Research Article / Systems and Computer Engineering
http://doi.org/10.15446/ing.investig.118469

# A Hybrid Voting Ensemble Model for the Efficient Sorting and Classification of Date Fruit Varieties

## Un modelo híbrido de votación por conjunto para la clasificación y ordenamiento eficiente de variedades de dátiles

Sofiane Abden [1], Mostefa Bendjima [2], and Soumia Benkrama [3]

**ABSTRACT**

Dates constitute one of Algeria's most important agricultural products, given their substantial health and economic advantages. Furthermore, they represent a vital export item beyond the hydrocarbon industry. The existing conventional techniques for classifying and sorting dates are ineffective, time-consuming, and labor-intensive, leading to a discrepancy between restricted exports and elevated production levels. This work presents an ensemble learning (EL) model that utilizes transfer learning (TL) strategies to overcome obstacles and improve date fruit classification. We assess the efficacy of four classifiers, *i.e.*, MobileNetV2, EfficientNet, DenseNet201, and an ensemble soft voting classifier that employs TL, with a dataset of 1619 photos representing 20 distinct varieties of Algerian dates. The dataset used is one of the greatest benchmarks for varietal diversity. The suggested hybrid model exhibits exceptional performance, with a validation accuracy of 99.07% and a classification accuracy of 99.93%. It establishes a new benchmark in agricultural technology by exceeding all assessed models in terms of precision, recall, and F1-score. These findings demonstrate the potential of this approach to completely transform date sorting and markedly improve agricultural production and efficiency.

**Keywords:** artificial intelligence, classification, machine learning, smart agriculture, date fruits

**RESUMEN**

Los dátiles constituyen uno de los productos agrícolas más importantes de Argelia debido a sus significativos beneficios para la salud y la economía. Además, representan un artículo de exportación vital más allá de la industria de los hidrocarburos. Las técnicas convencionales que existen para clasificar y ordenar dátiles son ineficaces, consumen mucho tiempo y requieren mucha mano de obra, lo que genera una discrepancia entre las exportaciones limitadas y los elevados niveles de producción. Este trabajo presenta un modelo de aprendizaje en conjunto (EL) que utiliza estrategias de aprendizaje por transferencia (TL) para superar obstáculos y mejorar la clasificación de frutos de dátiles. Evaluamos la eficacia de cuatro clasificadores, *i.e.*, MobileNetV2, EfficientNet y DenseNet201, y un clasificador de votación suave en conjunto que emplea TL, utilizando un conjunto de datos de 1619 imágenes que representan 20 variedades distintas de dátiles argelinos. El conjunto de datos utilizado es uno de los mayores referentes en diversidad varietal. El modelo híbrido propuesto presenta un rendimiento excepcional, con una precisión de validación del 99.07 % y una precisión de clasificación del 99.93 %, y establece un nuevo referente en la tecnología agrícola al superar a todos los modelos evaluados en términos de precisión, sensibilidad y puntuación F1. Estos hallazgos demuestran el potencial de este enfoque para transformar por completo la clasificación de dátiles y mejorar notablemente la producción y eficiencia agrícola.

**Palabras clave:** inteligencia artificial, clasificación, aprendizaje automático, agricultura inteligente, dátiles

**Received:** January 23rd, 2025
**Accepted:** May 14th, 2025

## Introduction

Agriculture is fundamental to economic development and essential for sustaining the increasing global population [1]. Plants are crucial to ecosystems and human existence since they serve as the major source of food and provide essential resources such as oxygen and medicine. Nevertheless, the agricultural industry faces considerable hurdles [2], including satisfying increasing food demands and maintaining crop health in the face of evolving environmental conditions. Agricultural innovation is crucial for enhancing food security and economic stability, especially when it comes to improving crop quality and resilience through adequate plant health management [3].

The date palm (*Phoenix dactylifera* L.), one of the world's oldest cultivated fruit trees, originating from the region of Mesopotamia, has been grown since ancient times. With a robust root system that can reach considerable depths in well-drained, sandy soils, this tree thrives in the hot, arid climates of the Middle East and North Africa [4]. Date palms are an essential crop in these regions, providing a vital source of food, income, and employment to millions. They hold not only economic value but also significant cultural and religious importance, making them a cornerstone of agricultural livelihoods in arid and semi-arid environments [5].

[1] Department of Mathematics and Computer Science, Faculty of Exact Sciences, Laboratory of Information Processing and Telecommunications (LTIT), Tahri Mohammed University of Bechar, 08000, Algeria. Affiliation: PhD student. Email: abdan.soufyane@univ-bechar.dz
[2] Department of Mathematics and Computer Science, Faculty of Exact Sciences, Laboratory of Information Processing and Telecommunications (LTIT), Tahri Mohammed University of Bechar, 08000, Algeria. Affiliation: Associate professor. Email: bendjima.mostefa@univ-bechar.dz
[3] Department of Mathematics and Computer Science, Faculty of Exact Sciences, Laboratory of Information Processing and Telecommunications (LTIT), Tahri Mohammed University of Bechar, 08000, Algeria. Affiliation: associate professor. Email: benkrama.soumia@univ-bechar.dz

Date palm fruits are a nutritious and affordable source of essential nutrients, providing numerous macro- and micro-elements. Rich in carbohydrates, particularly soluble sugars and dietary fiber [6], dates are an excellent source of energy. In addition to their high sugar content, they contain vital minerals, vitamins, and antioxidants, which contribute to their health benefits, such as supporting heart health and reducing the risk of diseases like cancer. This nutrient-dense profile makes dates an ideal, natural food with significant health-promoting properties [7].

