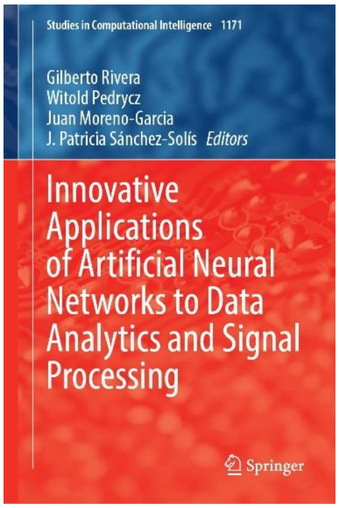
June 30, 2024

To whom it may concern:

By this letter, the editors certify that the acceptance of the following chapter was the result of a double-blind peer-review process:



Chapter title:	DANN: A Deep Attention Neural Network for Automatic Fruit Image Classification
Authors:	Abhik Ganguly, Rounak Chakraborty, Dipayan Ghosh, Pawan Kumar Singh, and Aimin Li
Book title:	Innovative Applications of Artificial Neural Networks to Data Analytics and Signal Processing
Editors:	Gilberto Rivera, Witold Pedrycz, Juan Moreno-Garcia, and J. Patricia Sánchez-Solís
Book series:	Studies in Computational Intelligence (Vol. 1171)
Publisher:	Springer Cham

Furthermore, the editorial review process complied with the publishing agreement stated by Springer in Contract Id. EF3420D3-F35B-474C-B2A8-6998BA07E03B. Accordingly, a piece of evidence of the peer-review process is enclosed. *Studies in Computational Intelligence* is currently indexed in Scopus, SCImago, zbMATH, DBLP, and WTI AG.

This contributed book was an editorial initiative of the Eurekas Community. Eurekas is an international and multidisciplinary scientific research network that joins professionals in mathematics, computer science, engineering, administration, economics, and social sciences. It was founded in 2008 and is currently integrating more than 60 research groups in more than 20 countries, mainly in America and Europe. The submitted chapters were accepted after a stringent review process by our collaborators worldwide, coordinated by the editors.

Please do not hesitate to contact me with any questions regarding this letter.

Sincerely yours,

Dr. Gilberto Rivera

EXECUTIVE SECRETARY OF THE EUREKAS COMMUNITY

COORDINATOR OF PUBLICATION PROJECTS

https://www.eurekascommunity.org
info@eurekascommunity.org
Tel. (+52) 667 244 4900

Paper Title: “DANN: A Deep Attention Neural Network for Automatic Fruit Image Classification”

We thank both the Editor and Reviewers for devoting their valuable time to analyze our manuscript and providing us comments for improvement. We have made all the necessary changes in the revised manuscript. The changes are marked in yellow color for easy detection and proper understanding. We hope that the modifications are on par with the reviewers’ expectations.

Review #1:

1. The research makes a valuable contribution, as the document is well-explained and easy to read, providing a comprehensive understanding for general audiences. From my perspective, the article provides compelling arguments for further study and serves as a foundation for new research.

Response: We thank the reviewer for such an insightful comment.

2. Tables: The tables should adopt a more formal and professional design to improve their appearance and readability. Consider refining the formatting, fonts, and colors to achieve a more sober aesthetic.

Response: We thank the reviewer for such an insightful comment. As per the reviewer’s suggestions, all the Tables in the updated manuscript have been reformatted to achieve a more sober aesthetic.

3. Figure 1: Enhance the quality of Figure 1 to ensure clarity and visual appeal. This may involve utilizing higher resolution images or graphics and ensuring proper formatting and labeling.

Response: We thank the reviewer for such an insightful comment. Following the reviewer's suggestion, the quality of Figure 1 has been enhanced in the modified version of the manuscript.

4. Figure 3: Increase the size of layer descriptions in Figure 3 to improve readability. Adjusting the font size or layout of the figure can accommodate larger text without overcrowding the image.

Response: We thank the reviewer for such an insightful comment. Please be noted that the size of layer descriptions provided in Figure 3 has been increased to improve readability in the updated manuscript.

5. Methodology: Provide a more explicit description of the care model implementation in the methodology section. Include details such as specific algorithms or techniques used, implementation steps, and relevant code snippets or diagrams to clarify the process.

Response: We thank the reviewer for such an insightful comment. As per the reviewer's suggestion, the implementation of our proposed DANN model has been elaborated in the modified version of the manuscript. Additionally, we have also provided the source codes of our model at the end of the updated manuscript.

6. DANN Algorithm: Offer a detailed explanation of the DANN algorithm to enhance reader understanding. Explain its key concepts, components, and differentiation from other algorithms.

Response: We thank the reviewer for such an insightful comment. Please be noted that a detailed understanding of the DANN model has been provided with the help of Figure 2 in the updated manuscript.

7. Hyperparameter Optimization: If hyperparameters like batch size, number of epochs, and dropout regularization are mentioned but not detailed in the document, consider adding a dedicated section on hyperparameter optimization. Describe the importance of optimizing these

parameters and provide insight into the optimization methods used in your study. By addressing these areas, the quality and clarity of the article can be significantly improved, making it more informative and engaging for readers.

Response: We thank the reviewer for such an insightful comment. All the discussion related to hyperparameter optimization of our proposed model has been provided as a separate paragraph in Section 4.3 of the updated manuscript.

Review #2:

1. Please avoid passive voice, and always double-check your paper grammar before submitting.

Response: We sincerely apologize for any grammatical errors in the paper. As per the reviewer's suggestions, we have checked and corrected passive voice problems and grammars as much as possible.

2. Use the same style for all tables. For example: Tables 4, 5, 6, and 8 have a different style than the other tables.

Response: We thank the reviewer for the meticulous observation.We have corrected styles of all the tables in the modified version of the manuscript.

3. Check the quality of the figures and their size. For example: In Figures 1 and 2, the text and size are not visible.

Response: We thank the learned reviewer for raising this concern. Please note that we have fixed the sizes of Figures 1 and 2 in the updated version.

4. Please reference all tables and figures used.

Response: We would like to thank the reviewer for raising this concern. As some of the tables and figures are generated form our observations as well as from our code, that is why we have not referenced all the tables and figures.

5. Are tables 9 and 11 the same as tables 5 and 6? If they are the same just reference.

Response: Thank you for the valuable suggestion. Following the reviewer's suggestion, we have corrected the reference of tables and reduced the redundancy.

6. You can include the concepts of the evaluated metrics in one section, instead of describing them in each subsection. As in the case of precision, recall, and F-1 score in sections 5.2.1 and 5.3.1.

Response: Thank you for your valuable comment. As we have discussed the results of different dataset in different sections, we would like to have separate metrics section for each dataset in separate section.

DANN: A Deep Attention Neural Network for Automatic Fruit Image Classification

Abstract. Classifying fruits is crucial in various fields, including agriculture, food processing, and computer vision. Typical fruit sorting often uses simple features along with basic machine learning models. It struggles with complex fruit images. Here, we introduce a deep learning model for fruit sorting. We propose a custom-designed CNN with an attention feature called DANN. Three standard datasets namely, Fruits-360, FIDS30, and FRUITSGB are used to evaluate the proposed DANN model. Testing shows that the proposed DANN model automatically classifies the fruit images with enhanced accuracies of 98.38%, 87.3% and 98% for the above mentioned three datasets respectively. We adjust settings like batch size and epochs for training. Compared to typical machine learning models, our CNN model with attention performs better. This research shows that adding attention to CNN improves the performance of fruit image classification. It helps in agriculture and food industries. Our model opens paths for more research with bigger datasets and different models. This work highlights deep learning and attention for good fruit classification. It helps in farming and food quality.

Keywords: Automatic fruit image classification, Deep attention neural network, Convolutional Neural Network, Fruits-360 dataset, FIDS30 dataset, FRUITSGB dataset

1 Introduction

Classifying fruits is important in many areas such as farming, food making, and computer vision. It helps with checking quality, grading, tracking, finding diseases, and estimating harvests in farming [1]. In the past, people used eye inspection and personal judgment to sort fruits which caused problems. Sorting fruits also helps shop workers know the price of each fruit. Deep learning, like Convolutional Neural Networks (CNN), has changed fruit classification by using data and complex networks to be more accurate. [2].

