

Multi-Model CNN-RNN-LSTM Based Fruit Recognition and Classification

Harmandeep Singh Gill^{1,*}, Osamah Ibrahim Khalaf², Youseef Alotaibi³, Saleh Alghamdi⁴ and Fawaz Allassery⁵

¹Department of Computer Science, Guru Arjan Dev Khalsa College Chohla Sahib, Punjab, 143408, India

²Al-Nahrain University, Al-Nahrain Nano-Renewable Energy Research Center, Baghdad, 964, Iraq

³Department of Computer Science, College of Computer and Information Systems, Umm Al-Qura University, Makkah, 21955, Saudi Arabia

⁴Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, 21944, Saudi Arabia

⁵Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif, 21944, Saudi Arabia

*Corresponding Author: Harmandeep Singh Gill. Email: profhdsgill@gmail.com

Received: 12 August 2021; Accepted: 18 November 2021

Abstract: Contemporary vision and pattern recognition issues such as image, face, fingerprint identification, and recognition, DNA sequencing, often have a large number of properties and classes. To handle such types of complex problems, one type of feature descriptor is not enough. To overcome these issues, this paper proposed a multi-model recognition and classification strategy using multi-feature fusion approaches. One of the growing topics in computer and machine vision is fruit and vegetable identification and categorization. A fruit identification system may be employed to assist customers and purchasers in identifying the species and quality of fruit. Using Convolution Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) deep learning applications, a multi-model fruit image identification system was created. For performance assessment in terms of accuracy analysis, the proposed framework is compared to ANFIS, RNN, CNN, and RNN-CNN. The motivation for adopting deep learning is that these models categorize pictures without the need for any intervention or process. The suggested fruit recognition method offers efficient and promising results, according to the findings of the experiments in terms of accuracy and F-measure performance analysis.

Keywords: CNN; RNN; LSTM; deep learning; fruit image; classification

1 Introduction

Fruits are an essential part of a balanced diet and provide many health advantages. Many kinds of fruits are available all year, while others are only available during certain seasons. Agriculture continues to be a major contributor to India's economy. Agricultural land accounts for 70% of the land in India. India is rated 3rd among the world's leading fruit producers. As a result, employing deep learning systems to classify fruits is advantageous for both marketers and consumers. Computer science and information



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

technology are becoming increasingly engaged in the agricultural industry in the current situation. To offer high-quality fruits to customers, artificial intelligence and various soft computing-based methods are utilized for fruit categorization.

Fruit categorization and identification software are essential since it aids in the improvement of fruit quality. In the market, recognizing a fruit is a difficult job. It's tough to categorize and price anything manually. Manually counting ripe fruits and evaluating their quality is a difficult job. Some significant issues with fruit production, marketing, and storage include increasing labor prices, a lack of trained employees, and lower storage costs, among others. The Soft computer vision system offers important information on the variety and quality of fruits, lowering costs, ensuring quality maintenance standards, and providing essential information. The main soft computing methods for creating computer-assisted recognition systems include artificial neural networks (ANN), fuzzy logic, and evolutionary computation models. ANN has the proclivity to learn from samples and classify previously unknown patterns.

One of the most recent trends in the field of computer vision is fruit identification and categorization. The number of features, kinds of features, selection of features from extracted data, and type of classifier employed all influence the accuracy of a fruit identification system. Fruit images taken in poor weather have lesser quality and conceal visible features. As a result, fruit image enhancement techniques are required to increase the nature and features of fruit images.

Fruit categorization and identification Fruit categorization has been identified as one of the burgeoning fields of computer science. The fruit classification system's accuracy is determined by the quality of the collected fruit images, the number of extracted features, the kinds of features, and the selection of optimum classification features from the retrieved features, as well as the type of classifier employed. Images were taken in poor weather reduce the visibility of the obtained fruit images and conceal key features.

Classification and recognition are effective tools for classifying fruit images and determining what kinds of fruits they include. On the other hand, color image classification methods suffer from poor visibility issues, particularly in low-light situations [1,2]. Image enhancement methods help to improve the visual superiority of images, making them more valuable for future vision applications [3].

In [4], the authors presented a method for counting and recognizing fruits from images in cluttered greenhouses. Bag of words data approach used to recognize peppers. A support vector window is allocated at the initial stage to provide better image information for recognition. Image identification and classification are important goals in image processing, and they have gotten a lot of attention from academics in recent years. In the form of shape, size, color, texture, and intensity, an image carries information about various scenes.

There is a possibility that some noise will be added to the obtained image during acquisition. This may be capturing, transferring, or saving. Reasons may be the improper shuttering speed of the camera during image capturing, inadequate light, and non-linear mapping of the image intensity. Precise classification of various fruits is an important topic in recent times [5]. Many researchers have done very much work in the automatic classification of fruits and vegetables. Automatic classification helps buyers and sellers in different ways. Seller identifies the variety and rate list from the classifier. Buyers automatically check the price list and variety of the specific fruit with the support of an automatic classification system, which is manually difficult to sort and buy.

The rest of the paper is described as Related work is discussed in Section 2. The proposed work methodology is discussed in Section 3. Image classification feature fusion is discussed in Section 4. Experimental results are discussed in Section 5. The conclusion is discussed in Section 6.

