# Deep learning stack learning for evaluating fruit

Divyanshu verma[a] ,Nidhi Singh[b] , Vinothini.R[c]
[a]Rajakumar.K
M.Tech-AI&ML, School of Computer Science and Engineering
Vellore Institute of Technology, Vellore – 632014, Tamil Nadu, India.
[b] Principal Scientist,
Computational Biology Laboratory
Vellore Institute of Technology
Vellore – 632014, Tamil Nadu, India
E-mail: : divyanshu.verma2022@vitstudent.ac.in, nidhi.singh2022@vitstudent.ac.in,
Vinothini.r2022@vitstudent.ac.in

**Abstract**

Stacking is a method that combines sub-model predictions with a learning algorithm. Identifying rotten fruits manually can be a challenging task, and ensuring consistency is crucial. Federated learning allows algorithms to be trained on decentralized devices without transferring local data. Deep learning involves using neural network simulation to extract features from a given dataset and make predictions. The image dataset may contain noise, which can be eliminated through SMV (Standard Normal Variate), and different pre-trained models are used to filter the image dataset. Stacking is applied to multiple models to achieve more accurate predictions. In this approach, 80% of the available data is used for training the model, while 20% is reserved for testing. The results show that the stacking and federated learning approach is highly accurate (>95%) and effective.

**Keywords:** Deep learning; Ensemble learning; SVM; Stacking; CNN; ResNet-50; VGG-19

## Introduction

The primary stage in post-harvest processes that plays a crucial role in determining the price competitiveness of agricultural produce, such as fruits and vegetables, is the grading process. It involves assessing the quality of the produce. During harvesting and storage, various factors can contribute to damages and losses, including mechanical damage, improper handling, hygiene issues, rotting, as well as unfavorable humidity and temperature conditions. To address these challenges, small-scale grading systems utilizing Artificial Intelligence (AI) have been developed specifically for farmers. These AI-powered systems aim to identify and classify the issues mentioned earlier, assisting farmers in improving their produce quality.

The food and beverage industry has witnessed a significant expansion in the application of AI to address unique challenges. Innovative start-ups and tech companies have developed machine learning algorithms to tackle these issues. One specific area where AI is employed is in reducing food waste through grading, where fruits are categorized based on their shape, color, and size. This approach optimizes various post-harvest operations, such as management and packaging. Traditionally, grading and sorting of agricultural produce relied on manual assessment by human experts who estimated factors like color, texture, and freshness. However, this method is labor-intensive, time-consuming, and susceptible to labor shortages during peak seasons. To overcome these challenges and reduce costs, AI-powered automated grading machines have become increasingly convenient. Deep learning techniques are utilized to identify fruit quality and freshness, enhancing the accuracy and speed of classification between fresh and rotten fruits. Various approaches, including Support Vector Machine (SVM), have been employed to handle high-dimensional data, prevent over-fitting, and achieve good generalization performance. Stack learning is also applied to train models for classifying fresh and bruised agricultural produce, improving overall efficiency.

Deep learning is a type of machine learning that utilizes multiple layers of artificial neural networks to learn from data. These layers in a deep neural network make predictions based on intricate features present in the input data. Convolutional Neural Networks (CNNs) are a specific type of deep neural network commonly used for image classification tasks. The convolutional layers in CNNs consist of learnable filters that are applied to the input data, resulting in a feature map. Pooling layers are then used to down-sample the feature map, reducing its spatial resolution while retaining important features. The combined output from the convolutional and pooling layers is fed into dense layers for the final prediction. CNNs have demonstrated remarkable performance in various tasks such as object detection, handwriting recognition, facial recognition, and other image recognition tasks. These capabilities make CNNs a promising approach in achieving accurate results in these areas.

Stack learning, also referred to as Stacked generalization, is an ensemble learning technique that involves training multiple base models or individual models on a dataset. The predictions made by these base models are then used as input features for another model called the stack model or meta model, which generates the final prediction. The concept behind stack learning is that different patterns in the data can be captured by the base models, and integrating their predictions as inputs to the meta model can improve overall prediction performance. This technique is applicable to both regression and classification tasks and finds utility in real-world scenarios where combining the predictions of multiple models can yield better results. In the context of agricultural sorting and grading systems, the integration of deep learning techniques holds significant potential for enhancing efficiency, improving the market competitiveness of agricultural produce, and minimizing post-harvest losses.