This study discusses the obstacles linked to methods of fruit categorization. Suggests an innovative deep learning approach using CNNs with an attention layer. Conventional techniques often depend on created features and basic learning algorithms, which might face challenges in generalizing across fruit varieties and differences, in shape, size and color. On the hand deep learning models have the ability to autonomously learn distinguishing features from raw pixel data allowing for resilient and precise categorization. The inspiration behind this investigation arises from the necessity, for fruit categorization methods of managing the complexities and subtleties found in real world fruit images. By using the power of CNNs, which are adept at learning hierarchical representations of visual features, we aim to develop a model capable of accurately identifying various fruits from their images.

Moreover by integrating an attention mechanism, into the CNN design we enhance the models complexity. This feature enables the model to concentrate on areas of the input images while filtering out distracting details. Such selective focus mirrors how the human visual system prioritizes attributes resulting in classification accuracy and interpretability. The study aims to accomplish two objectives; to create a CNN based fruit classification model that surpasses traditional methods in terms of accuracy and resilience and secondly to explore how incorporating attention mechanisms impacts classification performance. By meeting these objectives we aim to advance the development of automated fruit classification systems bringing benefits to sectors such, as agriculture, food processing and retail.

In the parts of this paper, we will talk about how we created the model we are suggesting share the results of our experiments and analyses and delve into what our findings mean. We'll also look at how our model stacks up against methods and suggest areas, for research in fruit classification and related fields. This study is about using learning and attention mechanisms to tackle practical issues, in fruit classification aiming to enhance efficiency and productivity in different areas.

1.1 Structure of the Chapter

The rest of this chapter is organized as follows:

Section 2. (Literature Review). An in-depth analysis of the research, on categorizing fruits showcasing the advancements in methods and the hurdles that deep learning models have tackled.
Section 3. (Dataset). Description of the dataset used for training and evaluation, outlining the diversity of fruit classes and the preprocessing steps applied.
Section 4. (Methodology). Detailed architecture of the CNN, the incorporation of attention mechanism, and the data augmentation techniques applied.
Section 5. (Results). Reports the final results of the experiments, including accuracy, precision and recall and comparisons between the models.
Section 6 (Discussion). Analyzes the results, discusses various observations.
Section 7 (Conclusion). Summarizes the key findings and contributions of the research, highlighting the scope of future work followed by acknowledgement and references.

The following sections delve into each aspect of the research, providing a comprehensive understanding of the methodologies, results, and implications of the study.

2. Literature Review

The automatic classification of fruits has piqued the interest of experts, across numerous fields such as agriculture, computer vision and machine learning. Traditional methods for fruit image classification often rely on handcrafted features and machine learning algorithms like Support Vector Machines (SVMs) and k Nearest Neighbors (k-NN). While these techniques have shown some effectiveness, they can struggle with the complexities and variations found in fruit images.

In years learning advancements, especially CNNs have transformed the landscape of image classification, including fruit categorization. CNNs are capable of extracting features from raw pixel data automatically eliminating the need, for manual feature engineering. This data centric approach has produced results in a range of image recognition tasks, including the identification of different types of fruits.

Numerous research efforts have solidified the efficiency of CNNs in the domain of fruit classification. For example Joseph et al. [3] proposed a CNN model with Tensorflow backend and achieved an accuracy of 94.35%. Chung et al. [4] proposed EfficientNet based architecture and got an accuracy of 95%. Similarly, Zhang et al. (2020) tried to use deep CNN architectures for fruit detection and classification in orchard environments, showcasing the potential of deep learning in agricultural applications [5]. Rathnayake et al. [6] used novel modified cascaded ANFIS algorithm to identify fruit 360 image and achieved classification accuracy of 98.36%. Seng and Mirisaee [7] proposed an alternative fruit detection method utilizing three key characteristics: color, shape, and size. Color was represented by the average RGB value, size by the area and perimeter measurements and shape by the roundness metric. When it comes to sorting fruits a major hurdle lies in the differences, in how fruits look, influenced by aspects, like their shape, size, color and any obstructions. To tackle this issue experts have explored methods to enhance the reliability and adaptability of CNN models. Tomar et al. tried another CNN model to get an accuracy of 95% [8]. Mundhana et al. [9] used a technique three stage Maturity grading in CNN to correctly classify the fruits .It achieved an accuracy of 90.24 %. Vijayalakshmi et al. [10] tried CNN model on banana species while achieving an impressive accuracy of 96.98%.

To artificially expand the variety of training data and improve model performance, data augmentation techniques like rotation, scaling, and flipping have been extensively employed. [11]. Bobde et al. [12] developed a deep learning based architecture using Keras and achieved an accuracy of 95%. Singh et al. [13] performed a multi-layered CNN architecture and secured an accuracy of 97%.

Incorporation of attention mechanism into CNN architectures is gaining momentum in recent years. These mechanisms enable the model to particularly focus on some specific areas of the input image removing noisy information. This selective attention has a similarity with human visual perception and it increases the classification performances, including image classification and object detection [14]. For example, Min et al. [15] proposed an attention-based CNN model for fruit classification that achieved superior performance compared to traditional CNN architectures.

Raghavendra et al. [16] proposed a deep learning architecture named as FruitNet-11 and achieved an accuracy of 96.15%. Kushwaha et al. [17] performed an optimized CNN model on the fruit 360 dataset. They achieved 96.88% accuracy.

An overview of previous studies has emphasized the promising potential of deep learning technologies, notably CNNs with attention properties, in advancing the area of fruit classification. In contrast to traditional approaches which have set the stage for this development, deep learning is a more adaptable and data-driven way to address complexities and variations associated with real world fruit images. The upcoming segments of this chapter will follow up on prior research work and introduce an original CNN model using an attention mechanism for the purpose of fruit classification in order to further improve accuracy and robustness within this domain

3 Datasets

3.1 Dataset Description

In this research, we utilized three primary datasets: the Fruits-360 dataset, FIDS30 and FRUITSGB datasets. These datasets comprise a diverse set of fruit images, capturing the minute details of different shapes, sizes, and colors commonly found in real-world scenarios.

Fruits-360 Dataset. A comprehensive collection of fruit images, spanning numerous classes is what the Fruits-360 dataset [18]. To prevent biases during model training, it includes a wide range of fruits with balanced distribution of these images across different classes. The dataset comprises diverse types of fruits such as apples, bananas oranges, berries and tropical fruits which represent most common agricultural and retail produce. Images in the set have different lightings, backgrounds and orientations that contribute to the challenges faced in real-world fruit classification. Some sample images for Fruits-360 dataset are shown in Table 1.

Image						
Label	Apple Red 2	Avocado	Banana	Blue Berry	Corn	Kaki

Image						
Label	Guava	Hazelnut	Lemon	Mulberry	Onion Red	Papaya

Table. 1. Fruits-360 dataset sample images

FIDS30 Dataset. In addition to the Fruits-360 dataset, we also utilized the FIDS30 dataset[19]. This dataset focuses specifically on different fruits, providing a specialized collection of images for a targeted classification task. It includes images of various fruits, lemons, and limes, along etc. Some Sample images of FIDS30 dataset are given in **Table 2**.

Table. 2. Sample images of FIDS30 dataset

Image					
Class	Bananas	Cherries	Grapes	Lemons	Oranges

FRUITSGB Dataset. After completing our work with the Fruits-360 dataset, we proceeded to analyze the FRUITSGB dataset[20] as part of our research. This dataset comprises images of various fruits, with the primary objective being the classification of these fruit images. All the images are labeled as good or bad. **Table 3** presents some sample images from the FRUITSGB dataset.

Table. 3. Some images of FRUITSGB dataset.

Image					
Class	Apple_Bad	Apple_Good	Banana_Bad	Banana_Good	Orange_Bad

3.2 Dataset Composition

There are total 131 fruit classes in the **Fruits-360** dataset. Each class consists of a distinct set of images. The **FRUITSGB** and **FIDS30** dataset consists of images of various citrus fruits and their corresponding leaves which are commonly found in agriculture.

3.3 Data Pre-processing

We conducted pre-processing steps to prepare the datasets for model training. The initial step involved resizing all images to a consistent dimension of 100x100 pixels. As the next step, we applied normalization to scale all pixel values of the image array between '0' and '1'. We implemented a stratified data splitting strategy for all the three datasets to mitigate overfitting. Following that, we partitioned all datasets into training, validation, and testing sets while preserving the class distribution within each subset. We utilized the training set for model development and the validation set for fine-tuning hyperparameters of the model. Finally, we used the testing set to evaluate the performance of the proposed DANN model.