Date palms hold great socioeconomic and environmental importance in many countries (*e.g.*, Algeria). They ranked first in Saharan agriculture, with more than 17 million trees planted and more than 800 varieties, such as *Deglet Nour* (considered to be the best and most famous variety worldwide), *feggus*, *aadam*, *lanjouhar*, and *cherka etc.* Algeria occupies an important place among the date-producing and exporting countries [8]. However, there is a considerable gap between its production and exports due to its reliance on traditional production and sorting methods.

Sorting and classifying date palm fruits is traditionally done manually, a process that is labor-intensive, time-consuming, and costly, often requiring skilled workers. To address these challenges, there is a growing need for automated systems that can perform these tasks efficiently and accurately [6]. As a result, many studies are exploring intelligent methods for the rapid and fully automated sorting of date palm fruits, aiming to improve speed and accuracy and reduce dependence on manual labor.

Advancements in AI-driven image classification have enabled a highly precise, cost-effective, and rapid evaluation of date palm fruits, even under challenging environmental conditions [3]. Consequently, industries are increasingly adopting computer vision techniques to automatically grade and sort dates based on features such as color, texture, and size, tasks that were previously performed manually. With numerous types of date fruits displaying subtle differences in color, shape, and fleshiness, quality and price are often assessed through physical attributes, while nutritional value is evaluated based on chemical composition and sensory qualities, largely influenced by the fruit's variety and ripeness [9].

The goal of this study was to leverage machine learning (ML), particularly TL and EL, to enhance agriculture by using images of date palm fruits to identify and sort twenty different local date varieties (*Ajina*, *Deglet*, *Tanslit*, *Hamraya*, *etc.*). This classification was achieved by extracting features such as color, size, and shape from the images, facilitating an efficient and precise sorting. Unlike previous works, which largely focused on individual ML or TL models using public datasets from other countries, this study addressed a gap by introducing EL, combining CNNs pretrained on a local dataset. Several challenges were encountered during this study, including the large number of species used and the great similarity between some of them. Moreover, the scarcity of information on palm diseases compared to those of other plants is likely due to their occurrence in specific geographical areas. The main contributions of this paper include the application of a hybrid method combining TL on MobileNet, EfficientNet, and DenseNet201, as well as our proposal of EL on the same TL models. The performance of our EL approach, based on a voting meta-classifier, was compared against the TL results using data collected from local markets of the Touggourt region, located in the south of Algeria.

## Related works

With a view to improve efficiency in the food and agricultural sector, this section describes previous studies that have used AI approaches, specifically ML and deep learning (DL), to automate the categorization and sorting of date fruit species by means of images.

The authors of [7] suggested using supervised and unsupervised DL networks to classify different types of date fruit. They achieved effective feature fusion and dimensionality reduction by fusing features from a CNN (VGG-F) and PCANet using discriminant correlation analysis (DCA). The accuracy of their method was 99.32% with VGG-F (CCA) and 98.20% with VGG-F (DCA). To test the system, the authors presented one of the largest benchmark datasets used to this effect, consisting of 20 date variants. The findings support the usefulness of fused features for better categorization, as well as the efficacy of DCA.

By applying a CNN in MATLAB R2015 to three different types of dates (Aseel, Karbalain, and Kupro), [10] developed a system for date recognition based on color, shape, and size feature datasets of collected photos, achieving a 97.2% accuracy.

[11] collected dataset a containing 3228 photos corresponding to 27 date classes, reporting their experimental results based on five stages. ML techniques were used in the first step, followed by a deep TL model utilizing DenseNet in the second stage, a number tree in the fourth stage, and fine-tuning to produce the model's optimal classification configurations. Regularization was applied in the fifth stage. The best test accuracy was 95.21%, and the validation accuracy reached 97.21%.

For automatic categorization and qualitative comparison, [12] employed three distinct ML techniques: k-nearest neighbors (KNNs), support vector machines (SVMs), and artificial neural networks (ANNs). Concerning the six most popular date fruit types in Oman, the combination of color, shape, and size factors helped to reach the best accuracy. The ANN classifier exhibited a maximum classification accuracy of 99.2% according to trial results.

The you-only-look-once (YOLO) technique was used in [13] to identify and categorize date fruits using DL. Three YOLO models, namely YOLOv5, YOLOv7, and YOLOv8, were trained using 1735 photos of nine different kinds of date fruit. Metrics including the F1-score, recall, and precision were used to assess performance. With a mean recall of 99,0%, a precision of 99.1%, and a mean average precision (mAP) of 99.4%, the results show that YOLOv8 attained great accuracy, implying that it efficiently recognizes and categorizes date fruits according to their surface quality, assisting in productivity optimization by considering uneven ripening over several harvests.

[14] classified seven date fruit types using three ML approaches. From 898 images, 34 features related to morphology, shape, and color were extracted. Logistic regression (LR) and ANN models achieved accuracies of 91.0 and 92.2%, respectively, while a stacked model

combining both increased accuracy to 92.8%. These results demonstrate the effectiveness of ML for date fruit classification.

In order to train a MobileNetV2-based model, [15] created a dataset with eight different types of date fruit. To improve accuracy, a number of pre-processing methods were used, including hybrid weight adjustment, decaying learning rates, picture augmentation, and model checkpointing. With a high classification accuracy of 99.0%, the MobileNetV2 model outperformed others using AlexNet, VGG16, InceptionV3, and ResNet, proving its greater efficacy for classifying date fruits.

In order to improve the model performance, [16] applied several GAN designs for both classification and augmentation, increased the number of training epochs, and broadened the dataset. A CycleGAN-augmented dataset combined with the ResNet152V2 model exhibited a maximum classification accuracy of 96.8%, while a CNN model reached 94.3%. ResNet152V2 and CNNs, on the other hand, achieved 83.0 and 75.0% accuracy, respectively, when using the original dataset.