**Literature Survey :**

The review of existing literature on imaging techniques and their applications in agriculture has presented substantial evaluations conducted over the past three decades. The review concentrated on different factors such as imaging platforms, sensors, analytical techniques, and methodologies based on literature analysis. The primary objective of this review is to provide agricultural researchers and practitioners with a deeper insight into the benefits and constraints of image processing in agricultural applications. It is anticipated that this extensive analysis will encourage the broader acceptance and utilization of this valuable technology within the agricultural sector.

[1] Sultana et al. (2022) published a paper presenting a dataset specifically designed for the classification of fresh and rotten fruits. The dataset consists of sixteen different fruit classes, comprising a total of 3,200 original images and 12,335 augmented images. The images were collected from various fruit shops and real fields with the assistance of a domain specialist. The study outlined a five-step process, including augmentation, image resizing, dataset splitting, model generation, and performance analysis, for effectively training the dataset using deep learning models. The availability of this dataset and the proposed steps can greatly contribute to the development of more accurate fruit recognition algorithms while reducing computation time.

[2] Knott et al. (2023) conducted research that introduces a streamlined machine learning method for assessing fruit quality based on images. The study specifically focuses on two types of fruit quality attributes: color-based assessment for banana ripeness and local features for detecting surface defects in apples. The proposed approach utilizes pre-trained vision transformers and employs transfer learning by combining pre-trained feature extractors with untrained classifiers. Remarkably, the approach achieves competitive classification accuracy while utilizing only one-third of the training samples compared to CNN-based models. The research findings indicate that the pre-trained DINO ViT embedding's outperform supervised CNN embedding in capturing crucial image features, underscoring the effectiveness of the proposed methodology for fruit quality assessment.

[3] In a study conducted by Kumar et al. (2023), the detection of rotting fruits in agriculture was identified as crucial in preventing contamination and reducing food waste. To address these concerns, the researchers developed a CNN model to automate the classification of fresh and rotten fruits. The proposed CNN model exhibited a remarkable accuracy rate of 97.14% in distinguishing between fresh and rotten fruits, surpassing existing methods. Automating the fruit classification process not only reduces the likelihood of human error but also contributes to the reduction of food waste. Future research can expand upon this study by incorporating a wider range of fruits and vegetables, thereby further enhancing its potential impact in minimizing food waste in today's society.

[4] Haque conducted research on leveraging deep learning techniques to improve the freshness of fruits and vegetables, with the goal of enhancing various aspects of food production, sorting, packaging, and delivery processes. They compared the performance of seven pre-trained CNN models with their own custom CNN-based image classification model called "FreshDNN." The custom model outperformed the pre-trained models across different performance metrics. It achieved an impressive training accuracy of 99.32% and a validation accuracy of 97.8%. The

custom model demonstrated superior precision, recall, and F1 score, except for VGG19. Additionally, their custom model excelled in terms of the number of parameters, training time, ROC-AUC score, computational cost, and space utilization. The results from five-fold cross-validation further validated the effectiveness of their model, providing promising outcomes.

[5] Bhavya and their colleagues developed a fruit quality prediction system utilizing deep learning techniques for agricultural purposes. They proposed two models: a customized CNN architecture and the transfer learning approach using the pre-trained VGG model. The models were trained and evaluated using a subset of the Fruits 360 dataset, which consisted of 10 different fruit types. The suggested model achieved impressive accuracy, with 99.39% accuracy on the training data and 99.99% accuracy on the validation data. The utilization of transfer learning further enhanced the classification accuracy. However, the model encountered issues with under-fitting and consumed significant time and storage resources. Despite these limitations, the results showcased the model's ability to effectively differentiate between fresh and rotten fruit, indicating its potential application in real-time farming scenarios.

[6] Kumar R. and his research team developed a machine vision system that efficiently evaluates the quality of pomegranate fruits. They extracted features from sample images in both the spatial domain and wavelet domain. Artificial neural networks (ANNs) and support vector machines (SVMs) were trained using these features. The results demonstrated that ANNs outperformed SVMs in accurately classifying pomegranate images, and wavelet features achieved higher accuracy compared to spatial domain features. The ANNs trained with wavelet features achieved an accuracy of 92.65%, highlighting their effectiveness in assessing fruit quality. The study suggests further analysis of different pomegranate cultivars and exploration of feature ranking and reduction techniques to enhance the system's performance.

## Materials and methods

### Dataset

### Data Acquisition

The analysis was conducted using a dataset comprising 10,901 fruit images. All images were captured at a resolution of 100x100 pixels. The process of acquiring these images presented various challenges, including issues related to lighting conditions such as sunlight, darkness, shadows, as well as variations in camera artifacts, pose, and lighting changes. To make the dataset more realistic, images from the same category were captured at different times and days. Detailed specifications of the images can be found in Table 1.