3.4 Data Augmentation

During training, we implemented data augmentation to address real-world image variations and mitigate overfitting. This step generated a diverse and representative dataset from the existing training images.

Methods included in this step:

- **Random Rotations:** Rotating the image by random angles we generated various viewing perspectives.
- **Horizontal Flips:** Horizontally mirroring the image we created natural variations in fruit orientation.
- **Brightness Adjustments:** Randomly changing brightness levels we created a range of lighting conditions.
- **Zoom and Shear :** Applying zoom and shear properties, we enhanced the data augmentation process for the fruit section of the Citrus Fruits and Leaves dataset.

According to Shorten and Khoshgoftaar (2019) [21], employing data augmentation techniques, especially on limited datasets, enhances model performance. Moreover, the study by Perez and Wang (2017) [22] demonstrates its effectiveness in addressing overfitting and enhancing resilience to real-world variations.

3.5 Dataset Statistics

The summary of the dataset statistics for the Fruits-360 dataset is shown in **Table 4** and representation of classes across the training, validation set, as well as testing sets.

The following sections delve into the methodologies used for model development, experimentation, and the interpretation of results, providing a holistic view of our approach to automatic fruit classification.

Table 4. Overview of the Fruits-360 dataset composition and distribution across subsets.

Subset	Number of Images	Number of image classes
Training	54154	131
Validation	13538	131
Testing	22688	131

Table 5 provides a summary of the dataset statistics for FIDS30 and representation of fruit image classes. In FIDS30 dataset we have only 971 images of 30 different fruit classes. So, in order to increase the size of the dataset we have applied the data augmentation technique and finally increased our dataset size. Then we divided the whole dataset into 3 parts. One part is for training, one for validation and one part for testing the accuracy. We have generated new instances of existing images by applying the property of shear, zoom, horizontal flip etc.

Table 5. Overview of the FIDS30 dataset used in the present work.

Subset	Number of Images	Number of image classes
Training	7717	30
Validation	858	30
Testing	969	30

Table 6 provides a summary of the dataset statistics for FRUITSGB and representation of fruit classes. We have 12,000 images of 12 different fruit classes and we have divided the all images into 3 parts. One part is for training, One part is for validation and for testing the accuracy, we have the testing part.

Table 6. Overview of the FRUITSGB dataset used in the present work.

Subset	Number of Images	Number of image classes
Training	8640	12
Validation	1200	12
Testing	2160	12

4 Methodology

4.1 Convolutional Neural Network (CNN)

CNNs, a popular type of deep learning model, excel at processing structured grid-like data, especially images [23, 24]. These networks comprise multiple layers: convolutional layers that identify features using learned filters, pooling layers that decrease spatial dimensions, and fully connected layers for classification tasks. CNNs leverage hierarchical patterns and spatial relationships within images, enabling them to learn complex representations and demonstrating state-of-the-art performance in image classification. Their design draws inspiration from the structure of the visual cortex in the brain, with neurons in deeper layers capturing high-level features. CNNs have revolutionized computer vision. They continue to be a cornerstone of modern AI, driving innovation in image analysis and understanding.

4.2 Transfer Learning

Transfer learning involves leveraging pre-trained models, often trained on massive datasets like ImageNet, transfer learning allows for fine-tuning these models to specific tasks. This approach effectively utilizes the learned features and parameters, even when limited labeled data is available. MobileNetV2 is a CNN architecture. It is specifically optimized for mobile and embedded vision applications [25]. It uses special types of convolutional layers and innovative techniques to perform efficiently on devices with limited resources. On the other side, VGG16 is a more complex CNN. It is famous for being straightforward and powerful [26]. This architecture usually arranges multiple convolutional layers, followed by Max-pooling layers that process fruit images and then decides what the fruit image represents. Despite being deep, VGG16 can be slow due to its large number of parameters. However, both MobileNetV2 and VGG16 provide useful choices for different computer vision tasks, depending on what the application needs.

4.3 Model Architecture of DANN

We developed an attention-based CNN model called DANN. It is specifically designed for fruit classification tasks. This design uses a sequence of convolutional layers that process images, along with Rectified linear unit (ReLU) activation function and Max-pooling layers in between. After the Max-pooling4, we have added an attention layer which helps to emphasize important regions of the feature maps. The convolutional layers typically find important features in the input images, while the Max Pooling layers reduce the size of these features to capture key patterns. **Fig. 1** illustrates the sequence of layers present in our proposed DANN model whereas **Fig. 2** provides an architectural overview.

The architecture is as follows:

Convolutional Layer 1: 16 filters, kernel size 2x2, ReLU activation
MaxPooling Layer 1: Pool size 2x2
Convolutional Layer 2: 32 filters, kernel size 2x2, ReLU activation

MaxPooling Layer 2: Pool size 2x2
Convolutional Layer 3: 64 filters, kernel size 2x2, ReLU activation
MaxPooling Layer 3: Pool size 2x2
Convolutional Layer 4: 128 filters, kernel size 2x2, ReLU activation
MaxPooling Layer 4: Pool size 2x2
Attention Layer :
- Attention layer applied after all the convolutional layers and max_pool_2d layers.
Dropout Layer: Dropout rate is 0.3
Flatten Layer : we added the Flatten Layer
Dense Layer 1:
- Units: 150
- Activation: ReLU
Dropout Layer: Rate: 0.4
Dense Layer 2 (Output Layer):
- Units: 131 (Number of classes)
- Activation: Softmax

In our proposed model architecture, we integrated the attention layer after the final convolutional layer. This weighting process enables the model to prioritize important regions while diminishing less relevant ones, thereby improving its capability to extract meaningful features from input images. By integrating attention, the model seeks to enhance its performance in tasks such as image classification by selectively focusing on relevant features while ignoring noise or irrelevant details. This approach boosts the model's accuracy, making it suitable for computer vision tasks.

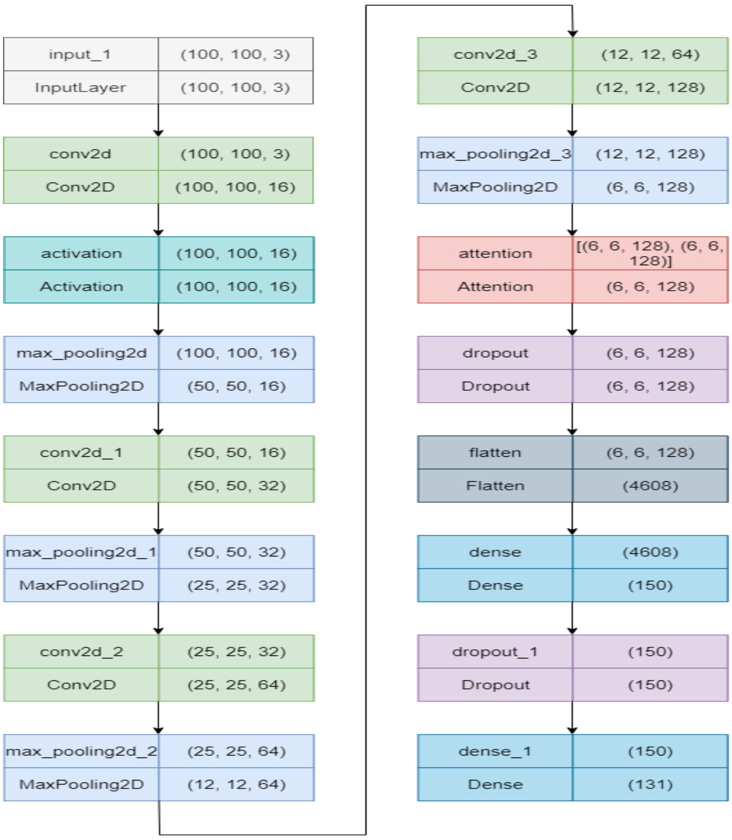


Fig. 1. **Sequence of layers of our proposed DANN model for automatic fruit classification.**