2 Related Work

Many recognition and classification systems can be found in the literature survey to automatically inspect the fruits for diseases detection, a maturity phase, and category recognition, etc. The method used by [6] to classify the bananas by hue channel and CIElab. The particle swarm optimization (PSO) algorithm involved tuning the fuzzy parameters. [7] used a multi-class kernel support vector machine (k-SVM) for the classification of fruits. SVM's were trained using 5-fold stratified cross-validation with the reduced feature vector. Color, texture, and shape features combination were used during the classification procedure. [8] recognize fruit form images deep learning to identify strengths and weaknesses of a classification scheme for standard datasets. In this work, 60 different fruit varieties are recognized from 38409 images. The essential condition of fruit recognition is to select a high-quality image dataset. [9] classify fruits by wavelet-entropy and a feed-forward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization algorithm. classification performance for 1653 fruit images from 18 different categories shows the superiority of the classification scheme with the state-of-the-art schemes. [10] automatically classify fruits and vegetables based on a unified approach by combining features for training and classification.

Convolution Neural Network (CNN) is a special multi-layered feed-forward unsupervised neural network designed to deal with image classification [11]. The convolution layer of the convolutional neural network (CNN) is also called the feature extraction layer. In the proposed work, fruit image features are extracted using CNN to originate coarse and fine labels. distinct optimal features are labeled to classify the fruits [12]. CNN is a multi-layered neural network, specifically used to extract image features without concerns about feature selection problems. CNN is good for extracting features immediately from collected images, while RNNs are superior for jobs where sequential modeling is more essential. Based on the above description, it makes sense to use a CNN for classification tasks such as sentiment classification, where sentiment is usually determined by a few key phrases, and RNNs for sequence modeling tasks such as language modeling, machine translation, or image captioning, where context dependencies must be modeled flexibly. CNN can learn to categorize a sentence or a paragraph, while RNNs are typically excellent at predicting what follows next in a sequence.

Recurrent Neural Networks (RNNs) are a kind of Neural Network in which the output from the previous step is used as input in the current phase. Sequence Classification, Sentiment Classification, and Image and Video Classification are the most common applications for RNNs. RNNs are a kind of artificial neural network in which nodes' connections form a directed graph in sequential order. It is essentially a chain of neural network pieces connected. Each one is sending a message to the next in line. An RNN is taught to identify patterns over time, while a CNN is taught to recognize patterns across space.

Hoch Reiter and Schmid Huber developed long short-term memory (LSTM) networks in 1997 as shown in Fig. 1, and they have set accuracy records in a variety of application areas. Automatic picture captioning was enhanced using LSTM and CNN. There have been attempts to speed LSTMs using hardware accelerators because to the compute and memory overheads of executing them.

2.1 Motivation

After the detailed analysis of the existing classification techniques, the following shortcomings are investigated:

1. Most existing classification approaches may suffer from spectral reflectance values.
2. Heterogeneous nature of images is also a major barrier during classification, which results in poor classification results.
3. Shape, color, texture, and intensity similarity among species of distinct fruits.

4. Extremely high variation in one type, which relies on the fruit maturity and ripeness stage.

This paper presents the multi-model CNN-RNN-LSTM deep learning framework for fruit image classification to overcome the above-mentioned shortcomings.

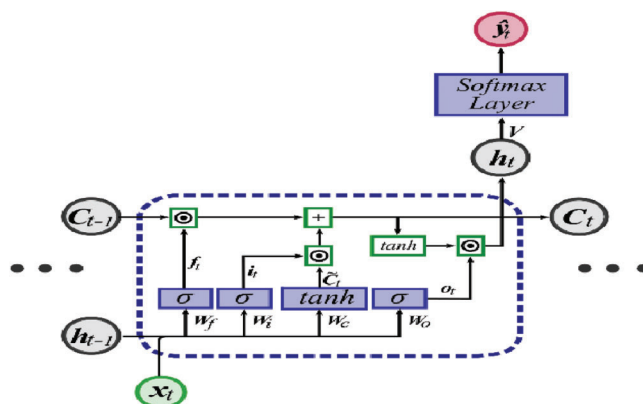


Figure 1: LSTM architecture

2.2 Contribution

The multi-model fruit classification model is developed using deep learning applications. In the proposed work, CNN is used to extract features with the help of convolution layers. Optimal features extracted are then labeled using RNN. RNN labels the extracted features based on the category of fruits. Labeling is done by the fine and coarse strategy of the recurrent model. LSTM is used to classify the fruits by optimal features extracted and selected by CNN and RNN.

The main contributions of the paper are as follows:

- Utilization of deep learning applications: CNN, RNN, LSTM to classify the fruit images,
- CNN, RNN, LSTM combined to develop a multi-model fruit recognition, generally used for recognition and classification of signals, texts, and speech. This work investigates all these approaches to perform recognition and classification of fruits.
- The experiment performed using the proposed approach obtained quite efficient and promising fruits classification results.

