Automating Date Fruit Sorting: A Multi-Modal Fusion and Deep Learning Approach

1st Ibtissam Boumaraf
Laboratory of LESIA
University of Mohamed khider
Biskra, Algeria
ibtissam.boumaraf@univ-biskra.dz

2nd Abdelhamid Djeffal
Laboratory of LESIA
University of Mohamed khider
Biskra, Algeria
abdelhamid.djeffal@univ-biskra.dz

3rd Abdelmalik Taleb-Ahmed
IEMN UMR CNRS 8520
UPHF
Valenciennes, France
abdelmalik.taleb-Ahmed@univ-valenciennes.fr

***Abstract*—Date fruit sorting plays a critical role in the agricultural and food processing industry, where precision and efficiency are of utmost importance. This paper introduces an innovative approach to automate date fruit sorting by utilizing a combination of multi-modal and deep learning techniques. We propose a novel method for categorizing date fruits into eight distinct classes, encompassing five grades of Deghlet Noor and three grades of Mech-Degla varieties. Our approach capitalizes on multi-modal fusion and deep learning techniques, specifically employing a late fusion method to integrate information from various sources. We comprehensively assess the performance of four deep learning models, namely CNN, VGG16, ResNet50, and MobileNet, across two distinct scenarios. Scenario I involves the classification of date fruits based on four different angle's views of the fruit, while Scenario II extends this approach by incorporating thermal images and weight metrics as additional modalities. Our results highlight the strengths of CNN in both scenarios, particularly its significant accuracy improvements in Scenario II when additional features are introduced, substantially enhancing overall classification performance. This research significantly contributes to the development of more efficient and accurate automated date fruit sorting systems, addressing the specific needs of the agricultural and food processing industries.**

***Index Terms*—Date fruit, Deep-Learning, Late fusion, Multi-modal, Thermal image, Weight.**

I. Introduction

The sorting of date fruits stands as a pivotal process within the date fruit industry, playing a vital role in ensuring product excellence and aligning with consumer preferences. Date fruits, celebrated for their nutritional value and economic importance, exhibit a wide array of types and grades. Yet, the accurate classification and sorting of these fruits pose substantial challenges due to their inherent diversity. Conventional sorting approaches, reliant on manual labor, visual examination, and human judgment, often struggle to meet the industry's escalating demands for precision and efficiency [1]. In response to these hurdles, advanced technologies, notably machine learning and deep learning, have emerged as revolutionary instruments, poised to enhance and modernize the date fruit sorting process [4], [7], [9]. Despite their success in various domains, the potential of CNNs in multi-modal fusion for date fruit classification remains underexplored. Some authors have utilized this technology to automate the sorting and grading of date fruits, moving away from labor-intensive manual procedures. This transition highlights the transformative potential of deep learning, with a focus on enhancing efficiency and fostering innovation.

In 2019, Nasiri et al. [8] conducted research with the aim of developing a method to predict the ripeness of healthy date fruits. They employed the VGG16 architecture and curated a dataset comprising four classes: Khalal, Rotab, Tamar, and defective dates. Their precision reached an impressive 98.49%. The study involved approximately 1300 date fruits from Southern Iran, utilized in both the training and validation phases. During the same year, Altaheri et al. [3] explored three different models (AlexNet, VGG-16, and a modified VGG-16) for real-time classification of date fruits. The fine-tuned VGG-16 model achieved accuracy of 99.01%, 97.25%, and 98.59% for the tasks of maturity, type, and harvesting decision classification, respectively. In 2020, Faisal et al. [5] utilized a dataset comprising approximately 8079 photos. Their system incorporated VGG-19, Inception-v3, and NASNet CNN types, capturing images from live video sources. The system then directed these images to the maturity level detection system. Remarkably, the suggested IHDS achieved maximum accuracy, F1 score, recall, and precision rates of 99.4%, 99.4%, 99.7%, and 99.7%, respectively. In 2022, Albarrak et al. [2] introduced "A Deep Learning-Based Model for Date Fruit Classification." Their dataset featured eight distinct types of dates from Saudi Arabia. Their study was founded on the MobileNetV2 architecture, featuring the addition of five layers to this architecture. This augmentation allowed the proposed model to predict and classify date fruit classes with an astounding estimated accuracy of 99.9%.