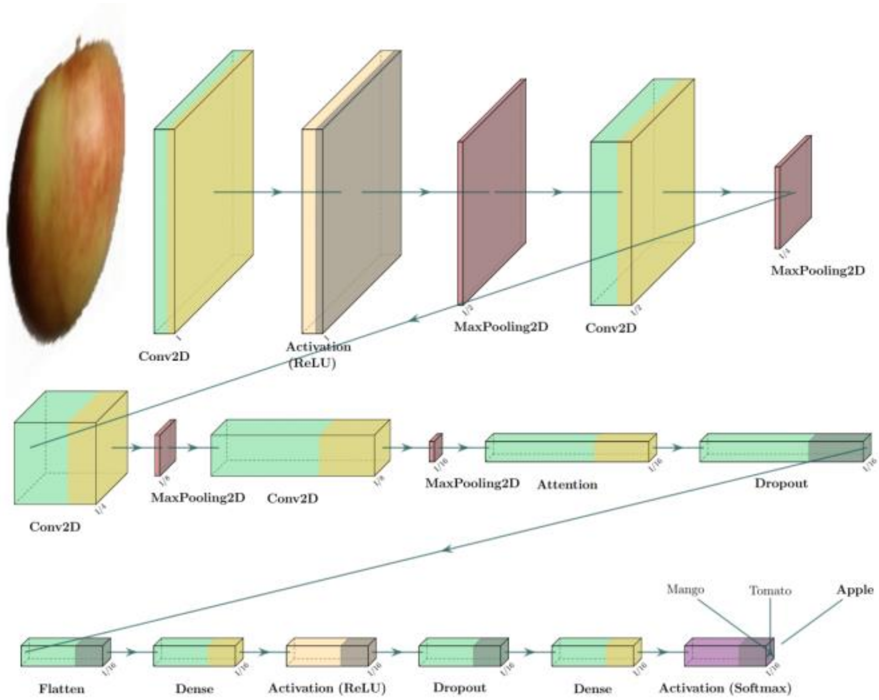


Fig. 2. Attention based CNN model architecture diagram.

Hyperparameter Optimization: The CNN architecture underwent training utilizing the Adam optimization algorithm coupled with a categorical cross-entropy loss function. The training phase has been iterated through 25 epochs, employing a batch size of 128. Early stopping has been integrated to cease training in case the validation loss stagnated for five successive epochs. The learning rate is configured at 0.001 to support steady convergence.

4.4 Model Evaluation

The proposed DANN model has been evaluated on the testing set, comprising 22,688 images from 131 fruit classes. We assessed classification performance using accuracy, precision, recall, and F1 score metrics. Confusion matrices were also produced to glean insights into the model's performance across individual classes.

5. Results Analysis

5.1 Results on Fruits-360 dataset

The results of the fruit classification on Fruits-360 dataset using our proposed DANN model are presented and analyzed below. Also, we have shown the results of two standard transfer learning models like VGG16 and MobileNetV2 on the Fruits-360 dataset. Here, the model with the attention layer outperformed transfer learning models.

5.1.1 Model Performance Metrics

Our DANN model exhibited performance on the fruit classification task, achieving an overall accuracy of 98.38% on the test dataset. The classification report offers a breakdown of key metrics for each fruit class, encompassing precision, recall, and F1-score. These metrics are calculated for each individual class, providing a thorough assessment of the model's performance across the entire range of categories [27].We have given the average precision and recall. Precision reflects the percentage of correctly predicted positive cases within all predictions labeled as positive. Conversely, recall measures the proportion of actual positive cases that the model correctly identified. Combining both metrics, the F1-score offers a balanced assessment of accuracy. Our model achieved high precision and recall across most fruit classes, signifying its capability for accurate classification with minimal errors [28]. However, for certain fruit classes, particularly those with similar visual characteristics (for example, Oranges and tangerines), the model exhibited slightly lower precision and recall values, suggesting potential challenges in distinguishing between closely related fruits. The overall results of DANN model on the Fruits-360 dataset is given in **Table 7**.

Table 7. Performance results (in terms of metrics) given by the proposed DANN model on the Fruits-360 dataset.

Model	Accuracy	Precision	Recall	F1-score
VGG16	95.98%	0.96	0.96	0.96
MobileNetV2	97.21%	0.99	0.88	0.93
Our Proposed DANN	**98.38%**	**0.99**	**0.98**	**0.98**

This high accuracy underscores the effectiveness of the CNN architecture in accurately categorizing fruit images.

5.1.2 Training and Validation Curves

The accuracy as well as loss curves during training and validation processes provide important information about how the CNN model learns over time. At the beginning, both the training and validation accuracy of our model showed a promising upward trend. However, as training continued, we observed minor fluctuations in validation accuracy and loss, hinting at potential overfitting or convergence issues [29]. We looked closely at these fluctuations to improve how we trained our model. This helped our model work better in real-world situations. Additionally, exploring regularization techniques or adjusting the learning rate may mitigate these issues and further optimize the model's performance. Here, we can see in **Fig. 3** that with the increase with training accuracy, test accuracy also increases and after running it for 25 epochs, we finally reached a position where training accuracy is nearly equal to testing accuracy. Also we can see that with the decrease in training loss testing loss decreases which is shown in **Fig. 4**.

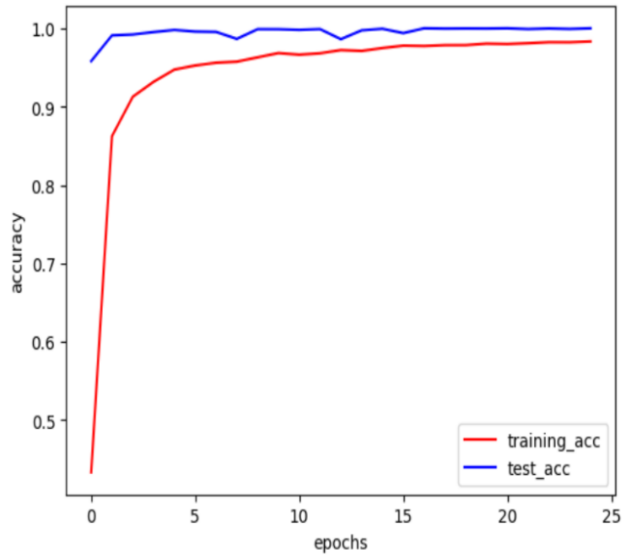


Fig. 3. Training accuracy versus test accuracy for our proposed DANN model

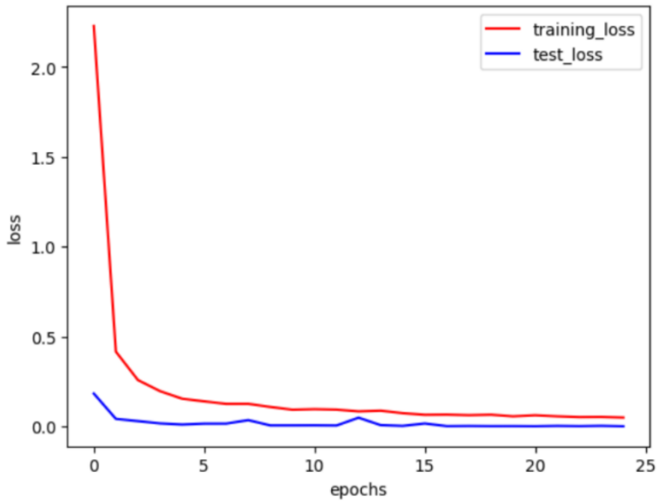


Fig. 4. Training loss versus test loss for our proposed DANN model.

5.1.3 Confusion Matrix Analysis

Analysis of the confusion matrix generated from the model predictions provides valuable insights into the classification errors made by the CNN model. The confusion matrix illustrates the model's effectiveness by displaying the count of true positives, true negatives, false positives, and false negatives for each fruit category [30].

Most fruits were classified correctly, as evidenced by the high values provided along the diagonals of the confusion matrix. However, some confusion was observed between visually similar fruit classes, such as oranges and tangerines, indicating potential ambiguity in distinguishing between closely related fruits. These findings highlight the importance of further refining the model's ability to differentiate between visually similar fruit categories. Confusion matrix of our model shown in **Fig. 5**. In a multi-class classification problem with more than two classes, each diagonal element represents the number of instances where the predicted label matches the true label for that specific class. The diagonal elements provide a quick overview of how well the classifier is performing for each class individually. Max instances of fruit classes like ‘walnut’, ‘water melon’, ‘Cherries’ etc. are giving us the positive true output.