3 Methodology

The proposed framework uses multi-modal deep learning based on fruit recognition. Fruit images are recognized using a multi-model classifier. A multi-model classifier is employed to train fine and coarse labels. The Convolution Neural Network is utilized to develop discriminative characteristics, and the Recurrent Neural Network (RNN) is utilized to develop sequential labels, and LSTM is involved in classifying the fruits by optimal extracted and labeled features.

In general, image recognition and classification rely on a combination of structural, statistical, and spectral approaches. Mean, variance, and entropy measures are primarily used during statistical measurements. For object recognition and identification, structural analysis is utilized. To recognize images based on intensity, texture, and color features, spectral approaches are involved. Sometimes, it is unusual to classify images using the same.

To address the aforementioned problems, this article uses CNN, RNN, and LSTM deep learning methods to classify and identify fruit images. The proposed simulation and experiment work for fruit recognition using deep learning applications is divided into the following phases.

3.1 Image Acquisition

For experiment and simulation work, the preliminary step is to acquire images. After image acquisition, images are selected and then set images size. For experiment work, it is essential to sort images according to size. For simulation, it is mandatory to set parameters for classification. Due to the heterogeneous nature of fruits, it is mandatory to classify them based on specific features. Intensity and texture classification is a challenging task, which is efficiently handled by CNN.

3.2 Pre-processing

Pre-processing performs at the abstract level on the fruit images. The main task is to improve the intensity levels to overcome the issues of unwanted distortions and to enhance those parts of the acquired image that are further used during features selection and extraction processing [13].

3.3 Feature-extraction

The best features are retrieved after pre-processing. The real numbers produced by applying some mathematical expression to image data are referred to as features. In recognition applications, choosing features is the most important job. Based on the performance of classifiers, optimal feature subset selection improves the recognition system's classification rate and accuracy.

3.4 Image Classification

A classification problem deals with recognizing a given input with one of the distinct classifiers. In the proposed study, deep learning applications are involved to classify the fruit images. Fruit image Classification procedure includes image enhancement, object detection, recognition, and classification.

3.5 Evaluation

Evaluation is the final step during image classification. Multi-model is tested for the fruit image classification process in the proposed work.

Following Fig. 2 shows the phases of the proposed image classification multi-model deep learning model.



Figure 2: Phases of the proposed work

4 Proposed Fruit Classification Multi-Model

We proposed CNN, RNN, and LSTM deep learning algorithms for the recognition of fruits. In recent times, deep learning becomes a popular recognition and classification scheme for images. Deep learning is an Artificial neural network unsupervised learning branch, emerged as a promising technique for image recognition and classification in recent times.

The primary aim of the proposed classification scheme is to label the hierarchical labels of the fruit images. CNN is mainly used to extract features and then label them by coarse and fine. RNN distinct the

extracted features by dynamic behavior development. LSTM classifies the fruit images by extracting and labeling fruits based on optimal features by CNN and RNN respectively. LSTM is combined with RNN by a memory cell to train the proposed classifier [14]. CNN consists of input, convolution, pooling, and fully connected layers. Fig. 3 shows the convolution neural network for image feature extraction.

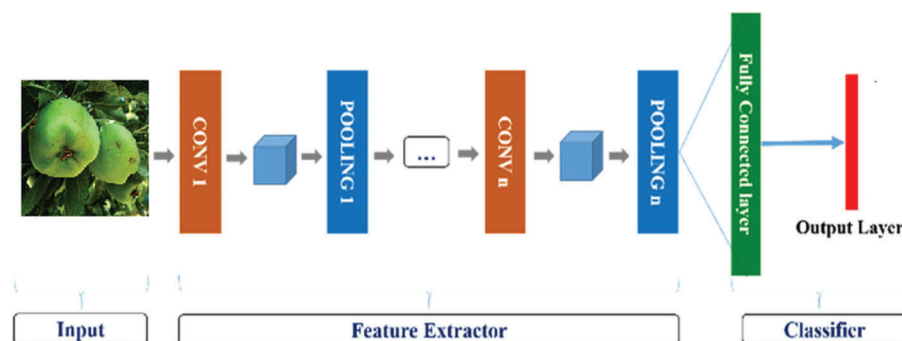


Figure 3: Convolution neural network feature extraction model

Feature extraction by convolution layer has several layers to extract different features. To extract features from available layers, the following equation is used:

$$x_j^i = f \left(\sum_{u \in M_j} \times k_j^i + b_j^i \right), \quad (1)$$

where, x_j^i denotes the feature map in the j^{th} layer of the network, k_j^i represents the j^{th} convolution filter, M_j is the set of features map.

A Convolution neural network-oriented generator is used to generate coarse and fine labels in the proposed study. We updated Y. Guo et al. [15] earlier work in by replacing the final layer of the traditional CNN structure with two layers (Fine and Coarse) and providing supervisory signals to the coarse and fine categories independently. CNN is unable to take advantage of the interaction between two supervisory signals on its own. In this scenario, a CNN-based developer will be unable to classify the hierarchical labels if the hierarchy has an adjustable size.

