Algerian Journal of Research and Technology
https://doi.org/10.58681/ajrt.25090105
A.J.R.T Volume 9 (N°1) (2025) 42– 49

FeatureExtractionandMachineLearningfor Classification Date Fruit

Ikram Kourtiche[1],Mostefa Bendjima[2],Mohammed El Amin Kourtiche[3]

[123]Tahri Mohamed University,MathematicsandComputerScienceDepartmentLaboratoryofTITBechar, Algeria

*Corresponding author.kourtiche.ikram@univ-bechar.dz;bendjima.mostefa@univ-bechar.dz;kourtiche.amin@univ-bechar.dz

Received. June 17, 2024. Accepted. December 06, 2024. Published. December 20, 2024.

DOI: https://doi.org/10.58681/ajrt.25090105

Abstract. Dates are important in many parts of the world, particularly in North Africa and the Middle East. As a highly nutritious fruit with strong demand in both local and international markets, the classification and quality control of dates play a crucial role in enhancing their commercial value. This work focuses on improving date fruit classification by applying data augmentation techniques to enrich the original dataset, and then we employed three pre-trained CNN models, ResNet50, EfficientNetB0, and DenseNet201, for feature extraction. The extracted features were then classified using traditional machine learning algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Random Forest (RF). The bestperformance was achieved using ResNet50 as a feature extractor withlogistic regressionforclassification, reaching anaccuracyof 97.42%.

Keywords.Date fruit, Classification, Feature extraction, Pre- trained CNN, Machine learning.

INTRODUCTION

In recent years, the rapid progress of artificial intelligence (AI) has brought about significant transformations across a wide range of sectors, including agriculture(Sukkasem et al., 2023).AI has become an indispensable tool for addressing complex agricultural challenges by offering innovative solutions that enhance both efficiency and sustainability on a global scale (Meghwanshi, 2023). Among these advancements, deep learning has been instrumental in revolutionizing various agricultural practices, including fruit classification (Gill andKhehra, 2022).One fruit that has garnered increasing attention in this context is the date, known for its high nutritional value, rich in carbohydrates, minerals, andvitamins, and recognized for its potential health benefits, such as reducing the risk of cancer and cardiovascular diseases.

Globally, date production is substantial, with an estimated annual output of approximately 8.46 million tons (Özaltin, 2024).
In recent years many studies have been published on the classification of date fruits:

A date fruit classification system was developed in (Khayer et al., 2021) toidentify six date types. Features were recognized by CNN models. Their dataset has 2246 images. Comparing the systemto MobileNetV1, Inception, and Resnet, MobileNetV1 had the highest accuracy (82.67%). In (Bichri et al., 2023) transfer learning was employed to classify images using the pre-trained models MobileNetV2,VGG 19 and ResNet50.The VGG19 model has achieved the best classification accuracy (95%) and highest overall accuracy compared to other models.

Altaheri et al. (Altaheri et al., 2019) introduced a machine vision framework for classifying date fruits according to their type, maturity, and harvest readiness in a natural orchard setting. This framework leverages deep convolutional neural networks (CNNs) and transfer learning to achieve high classification accuracy,utilizing a dataset of 8000 images. Notably, theframework achieved a type classification accuracy of 99.01%.

In (Özaltin, 2024), researchers evaluated various algorithms, including Decision Tree, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM), for classifying seven date varieties. the neural network model yielded the highest accuracy at 93.85%.

Alsirhani et al. (Alsirhani et al., 2023) presented a deep transfer learning approach for the classification of 27 distinct date varieties using a dataset of 3228 images. By fine-tuning a DenseNet201 model, the researchers attained a test accuracy of 95.21%.

A study conducted by (Al-Sabaawi et al., 2021) investigated a comprehensive dataset comprising 8,000 images of five distinct date fruit varieties.The performance of pre-trained deep learning models:GoogleNet, ResNet-50, DenseNet, and AlexNet, was evaluated on this dataset.The results indicate that ResNet-50 outperformed the other models, achieving an accuracy rate of 97.37%.

