Synergy: Cross-Disciplinary Journal of Digital Investigation
(ISSN 2995-4827) VOLUME 03 ISSUE 04, 2025

CLASSIFICATION OF DATE FRUITS USING DEEP LEARNING MODEL

Zahraa Jabbar Hussein
Department of Computer Science, College of Science for Women, University of Babylon, Iraq
zahraa.jabbar@uobabylon.edu.iq

Abstract:

Fruits are the most widely grown of all farm products worldwide and have the characteristic of possessing many varieties that differ from each other on the basis of their external characteristics such as color, length, diameter, and shape. These external attributes are equally critical in the definition of fruit types. However, identifying varieties of fruits using outside appearance only could require professional knowledge, and hence may consume lots of time and effort. The focus in the current study is to recognize various types of date fruits using artificial intelligence techniques. The varieties targeted are: Ajwa, Galaxy, Medjool, Meneifi, Nabtat Ali, Rutab, Shaishe, Sokari, and Sugaey. A total dataset of 1,658 images of the mentioned varieties was used in this research, which was obtained by means of a computer vision system (CVS). In contrast to conventional hand-crafted feature extraction-based work on color and shape, in this research study, a Convolutional Neural Network (CNN) model was employed for the direct classification of the images. The model indicated a very high accuracy of 98.8%, which validates the effectiveness and appropriateness of deep neural networks in classifying varieties of date fruit accurately.

Keywords: CNN, Artificial Intelligence, Deep learning, Fruits, Date.

Introduction

Date fruits are some of the most cultivated and economically important agricultural crops in the majority of Middle Eastern and African countries. They have cultural, dietary, and religious significance. As global production continues to rise—reaching over 8 million tons in 2017—the need for efficient classification and sorting systems becomes increasingly essential[1]. The physical similarity among many date varieties, including color, shape, and texture, makes manual classification a difficult and time-consuming task that often requires experienced labor[2]. Despite several studies focusing on the automation of fruit classification using machine learning techniques, limited research exists that addresses the classification of date fruits specifically. Furthermore, publicly available datasets for this purpose are scarce[3].

This research contributes to computer vision and agricultural engineering by leveraging deep learning techniques to classify dates varieties from image data. Nine types of dates were exemplified by a dataset consisting of 1,658 high-quality images and were categorized as Ajwa, Galaxy, Medjool, Meneifi, Nabtat Ali, Rutab, Shaishe, Sokari, and Sugaey. These images were captured in a controlled environment. A Convolutional Neural Network (CNN) model was developed and trained on this dataset using data augmentation strategies to improve generalization. This model achieved a very high accuracy in classification, at 98.8%, which proves the strength of CNNs in this context and their potential uses in real-world applications such as automated sorting machines or consumer and producer apps.

The most important aims of this study are the following:

1. To establish a precise and effective model of classification that can identify various types of dates across the globe.
2. To make the sorting and classification process more efficient using recent AI technologies to make it easier to market the date fruits.
3. In order to facilitate the use of artificial intelligence in farming, specifically for one of the globe's most valuable crops.
4. To verify the model's correctness and analyze the factors affecting date varieties' classification

Related Work

The field of classification of fruits and vegetables has enhanced significantly with advancements in computer vision and the use of Convolutional Neural Networks (CNNs)[4]. Beforehand, researches were dependent on manually segmenting images and extracting a great number of features, which was time-consuming and tended to demand small datasets[5]. For example, one study utilized 60 photos of potato chips boiled under varying temperatures, which were utilized to extract 1511 features and then 11 features selected for classification purposes via a decision tree[6]. Other work collected pictures of some fruits and vegetables with advanced imaging techniques for manual feature extraction and processing of images. Before CNNs became popular, these operations were carried out manually, which was challenging in achieving high classification performance.[7]

With the emergence of CNNs, it was possible to learn appropriate filters from images automatically, thus improving classification accuracy with reduced human intervention. However, feature understanding used by these models remains a challenge.[8]

In 2023, some scientists applied CNNs in classification of dates and other fruits. For example, scientists applied a light-weight model called DPXception for date variety classification based on images taken from orchards using outstanding classification accuracy[9].