RNN is a kind of artificial neural network that may hierarchically generate dynamic length labeling. between the edges, units form a directed cycle (see Fig. 4). It develops the temporal nature with dynamic sequences of varying sizes that refuse to vanish, as well as exploding gradient issues, as gradients must propagate down many layers of RNN. As a result, it is unable to create long-term dynamics. Due to their extremely dynamic nature, recurrent neural networks emerged, while multilayer feed-forward networks have static mappings. RNNs have been utilized in a variety of fields and have applications in associative memory, optimization, and generalization. RNNs are best for classifying time-series data because the feedback and current value are fed back into the network, and the output includes traces of values stored in the memory, which improves classification performance and provides better results than conventional feed-forward networks.

The LSTM architecture is a kind of RNN. LSTMs were created to represent temporal sequences, and RNNs' long-range dependencies and memory backup play a critical role, making them more accurate and effective than traditional RNNs as shown in Fig. 4. The technique is used after the data has been pre-processed, which includes the removal of undesirable, missing, and null signal values. Therefore, LSTM is used in the proposed framework to overcome the issues with RNN. LSTM provides a solution by

adding a memory cell (C) to encode knowledge at each step every time as shown in the figure. The working of the memory cell is controlled by an input gate (It), Forget gate (Ft), and Output gate (GT). Input read during classification is monitored by these gates. These control gates also help the LSTM to go through the input to a hidden state of deep learning unsupervised architecture input without any effect on the output.

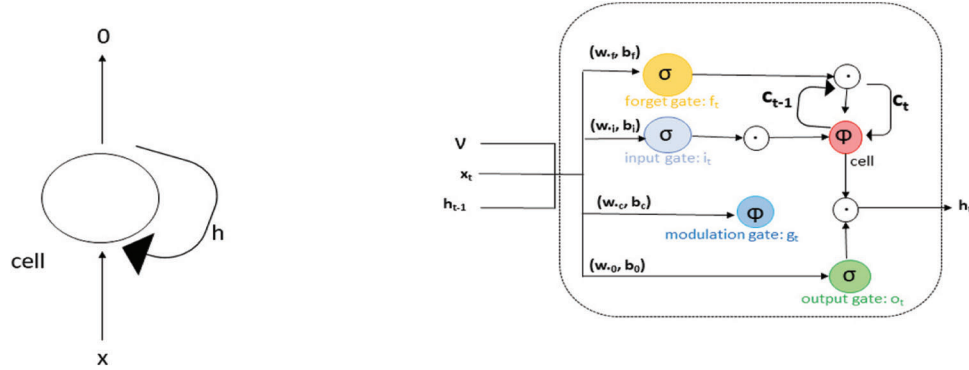


Figure 4: LSTM+RNN architecture

The gates and updating of LSTM at t are defined as:

$$i_t = s(W_{xi}x_t + W_{hi}h_{t-1} + W_{vi}v + b_i) \quad (2)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{vf}v + b_f) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{vo}v + b_o) \quad (4)$$

$$g_t = \gamma(W_{xc}x_t + W_{hc}h_{t-1} + W_{vc}v + b_c) \quad (5)$$

$$c_t = f_t \odot C_{t-1} + i_t \odot g_t \quad (6)$$

$$h_t = o_t \odot \gamma(c_t) \quad (7)$$

where \odot represents the product operation, Σ is the sigmoid function ($\sigma(x) = \frac{1}{1 + e^{-x}}$) and γ is the hyperbolic tangent function ($\gamma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$).

The definition for other symbols is: i_t , f_t , o_t , and g_t denote the input gate, forget gate, output gate, and input modulation gate, respectively. x , h , v , and c represent the input vector, hidden state, image visual feature, and memory cell, respectively.

During classification training of the multi-model, we use the soft-max function to optimize the assumptions about coarse and fine labels collectively. The proposed training model uses the following equation.

$$Loss = -\frac{1}{N} \sum_{i=1}^N \left(\sum_{t=1}^T \sum_{j=1}^{C+F} 1\{x_t^i = j\} \log p_j \right) \quad (8)$$

The main objective of the proposed technique is to develop hierarchical labels for fruit images. The labels are calculated in a coarse-to-fine pattern. To finish it, C (coarse) is combined with F (fine) as super-categories. These labels may provide useful information for predicting the appearance of finer labels [16].

Our proposed scheme defines the super categories of fruit image features along with training and classification. Classification has efficient results for different lengths without the need to design distinct classification networks for different feature categories.

In the proposed fruit model, CNN is used to extract optimal image features. Features are extracted from various fully connected layers of CNN. RNN is used to label the best-extracted features, and LSTM is then used to categorize the fruits based on the best features.

As shown in Fig. 5, we developed a novel multi-model classification scheme by combining fully connected layer CNN, soft-max, and classification layers RNN+LSTM [17].

