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

Abhik Ganguly1, Rounak Chakraborty, Dipayan Ghosh1, 1, Pawan Kumar Singh1,\*[0000-0002-9598-7981], Aimin Li2 [0000-0002-6983-2310]

1Department of Information Technology, Jadavpur University,

Jadavpur University Second Campus, Plot No. 8, Salt Lake Bypass, LB Block, Sector III, Salt Lake City, Kolkata-700106, West Bengal, India

2School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710048, P.R China.

[{ mail2dipayanghosh@gmail.com, abhik.g2002@gmail.com, rounak.chakraborty02@gmail.com, pawansingh.ju@gmail.com](about:blank), liaiminmail@gmail.com}

\*Corresponding author

**Abstract.** Classifying fruits is crucial in various fields, including agriculture, food processing, and computer vision. Conventional fruit classification techniques often depend on manually designed features and rudimentary machine learning algorithms, which may face challenges with intricate and varied fruit images. In this study, we present a deep learning-based method for automatic fruit classification. This method utilizes a custom-designed Convolutional Neural Network (CNN) incorporating an attention mechanism named as DANN. The datasets utilized in this research comprises a diverse collection of fruit images obtained from three benchmark datasets namely, Fruits-360, FIDS30 and FRUITSGB. Experimental results demonstrate the effectiveness of the proposed attention-based CNN model in accurately classifying fruit images. Training and validation processes involve optimizing hyperparameters such as batch size, number of epochs, and dropout regularization. Furthermore, comparison with baseline models showcases the superiority of the CNN with attention mechanisms in achieving higher classification accuracy. Our findings suggest that integrating attention mechanisms into CNN architectures can significantly improve fruit classification performance by selectively attending to relevant image regions. This research contributes to advancing the state-of-the-art in fruit classification and underscores the potential of deep learning techniques in addressing real-world challenges in agricultural and food-related industries. The proposed model offers promising avenues for future research, including exploring larger datasets, experimenting with different attention mechanisms, and adapting the approach to other object recognition tasks beyond fruit classification. Overall, this study underscores the importance of leveraging deep learning and attention mechanisms for robust and accurate fruit classification, with implications for various applications in agriculture, food quality assessment, and automated farming systems.

**Keywords:** Automatic fruit classification, Deep attention neural network, Fruits-360 dataset, FIDS30 dataset, FRUITSGB dataset

1. **Introduction**

Fruit classification plays a crucial role in numerous fields, including agriculture, food processing, and computer vision. Accurate classification of fruits not only aids in quality assessment and grading but also facilitates inventory management, disease detection, and yield estimation in agricultural settings [1]. Traditionally, fruit classification has relied on manual inspection and subjective judgment, leading to inconsistencies and inefficiencies.Also classifying a fruit allows grocery staff to quickly estimate its price .The emergence of deep learning, especially Convolutional Neural Networks (CNNs), has revolutionized fruit classification by enabling automated and data-driven approaches. These models leverage extensive collections of labeled data and intricate neural network architectures with multiple layers to achieve superior accuracy. [2].

This paper addresses the challenges associated with traditional fruit classification methods and proposes a novel deep learning approach utilizing CNNs with an attention mechanism. Traditional methods often rely on handcrafted features and shallow learning algorithms, which may struggle to generalize across diverse fruit types and variations in shape, size, and color. In contrast, deep learning models can automatically learn discriminative features from raw pixel data, enabling more robust and accurate classification. The motivation behind this research stems from the need for advanced fruit classification techniques that can handle the complexities and nuances present in real-world fruit images. By leveraging 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.

Furthermore, the incorporation of an attention mechanism into the CNN architecture adds an additional layer of sophistication to the model. Attention mechanisms allow the model to focus on relevant regions of the input images while suppressing irrelevant or noisy information. This selective attention mimics the human visual system's ability to prioritize important features, leading to improved classification performance and interpretability. The objectives of this study are twofold: first, to develop a CNN-based model for fruit classification that outperforms traditional methods in terms of accuracy and robustness, and second, to investigate the impact of integrating attention mechanisms on classification performance.

By achieving these goals, we strive to propel the development of automated fruit classification systems, ultimately benefiting various sectors like agriculture, food processing, and retail.

In the subsequent sections of this paper, we will discuss the methodology employed in developing the proposed model, present experimental results and analyses, and discuss the implications of our findings. Additionally, we will compare our model with existing approaches and outline potential avenues for future research in fruit classification and related fields. Overall, this research aims to harness the potential of deep learning and attention mechanisms to address real-world challenges in fruit classification, with implications for improving efficiency and productivity across various domains.

* 1. **Structure of the Chapter**

The remainder of this chapter is organized as follows:

**Section 2. (Literature Review).** Provides a comprehensive review of existing literature on fruit classification, highlighting the evolution of techniques and the challenges addressed by deep learning models.

**Section 3. (Dataset).** Describes the dataset used for training and evaluation, outlining the diversity of fruit classes and the preprocessing steps applied.

**Section 4. (Methodology).** Details the architecture of the CNN, the incorporation of Transformers, and the data augmentation techniques employed.

**Section 5. (Results).** Reports the findings of the experiments, including accuracy metrics and comparisons between the models.

**Section 6 (Discussion).** Analyzes the results, discusses observations, and explores the implications of the study.

**Section 7 (Conclusion).** Summarizes the key findings and contributions of the research, highlighting avenues for future work followed by acknowledgement and references.

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

**2. Literature Review**

Automatic fruit classification has been a subject of interest in various disciplines, including agriculture, computer vision, and machine learning.

Conventional fruit classification approaches frequently employ manually designed features and machine learning techniques including Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN).

