A. *Mulberry leaf dataset for image classification task*

This dataset consists of mulberry leaf images collected from various regions in Thailand (northeast, central, and northern), including five provinces from three different regions. It aims to address the challenges of classifying mulberry leaf images taken in natural environments, which are characterized by difficult inter-class similarities and differences in lighting and background conditions. This dataset, which consists of 5,262 annotated samples divided into ten categories: King Red, King White, Taiwan Maechor, Taiwan Strawberry, Black Austurkey, Black Australia, Chiang Mai 60, Buriram 60, Kamphaeng Saen 42, and Mixed Chiang Mai 60+Buriram 60 [28]. It provides useful resources for computer vision researchers who are interested in developing creative algorithms to address mulberry recognition issues.

The experts categorized the images into 10 different classes based on their distinct characteristics and traits. Therefore, this dataset is essential for creating new deep learning frameworks and image recognition systems specifically designed for the complexities of mulberry leaf classification. Figure 2 shows the presentation of the mulberry leaf dataset.

Fig 2. Mulberry leaf disease dataset sample images

B. *Deep Feature Extraction*

Deep feature extraction involves capturing abstract characteristics from data, notably images, through deep learning models. SqueezeNet, a convolutional neural network (CNN) architecture, is tailored for efficient computation and reduced memory usage without sacrificing accuracy. Achieving this, SqueezeNet integrates a "squeeze-and-excitation" (SE) module for channel-wise feature recalibration and replaces certain 3x3 convolutional layers with 1x1 convolutional layers to streamline model parameters.

In practical terms, SqueezeNet extracts deep features from input images, encapsulating essential image content characteristics. These features are pivotal for tasks like image classification, object detection, and image segmentation. In the referenced study, deep feature extraction using SqueezeNet was applied to extract significant features from Mulberry leaf disease images. Specifically, 1000 features were extracted from each image, facilitating subsequent disease classification employing machine learning methods such as SVM [29].

C. *Supervised Machine Learning Algorithm*

The model employed for the classification of Mulberry leaf disease utilizes a standard SVM type with a regularization parameter (C) set to 1.0, balancing the trade-off between maximizing the margin and minimizing classification error. The epsilon parameter (ε) is set to 0.1, controlling the margin around the decision boundary. Employing the Radial Basis Function (RBF) kernel with a formula of $\exp(-\text{auto}|x-y|^2)$, where 'auto' signifies automatic determination of the kernel coefficient, facilitates transforming the input data into a higher-dimensional space for improved classification. A numerical tolerance of 0.0001 dictates the threshold for stopping criteria during optimization, while the iteration limit is capped at 100, ensuring convergence within a specified number of iterations [30]. This SVM model [31, 32] aims to discern an optimal decision boundary for accurately classifying instances of Mulberry leaf disease within the dataset.

Modifying the parameters of the SVM model [33-35], such as the regularization parameter (C) or the kernel function, will inevitably alter the classification results, while employing different datasets with the same parameters may yield varied outcomes due to the inherent characteristics and distributions of the data (Figure 3).

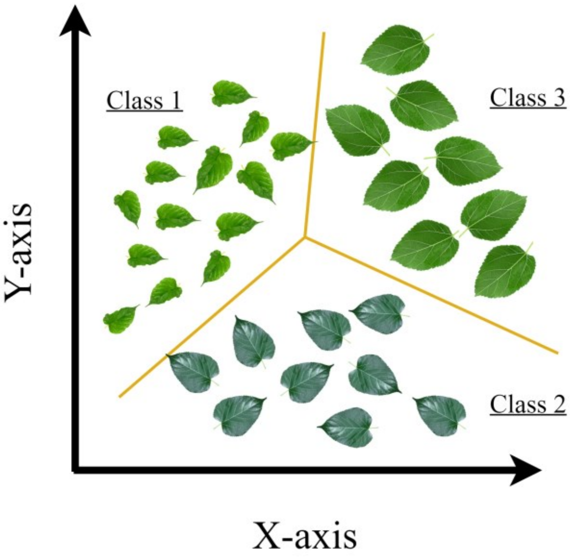


Fig 3. SVM diagram for multiclass classification

D. Confusion Matrix and Other Performance Metrics

In evaluating the classification results of Mulberry leaf disease using SVM, several common performance metrics were employed, including Area Under the Curve (AUC), Classification Accuracy (CA), F1 Score, Precision (Prec), Recall, and Matthews Correlation Coefficient (MCC), alongside the Confusion Matrix. The AUC provides a measure of the classifier's ability to distinguish between classes, with higher values indicating better performance.