The aforementioned studies mainly focus on ML and DL-based classification of date fruits utilizing image features. However, novel metaheuristic methods like ant colony optimization (ACO) and the whale optimization algorithm (WOA) have shown considerable promise in enhancing classification efficacy via automated feature selection and model optimization. These techniques, especially when combined with neural networks, have been utilized to optimize parameters like weights and biases, resulting in improved diagnosis accuracy in medical picture classification [17].

Considering the parallels in visual pattern recognition between medical imaging and agricultural classification, future research could investigate the utilization of bio-inspired optimization techniques to enhance EL efficacy, parameter tuning, and feature resilience in date fruit classification.

## Materials and methods

Google Colab was employed to perform date fruit classification studies via a soft ensemble technique. Utilizing its GPU and TPU capacities, together with up to 12 GB of RAM, Colab facilitated rapid model training and analysis. The dataset, located on Google Drive, was accessed within the Colab environment. Algorithms were formulated via libraries including Scikit-learn, NumPy, Pandas, Matplotlib, and Seaborn. The studies were conducted using an HP EliteBook G8 laptop equipped with an Intel Core i5-1165G7 CPU (2.6 GHz, four cores), 16 GB of RAM, and 256 GB of storage. A Python script was utilized to load pre-trained models, process the data, train each model, and merge the predictions through a soft ensemble method to assess performance.

Fig. 1 depicts a comprehensive hybrid approach using a voting classifier that incorporates three TL models: MobileNetV2, EfficientNet-B2, and DenseNet201. This method collects features including color, shape, and size from images to identify date fruit varieties. Its efficacy was evaluated against individual TL models alongside a test dataset from Algeria.

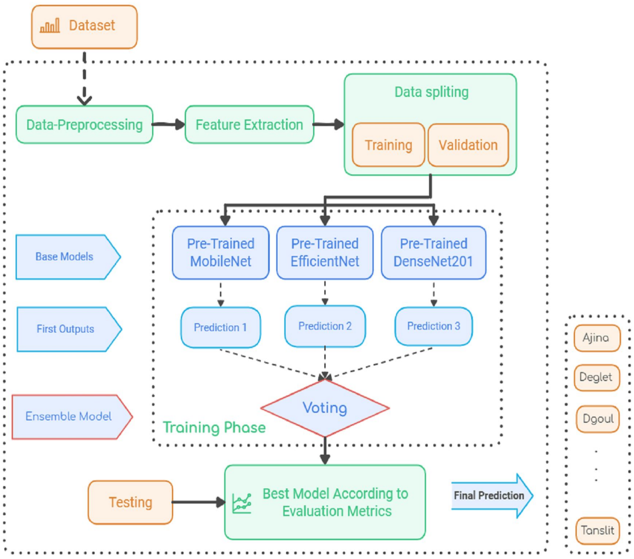


**Figure 1.** Flowchart of the proposed model
**Source:** Authors

### *Dataset description*

The model presented herein was trained and assessed using a public dataset from Kaggle [7], which comprises 1619 images from 20 distinct Algerian date fruit varieties, including Adam Deglet Nour, Ajina, Ghars, Bayd Hmam, Bouaarous, Deglet, Deglabayda, Degletkahla, Degletghabia, Dfarlgat, Dgoul, Litima, Loullou, Hamraya, Tarmount, Tanslit, Tantbucht, Techbeh Tati, Tinisin, and Tivisyaouin. These samples, sourced and captured from local markets in the Touggourt region in the southeast of Algeria, are presented in Table I, detailing the distribution across varieties. For image-based identification tasks, particularly during the training and evaluation phases, large datasets are essential to improving the accuracy of classification algorithms. This dataset was divided into training and validation sets to enhance the robustness of the model's evaluation. Each variety in the dataset exhibits distinct characteristics, such as shape, size, color, and hardness, further underscoring the need for a comprehensive dataset to accurately assess model performance.

The validation set comprised 20% of the complete date fruit dataset, whereas the training set included the remaining 80%. Each set contained 20 date fruit varieties. Examples of each category are presented in Fig. 2.

**Table I.** Distribution of date fruit samples across different varieties in the dataset

| Date variety | Number of samples |
|---|---|
| Adam Deglet Nour | 86 |
| Ajina | 85 |
| Ghars | 88 |
| Bayd Hmam | 87 |
| Bouaarous | 82 |
| Deglet | 38 |
| Degla Bayda | 95 |
| Deglet Kahla | 85 |
| Deglet Ghabia | 35 |
| Dfar Lgat | 86 |
| Dgoul | 103 |
| Litima | 85 |
| Loullou | 81 |
| Hamraya | 76 |
| Tarmount | 83 |
| Tanslit | 85 |
| Tantbucht | 76 |
| Techbeh Tati | 88 |
| Tinisin | 88 |
| Tivisyaouin | 87 |

**Source:** Authors

**Figure 2.** Date varieties, arranged from top to bottom and left to right: Adam Deglet Nour, Ajina, Ghars, BaydHmam, Bouaarous, Deglet, Deglabayda, Degletkahla, Degletghabia, Dfarlgat, Dgoul, Litima, Loullou, Hamraya, Tarmount, Tanslit, Tantbucht, Techbeh Tati, Tinisin, and Tivisyaouin
**Source:** Authors

## *Transfer learning (TL)*

An efficient ML technique called *TL* uses the knowledge gained from training a model on one task to improve performance on a different yet related one. For applications using little data, like plant disease classification, this makes it easier to reuse pre-trained models such as those developed on large datasets like ImageNet. By utilizing established patterns and information already in the pre-trained model, this technique reduces generalization error and training times. By sharing this knowledge, TL makes it possible to train DL models quickly and accurately, even with small datasets, which is typically an issue for DL techniques that require large amounts of data.