Another study used CNNs to classify the surface quality of dates as excellent and poor quality using 898 images in the dataset with a classification accuracy of 97%[10].

In addition, a comparative study of different CNN models in fruit and vegetable classification showed that deep learning techniques greatly improved accuracy and speed over traditional approaches[11], [12].

These recent studies indicate that CNNs hold promise for date and other fruit classification, with the potential to expand sorting and grading operations in agriculture, reduce dependence on human labor, and be more accurate in distinguishing between varieties.

Dataset

The dataset of images used in this study was accessed from Kaggle, a leading platform in the world in which open datasets commonly used in machine learning studies are shared. The dataset contains 1,658 quality images of various date fruits, which include: Shaishe, Galaxy, Medjool, Nabtat Ali, Meneifi, Rutab, Sokari, Ajwa, and Sugaey. The images were captured under a controlled condition for illumination, camera location, and distance and therefore were very compatible for computer vision-based classification. Figure 1 presents some example sample photographs of the nine date fruit types in the dataset, reflecting the visual difference in color, shape, and texture among the varieties.

Table 1 lists the split-up of the dataset by the nine date fruit classes, image quantity for each category, and size of the images. The above detail provides insight into the proportionality and configuration of the data utilized within the research of this work.

Table 1: Images Distribution by Date Fruit Type in the Dataset

Date Fruit Type	Number of Images	size
Ajwa	175	5184×3456
Galaxy	190	3456×2304
Medjool	135	3456×2304
Meneifi	232	3456×2304
Nabtat Ali	177	3456×2304
Rutab	146	3456×2304
Shaishe	171	3456×2304
Sokari	264	3456×2304
Sugaey	168	3456×2304

Figure 1: Sample images of the nine date fruit varieties in the dataset, showcasing the visual differences in color, shape, and texture among the type

Methodology

A set of nine date varieties were photographed under controlled conditions using the same light, background, and camera distance —Ajwa, Galaxy, Medjool, Meneifi, Nabtat Ali, Rutab, Shaishe, Sokari, and Sugaey—. All of the images were captured under the same conditions and then resized to the same 200×200 pixel resolution for consistency and compatibility for training classification models.

1- Pre-processing the Dataset

In this work, the images were first pre-processed by cropping out the white background to isolate the date object. After cropping, all images were resized to a fixed dimension of 200×200 pixels to ensure uniformity across the dataset. The RGB pixel values were then normalized to the [0, 1] range, which is a standard practice to improve neural network performance. The data was split into three sets: 70% for training, 20% for validation, and 10% for testing. This split enabled model evaluation to be conducted on unseen data and allowing efficient monitoring of training using the validation set.

2- proposed model

The suggested model was developed and run in Python through the Spyder IDE on the Anaconda platform. The model was trained and tested locally on Windows 10, 8 GB RAM, 500 GB SSD, and an integrated GPU. The system experimented with several CNN architectures using a varying number of convolutional and dense layers to find the most accurate and stable model for date type classification. Some hyperparameters were tuned, including dropout rate, batch size, image dimensions, number of filters, and patterns of neurons. The image dimension was fixed to 200×200 pixels, and data augmentation was applied to some phases in order to promote generalization.

Step by step, starting with a simple CNN model comprising two convolutional layers. First, the dropout was the hyperparameter tuned in order to prevent overfitting. Then the model was expanded to three dense layers and four convolutional layers, and other hyperparameters such as batch size, filter size, and number of neurons were also tuned. The input image was fixed at size 200×200 pixels, based on prior preprocessing steps. L2 regularization was also tried with different values in order to increase generalization.

Data augmentation techniques such as rotation, horizontal flip and vertical flip, zoom, random crop, brightness and contrast, and noise were utilized in order to help generalize the model and improve diversity of the dataset. The augmentation parameters were chosen so that they mimic real-world variation in date location in sorting. Batch size and L2 regularization values were tuned correspondingly in order to reduce over fitting. Additionally, certain optimizers and learning rates were also examined in order to further optimize training performance.