**Table 1.** List of fruit images

| S.NO | Fruit name | Image Count |
|---|---|---|
| 1. | Fresh Apple | 1693 |
| 2. | Fresh Banana | 1581 |
| 3. | Fresh Orange | 1466 |
| 4. | Rotten Apple | 2342 |
| 5. | Rotten Banana | 2224 |
| 6. | Rotten Orange | 1595 |

### Data Preparation

In this phase, the dataset consisting of 10,901 fruit images is divided into a training dataset and a test dataset, with a split of 70% and 30% respectively. To effectively analyze the information contained in the images' spatial and texture features, an efficient analytical method is needed. As a result, the data undergoes preprocessing to ensure its suitability for deep learning. This preprocessing step not only prepares the data but also enhances the accuracy and efficiency of the learning model.

## Experimental Procedure

### Methods

In this study, we employ federated learning as a method to introduce dataset heterogeneity and train multiple models. Our primary objective is to achieve the highest accuracy and precision in classifying rotten fruits. Initially, data augmentation is applied to expand the training dataset and enhance accuracy by increasing the quantity of training data. Data normalization is then performed to prevent over-fitting. The obtained data is subsequently divided into a 70:30 ratio for training and testing purposes. Next, we compare the accuracy and precision obtained from different models, namely ResNet, VGG19, and a simple CNN. Notably, the simple CNN model is trained using federated learning techniques to improve its accuracy and performance on any given dataset. The prediction results from all three models are combined in an ensemble model (meta model), which generates predictions by considering the inputs from all the models collectively. The results reveal an overall increase in accuracy through ensemble stacking compared to the individual base models.

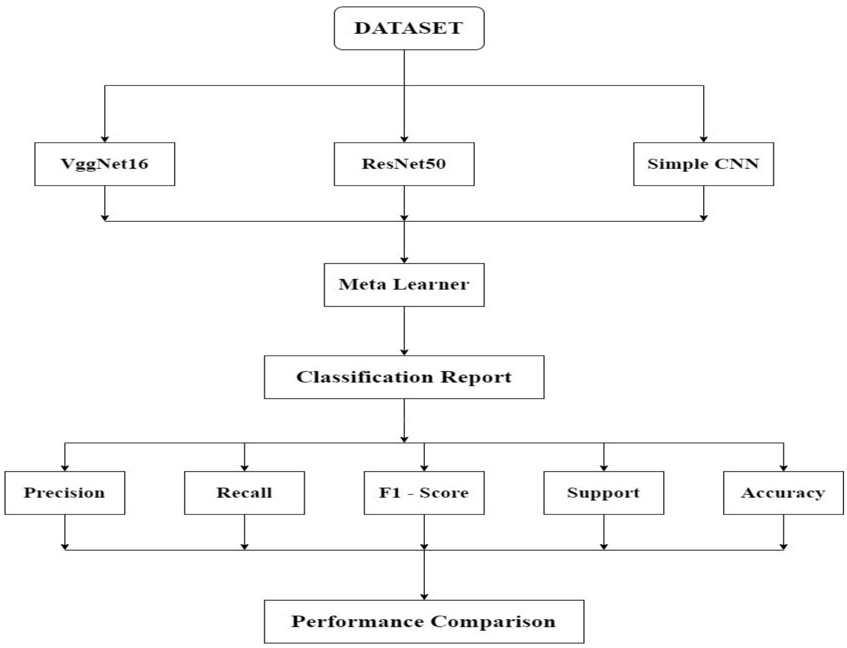


**Figure 1.** Flowchart of working set model

### A.VGG-19

VGG-19 surpasses VGG-16 in various aspects and is a convolutional neural network model with 19 layers. The model is constructed by stacking convolutions together. However, the model's depth is limited due to the issue of decreasing gradients, which poses challenges in training deep convolutional networks. Like the other models, VGG-19 was pre-trained on ImageNet, a dataset containing 1,000 different object categories, for classification tasks. In our study, we focused on feature extraction and amplification using VGG-19. The best results were achieved by solely utilizing feature extraction with VGG-19. In our specific case, retraining and fine-tuning the VGG-19 model did not yield significant improvements.