In our study, we propose a method for classifying date fruits using feature extraction from three pre-trained convolutional neural network models: ResNet50, DenseNet201, and EfficientNetB0. The features obtained from each model are then used as input for different machine learning algorithms, such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Random Forest (RF), to perform classification. The remainder of this paper is structured as follows: After the introduction, we describe materials and methods used in this study. The next section presents the experimental results, followed by a discussion of the findings. Finally, the conclusion is drawn in the last section.

MATERIAL AND METHODS

Pre-trainedmodels

1) *ResNet-50* a residual neural network has 50 layers and constructsa network bysequentiallystackingresidual blocks.Thearchitecturehas48convolutionallayers, oneMaxPooling layer, and one average pooling layer.ResNet-50 is a popular picture categorization system(Ahmed et al., 2023).
2) *EfficientNet-B0* is a convolutional neural network optimized forhighperformancewithfewer parameters. Ituses depthwise separable convolutions and squeeze-and-excitation (SE) modules, gradually decreasing spatial resolution and increasing channels. The architecture balances depth, width, and input resolution, achieving high accuracy and low computational cost(Ahmed et al., 2023).
3) *DenseNet201* is a convolutional neural network with direct feed-forward connections, which reduces gradient degradation and overfitting in deep learning applications. Its architecture enhances inputs at each layer, diminishes parameters, and elevates performance. DenseNet201, aversion of 201 layers, employs this compact architecture to develop models that are easy to train and exceptionally efficient (Dümen et al., 2024).

Classificationmethods

1) *Support Vector Machines (SVM):* is a highly esteemed traditional approach in machine learning, commonly used for both classification and regression tasks. It works by transforming data characteristics into higher dimensions to establish a boundary or hyperplane for classification. The SVM identifies a linear discriminant function that maximizes the margin between different classes of data. Support vectors, which are data points closest to the classification boundary, play a crucial role in defining this boundary. SVM is well- known for its accuracy and versatility, making it a popular choice in applications (Xian andNgadiran, 2021).
2) *Random Forest (RF):* The decision tree method is extensively employed for categorizing extensive datasets and identifying data that share common traits. It involves dividing the data into smaller subsets iteratively, culminating in the construction of a structured tree that includes both decision nodes and leaf nodes, yielding the final classificationoutcomes (Xian andNgadiran, 2021).
3) *k nearest neighbors (KNN):* The operational principle of the KNN classifier is direct and intuitive: it assigns categories to samples based on the classes of their nearest neighbors. This classification method, known as memory-based classification,requiresstoringtrainingsamplesinmemoryforreferenceduringanalysis(13)inthispaper;theparameter k is set to 9.
4) *Logistic Regression (LR):* is a commonly used statistical method for modeling the probability of a binary outcomebased on one or more explanatory variables. Its primary goal is to estimate the coefficients of a linear model that relates the logarithm of the odds (log-odds) to the independent variables (Koklu et al., 2021).

Dataset

This dataset, referred toas the Saudi Arabian Dataset, consists of 1658 images, each depicting one of nine date fruit varieties native to Saudi Arabia: Ajwa, Galaxy, Medjool, Nabtat Ali, Sokari, Rutab, Shaishe, Sugaey, and Meneifi, as shown in Fig.1.A dedicated setup was designed to photograph the nine different varieties. The imaging system included a Canon EOS 550D DSLR camera mounted with the flash turned on. Surrounding the subject, a 48 cm diameter ring light equipped with 240 LED bulbs operating at full brightness ensured even lighting. This ring light helped eliminate shadows by uniformly illuminating the date from all directions, while the camera's flash delivered an intense, focused burst of light to highlight the texture and surface characteristics of the date, such as its firmness or softness (Alhamdanand Howe, 2021).