Based on previous research findings, although their proposed methods achieved respectable accuracy, they may not consistently classify unseen fruit samples correctly. This limitation stems from the fact that classification primarily relies on visual features from one side of the date fruit, which may not capture all relevant information and details from all sides, potentially leading to reduced efficiency and precision. To address the limitation in date fruit classification, we propose integrating additional data through a multi-modal approach, combining information from various sources to enhance accuracy. Surprisingly, previous research lacks the use of fusion techniques like late fusion with multi-angles of date fruit, various sensors, and multiple modalities [6], [9].

In our study, we employed Convolutional Neural Networks (CNNs) along with the late fusion technique, combining features from various sources. We classified 853 date fruits using four distinct models, while also considering the findings of our other experiments to assess whether the use of multi-angles and diverse acquisition sensors influenced classification success. our contributions can be outlined as follows:

- We have curated a new dataset comprising eight distinct date varieties, including five classes of Deglet-Noor and three classes of Mech-Degla.
- This research introduces an innovative method to enhance date fruit sorting accuracy by integrating multi-modal data through late fusion techniques.
- The study explores two scenarios: ' Scenario I :Fusion Four angle of view' and 'Scenario II : includes additional data like thermal images and date fruit weight'.
- We utilize four well-established deep learning models—CNN, VGG16, ResNet50, and MobileNet—to process and analyze the multi-modal data.

The rest of this paper is structured as follows: Section II outlines the methods used in our proposed systems, Section III discusses experimental results and analysis, and we conclude in the final section.

II. MATERIAL AND METHODS

Our approach to automating date fruit sorting through a multi-modal fusion and deep learning method involves several key stages that collectively contribute to precise and efficient classification, as illustrated in Fig. 1. The foundation of our

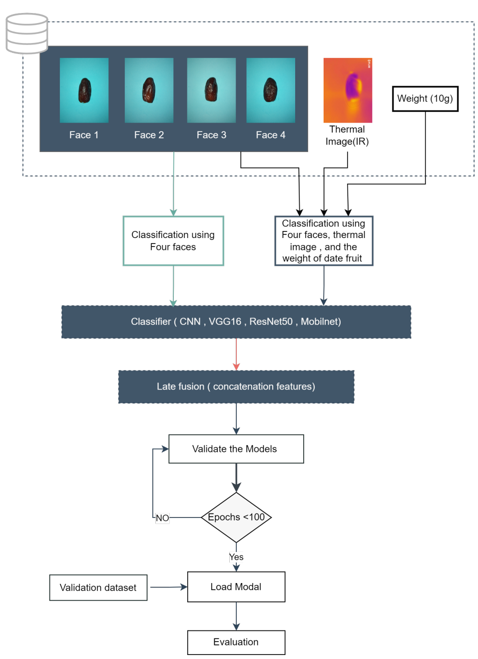

Fig. 1. Architecture diagram represent our proposed system

method lies in the data collection of two date fruit varieties: Deglet-Noor (divided into five quality classes) and Mech-Degla (Three quality classes). We acquired vital information from multiple sources, including high-resolution images captured from four distinct angles (right, left, front, and back) using an RGB camera. This four-angle photography process generated a comprehensive visual dataset. Furthermore, we utilized thermal infrared imaging to detect defects and assess date fruit quality, specifically using the FLIR ONE Gen 3 - Android (USB-C). We also recorded weight scale readings for each date fruit to enhance our dataset, as illustrated in Fig. 2.