In the final step, Gaussian noise layers were incorporated to make the model robust to input variation. Different noise levels were implemented at different layers, and the best one came out when adding a noise of 0.02 at the input layer. The final collection of models, augmented and non-augmented, with and without noise, were selected based on their performance and stability over the test and validation sets.

Each model was tested and trained several times to gain credible performance scores, and the best-performing models from each phase were retained for further comparison and analysis. As shown in Figure 2, the model suggested is a deep convolutional neural network employed specifically for the classification of nine date categories, optimized with data augmentation, hyperparameter tuning, and several stages of performance testing.

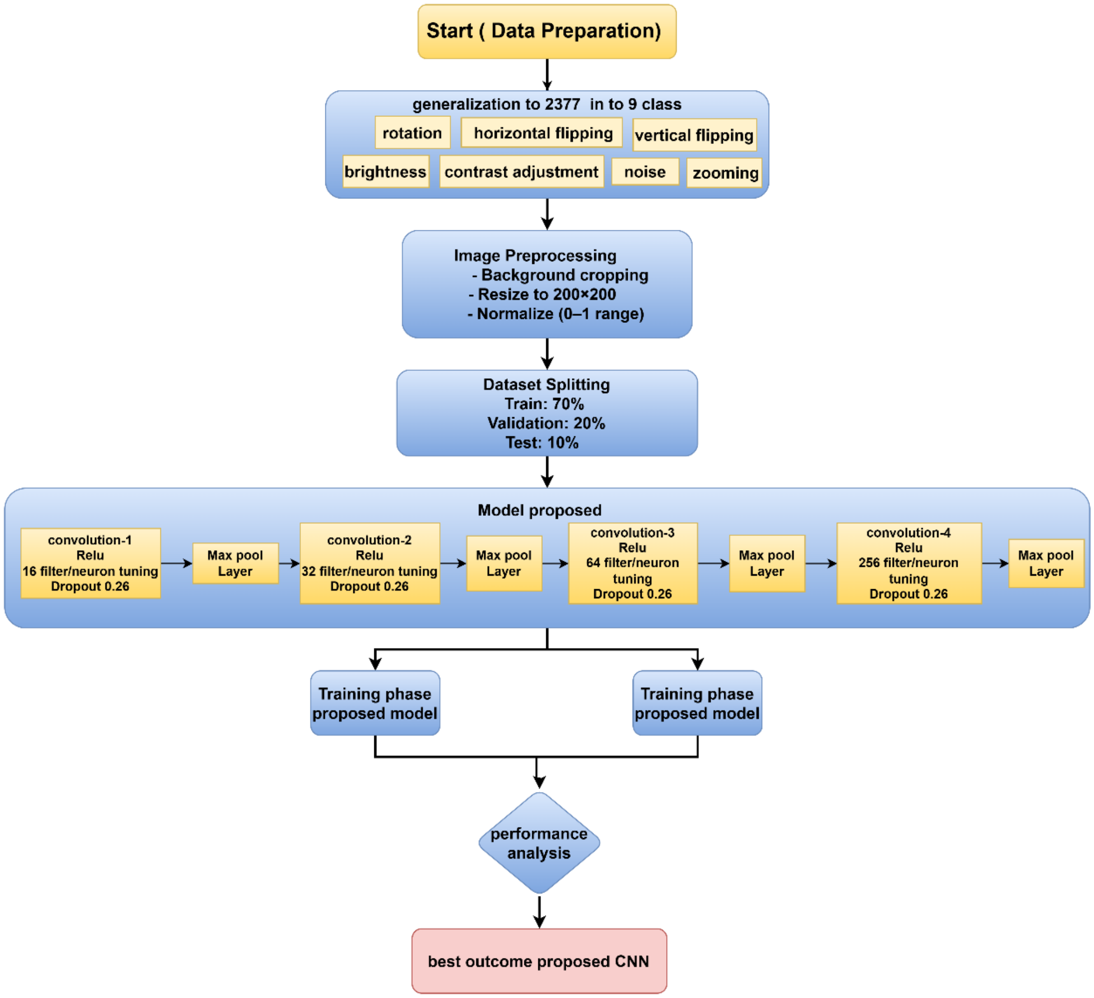

Figure 2: Illustration of the architecture of the proposed deep learning model used for classifying nine types of dates.