While these methods have been effective to some extent, they often struggle with the complexities and variations present in fruit images.

Recent advancements in deep learning, particularly CNNs, have revolutionized the field of image classification, including fruit classification. Convolutional Neural Networks (CNNs) possess the ability to automatically extract hierarchical feature representations directly from raw pixel data, rendering manual feature extraction obsolete. This data-driven method has demonstrated promising outcomes in diverse image recognition tasks, encompassing fruit classification.

Numerous research efforts have solidified the efficacy of Convolutional Neural Networks (CNNs) in the domain of fruit classification.. For example Joseph et al. proposed a CNN model with tensorflow backend and achieved an accuracy of 94.35% [3]. Chung et al. proposed EfficientNet based architecture and got an accuracy of 95% [4].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 used novel modified cascaded ANFIS algorithm to identify fruit 360 image and achieved accuracy of 98.36%[6] 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,.One of the key challenges in fruit classification is the variability in fruit appearance due to factors such as shape, size, color, and occlusions. To address this challenge, researchers have investigated various strategies for improving the robustness and generalization of CNN models.Tomar et al. tried another CNN model to get an accuracy of 95% [8].Mundhana et al. used a technique three stage Maturity grading in CNN to correctly classify the fruits .It achieved an accuracy of 90.24 %[9].Vijayalakshmi et al. tried CNN model on banana species (accuracy of 96.98%) [10]. To improve robustness and generalization of CNN models

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. developed a deep learning based architecture using keras and achieved an accuracy of 95% [12]. Singh et al. performed a multi layered CNN architecture and secured an accuracy of 97% [13].

In recent years, another research avenue gaining momentum is the incorporation of attention mechanisms into CNN architectures. These mechanisms enable the model to selectively focus on crucial areas of the input image, filtering out irrelevant or noisy information.This selective attention mimics human visual perception and has been shown to improve classification performance in multiple tasks, including image classification and object detection [14]. For example, Min et al. proposed an attention-based CNN model for fruit classification that achieved superior performance compared to traditional CNN architectures [15]. Mallapa et al. FruitNet based Deep learning architecture and achieved an accuracy of 96.15% [16]. Kushwala et al. performed an optimized CNN model on the fruit 360 dataset It achieved 96.88% accuracy [17].

A review of existing research underscores the promising potential of deep learning techniques, particularly CNNs with attention mechanisms, in advancing the field of fruit classification. While traditional methods have laid the foundation, deep learning offers a more flexible and data-driven approach that can adapt to the complexities and variations present in real-world fruit images. The following sections of this chapter we will build upon existing research and propose a novel CNN model with attention mechanism for fruit classification, aiming to further improve accuracy and robustness in this domain.

**3 Dataset**

The datasets which we have used in this study comprises a diverse set of fruit images, obtained from the Fruits-360 dataset[18]. It consists of various fruit classes with a substantial number of images per class. To prepare the dataset for model training, validation, and testing, we conducted data augmentation and then divided it accordingly.

**3.1 Dataset Description**

In this research, we utilized two primary datasets: the Fruits-360 dataset and the Citrus Fruits and Leaves Dataset. These datasets comprise a diverse set of fruit images, capturing the intricacies of different shapes, sizes, and colors commonly found in real-world scenarios.

### **Fruits-360 Dataset.** The Fruits-360 dataset is a comprehensive collection of fruit images spanning various classes[18]. It encompasses a wide array of fruits, ensuring a balanced distribution of images across classes to prevent biases during model training. A diverse range of fruits, including apples, bananas, oranges, berries, and tropical fruits, is included in the dataset, covering the spectrum of fruits commonly encountered in agricultural and retail environments. The images in the dataset exhibit variations in lighting conditions, backgrounds, and orientations, contributing to the challenges associated with real-world fruit classification tasks.Some sample images of Fruits-360 dataset are given in **Table 1**.

**Table. 1. Fruits-360 dataset sample images**

| Image |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Label | Apple Red 2 | Avocado | Banana | Blue Berry | Corn | Kaki |
| Image |  |  |  |  |  |  |
| Label | Guava | Hazelnut | Lemon | Mulberry | Onion Red | Papaya |

### **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. The dataset is curated to ensure a balanced representation of fruit classes, facilitating robust model training and evaluation.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.** In addition to the Fruits-360 dataset, we also utilized the FRUITSGB dataset[20]. This dataset focuses specifically on different fruits, providing a specialized collection of images for a targeted classification task. It includes images of various fruits along with Good or Bad labels. The dataset is curated to ensure a balanced representation of fruit classes, facilitating robust model training and evaluation.Some Sample images of FruitsGB dataset are given in **Table 3.**

**Table. 3. Sample images of FRUITSGB dataset.**

| Image |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | Apple\_Bad | Apple\_Good | Banana\_Bad | Banana\_Good | Orange\_Bad |

**3.2 Dataset Composition**

The Fruits-360 dataset comprises a total of 131 fruit classes, each represented by a distinct set of images. Similarly, the Citrus Fruits and Leaves Dataset includes images of various citrus fruits and their corresponding leaves, covering a range of citrus species commonly found in agricultural settings. Both datasets are meticulously curated to encompass a diverse array of fruits and leaves, enabling comprehensive model training and evaluation.

**3.3 Data Preprocessing**

To prepare the datasets for model training, we performed essential preprocessing steps. These steps included resizing all images to a consistent dimension of 100x100 pixels to ensure uniformity across the dataset. Additionally, we applied normalization to set all the pixel values of the image array between ‘0’ and ‘1’, enhancing convergence during the training process. To avoid overfitting and improve model generalization, we employed a stratified data splitting strategy for both datasets. We divided all the datasets into training, validation, and testing sets, maintaining the class distribution in each subset. A training set equipped the models, while a validation set fine-tuned their hyperparameters. Finally, the testing set offered an objective assessment of the models' performance.