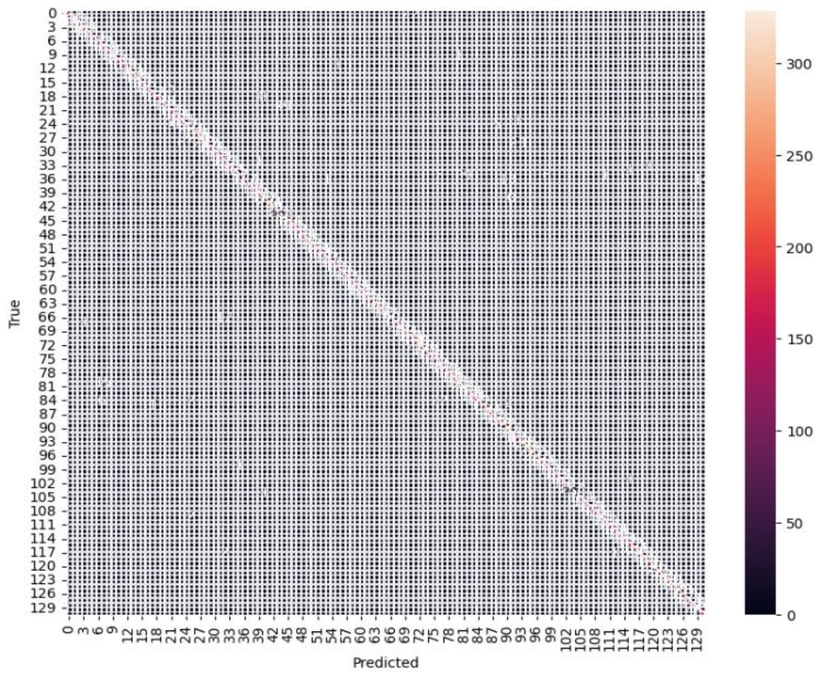


Fig. 5. Confusion matrix given by the proposed DANN model

5.1.4 Feature Visualization with Attention Mechanism

Our DANN model offers valuable insights into the regions of interest identified by the attention mechanism. Attention layer highlight the important regions of the input images that significantly contribute to the model's classification decisions [31].

Our DANN model corroborates the model's ability to learn discriminative representations and extract relevant features for accurate classification. Further analysis of our model provides valuable interpretability and enhances our understanding of the CNN model's decision-making process.

5.1.5 Comparison with Baseline Models

To give an assessment of the proposed DANN model's performance, a comparative analysis was conducted against baseline models reported in previous research papers. Specifically, we compared our DANN model with attention mechanisms against models from other studies which are widely cited for their effectiveness in fruit classification tasks. By applying our DANN model we have got accuracy of 98.38%. We have shown some results of some other models and it can easily be compared with our results. The comparison with some other models is shown in **Table 8**.

Table 8. Comparison of our proposed model with some recently developed fruit image classification models.

Author	Model	Year	Accuracy
Joseph et al. [3]	CNN model	2021	94.35%
Chung et al. [4]	EfficientNet	2019	95%
Mundhana et al. [9]	Maturity Grading	2021	90.24%
Bobde et al. [12]	CNN	2021	95%
Singh et al. [13]	CNN	2023	97%
Raghavendra et al. [16]	FruitNet	2022	96.15
Kushwala et al. [17]	optimized CNN model	2023	96.88%
Ghosh et al. (Proposed)	**DANN model**	**2024**	**98.38%**

5.1.6 Computational Efficiency

Our model achieved high accuracy with relatively quick training times, making it suitable for real-time or resource-limited applications [32]. Effective use of resources, like GPU acceleration and parallel processing, further boosted the scalability and practicality of our CNN model. These results show that using deep learning is efficient and scalable for classifying fruits.

5.2 Results on FRUITSGB Dataset

We also applied our proposed DANN model on another dataset named **FRUITSGB**, and all the outcomes are presented below. The **FRUITSGB** dataset, also known as the "Top Indian Fruits with Quality" dataset, is a publicly accessible set of images. It is designed to aid research in fruit quality classification. It specifically targets fruits of significant importance to the Indian market, filling a gap in existing datasets that frequently lack quality labels.

We used our attention-based CNN model on the **FRUITSGB** dataset after resizing all the images to 224x224 pixels. We achieved an accuracy of 98%. This shows how well the model can handle different fruit classification tasks. We utilized the model to identify fruits in the dataset. The last layer of our model consists of 12 nodes, each represents one of the 12 distinct classes.

Our study utilized the FRUITSGB dataset, which includes 12 different classes of popular Indian fruits, containing a total of 12,000 images. The dataset is partitioned into three subsets: training, testing, and validation, allowing for model training, evaluation, and performance assessment.

The training set comprises 12,000 images across 12 different classes, providing ample data for the model to learn and generalize patterns across various fruits varieties. Conversely, the testing set consisted of 2160 images across the same 12 classes, serving as an independent benchmark to evaluate the model's performance. Additionally, a validation set containing 1200 images, distributed among the 12 classes, was utilized to fine-tune model parameters and prevent overfitting. **Table 6** already provides the overview of the splitting of training, testing as well as validation.

Utilizing our DANN architecture, the research aimed to leverage advanced techniques for feature extraction and classification. As a result, this model provided us with a classification accuracy of 98%. We have got 0.98 precision, 0.98 recall and 0.98 f1-score.

5.2.1 Model Performance Metrics

Within machine learning, particularly classification tasks, three crucial evaluation metrics hold significant importance: precision, recall, and F1-score. **Table 9** presents the results for all three evaluation metrics produced by the proposed DANN model on FRUITSGB dataset.

Table 9 . Results in terms of all the evaluation metrics on FRUITSGB dataset

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-score**
DANN model	98	0.98	0.98	0.98

5.2.2 Training and Validation Curves

The training and validation curves show how the DANN model's performance changes over training epochs. These curves usually display both training and validation loss, as well as training and validation accuracy. These curves help us understand how well the model learns and applies what it learns to new situations [34].

Accuracy curves [35] are important for understanding how well the model performs, spotting overfitting, and adjusting hyperparameters effectively. They visually show the training and validation accuracies over epochs, helping with model selection, optimization, and understanding how dataset size affects learning. The provided figure shows the accuracy curve together with the training and testing losses. As training accuracy improves, we observe a similar increase in testing accuracy. This plot shows how well the model performs after completing 20 training epochs. During training, the model learned from training images. It verified its progress with validation images, and saved testing images for a final, unbiased evaluation of its capabilities. The accuracy curve for this model is given in **Fig. 6.**

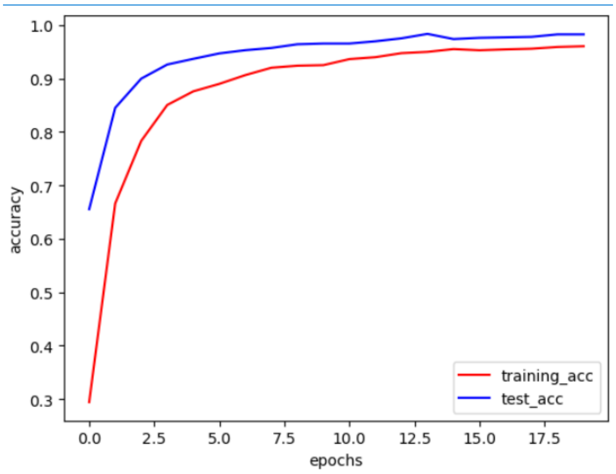


Fig 6. Accuracy curve of our proposed DANN model on FRUITSGB dataset.

Loss curves are essential for monitoring model performance during training. They reveal how well the model learns from data and converges to an optimal solution. A decreasing curve indicates improvement, while a flat or rising curve suggests adjustments are needed. Loss curves also help detect overfitting (training loss decreasing, validation loss increasing) and underfitting (both losses staying high). They guide optimizing performance and ensure effective learning. As training accuracy drops, testing accuracy follows, shown by the loss curve after 20 epochs. The loss curve for this model is given in **Fig. 7.**

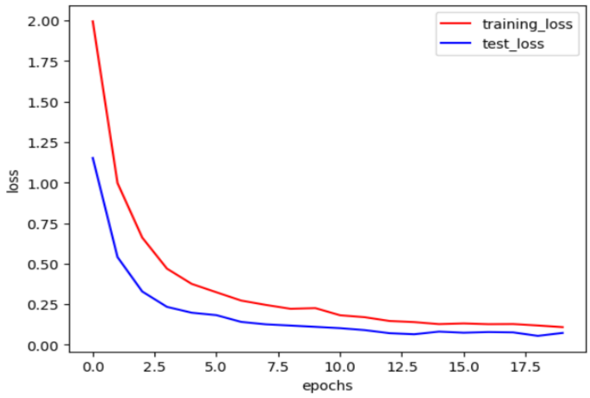


Fig 7. Loss curve of our proposed DANN model on FRUITSGB dataset.