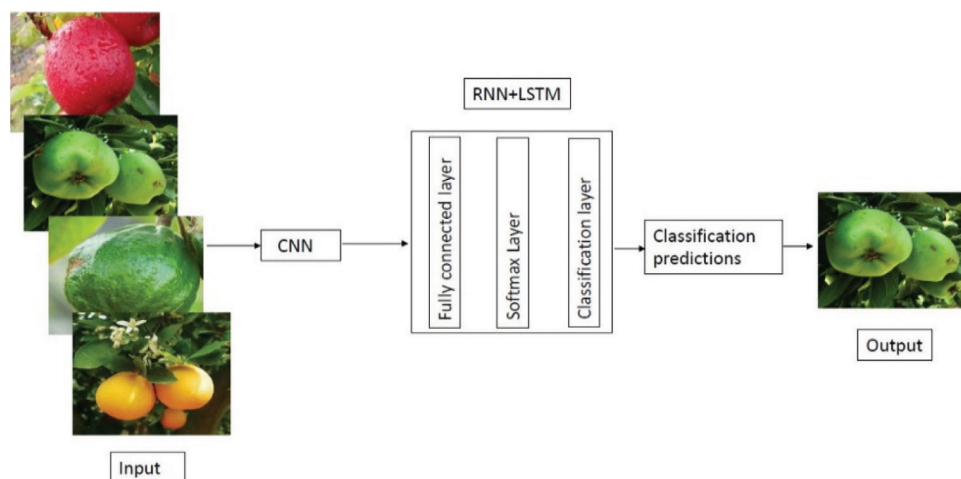


Figure 5: Proposed multi-model fruit classification model

4.1 Image Classification and Feature Fusion

The objective here is to improve the overall classification process by combining the set of all features. The divide and conquer approach is used here to avoid the feature mapping problem. The proposed approach is called multi-feature mapping. CNN, RNN, LSTM proposed model combines all the acquired, selected, and classification features.

As per the literature review, the proposed approach is novel and was not used before for feature classification and fusion. We train the classification model using three different classes. Class one is CNN oriented, where features are extracted. Class second is RNN oriented, where we divide the acquired features into categories to select the optimal and matched features. Finally, the class third is a classification based on LSTM. The complexity of the detection issue at test time and the variability contained within the training dataset determine the representational capacity of features retrieved by a CNN. It is anticipated that an intelligent integration of both modalities of characteristics would result in improved detection performance [18].

The method of feature fusion is extensively utilized in a variety of fields, including image recognition and classification. Feature fusion aims to extract the most discriminative information from many input characteristics while also removing redundant data. In the field of image classification, feature fusion methods are divided into two categories: serial feature fusion and parallel feature fusion [19].

Let X , Y , and Z represent three feature spaces, Ω represent the pattern sample space, and ξ represent a randomly chosen sample in Ω . Furthermore α , β and γ are the feature vectors of ξ where $\alpha \in X$, $\beta \in Y$, and $\gamma \in Z$, respectively. We use a weighted serial feature fusion technique to merge three feature vectors in the suggested

method, which modifies the serial feature fusion strategy. After normalization, we name the feature vectors of the global, appearance, and texture features f_1 , f_2 , and f_3 , respectively [20–23].

Then (10), where w_1 , w_2 , and w_3 are the weights of f_1 , f_2 , and f_3 , respectively, yields the fusion feature F .

$$F = \begin{bmatrix} W1 & f1 \\ W2 & f2 \\ W3 & f3 \end{bmatrix} \quad (9)$$

The values of weights w_1 , w_2 , and w_3 are set by one recognition layer of f_1 , f_2 , and f_3 , which are denoted by A_1 , A_2 , and A_3 , respectively. We calculate w_1 , w_2 , and w_3 using (10) to (12).

$$W1 = \frac{A1}{A1 + A2 + A3} \quad (10)$$

$$W2 = \frac{A2}{A1 + A2 + A3} \quad (11)$$

$$W3 = \frac{A3}{A1 + A2 + A3} \quad (12)$$

To evaluate the performance of the proposed fruit recognition and classification model, accuracy and f-measures are presented. Tab. 1 shows the performance in terms of accuracy, sensitivity analysis, precision, and recall of classification models for different fruit images (Tab. 1).

Table 1: Performance results of the proposed model

Image (size)	AC %	PR %	SE %	SP %
Red apple	98.51	99.07	94.57	96.89
Green apple	98.54	99.78	94.32	96.02
Golden apple	98.56	99.07	94.22	96.34
Guava	98.03	99.45	94.09	96.27
Orange	98.23	99.58	94.91	96.89

5 Experimental Results and Discussions

In the quest for finding the best fruit classification and recognition model, this paper analyzes various features for fruit recognition. Intensity, texture, shape-based, and size descriptors are used. For training and testing, existing classification schemes are compared with the proposed one. We perform our experiments on 10 fruit images. The performance analysis and quantitative analysis are discussed in the subsequent sections. The proposed scheme outperforms CNN, RNN, ANFIS, and RNN-CNN with the help of accuracy analysis and F-measure.

5.1 Performance Analysis

A multi-model fruit classification scheme has been implemented on Intel Core i3-4005U (1.7 GHz, 3MB L3 cache) processor and MATLAB tool 2013A to design the simulation environment. It has been observed that the experimental results of the proposed scheme outperform existing techniques during fruit classification.

5.1.1 Accuracy Analysis

When applied to data, accuracy is a performance metric of an image classification that represents the proportional number of times the created model is accurate. Some assess the accuracy of fruit data via image classification [24–28].