**3.4 Data Augmentation**

To increase the ability to handle real-world image variations and prevent overfitting, we implemented data augmentation during training. This crucial step involved strategically altering existing training images to create a diverse and representative dataset.

Our chosen methods included:

* **Random Rotations:** Images were rotated by random angles, simulating different viewing perspectives encountered in practice.
* **Horizontal Flips:** Images were mirrored horizontally, mimicking natural variations in fruit orientation.
* **Brightness Adjustments:** Brightness levels were randomly changed, reflecting varying lighting conditions.
* **Zoom and Shear :**We applied zoom and shear properties in the data augmentation process for the fruit section of the Citrus Fruits and Leaves dataset.

Research by Shorten and Khoshgoftaar (2019) reveals a notable performance boost in models when employing data augmentation techniques., particularly on limited datasets. Additionally, Perez & Wang (2017) highlight its effectiveness in mitigating overfitting and boosting robustness to real-world variations[21,22].

**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. **Table 5** provides a summary of the dataset statistics for FIDS30 and representation of classes. **Table 6** provides a summary of the dataset statistics for FRUITSGB and representation of classes.

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 Classes |
| --- | --- | --- |
| Training | 54154 | 131 |
| Validation | 13538 | 131 |
| Testing | 22688 | 131 |

**Table 5. Overview of the FIDS30 dataset**

| Subset | Number of Images | Number of classes |
| --- | --- | --- |
| All images | 971 | 30 |

**Table 6. Overview of the FRUITSGB dataset**

| Subset | Number of Images | Number of classes |
| --- | --- | --- |
| All images | 12000 | 12 |

**4 Methodology**

**4.1 Convolutional Neural Network (CNN) :**

Convolutional neural networks (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 lightweight convolutional neural network architecture optimized for mobile and embedded vision applications [25]. It features depth-wise separable convolutions and inverted residuals, enabling efficient model inference on resource-constrained devices. VGG16, on the other hand, is a deeper convolutional neural network known for its simplicity and effectiveness [26]. This architecture typically stacks multiple convolutional layers, followed by max-pooling layers, and concludes with fully connected layers.Despite its depth, VGG16 may suffer from computational inefficiency due to its large number of parameters. Nonetheless, both MobileNetV2 and VGG16 offer valuable options for various computer vision tasks, depending on the specific requirements of the application.

**4.3 Model Architecture**

We crafted a CNN framework named DANN, specialized for fruit categorization assignment. This architecture employs a sequence of convolutional layers, interspersed with rectified linear unit (ReLU) activation functions and max-pooling layers. The convolutional layers usually extract spatial features from given input images, while Max Pooling layers downsample the spatial dimensions to capture essential patterns. **Fig. 1** shows the sequence of layers present in our proposed DANN model whereas the architectural overview is presented in **Fig. 2**.

**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 the last convolutional layer and max\_pool\_2d layer

Dropout Layer: Dropout rate is 0.3

Flatten Layer : we added the Flatten Layer

Dense Layer 1:

- Units: 150

- Activation: ReLU

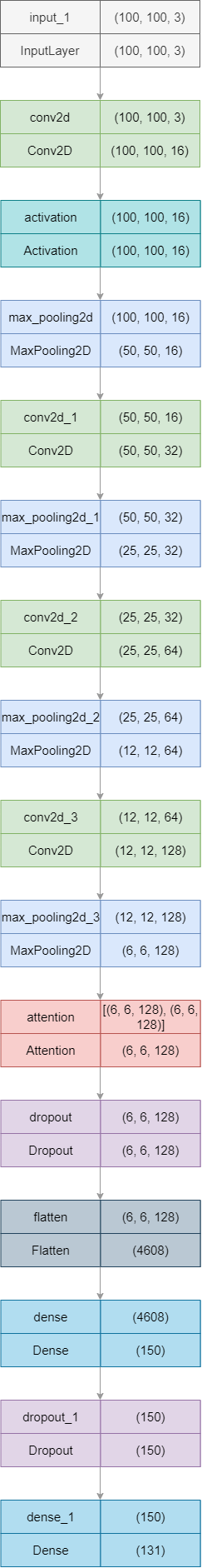
Dropout Layer: Rate: 0.4

Dense Layer 2 (Output Layer):

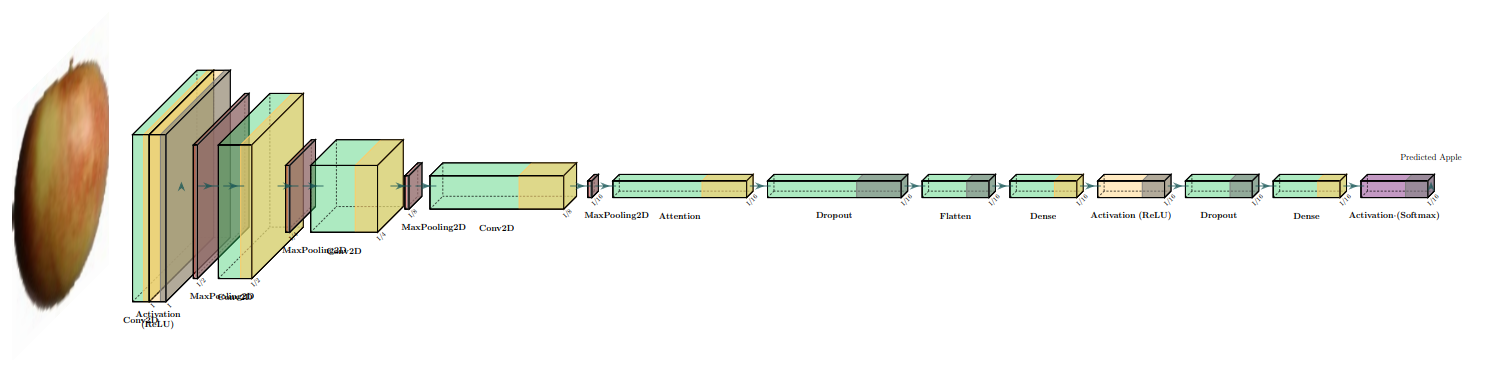
- Units: 131 (Number of classes)