For prediction, feature extraction, and fine-tuning purposes, this study used sophisticated pre-trained CNN models, including MobileNet, EfficientNet, and DenseNet201. These techniques, which will be further discussed in this document, were combined into ensembles to increase classification accuracy by utilizing the strengths of several models [18]. Each model was trained for 15 epochs with consistent parameters, including a dense layer with 512 units, ReLU activation, an input shape of 224 x 224 x 3, and a dropout rate of 0.5 to avoid overfitting. To preserve the learned features, the initial ten layers of each model were frozen. After the training and validation stage, these TL models, which are detailed in this section, were tested on our local dataset.

**MobileNetV2.** In 2017, Google introduced MobileNet, a DL-based framework with an emphasis on accuracy, efficiency, and small size for mobile devices. Building upon this framework, MobileNetV2, which was put forth in 2018, improved the architecture even further with a more effective layout tailored to mobile apps. In order to improve feature extraction, MobileNetV2 uses an inverted residual structure, with residual connections between bottleneck layers, in contrast to CNNs. By separating spatial and depth operations, its fundamental building element, the depth-wise separable convolution, lowers computational complexity and increases efficiency without compromising performance. A 32-filter convolutional layer is the first of 19 bottleneck layers in the design. Depth-wise and point-wise convolution layers are then added to capture nonlinearity and select features [19]. MobileNetV2 is especially well suited for resource constrained contexts because it offers greater accuracy with fewer parameters than its predecessor. In order to maximize speed and accuracy in mobile visual recognition tasks, the model's backbone consists of an average pooling layer, a conventional $3 \times 3$ convolution, and 17 inverted residual bottlenecks [20].

**EfficientNet-B2.** The EfficientNet family of DL network designs strike a balance between computing efficiency and accuracy to optimize picture classification. EfficientNet, created by in 2019, offers a compound scaling technique that balances computational cost and performance by consistently scaling the model's depth, width, and resolution. EfficientNet uses fewer parameters while achieving great accuracy in comparison with models such as ResNet and Inception.

The EfficientNet family comprises versions B0 through B7. Each model version retains effective resource use thanks to the compound scaling coefficients, which are found using a grid search. Because of this, EfficientNet is perfect for implementation on devices with limited resources, including smartphones and edge platforms, where it reliably produces high accuracy with low processing requirements.

Interestingly, EfficientNet's initial three iterations use input sizes of 224 x 224, ensuring interoperability with the other models in our ensemble approach. The main factor in our decision to use EfficientNet-B2 was its compatibility. EfficientNet has established itself as a leading option for effective DL applications by demonstrating good performance across image classification benchmarks such as ImageNet [21].

**DenseNet201.** As a member of the DenseNet family, this technique is an advanced DL architecture that uses its densely linked structure to address the vanishing gradient problem. DenseNet-201 has 201 levels and uses dense blocks, wherein each layer gives its outputs to later layers and receives inputs from all layers before it. The overall performance of the model is improved by this distinctive architecture, which encourages effective feature reuse and smooth gradient propagation.

In a variety of computer vision applications, such as semantic segmentation, object detection, and picture classification, DenseNet-201 has shown remarkable performance. This network is quite effective for complicated visual issues because of its deep architecture, which allows recording complex representations. For a variety of applications, academics and practitioners continue to favor DenseNet-201 because of its exceptional performance and versatility [22].

### *Ensemble learning (EL)*

A powerful ML method termed *ensemble modeling* blends several models to improve generalization and prediction accuracy. EL performs better than single models by means of fusion techniques to correct individual model faults. Three TL models, MobileNetV2, EfficientNet-B2, and DenseNet201, were used in this work to categorize Algerian date fruit types. A voting classifier was used to obtain the final predictions [23]. Google Colab was used to train each model on the same dataset, and performance metrics like recall, accuracy, precision, and the F1-score were used to conduct assessments. Models with superior validation performance were given more weight in weighted soft voting, ensuring that they had a greater impact on the ensemble's final predictions.

In order to improve classification accuracy, EL uses a weighted soft voting process to aggregate class probabilities from the base models, leveraging their advantages while mitigating their disadvantages. This method improves performance in sorting date fruits by reducing the impact of overfitting and making use of the varied knowledge representations of pre-trained models. Through the combination of the results of several models, EL demonstrates that it can handle tough classification problems by making the results more accurate and reliable [24].

### *Evaluation metrics*

Accuracy, precision, recall, the F1-score, and confusion matrices were used to assess the effectiveness of ML models for date fruit categorization. Using true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), *accuracy* is the percentage of correctly identified samples. *Recall*, also called *sensitivity*, is the percentage of correctly identified real positives (TP) compared to false negatives (FN). *Precision*, on the other hand, looks at how reliable positive predictions are by focusing on TP and FP. Finally, the *F1-score* is a harmonic mean of precision and recall that strikes a balance between the two metrics. The following mathematical formulas were used to measure accuracy, precision, recall, and F1-score:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

These metrics, as defined in Eqs. (1)-(4), provide a thorough examination of the models' classification performance. Furthermore, the classification results were interpreted and visualized in terms of TP, TN, FP, and FN using confusion matrices [21].

## Results and discussion

This section shows a detailed comparison of the proposed EL model's performance *vs.* that of other TL models using a local Algerian dataset. It also analyzes the training and validation curves of all models and evaluates additional metrics, such as precision, the F1-score, and recall.

This study utilized multiple ML models to sort and categorize various types of date fruit, showcasing the potential for automated sorting. To this effect, EL and three TL models were used on a dataset with 1619 photos divided into 20 classes. There is a notable paucity of research focusing on the application of EL to local date information, which is crucial for enhancing accuracy in real-world applications. By merging multiple TL models to enhance date fruit classification, this work fills this research gap. TL approaches employing pre-trained models, including EfficientNet-B2, MobileNet, and DenseNet201, were examined for categorizing date fruit varieties. Among these, the MobileNet model displayed remarkable efficiency during training, reaching a training accuracy of 98.46% and a validation accuracy of 88.54%. However, its performance on the local dataset was unsatisfactory, with a testing accuracy of 89%, the lowest among the models assessed. DenseNet201 achieved remarkable results, with training and validation accuracies of 99.81 and 98.14%, respectively, in addition to a testing accuracy of 98%. Similarly, the EfficientNet design performed well, obtaining training, validation, and testing accuracies of (97.98), (89.78), and (90%), respectively.