Fig.1. Samplesofdatefruitdatasetimages.

Dataaugmentation

A significant aspect of this study is the use of data augmentation to enhance model performance. Data augmentation is a crucial strategy in machine learning that involves artificially increasing the size and diversity of a dataset by applying various transformations to the existing data. In this context, several augmentation techniques were employed, including:

1) *Rescaling:*Theprocessofadjustingthesizeofimages.
2) *Random zoom:* Modifies the image scale to simulate varying distances.
3) *Flipping:*Involvesmirroringtheimagestocreate variations.
4) *Widthandheightshifts:*Slightlyrepositionthe images to account for different orientations.
5) *Randomrotations*:Rotatetheimagesatvariousangles.

These techniques are designed to improve the model's ability to generalize by exposing it to a wider variety of data representations. By augmenting the datasets in this manner,the study aims not only to enhance classification accuracy but also to ensure that the models can effectively recognize and differentiate between various date varieties. After applying data augmentation, the dataset comprises 3460 images ofSaudi Arabian date fruit.

After augmentation, the dataset was divided into two subsets: 80% for training and 20% for testing.

EXPERIMENTALRESULTSANDDISCUSSIONS

Evaluationmetricsused

In our study, we used several evaluations. These measures aim to evaluate the performance rate of our model. Precision, recall, f1-score, and accuracy were determined by quantifying the predicted classes based on the following quantities: the number of false negatives (FN), false positives (FP), true negatives (TN), and true positives (TP). The mathematical representation's definition is outlined below:

$$Accuracy = \frac{\mathrm{TP + TN}}{\mathrm{TP + TN + FN + FP}} (1)$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP + FP}} (2)$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP + FN}} (3)$$

$$\mathrm{f1 - score} = 2\mathrm{x}\frac{\mathrm{recall\ x\ precision}}{\mathrm{recall} + \mathrm{precision}} (4)$$

RESULTS

Our experiments are used on a computer with an Intel(R) Core(TM) i5-6300U CPU, with 8 GB of RAM, utilizing Kaggle, a cloud-based platformthat enables usersto write and execute Python code directly in their web browsers. Kaggle is particularly advantageous for machine learning, data analysis, and deep learning tasks, as it offers GPU support for acceleratedcomputation.Thisenvironmentfacilitatesufficient experimentation and model training by providing access to powerful resources and tools tailored for data science applications.

TheresultsofourstudyaresummarizedinTables 1, 2,and 3. The results in Table 1 indicate that the LR classifier achieved thehighestperformanceamongthecomparedalgorithms,with a testing accuracy of 97.42%, a recall of 97.42%, an F1-score of97.42%, and a precisionof97.44%.These results highlight the effectiveness of feature extraction using the Resnet50 Table 2 presentstheresultsafterfeatureextractionusing EfficentNetB0,wherethebestperformancewasachievedwith theLogisticRegression(LR)classifier,reachinganaccuracyof97.27%, and the table 3 shows the

results after feature extraction using DenseNet201, where the LR classifier obtained the highest accuracy of 96.84%.
Logistic Regression outperforms all other models across the three feature extraction methods (ResNet-50, EfficientNetB0, DenseNet201). Its strong performance is likely due to the extracted features being well-structured and linearly separable. LR remains a simple, efficient choice for this kind of classification task.

Table 1.PerformancemetricsforRenet-50extractedfeatures.

Models	Accuracy	Recall	Precision	F1-score
SVM	93.26%	93.26%	93.33%	93.25%
LR	**97.42%**	**97.42%**	**97.44%**	**97.42%**
KNN	85.51%	85.51%	86.28%	85.61%
RF	89.81%	89.81%	89.87%	89.79%

Table 2.PerformancemetricsforEfficientNetB0extractedfeatures.