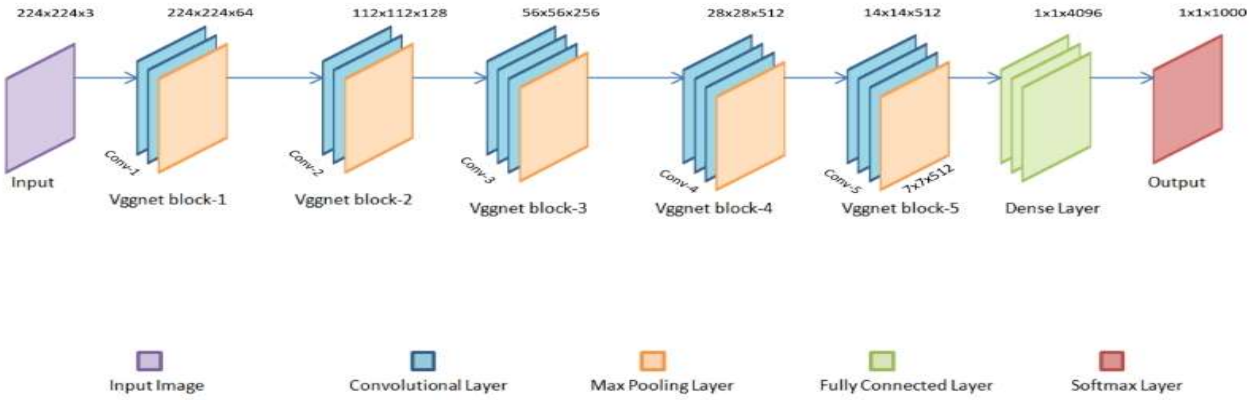


**Figure 2.** VGG-19 architecture

**B.ResNet-50**

The ResNet model was proposed as a solution to address the issue of decreasing gradients. It introduces skip connections, allowing the model to bypass certain layers and pass the residual information to the subsequent layers. This approach enables deeper training of convolutional neural network models. Among various modifications of the ResNet model, we chose ResNet50 for our study. From our experimentation, the optimal outcome was achieved by retraining approximately 40% of all the parameters in the ResNet50 model.

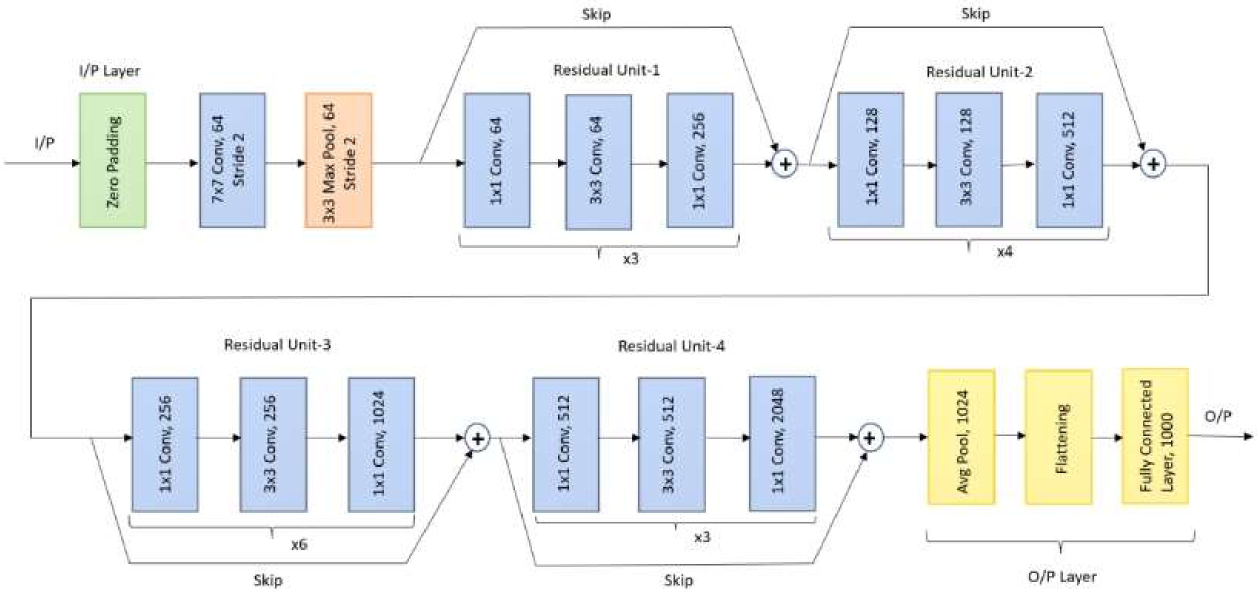


**Figure 3.** ResNet-50 architecture

## C. Simple CNN

This model consist of Conv2D layer with 32 filters, MaxPooling2D layer with pool size of (2, 2), Dropout layer with rate of 0.25, Conv2D layer with 64 filters, MaxPooling2D layer with pool size of (2, 2), Dropout layer with rate of 0.25, Conv2D layer with 128 filters, MaxPooling2D layer with pool size of (2, 2), Dropout layer with rate of 0.25, Flatten layer, two dense layers with 512 and 256 neurons, and a Softmax activation function. Compile model using Adam optimizer, categorical cross-entropy loss function, and accuracy metric.