5.2.3 Confusion Matrix Analysis

Analysis of the confusion matrix [36] generated from the model predictions provides valuable insights into the classification errors made by the CNN model. By visualizing the distribution of true positives, true negatives, false positives, and false negatives for each fruit class, the confusion matrix provides a clear picture of the model's performance. Confusion matrix for Our DANN model is given in **Fig. 8**. In a confusion matrix, the diagonal represents the instances where the predicted labels match the true labels. Each cell on the diagonal corresponds to a correct prediction for a particular class. Diagonal represents true positives. From the **Fig. 8,** we can see that some of the fruit classes have higher true positive rates. Max instances of fruit classes such as ‘Banana_Bad’, ‘Banana_Good’, ‘Guava_Good’ etc. are giving maximum true positive output.

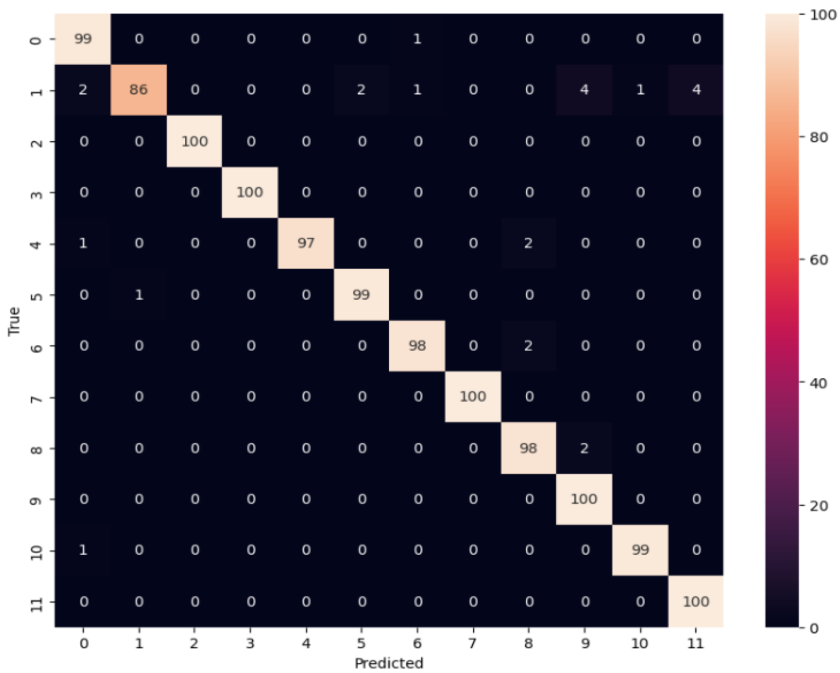


Fig 8. Confusion matrix for our proposed model on FRUITSGB dataset.

5.2.4 Comparison with Baseline models

Our proposed DANN model achieved an accuracy of 98% on the **FRUITSGB** dataset. While a direct comparison with all existing works is challenging due to the limited research conducted specifically on FRUITSGB, we can consider the results reported in "Visualization and Analysis of Transformer Attention" by Calderaro et al. [37], which achieved an accuracy of 96.0% on the larger Fruits-360 dataset. It is important to note that the **FRUITSGB** dataset, with its 12,000 images, is substantially smaller than the Fruits-360 dataset, which contains over 80,000 images. This difference in dataset size makes a direct comparison between the two accuracies less conclusive.

5.3 Results on FIDS30 dataset

We applied our CNN model to another dataset called FIDS30 dataset, and all the results are presented below. This dataset contains a minimal number of images, so we employed data augmentation techniques to increase its size.

Our proposed CNN model was applied on the FIDS30 dataset after resizing all the images to 224×224 pixels resulting in an accuracy of over 87.3%. This demonstrates the model's versatility and effectiveness across different fruit classification tasks. We have applied the model on the dataset for detection of fruits. Last layer of our model contains 30 units as we have 30 distinct fruit image classes to classify.

The research focused on the FIDS30 dataset distributed across 30 distinct fruit image classes. FIDS30 dataset initially had only 971 images of 30 distinct fruit classes. To increase the size of dataset, data augmentation techniques were employed where we have generated new images from the existing images by applying rotation, shear, zoom, horizontal flip and some other property by using ImageDatagenerator class. We have generated some new images from existing images and finally after data augmentation is performed our dataset size was over 9500 images. This process expanded the dataset's capacity to capture variations in the FIDS30 dataset ,crucial for training a robust model. Following data augmentation, the dataset was partitioned into three subsets: training, testing, and validation.

After data augmentation [34] is performed, the training set comprises 7717 images across 30 different classes, providing ample data for the model to learn and generalize patterns across various fruits varieties. Conversely, the testing set consisted of 969 images across the same 30 classes, serving as an independent benchmark to evaluate the model's performance. Additionally, a validation set containing 858 images, distributed among the 30 classes, was utilized to fine-tune model parameters and prevent overfitting. **Table 5** already provides the overview of the dataset splitting. Utilizing our DANN architecture, the research aimed to leverage advanced techniques for feature extraction and classification. And this model provided us with an accuracy of around 87.3%.

5.3.1 Model Performance Metrics

Within the realm of machine learning, specifically for tasks involving classification, three crucial metrics reign supreme: precision, recall, and F1-score. Precision, Recall and F1-score are the three most important measures in machine learning especially for classification tasks. The results for all three metrics given by the proposed DANN model are presented in **Table 10**.

Table 10 . Metrics analysis given by the proposed DANN model on FIDS30 dataset.

Model	**Accuracy(%)**	**Precision**	**Recall**	**F1-score**
DANN	**87.3**	**0.88**	**0.87**	**0.87**

5.3.2 Training and Validation Curves

The training and validation curves illustrate the performance of the attention-based CNN model over successive epochs. These curves typically display the training and validation loss, as well as training and validation accuracy, providing valuable insights into the model's convergence and generalization capabilities [33].

Accuracy curves are crucial for assessing model performance, detecting overfitting, and guiding hyperparameter tuning [35]. They visually display training and validation accuracies over epochs or iterations, aiding in model selection, optimization, and understanding dataset size effects on learning. Below the accuracy curve is shown and training loss and testing loss are labeled perfectly. With the increase of training accuracy we can see that testing accuracy is also increasing. The result of the curve between training accuracy and testing accuracy is shown after running the model for 25 epochs. We have used training images for training purposes and used validation images for the validation purpose. Testing images are not used to evaluate the model accuracy. The accuracy curve is illustrated in **Fig. 9.**

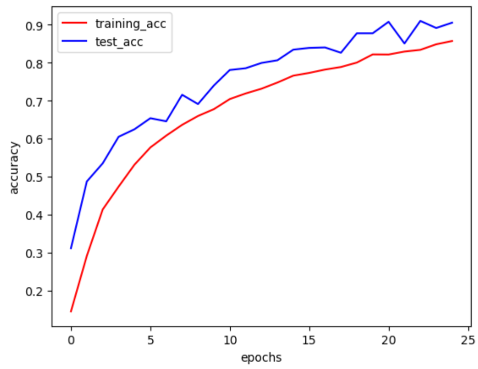


Fig 9. Accuracy curve for our proposed model on FIDS30 dataset.