In this work, the EL technique outperformed all individual models. Utilizing a weighted soft voting classifier, EL integrated the predictions of the TL models while allocating weights depending on each model's performance and variables collected from the date images, such as size, shape, and color. This technique obtained the maximum performance, with training and validation accuracies of 99,93% and 99.07%, respectively, after 15 epochs. Fig. 3 illustrates the training and validation accuracies of all models, whereas Fig. 4 provides a comprehensive analysis of performance measures, including precision, recall, the F1-score, and accuracy. This comparative phase is necessary to identify the optimal forecasting models based on their evaluation metrics.

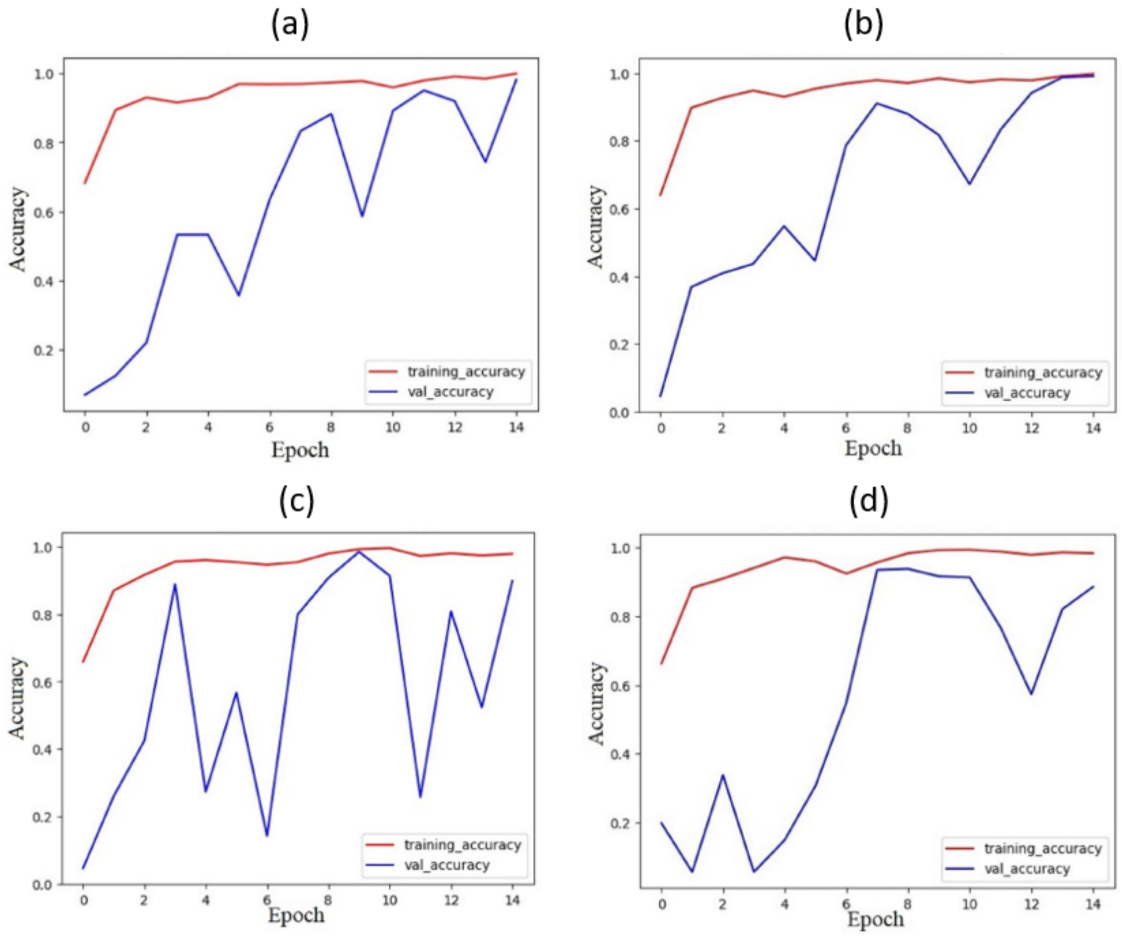


**Figure 3.** Training and validation accuracy: (a) EL, (b) EfficientNet, (c) DenseNet, (d) MobileNet
**Source:** Authors

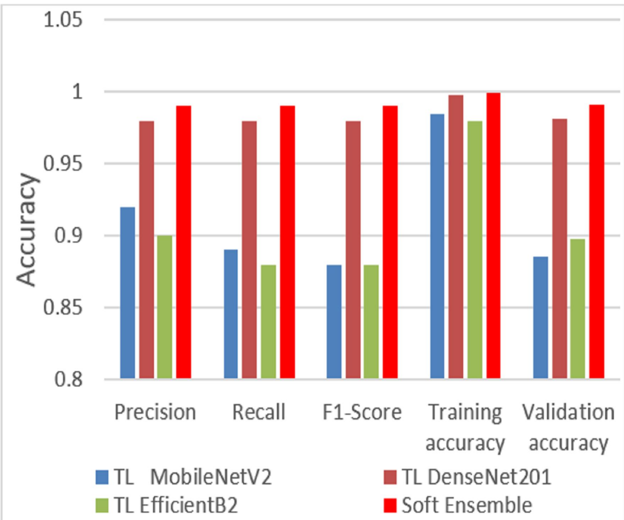


**Figure 4.** Model performance comparison in terms of evaluation metrics
**Source:** Authors

The EL model exhibited outstanding performance, attaining the highest testing accuracy (99%) among all experimental models. This outcome underscores its efficacy in precisely selecting and classifying local date fruit varieties. Fig. 5 displays the results of the confusion matrices for the test dataset, which includes 20 different date types, such as Ajina, Ghars, Deglet, Tanslit, and others. These results further underscore the efficacy of our approach in managing intricate categorization tasks across several date fruit classifications.