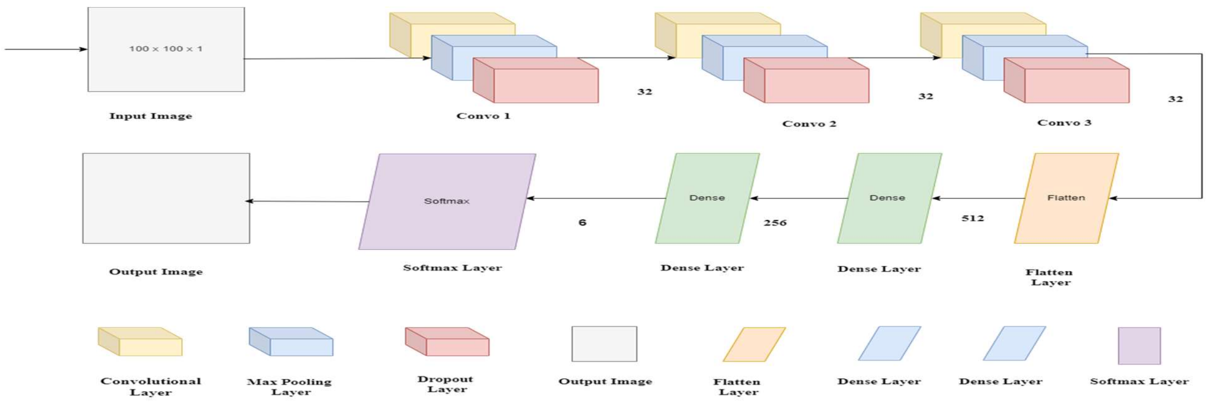


**Figure 4.** Simple CNN architecture

## Ensemble Learning

Ensemble learning is a machine learning approach that leverages multiple models to make predictions. The fundamental concept behind ensemble learning is that combining the predictions of multiple models can often yield improved overall performance compared to relying on a single model. Ensemble methods have gained widespread popularity and have demonstrated high effectiveness across various machine learning tasks.

**Stacking**

Stacking is an ensemble technique that operates at a meta-level by training multiple models and utilizing their predictions as input features for a meta-model (as shown in Figure 5). The meta-model then takes these predictions and combines them to produce the ultimate prediction. By leveraging the diverse characteristics of multiple models, stacking has the potential to enhance performance and achieve better results.

**The ensemble stacking process is outlined below:**

The training data should be split into at least two subgroups in the following manner: It is common practice to create a holdout set from the training data for evaluating the model, while the remaining data is utilized to train the foundational models.

Base model training: Train several base models using the provided training data. Each base model can be trained using a different algorithm or a specific set of hyper-parameters. Each model generates predictions for the holdout set.

Develop a meta-model by using the predictions of the base models as input features and the true labels from the holdout set as the target variable. Train the meta-model using this new dataset. The purpose of the meta-model is to learn how to effectively combine the predictions from the base models and generate the final prediction.

Make predictions using the ensemble: Once the meta-model is trained, utilize it to make predictions on new, unseen data. The base models generate individual predictions, which are subsequently fed into the meta-model. The meta-model then generates the final prediction based on these inputs.

**Federated Learning :**

Federated Learning is an innovative approach to machine learning that allows models to be trained on different data sources without the need for sharing raw data. It enables collaboration among multiple devices or clients to train a global model while ensuring that data remains stored locally on each device. This decentralized approach addresses concerns related to privacy and data ownership. In federated learning, models are trained locally on client devices, and their updates are aggregated to create a global model (as shown in Figure 5).

The process involves client selection, model distribution, local model training, model aggregation, and model updates. Federated learning offers several benefits, including privacy preservation, data ownership, efficiency, and scalability.