Loss curves are crucial for monitoring the performance of machine learning models during training. They provide insight into how well the model is learning from the training data and whether it is converging towards an optimal solution. A decreasing loss curve indicates that the model is improving its ability to make predictions, while a stagnant or increasing curve suggests that adjustments may be needed in the model architecture or training process. Loss curves also help detect issues such as overfitting (as the training loss diminishes and the validation loss escalates) or underfitting (when both training and validation losses remain high). Overall, loss curves serve as a guide for optimizing model performance and ensuring effective learning. As training accuracy decreases, testing accuracy similarly declines, as evidenced by the loss curve results obtained after running the model for 25 epochs. The loss curve is given in **Fig. 10.**

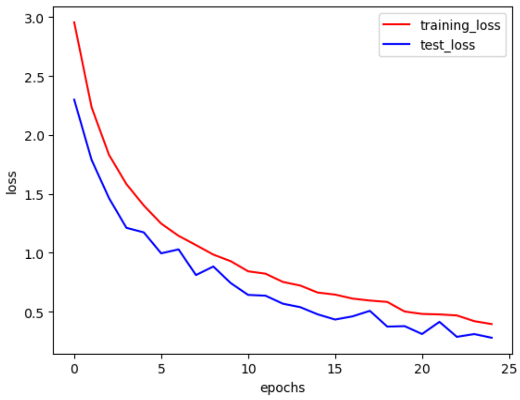


Fig 10. Loss curve for our proposed model on FIDS30 dataset.

5.3.3 Confusion Matrix Analysis

Analysis of the confusion matrix [36] generated from the model predictions provides valuable insights into the classification errors made by the CNN model. The confusion matrix visualizes the model's performance by showing the number of true positive, true negative, false positive, and false negative predictions for each fruit class. Confusion matrix of the dataset for our DANN model is shown in **Fig. 11**.

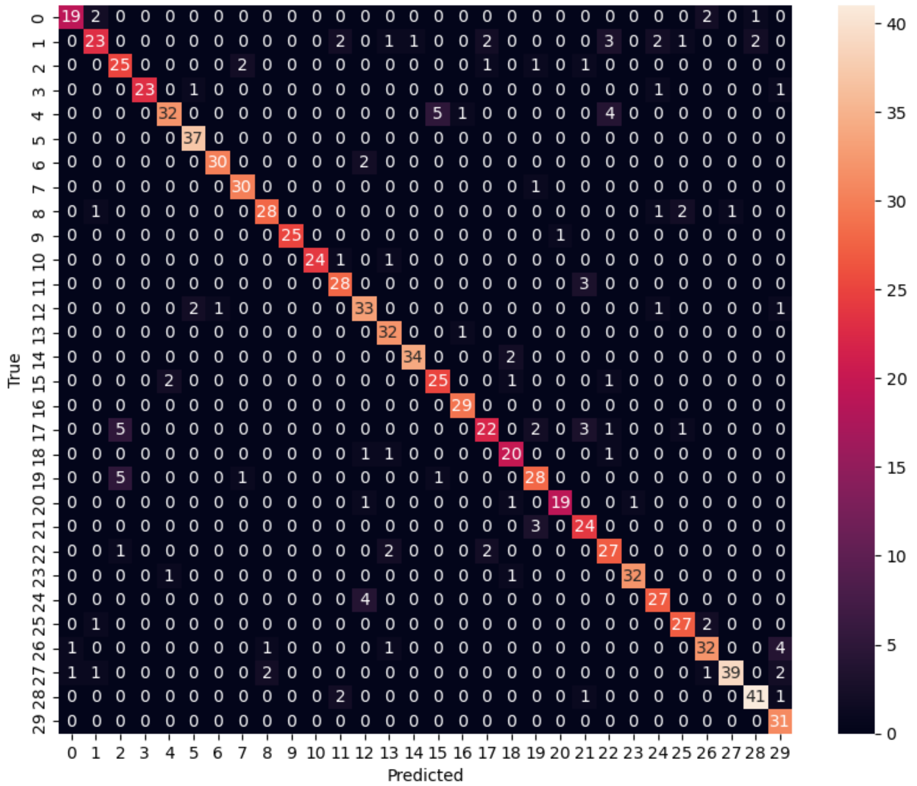


Fig 11. Confusion matrix for our proposed model on FIDS30 dataset

Fig.11. illustrates the confusion matrix which will help us compare between true and predicted values. The diagonal of the confusion matrix contains the true positives and true negatives, indicating the correct predictions made by the model for each class. By examining the values on the diagonal and comparing them with the total number of instances for each class, we can get insights into the model's accuracy for each class individually. Fruit classes such as ‘Grapes’ , ‘limes’, ‘raspberries’, etc. are giving the most true positive outputs.

5.3.4 Comparison with Baseline models

It's important to note that the FIDS30 dataset might not be readily available or well-documented, making it difficult to find published research papers referencing it explicitly. However, FIDS30 is very smaller as compared to Fruits-360 dataset and we tried our best to get the best accuracy by applying different techniques like data augmentation to increase the size of the dataset and finally reached an accuracy of around 87.3%. In future, we will try to make some changes so that we can get a better accuracy as well as better instances of accuracy and loss curves.

6 . Discussion of Findings

The results obtained from the experimental evaluation provide valuable insights into the performance and capabilities of the DANN model for fruit classification. These findings not only contribute to advancing the field of computer vision but also have significant implications for various domains, including agriculture, food quality assessment, and automated systems. In this section, we delve into a detailed discussion of the key findings and their implications.

6.1 Performance Evaluation

Through the Fruits-360 data set, our DANN model managed a remarkable global accuracy of 98.38% hence implying that it is efficient in classifying pictures of fruits correctly. The high precision rating implies that deep learning techniques, with specific reference to CNN architectures are appropriate for addressing intricate classifications. In this way, CNNs can provide a good segmentation in fruit classification based on their visual appearance using raw pixel information. Additionally, MobileNetV2 and VGG16 models were tested on the Fruits 360 dataset yielding accuracies of 97.21% and 95.98%, respectively. These findings mean that MobileNetV2 and VGG16 architectures can be used to classify fruits as well though at slightly lower precision than our specific DANN design model made for us by ourselves.

6.2 Interpretability and Explainability

The ability to understand deep learning models, particularly CNNs, remains crucial for their application in practice, specifically concerning domains like agriculture and food quality assessment requiring trust and transparency. Though the CNN model performed incredibly well, it is important to know how it makes its decisions and locate areas of interests on input images that will enable trust building and understanding classification outcomes better.

The interpretation by attention maps helps to expose those parts of the input image which have an impact on the predictions made by a model. Stakeholders can get insights into the characteristics or features driving classification decisions by visualizing where a model concentrates its attention. This not only increases confidence in the model but also helps domain experts comprehend what causes classification outcomes below.

6.3 Computational Efficiency and Scalability

It is crucial to have efficient resource utilization, such as GPU acceleration and parallel processing, for scaling deep learning models in real-world applications. The best performing model was our attention-based CNN model with fast training times and efficient resource utilization. Nevertheless, there are benefits of using models like MobileNetV2. Although it may not have been the top performer, MobileNetV2 has a lightweight architecture and computational efficiency that make it suitable for deployment in resource-constrained environments or real-time applications especially on edge devices or platforms with limited computational resources.

6.4 Challenges and Future Directions

However, several obstacles and possibilities for future research must be addressed although the results are very promising. There is need to continue studying about model interpretability, class imbalance, and dataset biases so as to improve the performance and generalization ability of the models. These problems can however be solved by means of strategies like data augmentation techniques, class weighted loss functions and a more interpretable CNN model.

Similarly, solutions to improve fruit classification performance especially in situations where there is limited labelled data or when deploying the model on new domains with a domain shift may involve multimodal data fusion, transfer learning methods as well as domain adaptation strategies.

In order to maximize the potential of deep learning in addressing real-life issues around agriculture and food security, it is necessary for researchers to join hands with industry players and policy formulators. By combining advanced AI technologies with multidisciplinary collaborations we can create novel approaches that transform farming systems, enhance food quality evaluation system and contribute towards worldwide moves aimed at sustainable agriculture and food production.

The model with the attention mechanism outperformed both certain transfer learning models and other research works. Future work involves exploring advanced data augmentation techniques specific to citrus fruits, optimizing attention-based CNN model architecture, and using ensemble learning methods for higher accuracy. Analyzing misclassifications, integrating domain-specific knowledge, and deploying semi-supervised or active learning techniques are crucial for enhancing model robustness, generalization, and performance without extensive labeled data. In future we will try to implement some new model like fuzzy rank-based fusion of CNN model to check if we can improve the accuracy[37].We have also plan to integrate ensemble CNN model to improve our model accuracy further[38].