Results

This section presents the results obtained from the different development stages, highlighting the four most significant models. It was observed that models with only 2 convolutional layers were insufficient in learning deep features, while architectures with 8 convolutional layers significantly increased training time without a notable improvement in performance compared to 4-layer models. Therefore, the focus was placed on optimizing 4-layer CNN architectures. Table 1 summarizes the classification accuracy achieved by the selected models on the test set, considering different data processing techniques applied during training. Show Table 2. Proposed Model Architecture.

1- Custom-4L Model Architecture Overview:

The proposed Custom-4L model is a Convolutional Neural Network (CNN) composed of 4 convolutional layers, designed for image classification tasks. It is performing resizing on input images to 200×200 pixels and has been trained for 20 epochs in batch size 32. Optimization has been done using SGD with learning rate 0.001. L2 regularization of strength 0.0005 has been used to prevent overfitting in convolutional layers.

The model employs a progressive filter structure of (16, 32, 64, 256) filters with kernel sizes of (5×5, 5×5, 3×3, 3×3). Max pooling (2×2) is applied after the 1st, 2nd, and 4th convolutional layers. The output of the convolutional blocks is passed through three fully connected (FC) layers, each containing 1024 neurons, followed by a final FC layer with 9 neurons representing the output classes. A dropout of 26% is used in the FC layers (excluding the last one) to further reduce overfitting.

The model achieves a strong test accuracy of 95%, and when combined with a Softmax classifier, the classification performance improves to 98.8%, indicating that the model generalizes well and effectively distinguishes between the 9 classes in the dataset.

Table 2. Proposed Model Architecture

Model name	Custom-4L
Batch size	32
Picture size	200 × 200
Epochs	20
L2 value	0.0005
Accuracy	95%
Optimizer and lr	SGD, lr = 0.001
Filter numbers	(16, 32, 64, 256)
Filter sizes	(5 × 5, 5 × 5, 3 × 3, 3 × 3)
Max pooling (2 × 2)	(y, y, n, y)
Neurons	(1024, 1024, 1024, 9)
Dropout on FC	(26%, 26%, 26%, 0%)
Classification (CNN+ Softmax)	98.8%

The performance curves illustrated in the figure demonstrate the training and validation accuracy and loss over 23 epochs. The precision of the graph illustrates steady enhancement in validation and training performance to over 98.9%, an indication of effective learning. Conversely, the loss curve illustrates a sharp decline at the start of the epochs followed by stability and extremely well-matching consensus among the training and validation loss. The close resemblance between these two sets is the testament that the model generalizes excellently without having any over fitting.

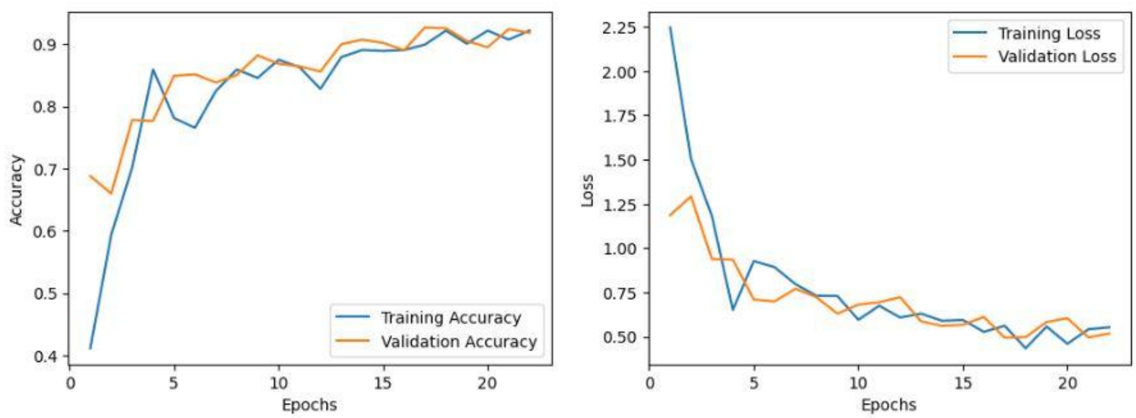

Figure 3: Demonstration of the validation and training accuracy and loss curve of the suggested CNN model across 20 epochs and signifies the proper learning and generalization capacity of the model.