Classification Accuracy (CA) measures the proportion of correctly classified instances out of the total. The F1 Score, which combines Precision and Recall, offers a balance between them, with higher values indicating better overall performance. Precision (Prec) quantifies the ratio of correctly predicted positive observations to the total predicted positives, emphasizing the classifier's ability to avoid false positives. Recall assesses the ability of the model to capture all positive instances, calculating the ratio of correctly predicted positive observations to all actual positives. Matthews Correlation Coefficient (MCC) considers true and false positives and negatives, providing a balanced measure even for imbalanced classes.

Lastly, the Confusion Matrix offers a tabular representation of actual versus predicted class labels, facilitating a detailed examination of classification performance across different categories. Together, these metrics offer a comprehensive evaluation of the SVM model's efficacy in classifying Mulberry leaf disease, enabling insights into its strengths and limitations [36-38].

III. RESULTS AND DISCUSSION

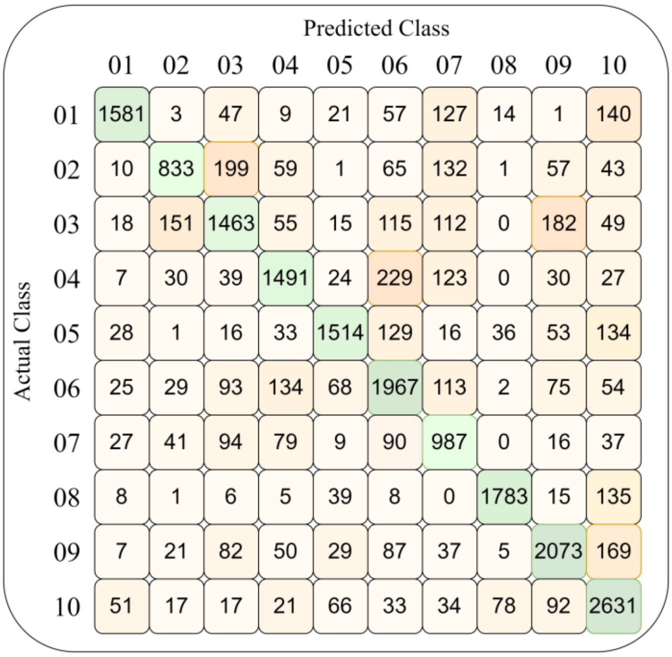
The classification results for Mulberry leaf diseases using the SVM model and a Stratified Shuffle Split sampling technique, comprising 20 random samples with 80% of the data, revealed promising performance metrics. The SVM model achieved an impressive Area Under the Curve (AUC) score of 0.973, indicating excellent discriminative ability between the disease classes. Despite a moderate Classification Accuracy (CA) of 0.775, the model demonstrated balanced performance, as reflected in the F1 score of 0.776, Precision of 0.780, Recall of 0.775, and Matthews Correlation Coefficient (MCC) of 0.749. The results are presented in Table I. These metrics collectively suggest that while the SVM model may not achieve perfect accuracy, it effectively balances the trade-off between correctly classifying instances of Mulberry leaf diseases while minimizing false positives and false negatives.

TABLE I. MULBERRY LEAF CLASSIFICATION EVALUATION RESULTS

Model	SVM
Accuracy (%)	77.5%
Precision	0.780
Recall	*0.775*
F1-Score	*0.776*
AUC	*0.973*
MCC	*0.749*

The results indicate that the SVM model, trained using a Stratified Shuffle Split sampling technique, exhibits robust performance in classifying Mulberry leaf diseases. The high AUC score suggests that the model effectively separates diseased and healthy instances, while the balanced F1 score, Precision, Recall, and MCC indicate its ability to maintain equilibrium between correctly identifying positive cases and avoiding misclassifications. However, the moderate Classification Accuracy implies that there is still room for improvement, possibly through feature engineering or fine-tuning model parameters. Overall, these findings underscore the SVM model's potential as a reliable tool for Mulberry leaf disease classification, laying a solid foundation for further refinement and application in agricultural contexts.

The confusion matrix depicting the classification results is shown in Figure 4.



Predicted Class

Actual Class

	01	02	03	04	05	06	07	08	09	10
01	1581	3	47	9	21	57	127	14	1	140
02	10	833	199	59	1	65	132	1	57	43
03	18	151	1463	55	15	115	112	0	182	49
04	7	30	39	1491	24	229	123	0	30	27
05	28	1	16	33	1514	129	16	36	53	134
06	25	29	93	134	68	1967	113	2	75	54
07	27	41	94	79	9	90	987	0	16	37
08	8	1	6	5	39	8	0	1783	15	135
09	7	21	82	50	29	87	37	5	2073	169
10	51	17	17	21	66	33	34	78	92	2631