- Activation: Softmax

In this model architecture, the attention mechanism is integrated after the final convolutional layer. This weighting process allows the model to focus on important regions while downplaying less relevant ones, enhancing its ability to extract meaningful features from input images. By incorporating attention, the model aims to improve its performance in tasks such as image classification by selectively attending to relevant features and ignoring noise or irrelevant details. This approach enhances the model's accuracy and robustness in real-world applications, making it well-suited for various 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.**

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

**4.4 Model Evaluation**

The models are 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 model exhibited performance on the fruit classification task, achieving an overall accuracy of 98.06% on the test dataset.Fluctuations in model performance, such as variations in accuracy between different runs of the same model, are normal in machine learning.In general, a difference of a few percentage points (e.g., 1-2%) in accuracy between runs could be considered a slight difference.If we run a model multiple times every time we will not get the same accuracy , it will vary 1-2% .

However, this is a subjective measure and can vary based on the specific context.

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 (e.g Oranges and tangerines), the model exhibited slightly lower precision and recall values, suggesting potential challenges in distinguishing between closely related fruits.The results of DANN model on the Fruits-360 dataset is given in **Table 7**.

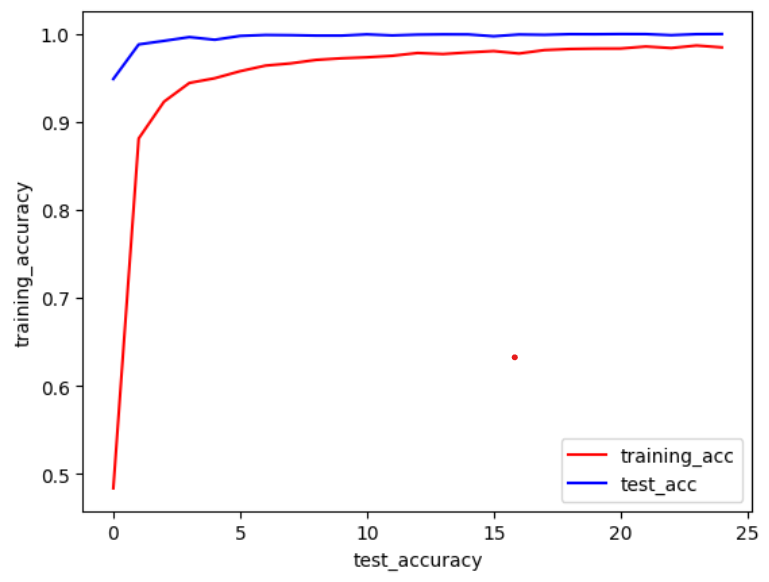
**Table 7. Results of metrics on the Fruits-360 dataset**

| **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** | **98.06%** | **0.98** | **0.98** | **0.98** |

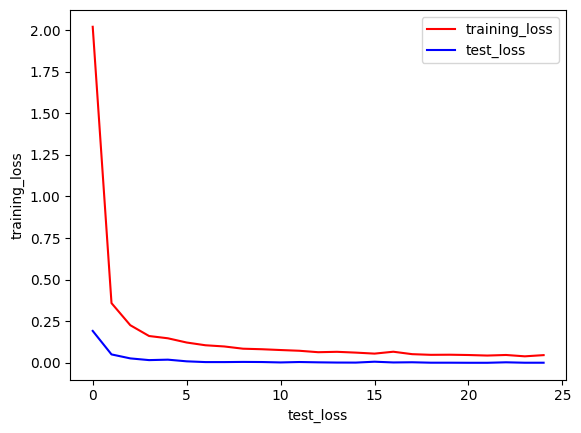
This high accuracy underscores the effectiveness of the CNN architecture in accurately categorizing fruit images.

## 5.1.2 Training and Validation Curves

The training as well as validation accuracy and loss curves offer valuable insights into the learning dynamics of the CNN model throughout the training epochs. Initially, both training and validation accuracy showed a promising upward trend, signaling effective learning and model convergence. However, as training progressed, slight fluctuations in validation accuracy and loss emerged, indicating potential overfitting or convergence challenges [29]. A deeper examination of these fluctuations is essential to fine-tune the training strategy and hyperparameter settings, ensuring enhanced generalization performance and robustness in real-world applications. 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 vs test accuracy for proposed DANN model**