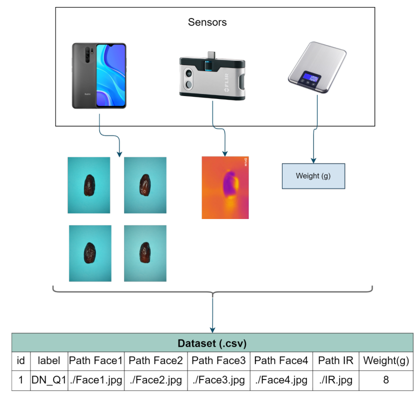

Fig. 2. Data Collection Process Overview

Following data collection, our approach diverges into two distinct scenarios for classification. In the first scenario, we rely on the fusion of four angles of views of the date as illustrated in Fig. 3.

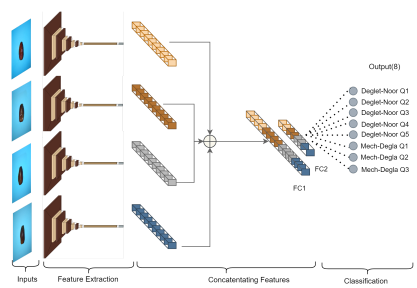

Fig. 3. Illustration showcasing the fusion of multiple angles of date fruit.

In the second scenario as in Fig. 4, we enrich the dataset by incorporating additional sensor data, such as thermal images and weight scale measurements. This comprehensive approach allows us to explore the impact of multi-modal fusion on date fruit classification accuracy. Each scenario is examined using four models (CNN, VGG16, ResNet50, and MobileNet). Each view's angle of the date and thermal image, undergoes feature extraction using these models. After this step, we obtain five collections of feature data. Following feature extraction by these deep learning models, we concatenate the extracted features with the weight scale data of each date to create a fused representation of the date fruit.

Our classification process encompasses training and validation phases, executed over 100 epochs to ensure model

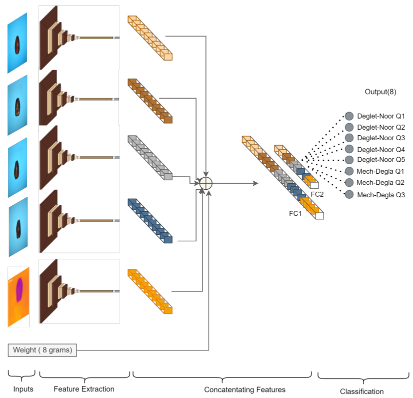

Fig. 4. Illustration showcasing the fusion of multiple angles of date fruit with the Thermal image , and weight metric.

convergence. The training data allows our models to learn and recognize patterns within the multi-modal dataset, while the validation data assesses model performance and generalization.

We evaluate the performance of our proposed method using various metrics, including accuracy, precision, recall, and F1 score. These metrics offer insights into classification accuracy and the overall effectiveness of our method.

A. Using basic CNN

Each of the two scenarios involves the use of the Basic CNN, which comprises five identical sub-CNNs, each defined as 'classifier_F1,' 'classifier_F2,' 'classifier_F3,' 'classifier_F4,' and 'classifier_IR.' Each sub-CNN follows a similar structure:

- Convolutional Layers: Each sub-CNN includes four sets of Conv2D layers with 64, 64, 128, and 128 filters, respectively, using a (3,3) kernel size and ReLU activation.
- MaxPooling Layers: Corresponding to these Convolutional Layers, each sub-CNN includes four MaxPooling2D layers with a (2,2) pool size.
- Flattening Layer: Each sub-CNN features a Flattening Layer to prepare the output for fully connected layers.
- Dense Layers: Each sub-CNN includes a dense layer with 512 units and ReLU activation.

The architecture then combines the outputs of these five sub-CNN with a tabular input (Weight) and passes them through a series of dense layers:

- Concatenation: The outputs of the five sub-CNNs and the tabular input are concatenated.
- Dense Layer: A dense layer with the number of units equal to the concatenated shape and ReLU activation.
- Output Layer: The architecture includes a dense output layer with eight classes and softmax activation.