Figure 4. Mulberry leaf diseases confusion matrix

IV. CONCLUSIONS

This research presents a rigorous investigation into the classification of Mulberry leaf diseases using Support Vector Machine (SVM) modeling. Through meticulous analysis, our findings underscore the considerable potential of SVM-based methodologies in accurately discerning between diseased Mulberry leaves, thereby facilitating effective disease identification crucial for informed agricultural management decisions. The demonstrated efficacy of SVM highlights its relevance as a sophisticated tool within the realm of agricultural research and practice, offering avenues for improved disease surveillance and mitigation strategies. Moving forward, continued endeavors toward model refinement and the integration of diverse datasets hold promise for advancing the accuracy and robustness of disease classification frameworks. As such, this study contributes valuable insights to the scholarly discourse on leveraging machine learning techniques for agricultural innovation, offering practical implications for sustainable crop protection and food security initiatives.

ACKNOWLEDGMENT

The Selcuk University Scientific Research Coordinator is to be sincerely thanked for their unwavering support and guidance during this research.

DATA AVAILABILITY

The dataset pertinent to this study is accessible through the following hyperlink:
https://data.mendeley.com/datasets/ds45yy9jrc/3

REFERENCES

[1] K. Tutuncu, I. Cinar, R. Kursun, and M. Koklu, "Edible and poisonous mushrooms classification by machine learning algorithms," in *2022 11th Mediterranean Conference on Embedded Computing (MECO)*, 2022: IEEE, pp. 1-4, doi: 10.1109/MECO55406.2022.9797212.

[2] R. Kursun, K. K. Bastas, and M. Koklu, "Segmentation of dry bean (Phaseolus vulgaris L.) leaf disease images with U-Net and classification using deep learning algorithms," *European Food Research and Technology,* vol. 249, no. 10, pp. 2543-2558, 2023, doi: 10.1007/s00217-023-04319-5.

[3] O. Ozaltin, "Date Fruit Classification by Using Image Features Based on Machine Learning Algorithms," *Research in Agricultural Sciences,* vol. 55, no. 1, pp. 26-35, 2024.

[4] N. Kumar *et al.*, "Leafsnap: A computer vision system for automatic plant species identification," in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part II 12*, 2012: Springer, pp. 502-516.

[5] A. Ghosh and P. Roy, "An automated model for leaf image-based plant recognition: an optimal feature-based machine learning approach," *Innovations in Systems and Software Engineering,* pp. 1-14, 2022.

[6] S. S. Chouhan, U. P. Singh, A. Kaul, and S. Jain, "A data repository of leaf images: Practice towards plant conservation with plant pathology," in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, 2019: IEEE, pp. 700-707.

[7] S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain. *A Database of Leaf Images: Practice towards Plant Conservation with Plant Pathology*, doi: 10.17632/hb74ynkjcn.5.

[8] M. N. Ornek and H. Kahramanlı Ornek, "Developing a deep neural network model for predicting carrots volume," *Journal of Food Measurement and Characterization,* vol. 15, no. 4, pp. 3471-3479, 2021.

[9] Y. R. Erkan and H. KAHRAMANLI Ornek, "Mushroom species detection using image processing techniques," *International Journal of Engineering and Innovative Research,* vol. 1, no. 2, pp. 71-83, 2019.

[10] E. Ropelewska, K. Sabanci, and M. F. Aslan, "Distinguishing Pickled and Fresh Cucumber Slices Using Digital Image Processing and Machine Learning," in *Biology and Life Sciences Forum*, 2022, vol. 16, no. 1: MDPI, p. 1, doi: 10.3390/IECHo2022-12477.

[11] E. Ropelewska, W. Popińska, K. Sabanci, and M. F. Aslan, "Flesh of pumpkin from ecological farming as part of fruit suitable for non-destructive cultivar classification using computer vision," *European Food Research and Technology,* vol. 248, no. 3, pp. 893-898, 2022, doi: 10.1007/s00217-021-03935-3.

[12] B. Gencturk *et al.*, "Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models," *European Food Research and Technology,* vol. 250, no. 1, pp. 97-110, 2024, doi: 10.1007/s00217-023-04369-9.

[13] H. Isik *et al.*, "Maize seeds forecasting with hybrid directional and bi‐directional long short‐term memory models," *Food Science & Nutrition,* vol. 12, no. 2, pp. 786-803, 2024, doi: 10.1002/fsn3.3783.

[14] I. Cinar, "Detection of Chicken Diseases from Fecal Images with the Pre-Trained Places365-GoogLeNet Model," in *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 2023, vol. 1: IEEE, pp. 752-758, doi: 10.1109/IDAACS58523.2023.10348829.