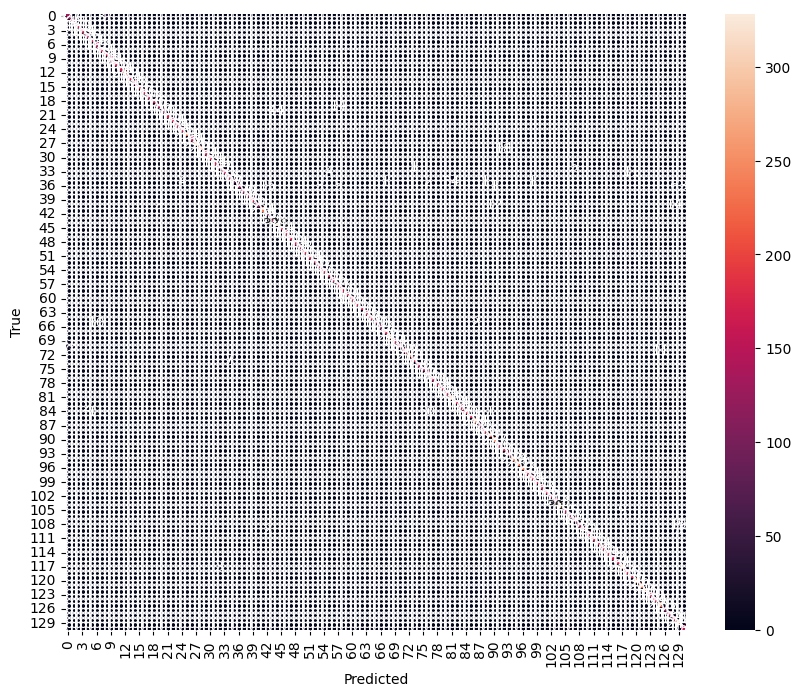


**Fig. 4.   Training loss vs Test loss for proposed DANN model.**

## 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**.



**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 maps 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 attention maps 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 98.06% . 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 or our proposed model with proposed models of other studies**

| **Model** | **Author** | **Year** | **Accuracy** |
| --- | --- | --- | --- |
| CNN model | Joseph et al. | 2021 | 94.35% |
| EfficientNet | Chung et al. | 2019 | 95% |
| optimized CNN model | Kushwala et al. | 2023 | 96.88% |
| FruitNet | Mallappa et al. | 2022 | 96.15 |
| CNN | Singh et al. | 2023 | 97% |
| Maturity Grading | Mundhana et al. | 2021 | 90.24% |
| CNN | Bobde et al. | 2021 | 95% |
| **DANN model** | **Ghosh et al. (Proposed)** | **2024** | **98.06%** |

## 5.1.6   Computational Efficiency

The model achieved high accuracy with relatively fast training times, indicating its suitability for real-time or resource-constrained applications [32].

Efficient resource utilization, including GPU acceleration and parallel processing, further enhances the scalability and practical utility of the CNN model. These findings highlight the computational efficiency and scalability of deep learning approaches for fruit classification tasks.

**5.2 Results on FRUITSGB Dataset**

We Applied our proposed DANN model with an attention layer to another dataset called **FRUITSGB** dataset, and all the results are presented below. The **FRUITSGB** dataset, also referred to as the "Top Indian Fruits with Quality'' dataset, is a publicly available collection of images aimed at supporting research in fruit quality classification. It specifically focuses on fruits of high importance to the Indian market, addressing a gap in existing datasets that often lack quality labels.

Our proposed attention-based CNN model was applied on the **FRUITSGB** dataset after resizing all the images to 224\*224 pixels resulting in an accuracy of 98.67%. This highlights the model's adaptability and efficacy in handling diverse fruit classification challenges. We employed the model on the dataset for fruit identification. The final layer of our model incorporates 12 nodes, corresponding to the 12 distinct classes it can classify.

Our research utilized the **FRUITSGB** dataset, encompassing 12 distinct classes of popular Indian fruits. While the initial **FRUITSGB** dataset comprised 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 9** provides the overview of the splitting of training , testing as well as validation.

**Tabel 9. Overview of FRUITSGB dataset composition and distribution across subset after data augmentation is performed.**

| **Subset** | **Number of Images** | **Number of classes** |
| --- | --- | --- |
| Training | 8640 | 12 |
| Validation | 1200 | 12 |
| Testing | 2160 | 12 |

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 98.67% . 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.

* **Precision** acts as a **gauge for the correctness of positive predictions**. It reveals the **proportion of truly positive cases** amongst all instances labeled as positive.
* **Recall**, also known as **sensitivity**, assesses the model's ability to **capture all relevant instances**. It calculates the **ratio of correctly predicted positive cases** to the total number of true positives.
* **F1-score**, derived from the **harmonic mean of precision and recall**, offers a **balanced perspective** on model performance. This metric is particularly valuable in scenarios with **class imbalance**, where the distribution of classes is uneven.

These metrics are instrumental in refining models to achieve optimal effectiveness across various applications. The results for all three metrics are presented in **Table 10**

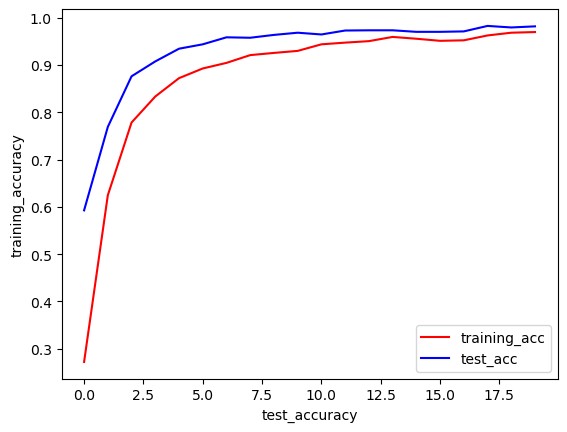
**Table 10 . Metrics analysis of result on FRUITSGB dataset**

| **Model** | **Accuracy(%)** | **Precision** | **Recall** | **F1-score** |
| --- | --- | --- | --- | --- |
| **DANN model** | 98.67 | 0.99 | 0.99 | 0.99 |