Mathmatically,

$$\mathrm{Wt+1} = \Sigma_{k=1} {}^{n_k}/_{n}\, Wt + 1$$

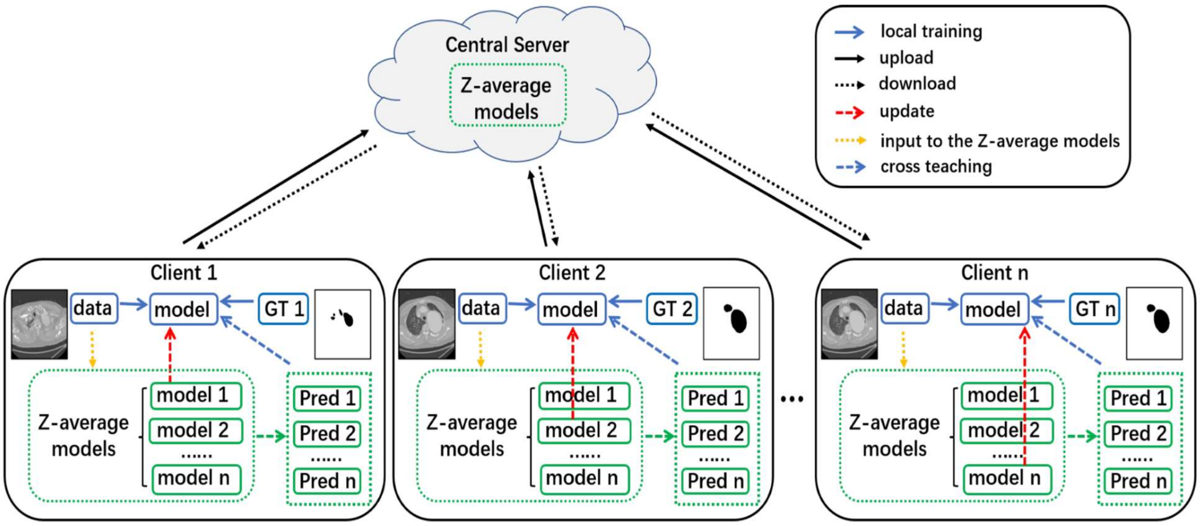

**Figure 5** : Block Diagram of stacking and ensemble learning model

The framework of this study involves a process where all clients train their models locally and submit them to a global server after five training iterations. The global server aggregates the models from three clients to create Z-average models. Each client then downloads the Z-average models and updates its local model accordingly. In the subsequent cross-training epoch, the predictions from the Z-average models are combined with the local ground truth to train the local model further.

**Table 2.** Accuracy of different models for all samples

| Fruit | Model | Precision | Recall | F1-score | Support |
|---|---|---|---|---|---|
| **Fresh Apples** | ResNet50 | 0.65 | 0.92 | 0.76 | 516 |
| | VggNet | 0.72 | 0.85 | 0.78 | 440 |
| | Simple CNN | 0.61 | 0.89 | 0.72 | 356 |
| | Ensemble | 0.89 | 0.88 | 0.88 | 516 |
| **Fresh Banana** | ResNet50 | 0.69 | 1.00 | 0.82 | 471 |
| | VggNet | 0.94 | 0.94 | 0.94 | 472 |
| | Simple CNN | 0.79 | 0.99 | 0.88 | 374 |
| | Ensemble | 0.93 | 0.98 | 0.95 | 471 |
| **Fresh Oranges** | ResNet50 | 0.86 | 0.82 | 0.84 | 419 |
| | VggNet | 0.86 | 0.82 | 0.84 | 438 |
| | Simple CNN | 0.73 | 0.84 | 0.78 | 363 |

| | | | | | |
|---|---|---|---|---|---|
| | Ensemble | 0.93 | 0.88 | 0.90 | 419 |
| **Rotten Apples** | ResNet50 | 0.94 | 0.46 | 0.61 | 700 |
| | VggNet | 0.71 | 0.79 | 0.75 | 635 |
| | Simple CNN | 0.87 | 0.75 | 0.80 | 815 |
| | Ensemble | 0.90 | 0.78 | 0.84 | 700 |
| **Rotten Banana** | ResNet50 | 1.00 | 0.82 | 0.90 | 678 |
| | VggNet | 0.89 | 0.99 | 0.94 | 610 |
| | Simple CNN | 0.99 | 0.94 | 0.96 | 714 |
| | Ensemble | 0.99 | 0.96 | 0.97 | 678 |
| **Rotten Oranges** | ResNet50 | 0.69 | 0.80 | 0.74 | 487 |
| | VggNet | 0.83 | 0.60 | 0.69 | 676 |
| | Simple CNN | 0.86 | 0.64 | 0.73 | 649 |
| | Ensemble | 0.73 | 0.89 | 0.80 | 487 |