[15] J. Hang, D. Zhang, P. Chen, J. Zhang, and B. Wang, "Classification of plant leaf diseases based on improved convolutional neural network," *Sensors,* vol. 19, no. 19, p. 4161, 2019.

[16] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Computers & Electrical Engineering,* vol. 76, pp. 323-338, 2019.

[17] R. Kursun, E. T. Yasin, and M. Koklu, "The Effectiveness of Deep Learning Methods on Groundnut Disease Detection," doi: 10.58190/icat.2023.11.

[18] R. Kursun, E. T. Yasin, and M. Koklu, "Machine Learning-Based Classification of Infected Date Palm Leaves Caused by Dubas Insects: A Comparative Analysis of Feature Extraction Methods and Classification Algorithms," in *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2023: IEEE, pp. 1-6, doi: 10.1109/ASYU58738.2023.10296641.

[19] A. Yasar, A. Golcuk, and O. F. Sari, "Classification of bread wheat varieties with a combination of deep learning approach," *European Food Research and Technology,* vol. 250, no. 1, pp. 181-189, 2024, doi: 10.1007/s00217-023-04375-x.

[20] T. Chompookham and O. Surinta, "Ensemble methods with deep convolutional neural networks for plant leaf recognition," *ICIC Express Letters,* vol. 15, no. 6, pp. 553-565, 2021.

[21] P. Enkvetchakul and O. Surinta, "Effective data augmentation and training techniques for improving deep learning in plant leaf disease recognition," *Applied Science and Engineering Progress,* vol. 15, no. 3, p. 3810, 2021.

[22] E. Yasin and M. KOKLU, "Utilizing Random Forests for the Classification of Pudina Leaves through Feature Extraction with InceptionV3 and VGG19," in *Proceedings of the International Conference on New Trends in Applied Sciences*, 2023, vol. 1, pp. 1-8.

[23] E. T. Yasin, R. Kursun, and M. Koklu, "Deep learning-based classification of black gram plant leaf diseases: A comparative study."

[24] B. Pakruddin and R. Hemavathy, "A Comprehensive Standardized Dataset of Numerous Pomegranate Fruit Diseases for Deep Learning," *Data in Brief,* p. 110284, 2024.

[25] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE transactions on geoscience and remote sensing,* vol. 54, no. 10, pp. 6232-6251, 2016, doi: 10.1109/TGRS.2016.2584107.

[26] t. Chom and o. surinta. *Mulberry Leaf Dataset*, doi: 10.17632/ds45yy9jrc.3.

[27] I. Cinar and M. Koklu, "Determination of effective and specific physical features of rice varieties by computer vision in exterior quality inspection," *Selcuk Journal of Agriculture and Food Sciences,* vol. 35, no. 3, pp. 229-243, 2021.

[28] T. Choompookham, E. Okafor, and O. Surinta, "Mulberry leaf dataset for image classification task," *Data in Brief,* p. 110281, 2024, doi: 10.1016/j.dib.2024.110281.

[29] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," *arXiv preprint arXiv:1602.07360,* 2016.

[30] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd international conference on computing for sustainable global development (INDIACom)*, 2016: Ieee, pp. 1310-1315.

[31] I. Cinar, "Disaster Detection Using Machine Learning Methods with Deep Features," *disasters,* vol. 19, p. 21, 2023.

[32] A. Yasar, "Analysis of selected deep features with CNN-SVM-based for bread wheat seed classification," *European Food Research and Technology,* pp. 1-11, 2024, doi: 10.1007/s00217-024-04488-x.

[33] E. T. Yasin, I. A. Ozkan, and M. Koklu, "Detection of fish freshness using artificial intelligence methods," *European Food Research and Technology,* vol. 249, no. 8, pp. 1979-1990, 2023, doi: 10.1007/s00217-023-04271-4.

[34] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications,* vol. 13, no. 4, pp. 18-28, 1998, doi: 10.1109/5254.708428.

[35] D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine learning*: Elsevier, 2020, pp. 101-121.

[36] K. Sabanci, M. Koklu, and M. F. Unlersen, "Classification of Siirt and long type pistachios (pistacia vera l.) by artificial neural networks," *International Journal of Intelligent Systems and Applications in Engineering,* vol. 3, no. 2, pp. 86-89, 2015, doi: 10.18201/ijisae.74573.

[37] I. Ozkan, M. Koklu, and R. Saraçoğlu, "Classification of pistachio species using improved k-NN classifier," *Health,* vol. 23, p. e2021044, 2021, doi: 10.23751/pn.v23i2.9686.

[38] M. Koklu and K. Tutuncu, "Tree based classification methods for occupancy detection," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 675, no. 1: IOP Publishing, p. 012032, doi: 10.1088/1757-899X/675/1/012032.