In the same way, the number of 0's determines the error rate. The accuracy can be computed as:

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (13)$$

where $T_p + T_n + F_p + F_n$ defines true positive, true negative, false positive, and false negative, respectively. Therefore, accuracy ranges from 0 to 100. The proposed scheme has more accuracy as compared to the existing schemes.

Fig. 6 represents the accuracy analysis of the multi-model classifier and existing fruit classification schemes. It has been observed that the proposed technique outperforms others because proposed scheme has a better accuracy rate.

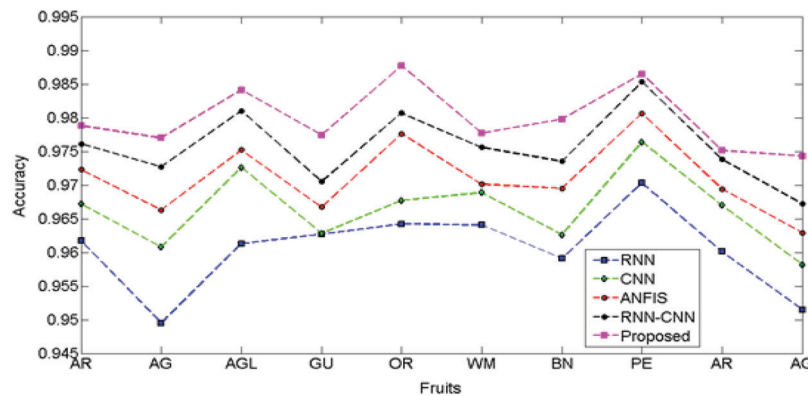


Figure 6: Accuracy analysis

5.1.2 F-Measure

F-measure is used to evaluate the accuracy of the proposed classification scheme as shown in Fig. 7. F-measure consists of p (precision) and r (recall). Matching positive results are defined by p and dividing the number of all positive results by r , which is the number of all positive matching results. F-measure is harmonic mean for all p and r values.

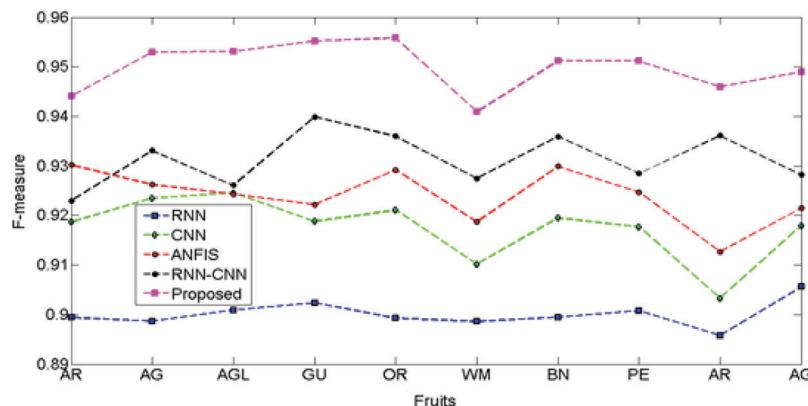


Figure 7: F-measure

Here 1 value for F measure means best classifier result and 0 means worst classification results. Precision is the fraction of relevant values among the obtained values and recall is the fraction of relevant values that have been obtained from the total obtained classification values. Precise is positive value and recall is called sensitivity during the classification process.

F-measure is defined as:

$$F \text{ measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

Precision and recall are defined as:

$$\text{Precision} = \frac{T_P}{T_P + F_P} \quad (15)$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \quad (16)$$

Sensitivity

Sensitivity is true positive rate having best classification values as shown in Fig. 8 and is defined as:

$$\text{Specificity} = \frac{T_P}{T_P + F_N} \quad (17)$$

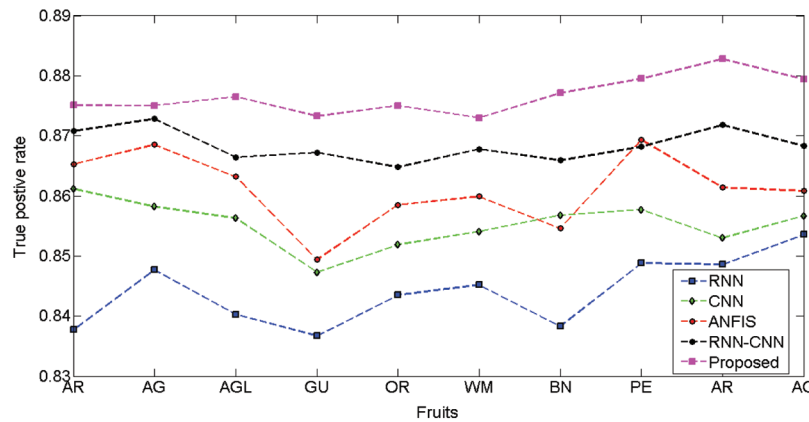


Figure 8: True positive rate (sensitivity)

True positive (TP), means obtained values are correct and classifier rate is best and classification is correctly done (best). False-positive (FP) means values are not correct but classification results are still good (average). False-negative (FN) means values are correct but the classification is poor (worst).