Models	Accuracy	Recall	Precision	F1-score
SVM	94.26%	94.26%	94.37%	94.26%
LR	**97.27%**	**97.27%**	**97.32%**	**97.28%**
KNN	88.24%	88.24%	88.63%	88.29%
RF	90.67%	90.70%	90.67%	90.63%

Table 3.Performancemetricsfordensenet201extractedfeatures.

Models	Accuracy	Recall	Precision	F1-score
SVM	94.12%	94.12%	94.13%	94.11%
LR	**96.84%**	**96.84%**	**96.91%**	**96.85%**
KNN	88.52%	88.52%	89.11%	88.52%
RF	93.11%	93.11%	93.24%	93.52%

Fig. 2, 3,and4showtheconfusion matricesfor thebest-performingmethodsusingfeaturesextractedwith Resnet50, EfficentNetB0 , and densenet201, respectively.

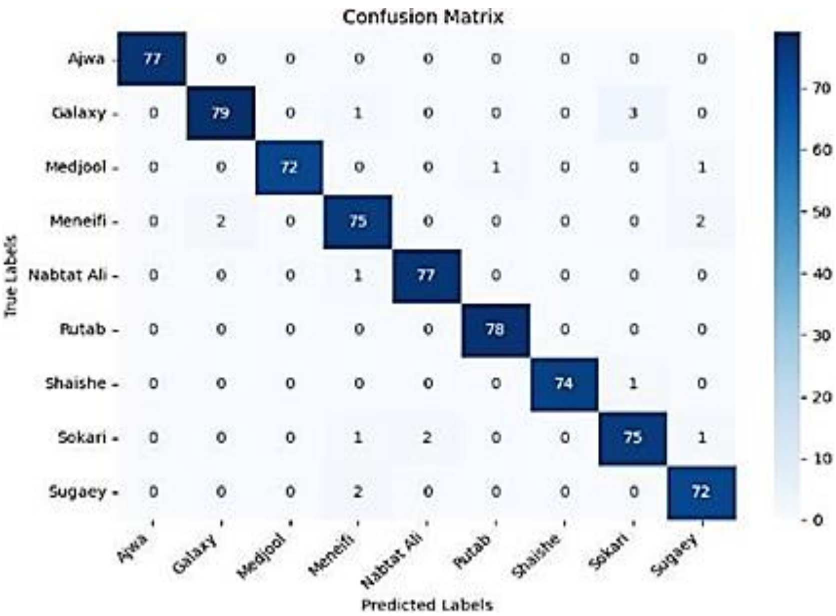

Fig.2. ConfusionmatrixforLRusingResnet-50.

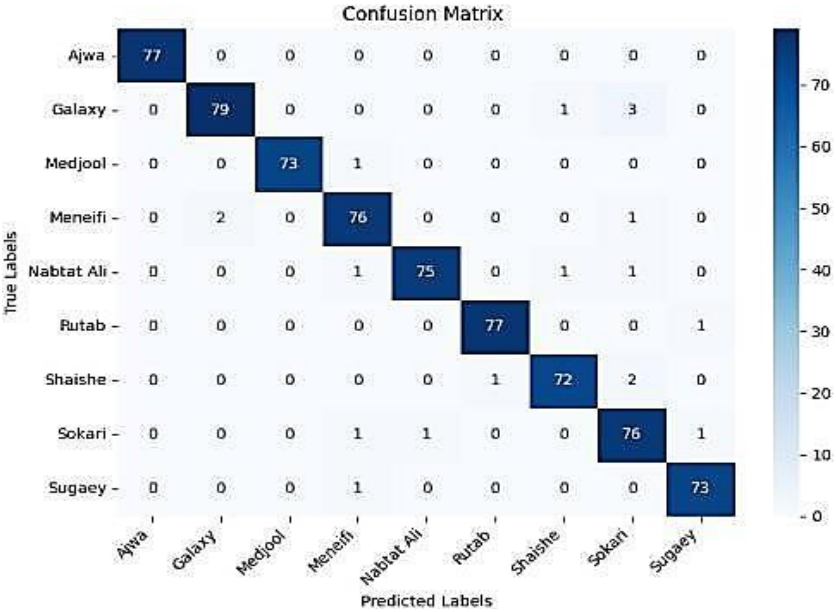

Fig.3. ConfusionmatrixforLRusingEfficientNetB0.