7. Conclusion

In this research, we embarked on an exploration of advanced deep learning techniques for fruit classification, with a primary focus on Convolutional Neural Networks (CNN) and their integration with attention mechanisms. Through extensive experimentation and analysis, this chapter has provided valuable understanding of the strengths and limitations of these models for fruit image recognition

7.1 Key Findings

Attention layer-based CNN. The CNN model with attention layer demonstrated commendable performance in capturing spatial features from fruit images, showcasing its capability to depict visual patterns. However, we faced challenges when distinguishing visually similar kind of fruits, stressing on the importance of leveraging both local and global context information for precise and accurate classification.

7.2 Implications and Contributions

Robustness and Sensitivity. It was also noted that the attention model is robust to perturbations, showing its good generalization capacity to slightly modified input data. The sensitivity analysis provided insights into how the models could withstand the real-life cases of variations in input data.

Model Interpretability. The model's attention mechanisms are more interpretable relative to typical classic CNNs in many instances. This interpretability attribute is essential for applications where user trust in and understanding of a model's logic are needed.

7.3 Potential Paths for Future Exploration

Model Optimization. The high performance of the attention model suggests that there is potential for optimization techniques used in modeling it. Further works could involve improving attention mechanisms, exploring alternative architectures, or seeking ways of using transfer learning to improve both speed and accuracy

Real-world Deployment. How the model would be deployed in actual environments should not be overlooked. In future, researchers ought to strike a balance between how complex a model should be with respect to computational efficiency so as not hinder implementation in low resource settings.

7.4 Concluding Remarks

To sum up, this study has given important contributions to the domain of fruit classification by making use of the deep learning models. The hybrid model is a combination of CNNs with attention layers as a way to address the issues brought about by visually complex and similar fruit classes. Therefore, such findings will help in laying a foundation for future improvement in image recognition and show practical implications for application in Agriculture, retailing etc.

As deep learning landscape keeps on changing, the insights from this research will contribute towards future efforts aimed at creating more accurate, robust and interpretable models for various image classification tasks.

Acknowledgments

We extend our heartfelt appreciation to the individuals whose support and contributions have been invaluable in the completion of this research on fruit classification using advanced deep learning techniques.

This work was partially supported by Natural Science Basic Research Program of Shaanxi (Program No. 2024JC-YBMS-484 to Aimin Li).

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article. We have used only publicly available datasets for experimentation.

Code Availability Statement: The source codes related to this work can be found at: https://github.com/dip122/Fruit_classification_DL_image.

References

1. Rocha A, Hauagge D C, Wainer J, Goldenstein S 2010 Automatic fruit and vegetable classification from images Comput. Electron 70 96–104.

2. Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. *sensors*, *16*(8), 1222.

3. Joseph, J. L., Kumar, V. A., & Mathew, S. P. (2021). Fruit classification using deep learning. In *Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2021* (pp. 807-817). Springer Singapore.

4. Chung et al. A fruits recognition system based on a modern deep learning technique 2019 *J. Phys.: Conf. Ser.* 1327 012050

5. Zhang, Y. D., Dong, Z., Chen, X., Jia, W., Du, S., Muhammad, K., & Wang, S. H. (2019). Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. *Multimedia Tools and Applications*, *78*, 3613-3632. .

6. Rathnayake, N., Rathnayake, U., Dang, T. L., & Hoshino, Y. (2022). An efficient automatic fruit-360 image identification and recognition using a novel modified cascaded-ANFIS algorithm. *Sensors*, *22*(12), 4401.

7. Seng et al. A new method for the fruit recognition system. In Proceedings of the 2009 International Conference on Electrical Engineering and Informatics, Bangi, Malaysia, 5–7 August 2009; Volume 1, pp. 130–134.

8. Tomar, H. (2020). *Multi-class image classification of fruits and vegetables using transfer learning techniques* (Doctoral dissertation, Dublin Business School).

9. Mundhada, H., Sanagavarapu, S., Sood, S., Damdoo, R., & Kalyani, K. (2023). Fruit Detection and Three-Stage Maturity Grading Using CNN. *International Journal of Next-Generation Computing*, *14*(1).

10. Vijayalakshmi, M., & Peter, V. J. (2021). CNN based approach for identifying banana species from fruits. *International Journal of Information Technology*, *13*(1), 27-32.

11. Shen, R., Bubeck, S., & Gunasekar, S. (2022, June). Data augmentation as feature manipulation. In *International conference on machine learning* (pp. 19773-19808). PMLR.

12. Bobde, S., Jaiswal, S., Kulkarni, P., Patil, O., Khode, P., & Jha, R. (2021, October). Fruit Quality Recognition using Deep Learning Algorithm. In *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)* (pp. 1-5). IEEE.

13. Singh, P. K., Parashar, N., & Pachauri, R. (2023, July). Fruits and vegetables classification using multi-layered CNN architecture. In *AIP Conference Proceedings* (Vol. 2721, No. 1). AIP Publishing.

14. Xin, M., & Wang, Y. (2019). Research on image classification model based on deep convolution neural network. *EURASIP Journal on Image and Video Processing*, *2019*, 1-11.

15. Min et al.Vision-based fruit recognition via multi-scale attention CNN, Computers and Electronics in Agriculture,Volume 210,2023,107911,ISSN 0168-1699,

16. Raghavendra, & Mallappa, S. (2022). Multi-fruit Classification Using a New FruitNet-11 Based on Deep Convolutional Neural Network. In *Proceedings of the International Conference on Paradigms of Communication, Computing and Data Sciences: PCCDS 2021* (pp. 621-628). Springer Singapore.

17. A. Kushwaha, "Fruit Classification Using Optimized CNN," 2023 International Conference on IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2023, pp. 1-5, doi: 10.1109/ICICAT57735.2023.10263596.

18. Mureşan, H.; Oltean, M. Fruit recognition from images using deep learning. *Acta Univ. Sapientiae Inform.* 2018, *10*, 26–42.

19. Marko Škrjanec (2013) *Computer vision based automatic fruit recognition*. EngD thesis, Faculty of Computer and Information Science, University of Ljubljana, Slovenia.

20. Meshram, V., Thanomliang, K., Ruangkan, S., Chumchu, P., & Patil, K. (2020). Fruitsgb: top Indian fruits with quality. *IEEE Dataport*.

21. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60.

22. .Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep convolutional neural networks. arXiv preprint arXiv:1711.04340.

23. Yao G, Lei T, Zhong J. A review of convolutional-neural-network-based action recognition. Pattern Recogn Lett. 2019; 118:14–22.

24. Dhillon A, Verma GK. Convolutional neural network: a review of models, methodologies and applications to object detection. Prog Artif Intell. 2020;9(2):85–112.

25. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4510-4520).

26. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

27. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427-437.

28. Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37-63.

29. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). MIT press Cambridge.

30. Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37-63.

31. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2921-2929).

32. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252.

33. Ghosh, S., Kim, S., Ijaz, M. F., Singh, P. K., & Mahmud, M. (2022). Classification of mental stress from wearable physiological sensors using image-encoding-based deep neural network. *Biosensors*, *12*(12), 1153.

34. Sahoo, K. K., Ghosh, R., Mallik, S., Roy, A., Singh, P. K., & Zhao, Z. (2023). Wrapper-based deep feature optimization for activity recognition in the wearable sensor networks of healthcare systems. *Scientific Reports*, *13*(1), 965.

35. Chattopadhyay, S., Dey, A., Singh, P. K., Oliva, D., Cuevas, E., & Sarkar, R. (2022). MTRRE-Net: A deep learning model for detection of breast cancer from histopathological images. *Computers in Biology and Medicine*, *150*, 106155.

36. Biswas, M., Sahu, M., Agrebi, M., Singh, P. K., & Badr, Y. (2023). Speech Emotion Recognition Using Deep CNNs Trained on Log-Frequency Spectrograms. In *Innovations in Machine and Deep Learning: Case Studies and Applications* (pp. 83-108). Cham: Springer Nature Switzerland.

37. Kundu, R., Basak, H., Singh, P. K., Ahmadian, A., Ferrara, M., & Sarkar, R. (2021). Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans. *Scientific reports*, *11*(1), 14133.

38. Banerjee, A., Sarkar, A., Roy, S., Singh, P. K., & Sarkar, R. (2022). COVID-19 chest X-ray detection through blending ensemble of CNN snapshots. *Biomedical Signal Processing and Control*, *78*, 104000.