Specificity

Specificity is true negative rate having average classification values as shown in Fig. 9 and is defined as:

$$\text{Specificity} = \frac{T_N}{T_N + F_P} \quad (18)$$

True negative (TN) means values are not correct and classification is not correctly defined (worst). False-positive (FP) means values are not correct but classification results are still good (average). False-negative (FN) means values are correct but the classification is poor (worst).

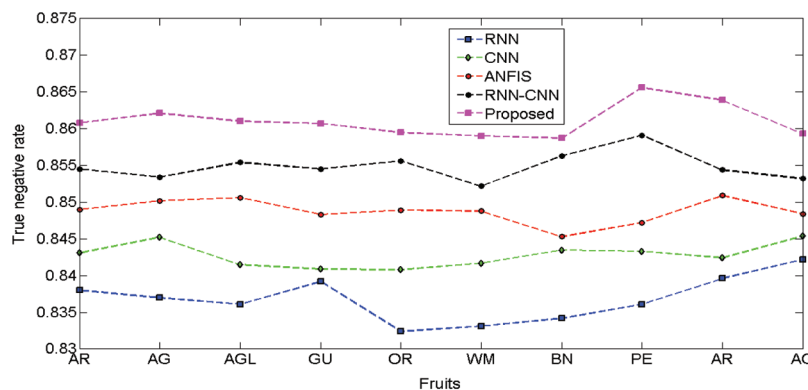


Figure 9: True negative rate (specificity)

5.2 Discussion

Our work aimed to recognize fruits automatically from images. During recognition, the main challenge is that fruits have heterogeneous nature due to color, intensity, and texture features. In the earlier research related to fruit recognition and classification, several schemes were adapted. Each scheme has some limitations, some approaches are appropriate during the initial classification of fruits based on shape and size features but failed during texture recognition. Traditional fruit recognition approached failed during the grading based on color and intensity levels.

To recognize fruit by setting benchmarks is failed during bad weather of environment. Seasonal changes are also the major barriers to recognition and classification based on pre-planned methods. To overcome these issues, deep learning can adopt the dynamic approach based on training features to classify based on best-matched features.

In the proposed research project, deep learning applications: CNN, RNN, and LSTM are utilized to recognize the fruits. Fruits are recognized in their intensity, texture, and shape features. The major benefit of using the CNN approach for feature extraction does not require and hand-crafted feature extraction policy. CNN utilizes different filters to extract the potential image features automatically.

RNN can label the automatic features quite efficiently. RNN uses the coarse and fine label facility to recognize optimal features impressively. Train the deep learning fruit recognition model robustly.

The multi-model CNN-RNN-LSTM fruit classification model is compared with previous studies on fruit and vegetable classification schemes. The proposed scheme has strong adaptability to the variances during the features extraction and classification process.

6 Conclusion and Future Work

Sometimes, when handling complex classification problems, one feature descriptor is not enough to justify the outcomes. Therefore, efficient and optimal feature fusion approaches become necessary. Some classifiers are quite justified for normal classifications but do not properly handle the fruit classification due to their diverse nature. This paper introduces a novel fruit recognition and classification strategy, namely the multi-model CNN-RNN-LSTM approach.

In this paper, we proposed to ensemble CNN, RNN, and LSTM deep learning algorithms to classify the fruits. Fruit classification is an important task in fruit recognition and classification. CNN extracts the image features through different convolution layers. RNN labeled the distinct features and finally, LSTM classify the fruits based on optimal extracted features. Extensive experiments have been done through 10 fruit images by proposed and existing classification techniques such as CNN, RNN, ANFIS, and RNN-CNN on fruit

images. It has been concluded that the proposed classification technique outperforms existing image classification techniques in terms of accuracy analysis and F-measure.

Future work will focus on exploring more functions for growth monitoring, ripeness detection. Moreover, future work will investigate automatic grading and labeling of fruits in the A, B, C, D categories.

Acknowledgement: We deeply acknowledge Taif University for supporting this study through Taif University Researchers Supporting Project Number (TURSP-2020/150), Taif University, Taif, Saudi Arabia.