## 5.2.2     Training and Validation Curves

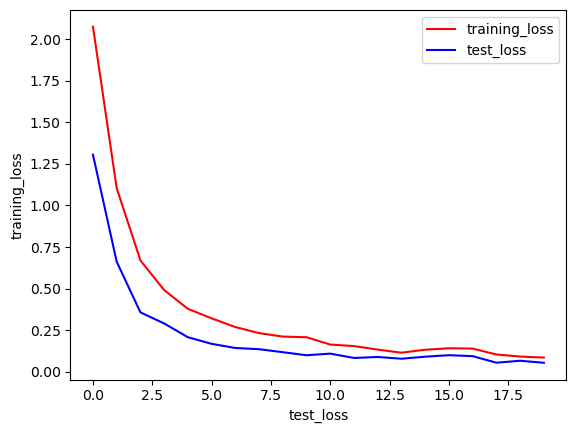
The **training and validation curves** visualize the **performance trajectory** of the DANN model across **training epochs**. These curves typically depict both **loss** (training and validation) and **accuracy** (training and validation), offering valuable insights into the model's ability to **converge** and **generalize** effectively [34].

**Accuracy curves** [35] play a crucial role in **evaluating model performance**, **identifying overfitting**, and guiding the process of **hyperparameter tuning**. They provide a **visual representation** of training and validation accuracies over epochs, facilitating **model selection**, **optimization**, and understanding the **impact of dataset size on learning**. The provided figure displays the **accuracy curve** alongside the **training and testing losses**. As **training accuracy increases**, we observe a corresponding **upward trend in testing accuracy**.This plot depicts the model's performance following 25 training epochs. During training, the model utilized training images, validated its learning with validation images, and reserved testing images for a final, unbiased evaluation of its capabilities.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.**

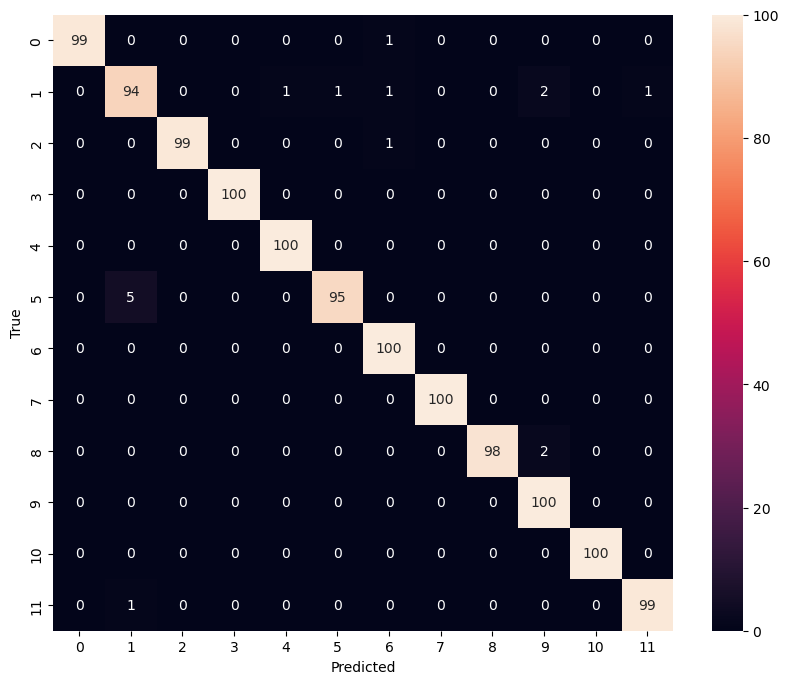


**Fig 7. Loss curve of our proposed DANN model on FRUITSGB Dataset**

## 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**.

****

**Fig 8. confusion matrix for our proposed model on FRUITSGB dataset**

**5.2.4 Comparison With Baseline model**

Our proposed DANN model achieved an accuracy of 98.67% 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 83%. 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 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 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 11** provides the overview of the dataset splitting.

**Tabel 11. Overview of FIDS30 dataset composition and distribution across subset after data augmentation is performed.**

| **Subset** | **Number of Images** | **Number of classes** |
| --- | --- | --- |
| Training | 7717 | 30 |
| Validation | 858 | 30 |
| Testing | 969 | 30 |

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 83% .

## 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** acts as a **sharp pointer** towards the **exactitude of positive predictions**, revealing the **proportion of truly positive cases** amongst all instances labeled as positive.

**Recall**, also known as **sensitivity**, acts as a **spotlight**, illuminating the model's ability to **capture all relevant instances**. It calculates the **ratio of correctly predicted positive cases** to the total number of true positives.

The **F1-score**, derived from a harmonious blend of precision and recall, offers a **balanced perspective** on a model's performance. This metric is particularly valuable in scenarios with **class imbalance**, where the distribution of classes is uneven.

These metrics play a critical role in fine-tuning models to achieve optimal performance across various applications. Results of all three metrics are represented in **Table 12**.