B. Using Pre-trained models

Each of the two scenarios involves the use of Three types of Pre-trained models. We have selected the most renowned methods, namely VGG16, ResNet50, and MobileNet, and conducted experimental comparisons This architecture consists of five sub-models, each using the pre-trained network as a base model for feature extraction. Each sub-model ('classifier_F1,' 'classifier_F2,' 'classifier_F3,' 'classifier_F4,' and 'classifier_IR.'.) follows a similar structure:

- Base Model: pre-trained model (eg: VGG16) is used as the base model with weights pre-trained on ImageNet data.
- The layers in each base model are made trainable for fine-tuning.
- Global Average Pooling: is added to the base model's output.
- Dropout Layer: Dropout is applied to prevent overfitting.

The tabular data ('Weight') is combined with the output of the five sub-models using concatenation. The combined features are then passed through a series of dense layers, including two dense layers with ReLU activation and dropout layers to prevent overfitting. Finally, the architecture outputs a softmax activation layer to classify the input data into eight classes.

All the steps described above are part of the architecture for the first scenario. However, in the first scenario, we do not include the fifth sub-CNN models or the weight feature.

III. Results and Discusion

This section presents the results of our experiments comparing the performance of the two scenario proposed with the four different model: basic CNN, VGG16, Resnet50, and Mobilenet . We discuss the training and validation processes for each model of two scenario. The experiments were conducted on Processeur : Intel Xeon (R) E5-2660 v3 @ de 2.60 GHz x 20, 64 GB de RAM, 2 TB HDD, RedHat Enterprise Linux Server 7.2, 64 bit. We utilized our self-made dataset, which consisted of images of the Deglet Noor and Mech Degla varieties of date fruit. The dataset included 80% for training, 20% for validation and test. According to our exprimental which involved and summarized in Table I, This table provides a comprehensive evaluation of four different machine learning models (CNN, VGG16, ResNet50, and MobileNet) across two distinct scenarios (Scenario I and Scenario II). Performance metrics including accuracy, precision, F1-score, recall, training accuracy, and validation accuracy. In Scenario I, four different deep learning models were evaluated for their performance in classifying date fruit images. CNN achieved the highest accuracy at 80% on the validation dataset, with a precision score of 73% and a balanced F1-score of 77%. This suggests a good trade-off between precision and recall, and a recall score of 80% indicates its ability to capture actual positive cases effectively.

VGG16, ResNet50, and MobileNet also demonstrated reasonable performance in Scenario I, with accuracy ranging from 65% to 76%, precision rates of 62% to 70%, and balanced F1-scores of 61% to 71%. These models showed varying levels of effectiveness in distinguishing true positive cases from false positives and capturing actual positive cases.

TABLE I
PERFORMANCE METRICS FOR FOUR MODELS OF TWO SCENARIOS

Model Type	Scenario	Accuracy	Precision	F1-score	Recall	Accuracy(train)	Accuracy(val)
CNN	Scenario I [a]	80	73	77	80	100	80
	Scenario II [b]	89	80	88	90	100	89
VGG16	Scenario I	75	70	71	75	79	75
	Scenario II	82	83	79	82	86	82
ResNet50	Scenario I	76	70	71	76	90	76
	Scenario II	78	67	75	78	93	78
MobileNet	Scenario I	65	62	61	65	66	65
	Scenario II	74	69	70	74	76	74

[a]Scenario I: Using four views of date fruit. [b] Scenario II: Using four views of date fruit, thermal image, and weight.

In Scenario II, additional features such as thermal images and weight were incorporated into the evaluation. CNN significantly improved its performance with an accuracy of 89%, but a perfect training accuracy of 100% suggested potential overfitting. VGG16 maintained strong performance with an accuracy of 82%, excelling in precision (83%) and demonstrating effective generalization.

ResNet50 continued to perform well in Scenario II with an accuracy of 78%, and MobileNet also improved with a 74% accuracy. These models showed good trade-offs between precision and recall, indicating their ability to classify date fruit images with additional features.

The results suggest that different models have varying levels of performance in classifying date fruit. CNN and VGG16 performed well in both scenarios, with CNN showing significant improvement when additional features were incorporated. ResNet50 also demonstrated consistent performance, while MobileNet showed reasonable results, especially in Scenario II. Our findings indicate that the inclusion of thermal images and weight metrics can significantly improve performance. The results of the paper demonstrate the power of the late fusion approach for combining multiple sources of information for the date fruit sorting process but also highlight the importance of selecting an appropriate model architecture.