Our approach exhibited a remarkable accuracy of 99.93%, surpassing previous research, as illustrated in Table II, while effectively tackling the complexities associated with a local dataset of 20 varieties, unlike some studies that depended on a limited number of species [10, 16]. Furthermore, only a small number of research studies, such as [14], have used EL stacking, but their accuracy ranges from 91.0 to 92.8%. Furthermore, [11], which included TL and covered a larger total of 27 classes, achieved an accuracy of 97.21%, lower than our model. The study most comparable to ours regarding the database is [7], despite minor discrepancies in the conclusions while employing significantly different methodologies. The implementation of weighted soft voting and the foundational pre-trained models in our EL model differentiates our research from others, constituting an innovative approach to harnessing the advantages of several selected TL models for enhanced outcomes. It offers a more resilient solution, improving overall performance while mitigating the danger of overfitting and misclassification. These findings corroborate our notion that EL will surpass individual models, especially when utilized on local datasets containing similar date varieties. The results demonstrate that incorporating several models for feature extraction can markedly enhance classification outcomes.

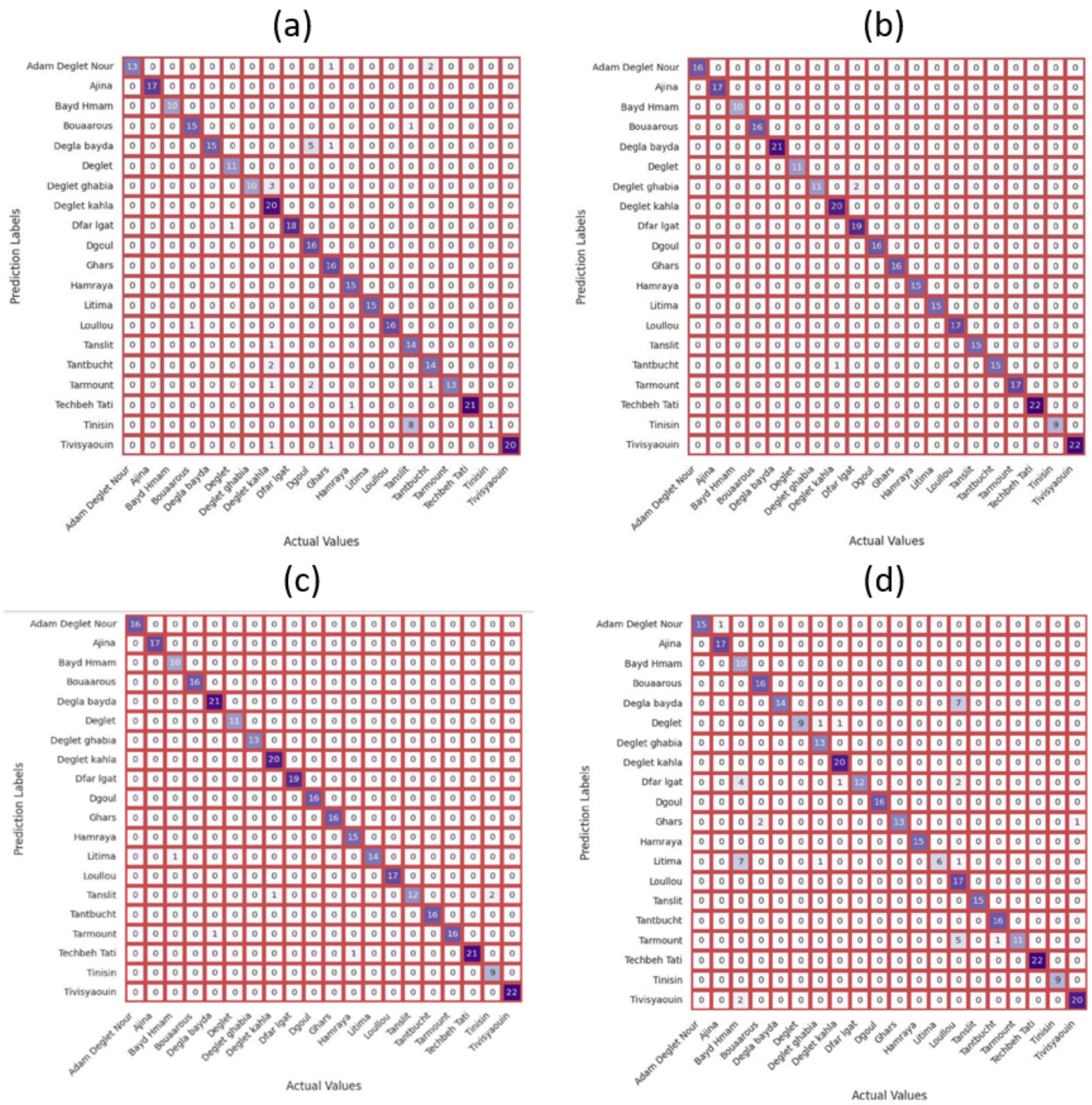

**Figure 5.** Confusion matrices of the models: (a) EL, (b) EfficientNet, (c) DenseNet, (d) MobileNet
**Source:** Authors

This work's ramifications go beyond the particular challenge of classifying date fruits. The EL model's effectiveness shows how it can be applied to various agricultural problems, providing opportunities to enhance crop management techniques, including date sorting and quality evaluation.