**Results and Discussions :**

In this research, we utilize the federated learning approach to train multiple models, introducing heterogeneity in the dataset. The dataset used consists of 10,901 images of fresh and rotten fruits, which are divided into 70% for training and 30% for testing. Various models, including Simple CNN, VggNet, ResNet, and ensemble learning, are employed to assess fruit quality. Preprocessing techniques are applied to enhance the accuracy and efficiency of the learning models. Data augmentation is performed to increase the number of training data samples, and data normalization helps avoid over-fitting. The performance of ResNet50, VggNet19, and Simple CNN models are evaluated in terms of training and prediction. Additionally, a federated learning approach is applied to train the Simple CNN model for improved accuracy. The prediction results from these models are combined in an ensemble model, resulting in higher accuracy compared to the individual base models. The overall accuracy of ResNet50, VggNet19, Simple CNN, and Ensemble models is reported as 78%, 82.18%, 82.05%, and 89%, respectively.

| Model | Loss | Accuracy (%) |
|---|---|---|
| Simple CNN | 0.5037 | 82.05% |
| VggNet | 0.4802 | 82.18% |
| ResNet50 | 0.8070 | 78% |
| Ensemble | 0.267 | 89% |

**Table 3.** The accuracy and loss value obtained by different models

### VggNet

According to Table 2, the VggNet model achieves the highest F1-scores for fresh bananas and rotten bananas, with scores of 0.94. However, it performs poorly for fresh apples and rotten oranges, with F1-scores of 0.78 and 0.69, respectively. The VggNet model demonstrates accuracy above 84% for all other fruits. The overall accuracy of the VggNet model is reported as 82.18%, with a loss value of 0.4802. Compared to other existing models, the VggNet model outperforms ResNet50 but falls short of the performance achieved by the Ensemble model.

### ResNet-50

Based on the results presented in Table 3, ResNet50 exhibits the lowest accuracy among the four models evaluated. It demonstrates good performance for rotten bananas and fresh oranges, achieving F1-scores of 0.90 and 0.84, respectively. However, when compared to the other models, ResNet50 performs poorly in terms of F1-score, recall, and loss value. The overall accuracy of the ResNet model is reported as 78%, with a loss value of 0.8070. Although it performs well in classifying both fresh and rotten fruits, the ResNet model is unable to surpass the performance of the other existing models in terms of overall accuracy. Figure 6 visually illustrates the classification accuracy of the ResNet model, considering both accuracy and loss value.

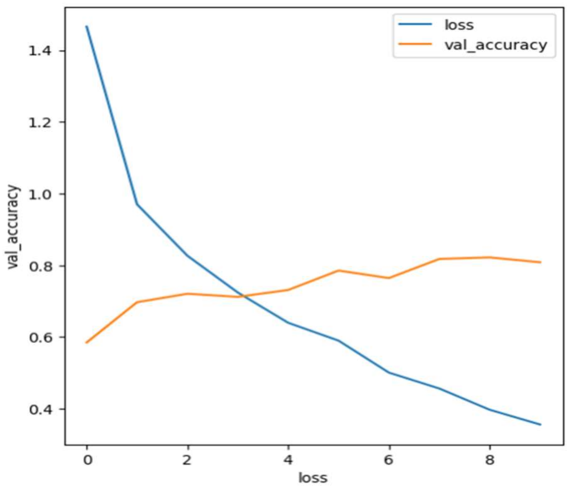


**Figure 6.** VggNet classification accuracies.

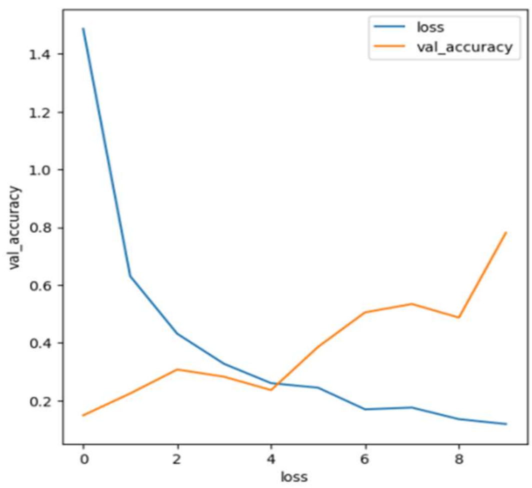


**Figure 7.** ResNet classification accuracies.

### Ensemble Learning

As indicated in Table 3, the ensemble model achieves the highest accuracy compared to all the other models. Specifically, Table 2 shows that the ensemble model performs well in classifying fresh bananas and rotten bananas, achieving F1-scores of 0.95 and 0.97, respectively. However, it exhibits lower performance in the case of fresh apples and rotten oranges, with F1-scores of 0.88 and 0.80, respectively. Except for these fruits, the ensemble model demonstrates accuracy exceeding 90% for all other fruits. In comparison to the other existing models (ResNet50, VggNet, and Simple CNN), the ensemble model outperforms them. With an overall accuracy of 89% and a loss value of 0.0676, the ensemble model surpasses the other three models in terms of overall accuracy and performs well in classifying all types of fruits