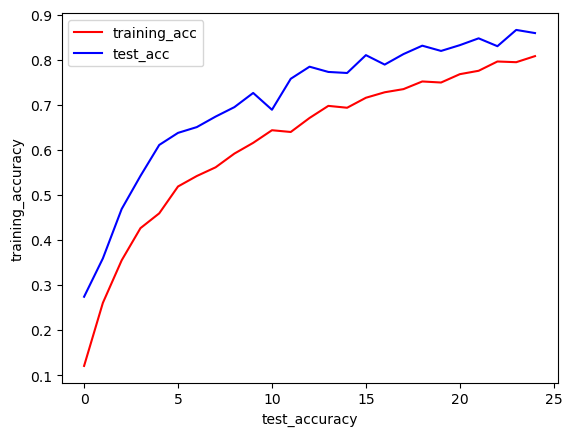
**Table 12 . Metrics analysis of result on FIDS30 dataset**

| **Model** | **Accuracy(%)** | **Precision** | **Recall** | **F1-score** |
| --- | --- | --- | --- | --- |
| **DANN** | 83 | 0.84 | 0.83 | 0.83 |

## 5.3.2     Training and Validation Curves

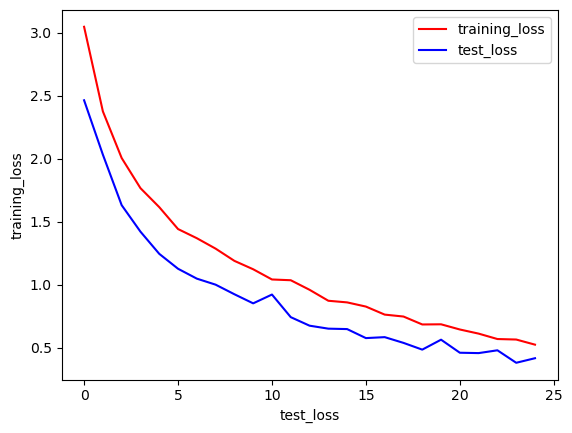
The training and validation curves illustrate the performance of the attention-based CNN model over successive epochs. These curves typically display the training and validation loss, as well as training and validation accuracy, providing valuable insights into the model's convergence and generalization 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.We can see the accuracy curve in **Fig. 9.**



**Fig 9. Accuracy Curve for our proposed model on FIDS30 dataset**

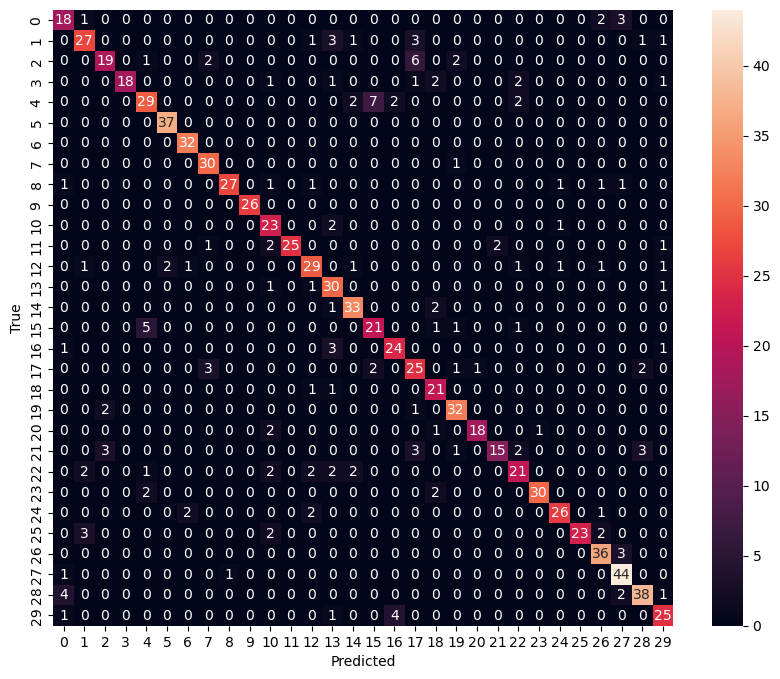
Loss curves are crucial for monitoring the performance of machine learning models during training. They provide insight into how well the model is learning from the training data and whether it is converging towards an optimal solution. A decreasing loss curve indicates that the model is improving its 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**

Above figure is the heatmap of the confusion matrix which will help us compare between true and predicted values .

**5.3.4 Comparison With Baseline model**

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 83%.

**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**

Our DANN model achieved an impressive overall accuracy of 98.06% on the **Fruits-360** dataset, indicating its efficacy in accurately categorizing fruit images. This high accuracy rate underscores the effectiveness of deep learning approaches, particularly CNN architectures, in solving complex classification tasks. By leveraging hierarchical feature representations learned from raw pixel data, CNNs demonstrate remarkable performance in distinguishing between different fruit categories based on their visual characteristics.

Furthermore, the MobileNetV2 and VGG16 models were also applied to the Fruits 360 dataset, yielding accuracies of 97.21% and 95.98%, respectively. These results indicate that both MobileNetV2 and VGG16 are capable architectures for fruit classification, albeit with slightly lower accuracies compared to the custom our DANN model.

**6.2 Interpretability and Explainability**

Interpretability of deep learning models, especially CNNs, remains a crucial aspect for real-world applications, particularly in domains such as agriculture and food quality assessment, where trust and transparency are paramount. While the CNN model demonstrated exceptional performance, understanding its decision-making process and identifying regions of interest in input images are essential for building trust and gaining insights into classification outcomes.

The visualization of attention maps provides valuable interpretability by highlighting the regions of input images that significantly contribute to the model's predictions. By visualizing where the model focuses its attention, stakeholders can gain insights into the features and characteristics that drive classification decisions. This not only enhances trust in the model but also facilitates domain experts' understanding of the underlying factors influencing classification outcomes.

**6.3 Computational Efficiency and Scalability**

The CNN model demonstrated fast training times and efficient resource utilization.

Efficient resource utilization, such as GPU acceleration and parallel processing, is crucial for scaling deep learning models in real-world applications. Our attention-based CNN model emerged as the top performer, showcasing fast training times and efficient resource utilization. Despite this, it's worth noting the advantages of models like MobileNetV2. While MobileNetV2 may not have been our best performer, its lightweight architecture and computational efficiency position it as a viable option 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**

Despite the promising results, several challenges and opportunities for future research remain. Class imbalance, dataset biases, and model interpretability are critical areas that warrant further investigation to enhance the model's performance and generalization capabilities. Strategies such as data augmentation, class-weighted loss functions, and interpretability techniques can mitigate these challenges and improve the robustness of the CNN model.