Funding Statement: This research is funded by Taif University, TURSP-2020/150.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] L. Zheng, H. Shi and M. Gu, "Infrared traffic image enhancement algorithm based on dark channel prior and gamma correction," *Modern Physics Letters B*, vol. 31, no. 19–21, pp. 1740044, 2017.
- [2] D. Singh and V. Kumar, "Comprehensive survey on haze removal techniques," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9595–9620, 2018.
- [3] D. Singh and V. Kumar, "Single image haze removal using integrated dark and bright channel prior," *Modern Physics Letters B*, vol. 32, no. 4, pp. 850051, 2018.
- [4] Y. Song, C. A. Glasbey, G. W. Horgan, G. Polder, J. A. Dieleman *et al.*, "Automatic fruit recognition and counting from multiple images," *Biosystems Engineering*, vol. 118, pp. 203–215, 2014.
- [5] M. Hossain, M. Shamim, M. Al-Hammadi and M. Ghulam, "Automatic fruit classification using deep learning for industrial applications," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1027–1034, 2018.
- [6] S. Marimuthu and M. M. Roomi, "Particle swarm optimized fuzzy model for the classification of banana ripeness," *IEEE Sensors Journal*, vol. 17, no. 15, pp. 4903–4915, 2017.
- [7] Y. Zhang and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine," *Sensors*, vol. 12, no. 9, pp. 12489–12505, 2012.
- [8] M. Oltean, "Fruit recognition from images using deep learning," *Acta Universitatis Sapientiae, Informatica*, vol. 17, no. 12, pp. 00580, 2017.
- [9] S. Wang, Y. Zhang, J. Genlin, Y. Jiquan, W. Jianguo *et al.*, "Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization," *Entropy*, vol. 17, no. 8, pp. 5711–5728, 2018.
- [10] A. Rocha, D. C. Hauagge, W. Jacques and G. Siome, "Automatic fruit and vegetable classification from images," *Computers and Electronics in Agriculture*, vol. 70, no. 1, pp. 96–104, 2010.
- [11] Z. Tang, Y. Jialing, L. Zhe and Q. Fang, "Grape disease image classification based on lightweight convolution neural networks and channel-wise attention," *Computers and Electronics in Agriculture*, vol. 178, pp. 105735, 2010.
- [12] H. S. Gill and B. S. Khehra, "Efficient image classification technique for weather degraded fruit images," *IET Image Processing*, vol. 14, no. 14, pp. 3463–3470, 2020.
- [13] H. S. Gill and B. S. Khehra, "Visibility enhancement of color images using type-II fuzzy membership function," *Modern Physics Letters B*, vol. 32, no. 11, pp. 1850130, 2018.
- [14] H. S. Gill and B. S. Khehra, "An integrated approach using CNN-RNN-LSTM for classification of fruit images," *Materials Today: Proc.*, pp. 1–5, 2021. <https://doi.org/10.1016/j.matpr.2021.06.016>.
- [15] Y. Guo, Y. Liu, M. Erwin and S. Michael, "CNN-RNN: A large-scale hierarchical image classification framework," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 10251–10271, 2018.
- [16] S. Singh, S. K. Pandey, U. Pawar and R. R. Janghel, "Classification of ECG arrhythmia using recurrent neural networks," *Procedia Computer Science*, vol. 132, pp. 1290–1297, 2018.

- [17] H. S. Gill and B. S. Khehra, "Hybrid classifier model for fruit classification," *Multimedia Tools and Applications*, vol. 80, pp. 1–36, 2021.
- [18] E. Hassan, Y. Khalil and I. Ahmad, "Learning feature fusion in deep learning-based object detector," *Journal of Engineering*, vol. 20, pp. 1–20, 2020. <https://doi.org/10.1155/2020/7286187>.
- [19] F. Yang, M. Zheng and X. Mei, "Image classification with superpixels and feature fusion method," *Journal of Electronic Science and Technology*, vol. 19, no. 1, pp. 100096, 2021.
- [20] O. I. Khalaf, "Preface: Smart solutions in mathematical engineering and sciences theory," *Mathematics in Engineering, Science and Aerospace*, vol. 12, no. 1, pp. 1–4, 2021.
- [21] N. A. Khan, O. I. Khalaf, C. A. T. Romero, M. Sulaiman and M. A. Bakar, "Application of euler neural networks with soft computing paradigm to solve nonlinear problems arising in heat transfer," *Entropy*, vol. 23, no. 8, pp. 1053, 2021.
- [22] J. Nishant, D. Prashar, O. I. Khalaf, Y. Alotaibi, A. Alsufyani *et al.*, "Blockchain-based crop insurance: A decentralized insurance system for modernization of Indian farmers," *Sustainability*, vol. 13, no. 16, pp. 8921, 2021.
- [23] O. I. Khalaf, C. A. Tavera Romero, A. A. J. Pazhani and G. Vinuja, "VLSI implementation of a high-performance nonlinear image scaling algorithm," *Journal of Healthcare Engineering*, vol. 21, pp. 01–10, 2021.
- [24] O. I. Khalaf and G. M. Abdulsahib, "An improved efficient bandwidth allocation using TCP connection for switched network," *Journal of Applied Science and Engineering*, vol. 24, no. 5, pp. 735–741, 2021.
- [25] C. A. T. Romero, J. H. Ortiz, O. I. Khalaf and W. M. Ortega, "Software architecture for planning educational scenarios by applying an agile methodology," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 8, pp. 132–144, 2021.
- [26] O. I. Khalaf, "Preface: Smart solutions in mathematical engineering and sciences theory," *Mathematics in Engineering, Science and Aerospace*, vol. 12, no. 1, pp. 1–4, 2021.
- [27] O. I. Khalaf, M. Sokiyna, Y. Alotaibi, A. Alsufyani and S. Alghamdi, "Web attack detection using the input validation method: Dpda theory," *Computers, Materials & Continua*, vol. 68, no. 3, pp. 3167–3184, 2021.
- [28] G. Suryanarayana, K. Chandran, O. I. Khalaf, Y. Alotaibi, A. Alsufyani *et al.*, "Accurate magnetic resonance image super-resolution using deep networks and Gaussian filtering in the stationary wavelet domain," *IEEE Access*, vol. 9, pp. 71406–71417, 2021.