Future studies should concentrate on creating mobile applications for real-time classification, so that farmers can quickly implement these cutting-edge methods. Furthermore, the date industry's efficiency might be greatly improved by incorporating this model into automated sorting systems, which would increase productivity and expand export opportunities. These developments would allow the public and even novice farmers to recognize date varieties, evaluate their quality, and accurately determine their price or market value. Moreover, investigating alternative learning paradigms like self-supervised learning (SSL) could diminish the reliance on extensive labeled datasets, facilitating more scalable and adaptive classification systems in various agricultural contexts.

## Conclusions

In this study, we proposed an EL model capable of classifying 20 popular varieties of date fruits in Algeria. Additionally, three TL techniques were evaluated for the dataset, *i.e.*, TL with EfficientNetB2, MobileNetV2, and Densenet201, in addition to a soft voting EL approach. The results demonstrate the potential of these models to assist farmers in effectively classifying date varieties, contributing to improved agricultural practices, enhanced marketing, and sustainable date production.

Among the TL models, EfficientNetB2 achieved the lowest accuracy (97.98%), while DenseNet201 and MobileNet exhibited the highest values (99.81 and 98.46%, respectively). The EL model, leveraging a soft voting classifier, outperformed all individual models, achieving a remarkable accuracy of 99.93%. This highlights the effectiveness of combining multiple high-performance TL models to enhance classification accuracy.

**Table II.** Accuracy comparison of the proposed approach *vs.* previous literature

| Ref | Number of images/classes | Dataset source | Techniques | Best accuracy (%) |
|---|---|---|---|---|
| [7] | 1619 images, 20 varieties | Manually collected in Touggourt, Algeria | Supervised (VGG-F) | 98.20 |
| | | | Unsupervised (PCANet) | 99.32 |
| [10] | 500 images, 3 varieties | Manually collected, Pakistan | CNN | 97.20<br>95.21 |
| [11] | 3228 images, 27 varieties | Farms and shops | ML, TL (DenseNet), Fine-tuned | 97.21 |
| [12] | 6 varieties | AL-Dhahirah, Oman | ANN, SVM, KNN | 99.20 |
| [13] | 1735 images, 9 varieties | Kaggle (Saudi dataset) | YOLOv5<br>YOLOv7 | 98.50<br>99.50 |
| [14] | 898 images, 7 varieties | Kaggle (combined dataset) | LR<br>ANN<br>Stacking | 91.00<br>92.20<br>92.80 |
| [15] | 1750 images, 8 varieties | Saudi Arabia (manual) | MobileNetV2 | 99.00 |
| [16] | 628 images, 3 varieties | Online (augmented) | ResNet152V2<br>CNN | 96.80<br>94.30 |
| Our EL Model | 20 varieties (Algeria) | From [7] | DenseNet201, EfficientNet-B2, MobileNetV2 | 99.93 |
| | | | | 99.07 |

**Source:** Authors

The findings align with and extend previous research, showcasing the value of EL in agricultural applications and its ability to address challenges such as limited local datasets and the similarity between some varieties. Our EL model, applied to a local database, represents a significant advancement in date fruit classification. It demonstrates the potential for integrating computer vision systems into agricultural processes such as automatic date sorting in packaging factories. Furthermore, the proposed approach can contribute to more efficient management in the agricultural sector.

Future research ought to concentrate on enhancing these models and creating practical instruments, such as mobile applications, for the real-time classification of date fruits. Furthermore, investigating alternative methods would enhance the applicability of the method across various agricultural environments. We recommend the inclusion of developing technologies such as generative adversarial networks (GANs) and transformers. Furthermore, SSL signifies a promising avenue, particularly in contexts with a scarcity of labeled data. Utilizing SSL techniques for preliminary tasks like image inpainting, contrastive learning, or clustering-based representation learning allows models to develop robust feature representations before fine-tuning, thereby diminishing reliance on extensive manual annotation and enhancing scalability in agricultural machine vision systems.

## CRediT author statement

**Sofiane Abden:** Conceptualization, methodology, software, data collection, writing (original draft, review, and editing), and formal analysis. **Mostefa Bendjima:** investigation, validation, supervision, project administration, writing (review and editing), and approval of the final version. **Soumia Benkrama:** investigation, validation, supervision, project administration, and writing (review and editing).

## References


[1] S. Abden, M. Bendjima, and S. Benkrama, "Deep learning model based on VGG16 for tomato leaf diseases detection and categorization," in *Proc. 2nd Int. Conf. Electr. Eng. Autom. Control* (*ICEEAC*), IEEE, May 2024, pp. 1–6. https://doi.org/10.1109/ICEEAC61226.2024.10576347

[2] J. Rashid, I. Khan, I. Ahmed Abbasi, M. Rizwan Saeed, M. Saddique, and M. Abbas, "A hybrid deep learning approach to classify the plant leaf species," *Comput. Mater. Contin.*, vol. 76, no. 3, pp. 3897–3920, 2023. https://doi.org/10.32604/cmc.2023.040356

[3] A. A. Al-Kharaz, A. B. A. Alwahhab, and V. Sabeeh, "Innovative date fruit classifier based on scatter wavelet and stacking ensemble," *HighTech Innov. J.*, vol. 5, no. 2, pp. 361–381, Jun. 2024. https://doi.org/10.28991/HIJ-2024-05-02-010

[4] Y. Al-Mulla, A. Ali, and K. Parimi, "Detection and analysis of dubas-infested date palm trees using deep learning, remote sensing, and GIS techniques in Wadi Bani Kharus," *Sustainability*, vol. 15, no. 19, art. 14045, Sep. 2023. https://doi.org/10.3390/su151914045