Confusion matrix is also one of the most crucial testing tools that reveal the amount of true and false predictions for each class, and how to identify the weaknesses and strengths of a classification model.

Confusion matrix for classification of 9 classes of Date Fruits is depicted through figure 4, where below figure depicts confusion matrix for classification of 9 classes of Date Fruits, where horizontal axis is considered for predicted labels and vertical axis for true labels. Table 3 also presents comparison of the given model with other CNN architectures for classification of date fruit.

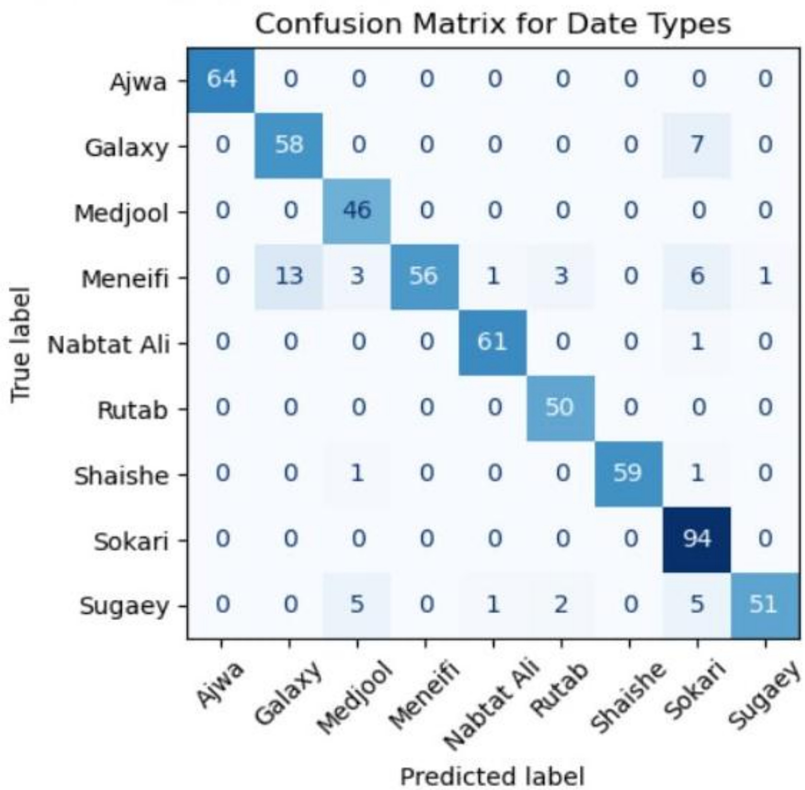

Figure 4: Date Fruits Classification Confusion Matrix

Table 3: Comparison of the presented model with various CNN architectures for date fruits classification

Algorithms	Accuracy %	Precision %	Recall %	F1-score %
LeNet-5	97.1	96.1	96.6	96.9
AlexNet	89.9	89.7	89.4	89.7
VGG-16	89.5	89	89.1	89.6
ResNet-50	78.9	78.2	78.1	78.9
DenseNet-121	90.1	89.9	89.7	89.5
MobileNet	91.7	91.1	91	91.6
Proposed model	98.8	98.1	98.5	98.7

Conclusion

This study suggested a tailored 4-layer CNN framework for classifying images with the size of 200×200 medical images. The model did a good classification accuracy of 98.8% when the Softmax classifier was utilized. Numerous approaches such as L2 regularization and dropout were utilized to enhance the generalization. The model acquired a balance in training and validation accuracy by viewing the performance graphs. Augmentation methods like rotation and flipping improved the model's robustness. The 4-layer model, in contrast to deep networks, had a satisfactory trade-off between performance and training time. Overall, the proposed system was effective and efficient in application to medical image analysis.