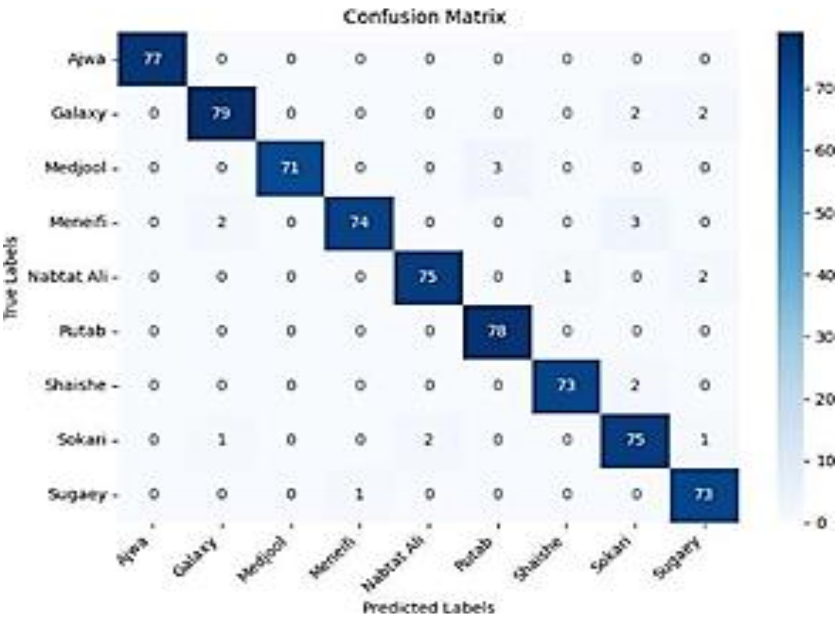

Fig.4. ConfusionmatrixforLRusingDenseNet201

Our proposed study is evaluated against several recent state- of-the-art techniques, as presented in Table 5 demonstrates superior performance using a dataset with 9 different types of date fruit.

Table 5.Comparison withstate oftheartmethods.

Ref	Technique	Date type	Best Accuracy
Khayer et al., 2021	Various pre-trained models(MobileNet,Inception, and Resnet)	6	MobileNetV182.67%
Al-Sabaawi et al., 2021	GoogleNet,ResNet50,DenseNet andAlexNet	5	ResNet5097.37%
Koklu et al., 2021	StackingmodelcreatedbycombiningLR andANN	7	92.80%
Magsi et al., 2019	Features extraction+combination ofseveral hidden layers	3	97.20%
Nasiri et al., 2019	VGG16	4	96.98%
Our Study	Feature extractionusing ResNet50,DenseNet201, EfficientNetB0 andseveralmachinelearningalgorithms	9	FeatureextractionusingResnet50+LR**97.42%**

CONCLUSION

The objective of this study was to develop a classification system for date fruits byutilizing featureextraction fromthree pre-trained convolutional neural network models: ResNet50, EfficientNetB0, and DenseNet201. The extracted featureswere subsequently classified using traditional machinelearning algorithms, including Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and K-Nearest Neighbors (KNN). This research aims to support and improve agricultural practices related to date fruit classification.

For future work, we plan to apply this approach to other agricultural products, with the aim of improving classification accuracy.