[5] M. Safran, W. Alrajhi, and S. Alfarhood, "DPXception: A lightweight CNN for image-based date palm species classification," *Front. Plant Sci.*, vol. 14, pp. 1–14, Jan. 2024. https://doi.org/10.3389/fpls.2023.1281724

[6] P. Rybacki *et al.*, "Convolutional neural network (CNN) model for the classification of varieties of date palm fruits (*Phoenix dactylifera L.*)," *Sensors*, vol. 24, no. 2, art. 558, Jan. 2024. https://doi.org/10.3390/s24020558

[7] O. Aiadi, B. Khaldi, M. L. Kherfi, M. L. Mekhalfi, and A. Alharbi, "Date fruit sorting based on deep learning and discriminant correlation analysis," *IEEE Access*, vol. 10, pp. 79655–79668, 2022. https://doi.org/10.1109/ACCESS.2022.3194550

[8] S. Khirani, H. Boutaj, C. El Modafar, and A. O. E. Khelil, "Arbuscular mycorrhizal fungi associated with date palm in Ouargla region (southeastern Algeria)," *Plant Cell Biotechnol. Mol. Biol.*, vol. 21, no. 45–46, pp. 15–28, 2020. https://www.researchgate.net/publication/344712109

[9] M. E. H. Chowdhury *et al.*, "Automatic and reliable leaf disease detection using deep learning techniques," *AgriEngineering*, vol. 3, no. 2, pp. 294–312, May 2021. https://doi.org/10.3390/agriengineering3020020

[10] A. Magsi, J. Ahmed Mahar, and S. H. Danwar, "Date fruit recognition using feature extraction techniques and deep convolutional neural network," *Indian J. Sci. Technol.*, vol. 12, no. 32, pp. 1–12, Aug. 2019. https://doi.org/10.17485/ijst/2019/v12i32/146441

[11] A. Alsirhani, M. H. Siddiqi, A. M. Mostafa, M. Ezz, and A. A. Mahmoud, "A novel classification model of date fruit dataset using deep transfer learning," *Electronics*, vol. 12, no. 3, art. 665, Jan. 2023. https://doi.org/10.3390/electronics12030665

[12] L. Khriji, A. Chiheb, and M. Awadalla, "Artificial intelligent techniques for palm date varieties classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, pp. 489–495, 2020. https://doi.org/10.14569/IJACSA.2020.0110958

[13] A. Almutairi, J. Alharbi, S. Alharbi, H. F. Alhasson, S. S. Alharbi, and S. Habib, "Date fruit detection and classification based on its variety using deep learning technology," *IEEE Access*, vol. 12, pp. 190666-190677, 2024. https://doi.org/10.1109/ACCESS.2024.3433485

[14] M. Koklu, R. Kursun, Y. S. Taspinar, and I. Cinar, "Classification of date fruits into genetic varieties using image analysis," *Math. Probl. Eng.*, vol. 2021, pp. 1–13, Nov. 2021. https://doi.org/10.1155/2021/4793293

[15] K. Albarrak, Y. Gulzar, Y. Hamid, A. Mehmood, and A. B. Soomro, "A deep learning-based model for date fruit classification," *Sustainability*, vol. 14, no. 10, art. 6339, May 2022. https://doi.org/10.3390/su14106339

[16] D. M. Ibrahim and N. M. Elshennawy, "Improving date fruit classification using CycleGAN-generated dataset," *Comput. Model. Eng. Sci.*, vol. 131, no. 1, pp. 331–348, 2022. https://doi.org/10.32604/cmes.2022.016419

[17] Y. A. Mukhlif *et al.*, "Ant colony and whale optimization algorithms aided by neural networks for optimum skin lesion diagnosis: A thorough review," *Mathematics*, vol. 12, no. 7, art. 1049, Mar. 2024. https://doi.org/10.3390/math12071049

[18] D. Chen, X. Qi, Y. Zheng, Y. Lu, Y. Huang, and Z. Li, "Synthetic data augmentation by diffusion probabilistic models to enhance weed recognition," *Comput. Electron. Agric.*, vol. 216, art. 108517, Jan. 2024. https://doi.org/10.1016/j.compag.2023.108517

[19] J. Kotwal, R. Kashyap, P. M. Shafi, and V. Kimbahune, "Enhanced leaf disease detection: UNet for segmentation and optimized efficientnet for disease classification," *Softw. Impacts*, vol. 22, p. 100701, Nov. 2024. https://doi.org/10.1016/j.simpa.2024.100701

[20] Z. Li *et al.*, "Enhancing tea leaf disease identification with lightweight MobileNetV2," *Comput. Mater. Contin.*, vol. 80, no. 1, pp. 679–694, 2024. https://doi.org/10.32604/cmc.2024.051526

[21] T. O'Halloran, G. Obaido, B. Otegbade, and I. D. Mienye, "A deep learning approach for maize lethal necrosis and maize streak virus disease detection," *Mach. Learn. with Appl.*, vol. 16, art. 100556, Jun. 2024. https://doi.org/10.1016/j.mlwa.2024.100556

[22] Zia-ur-Rehman *et al.*, "Classification of Alzheimer disease using DenseNet-201 based on deep transfer learning technique," *PLoS One*, vol. 19, no. 9, art. e0304995, Sep. 2024. https://doi.org/10.1371/journal.pone.0304995

[23] M. Azim Mim, N. Majadi, and P. Mazumder, "A soft voting ensemble learning approach for credit card fraud detection," *Heliyon*, vol. 10, no. 3, art. e25466, Feb. 2024. https://doi.org/10.1016/j.heliyon.2024.e25466

[24] H. Bakır, "VoteDroid: A new ensemble voting classifier for malware detection based on fine-tuned deep learning models," *Multimed. Tools Appl.*, vol. 84, pp. 10923{-10944, May 2024. https://doi.org/10.1007/s11042-024-19390-7