Reference

1. F. Moza Villalobos, J. Natividad Villanueva, and B. Meneses Claudio, “Use of Convolutional Neural Networks (CNN) to recognize the quality of oranges in Peru by 2023,” *Data and Metadata*, vol. 2, p. 175, Dec. 2023, doi: 10.56294/dm2023175.
2. L. Gillani Fahad, S. Fahad Tahir, U. Rasheed, H. Saqib, M. Hassan, and H. Alquhayz, “Fruits and Vegetables Freshness Categorization Using Deep Learning,” *Computers, Materials & Continua*, vol. 71, no. 3, pp. 5083–5098, 2022, doi: 10.32604/cmc.2022.023357.
3. S. Ghazal, W. S. Qureshi, U. S. Khan, J. Iqbal, N. Rashid, and M. I. Tiwana, “Analysis of visual features and classifiers for Fruit classification problem,” *Comput Electron Agric*, vol. 187, p. 106267, Aug. 2021, doi: 10.1016/j.compag.2021.106267.
4. N. Wijaya, S. H. Mulyani, and Y. W. Anggraini, “DeepFruits: efficient citrus type classification using the CNN,” *Iran Journal of Computer Science*, vol. 6, no. 1, pp. 21–27, Mar. 2023, doi: 10.1007/s42044-022-00117-6.
5. J. Jayanth, M. Mahadevaswamy, and M. Shivakumar, “Fruit Quality Identification and Classification by Convolutional Neural Network,” *SN Comput Sci*, vol. 4, no. 3, p. 220, Feb. 2023, doi: 10.1007/s42979-022-01527-w.
6. H. S. Gill, G. Murugesan, A. Mehbodniya, G. Sekhar Sajja, G. Gupta, and A. Bhatt, “Fruit type classification using deep learning and feature fusion,” *Comput Electron Agric*, vol. 211, p. 107990, Aug. 2023, doi: 10.1016/j.compag.2023.107990.
7. F. Salim, F. Saeed, S. Basurra, S. N. Qasem, and T. Al-Hadhrami, “DenseNet-201 and Xception Pre-Trained Deep Learning Models for Fruit Recognition,” *Electronics (Basel)*, vol. 12, no. 14, p. 3132, Jul. 2023, doi: 10.3390/electronics12143132.
8. F. Yu *et al.*, “Progress in the Application of CNN-Based Image Classification and Recognition in Whole Crop Growth Cycles,” *Remote Sens (Basel)*, vol. 15, no. 12, p. 2988, Jun. 2023, doi: 10.3390/rs15122988.
9. M. Safran, W. Alrajhi, and S. Alfarhood, “DPXception: a lightweight CNN for image-based date palm species classification,” *Front Plant Sci*, vol. 14, 2023, doi: 10.3389/fpls.2023.1281724.
10. M. Almomen, M. Al-Saeed, and H. F. Ahmad, “Date Fruit Classification Based on Surface Quality Using Convolutional Neural Network Models,” *Applied Sciences*, vol. 13, no. 13, p. 7821, Jul. 2023, doi: 10.3390/app13137821.
11. Y. Chen, S. Pan, and H. Wang, “Performance comparison of different convolutional neural networks for vegetable and fruit recognition,” *Applied and Computational Engineering*, vol. 5, no. 1, pp. 593–602, May 2023, doi: 10.54254/2755-2721/5/20230652.
12. Z. J. Hussein, A. M. Hussein, G. I. Maki, and H. Q. Gheni, “Improved Model for Skin Illnesses Classification Utilizing Gray-Level Co-occurrence Matrix and Convolution Neural Network,” *Journal of Advances in Information Technology*, vol. 14, no. 6, pp. 1273–1279, 2023, doi: 10.12720/jait.14.6.1273-1279.