IV. CONCLUSION

The automation of date fruit sorting is of critical significance in modern agriculture and the food industry. It not only enhances efficiency but also ensures consistent quality and precision throughout the sorting process. Our research has unveiled an innovative and efficient approach to meet this critical need. The demands for precision and efficiency in date fruit sorting are effectively addressed through our multi-modal fusion and deep learning techniques.

We have taken a comprehensive approach to date fruit quality assessment by categorizing date fruits into eight distinct classes. These classes encompass five grades of Deghlet Noor and three grades of Mech-Degla varieties. Our method transcends traditional sorting processes, providing a holistic evaluation of various date fruit characteristics.

Our approach has demonstrated its prowess through rigorous evaluation. We conducted evaluations on four deep learning models—CNN, VGG16, ResNet50, and MobileNet—in two distinct scenarios (Scenario I: combining four images of multi-angle of date fruit, Scenario II: incorporating thermal images and weight metric with previous four images). Notably, CNN exhibited remarkable accuracy improvements in Scenario II, where thermal images and weight metrics were included. With accuracy, precision, F1-score, recall, and training accuracy reaching 89%, 80%, 88%, 90%, and 100%, respectively, our approach has significantly enhanced overall classification performance.

In conclusion, our research not only addresses the immediate need for more efficient and accurate automated date fruit sorting systems but also provides a comprehensive assessment of all features that determine date fruit quality. Furthermore, it paves the way for further innovations in the agricultural sector, ensuring the quality and consistency of date fruit production.

REFERENCES

[1] Al Ohali, Y. (2011): Computer vision-based date fruit grading system: Design and implementation. Journal Of King Saud University - Computer and Information Sciences. 23, 29-36, https://www.sciencedirect.com/science/article/pii/S1319157810000054.

[2] Albarrak, K., Gulzar, Y., Hamid, Y., Mehmood, A., Soomro, A. B. (2022). A Deep Learning-Based Model for Date Fruit Classification. Sustainability, 14(10). Retrieved from https://www.mdpi.com/2071-1050/14/10/6339

[3] Altaheri, Hamdi, Mansour Alsulaiman, and Ghulam Muhammad (2019): Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep Learning. IEEE Access 7 (September): 117115–33. https://doi.org/10.1109/ACCESS.2019.2936536.

[4] Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., Lladó, X. (2018): Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artificial Intelligence in Medicine. Doi: 10.1016/j.artmed.2018.08.008.

[5] Faisal, M., Alsulaiman, M., Arafah, M., Mekhtiche, M. A. (2020). IHDS: Intelligent Harvesting Decision System for Date Fruit Based on Maturity Stage Using Deep Learning and Computer Vision. in IEEE Access, 8, 67985-16799.

[6] Farahnakian, Fahimeh, and Jukka Heikkonen. (2020): Deep Learning Based Multi-Modal Fusion Architectures for Maritime Vessel Detection, Remote Sensing 12, no. 16: 2509. https://doi.org/10.3390/rs12162509.

[7] Jia, P., Zhang, M., Yu, W., Shen, F., Shen, Y. (2016): Convolutional neural network based classification for hyperspectral data. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). doi:10.1109/igarss.2016.7730323.

[8] Nasiri Amin, Amin Taheri-Garavand, and Yu-Dong Zhang (2019): Image-Based Deep Learning Automated Sorting of Date Fruit. Postharvest Biology and Technology 153: 133–41. https://doi.org/https://doi.org/10.1016/j.postharvbio.2019.04.003.

[9] Sören Richard Stahlschmidt, Benjamin Ulfenborg, Jane Synnergren, (2022): Multimodal deep learning for biomedical data fusion: a review, Briefings in Bioinformatics, Volume 23, Issue 2, https://doi.org/10.1093/bib/bbab569.