REFERENCES

Ahmed, N., Rahman, M., & Ishrak, F. (2024). *Comparative performance analysis of transformer-based pre-trained models for detecting keratoconus disease*. arXiv. https://doi.org/10.48550/arXiv.2408.09005

Al-Sabaawi, A., Hasan, R. I., Fadhel, M. A., Al-Shamma, O., &Alzubaidi, L. (2021). Employment of Pre-trained Deep Learning Models for Date Classification: A Comparative Study. In *Intelligent Systems Design and Applications* (pp. 181–189). Springer.https://doi.org/10.1007/978-3-030-71187-0_17

Alhamdan, W., & Howe, J. M. (2021). Classification of date fruits in a controlled environment using convolutional neural networks. In *Advanced Machine Learning Technologies and Applications*, vol. 9(1), pp. 154–163. Springer.https://doi.org/10.1007/978-3-030-69717-4_16

Altaheri, H., Alsulaiman, M., & Muhammad, G. (2019). Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep Learning. *IEEE Access*, 7, 117115–117133. https://doi.org/10.1109/ACCESS.2019.2936536

Alsirhani, A., Siddiqi, M. H., Mostafa, A. M., Ezz, M., & Mahmoud, A. A. (2023).A Novel Classification Model of Date Fruit Dataset Using Deep Transfer Learning.*Electronics*, 12(3), 665.https://doi.org/10.3390/electronics12030665

Bayram, H. Y., Bingol, H., &Alatas, B. (2022).Hybrid Deep Model for Automated Detection of Tomato Leaf Diseases.*Traitement du Signal*, 39(5), 1781–1787. https://doi.org/10.18280/ts.390537

Bichri, H., Chergui, A., &Hain, M. (2023). Image Classification with Transfer Learning Using a Custom Dataset: Comparative Study. *Procedia Computer Science*, 220, 48–54. https://doi.org/10.1016/j.procs.2023.03.009

Dümen, S., KavalcıYılmaz, E., Adem, K., &Avaroglu, E. (2024). Performance of vision transformer and swin transformer models for lemon quality classification in fruit juice factories. *European Food Research and Technology*, 250(9), 2291–2302.https://doi.org/10.1007/s00217-024-04537-5

Gill, H. S., &Khehra, B. S. (2022).An integrated approach using CNN-RNN-LSTM for classification of fruit images.*Materials Today: Proceedings*, 51, 591–595. https://doi.org/10.1016/j.matpr.2021.06.016

Khayer, Md. A., Hasan, Md. S., &Sattar, A. (2021). Arabian Date Classification using CNN Algorithm with Various Pre-Trained Models. In *Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 1431–1436). IEEE.https://doi.org/10.1109/ICICV50876.2021.9388413

Koklu, M., Kursun, R., Taspinar, Y. S., &Cinar, I. (2021).Classification of Date Fruits into Genetic Varieties Using Image Analysis.*Mathematical Problems in Engineering*, 2021, 1–13. https://doi.org/10.1155/2021/4793293

Magsi, A., Mahar, J. A., &Danwar, S. H. (2019). Date Fruit Recognition using Feature Extraction Techniques and Deep Convolutional Neural Network. *Indian Journal of Science and Technology*, 12(32), 1–12. https://doi.org/10.17485/ijst/2019/v12i32/146441

Meghwanshi, S. (2023). ARTIFICIAL INTELLIGENCE IN AGRICULTURE: A REVIEW. *Open Access*, 6(3). *(DOI non disponible dans la source fournie)*

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

Özaltin, Ö.,& Department of Mathematics, Atatürk University. (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

Sukkasem, S., Jitsakul, W., &Meesad, P. (2023).Fruit Classification with Deep Transfer Learning using Image Processing.In *7th International Conference on Information Technology (InCIT)* (pp. 464–469).IEEE.https://doi.org/10.1109/InCIT60207.2023.10413036

Xian, T. S., &Ngadiran, R. (2021). Plant Diseases Classification using Machine Learning.*Journal of Physics: ConferenceSeries*, 1962(1), 012024. https://doi.org/10.1088/1742-6596/1962/1/012024