**Simple CNN**

Based on Table 3, the Simple CNN model achieves the highest average accuracy among the four models. Specifically, it demonstrates good performance in classifying rotten bananas and fresh bananas, with F1-scores of 0.96 and 0.88, respectively. In comparison to the other three models, Simple CNN outperforms ResNet50 in terms of F1-Score, recall, and loss value. However, it falls short in comparison to VggNet19 and the Ensemble model. The overall accuracy of the Simple CNN model is 82.05%, with a loss value of 0.5037. Nevertheless, it is unable to surpass the other existing models in terms of overall accuracy, although it performs well in classifying both fresh and rotten fruits.

**Conclusion :**

This research focuses on using deep learning techniques to accurately classify the quality and freshness of fruits, specifically distinguishing between fresh and rotten ones. Various methods, including support vector machines (SVM), have been explored for this purpose. The dataset used in this study consists of 10,901 fruit images captured at a resolution of 100 x 100 pixels. The primary goal is to achieve highly accurate classification of rotten fruits. To enhance the training dataset and optimize the learning model's performance, data preprocessing techniques are applied to reduce redundancy and increase accuracy. The results indicate the promising potential of the proposed models for this application. Three different models, namely VggNet19, ResNet50, and Simple CNN, along with ensemble learning, are evaluated for their classification performance. Federated learning is employed, allowing multiple models to be trained on heterogeneous datasets. The predictions from these models are then combined in the ensemble model to make accurate test case predictions. The ensemble model achieves the highest accuracy of 89%, outperforming the other developed models in terms of classification performance. VggNet19 is particularly useful for assessing fruit quality due to its overall accuracy. However, the fruit quality assessment research community still faces significant challenges and obstacles in this domain.

**Acknowledgments:**

The authors would like to express their gratitude to S. Rajakumar, an Associate Professor at the School of Computer Science and Engineering, for providing valuable encouragement and support during the development of the original version of this document.

**Conflict of Interest Statement:**

The authors affirm that there are no conflicts of interest.

**References**

[1] Sultana, Nusrat et al. “An extensive dataset for successful recognition of fresh and rotten fruits.” Data in brief vol. 44 108552. 24 Aug. 2022, doi:10.1016/j.dib.2022.108552

[2] Manuel Knott, Fernando Perez-Cruz, Thijs Defraeye, Facilitated machine learning for image-based fruit quality assessment, Journal of Food Engineering, Volume 345, 2023, 111401, ISSN 0260-8774, https://doi.org/10.1016/j.jfoodeng.2022.111401.

[3] T. Bharath Kumar, Deepak Prashar, Gayatri Vaidya, Vipin Kumar, S. Deva Kumar, F. Sammy, "A Novel Model to Detect and Classify Fresh and Damaged Fruits to Reduce Food Waste Using a Deep Learning Technique", Journal of Food Quality, vol. 2022, Article ID 4661108, 8 pages, 2022. https://doi.org/10.1155/2022/4661108

[4] Haque et al, (2022).Fruit and vegetable freshness detection using deep learning. Brac University, http://hdl.handle.net/10361/17930

[5] Bhavya K. R and S. Pravinth Raja, “Fruit Quality Prediction using Deep Learning Strategies for Agriculture”, Int J Intell Syst Appl Eng, vol. 11, no. 2s, pp. 301–310, Feb. 2023

[6] Kumar R. A.,Rajpurohit V. and Jirage B. (2018) Pomegranate Fruit Quality Assessment Using Machine Intelligence and Wavelet Features. Journal of Horticultural Research, Vol.26 (Issue 1), pp. 53-60. https://doi.org/10.2478/johr-2018-0006

[7] Raikar, M.M., Meena, S.M., Kuchanur, C., Girraddi, S. and Benagi, P., 2020. Classification and Grading of Okra-ladies finger using Deep Learning. Procedia computer science, 171, pp.2380-2389.

[8] Singh, A., Vaidya, G., Jagota, V., Darko, D.A., Agarwal, R.K., Debnath, S. and Potrich, E., 2022. Recent advancement in postharvest loss mitigation and quality management of fruits and vegetables using machine learning frameworks. Journal of Food Quality, 2022, pp.1-9.