Furthermore, exploring multimodal data fusion, transfer learning techniques, and domain adaptation strategies can further enhance fruit classification performance, particularly in scenarios where limited labeled data is available or when deploying the model in new environments with domain shifts.

Collaborative efforts between researchers, industry stakeholders, and policymakers are essential to realize the full potential of deep learning in addressing real-world challenges in agriculture and food security. By leveraging advanced AI technologies and interdisciplinary collaborations, we can develop innovative solutions that revolutionize agricultural practices, improve food quality assessment, and contribute to global efforts towards 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..

# 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 discern visual patterns. However, challenges arose when distinguishing visually similar fruits, emphasizing the importance of leveraging both local and global context information for precise classification.

## 7.2 Implications and Contributions

### **Robustness and Sensitivity.** The attention model demonstrated robustness to perturbations, showcasing their ability to generalize well to slightly modified input data. Sensitivity analysis provided valuable insights into the models' resilience, reaffirming their potential for deployment in real-world scenarios where variations in input data are inevitable.

### **Model Interpretability.** The model's attention mechanisms offered a level of interpretability which may be slightly better than normal traditional CNNs in many cases. This interpretability aspect is crucial for applications requiring user trust and understanding of the model's reasoning.

## 7.3 Potential Paths for Future Exploration

**Model Optimization.** The attention model's superior performance prompts further exploration into model optimization strategies. Future research may delve into refining attention mechanisms, exploring different architectures, or leveraging transfer learning techniques to enhance both efficiency and accuracy.

### **Real-world Deployment.** Consideration for the deployment of the model in real-world environments is paramount. Moving forward, researchers should strive to strike a balance between model complexity and computational efficiency, ensuring the feasibility of implementation in resource-constrained settings.

## 7.4 Concluding Remarks

In conclusion, this research has contributed valuable insights to The realm of fruit classification utilizing deep learning models. The hybrid approach, combining CNNs with attention layers, emerges as a promising solution for addressing the challenges associated with visually complex and similar fruit classes. The findings presented here lay the groundwork for continued advancements in image recognition and offer practical implications for applications in agriculture, retail, and beyond.

As the landscape of deep learning undergoes continuous transformation, the lessons learned from this research will inform future endeavors in developing more accurate, robust, and interpretable models for diverse 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.

**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. I Sa, Z Ge, F Dayoub, B Upcroft T. Perez, and C McCool 2016 Deepfruits: A fruit detection system using deep neural networks Sensors 16(8) 1222.
3. Joseph et al. Fruit Classification Using Deep Learning LNEE 756
4. Chung et al. A fruits recognition system based on a modern deep learning technique 2019 *J. Phys.: Conf. Ser.* 1327 012050
5. Y.-D. Zhang et al. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. *Multimedia Tools Appl.*
6. Rathnayake et al. An Efficient Automatic Fruit-360 Image Identification and Recognition Using a Novel Modified Cascaded-ANFIS Algorithm Sensors 2022, 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 et al. Multiclass Image classification of fruits and vegetables using transfer learning techniques
9. Mundhana et al. Fruit Detection and Three-Stage Maturity Grading Using CNN.
10. Vijayalashmi et al. CNN based approach for identifying banana species from fruits Int. j. inf. tecnol. 13, 27–32 (2021).
11. Shen et all Data Augmentation as Feature Manipulation [arXiv:2203.01572v2](https://arxiv.org/abs/2203.01572v2)
12. Bobde et al. "Fruit Quality Recognition using Deep Learning Algorithm," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Pune, India, 2021, pp. 1-5
13. Singh et al. Fruits and vegetables classification using multi-layered CNN architecture AIP Conf. Proc. 2721, 070044 (2023).
14. Wang et al. Research on image classification model based on deep convolutional neural network J Image Video Proc. 2019, 40 (2019).
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,

1. Mallapa et al. Multi-fruit Classification Using a New FruitNet-11 Based on Deep Convolutional Neural Network
2. 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.
3. Mureşan, H.; Oltean, M. Fruit recognition from images using deep learning. *Acta Univ. Sapientiae Inform.* 2018, *10*, 26–42.
4. Marko Škrjanec (2013) *Computer vision based automatic fruit recognition*. EngD thesis, Faculty of Computer and Information Science, University of Ljubljana, Slovenia.
5. Vishal Meshram, Koravat Thanomliang, Supawadee Ruangkan, Prawit Chumchu, Kailas Patil, July 8, 2020, "FruitsGB: Top Indian Fruits with quality ", IEEE Dataport
6. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60.
7. .Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep convolutional neural networks. arXiv preprint arXiv:1711.04340.
8. Yao G, Lei T, Zhong J. A review of convolutional-neural-network-based action recognition. Pattern Recogn Lett. 2019;118:14–22.
9. 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.
10. 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).
11. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
12. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427-437.
13. 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.
14. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). MIT press Cambridge.
15. 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.
16. 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).
17. 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.
18. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60.
19. Perez, L., & Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv preprint arXiv:1712.04621.
20. Brownlee, J. (2020). How to Develop a CNN for Time Series Forecasting. Machine Learning Mastery.
21. Brownlee, J. (2019). Machine Learning Mastery.
22. Calderaro, S., et al. "Visualization and Analysis of Transformer Attention." arXiv preprint arXiv:2201.07302 (2022).