Transfer Learning Method to Determine the Freshness of Fruits and Vegetables from Image Dataset

Abstract—**Fruits and vegetables are valuable in our life due to their strong nutritional and therapeutic value. The process of manually determining the freshness of fruits and vegetables using human vision is time-consuming. In this study, we have suggested a model that will determine the freshness of fruits and vegetables. With the help of our model, we can quickly determine which fruits and vegetables we should consume before they go stale. It contributes in minimizing the financial loss, brought on by wasted fruits and vegetables. MobileNetV2 and InceptionV3 are pre-trained convolutional neural networks which are fine-tuned using the transfer learning approach. These models are used in our research to help categorize the freshness of fruits and vegetables. To boost the amount of data, we built a new dataset by combining the two Kaggle datasets named Fruits fresh and rotten for classification and Fresh and Stale Images of Fruits and Vegetables. Among MobileNetV2 and InceptionV3 models, MobileNetV2 scored better with an accuracy of 98% to detect the freshness of fruits and vegetables.**

Index Terms—**Fruit Freshness, Image Dataset, Transfer Learning, MobileNetV2, InceptionV3, ImageNet, TensorFlow Hub**

I. INTRODUCTION

Agriculture is the foundation of the global economy. According to the World Bank [1], in 2018 it contributed around 4% of the world's GDP and approximately 45% of the GDP of the least developed countries of the world. Different countries like France, Mexico, Brazil, India, USA, Germany, Japan, China, Russia, and Turkey are the top 10 agriculture-producing counties in the world [2]. Bangladesh is an agricultural nation consisting about 22.7 million people involved in agriculture. It is the majority of people in our nation that are directly or indirectly dependent on agriculture. By the end of 2022, Bangladesh is predicted to have a GDP from agriculture of 12291.00 Million BDT [3].

A variety of nations around the world import and export agricultural products. Additionally, as people consume fruits and vegetables very often, their freshness is crucial for maintaining general welfare. Moreover, if fresh fruits come into contact with rotten ones, fresh fruits can be harmed as well, which will impede the nation's economic progress. So, in many aspects of people's lives, the ability to identify fruit freshness is essential.

Using advanced technologies, several researchers in the agricultural industry are working to identify and detect different plant diseases. Some common fruit diseases are apple scab, powdery mildews, Sooty Blotch, Bitter Rot, Grape black rot, etc. These diseases cause great losses to the fruit industry around the world. Various researchers applied different Machine learning, deep learning, and transfer learning models to classify these diseases from different fruits and vegetables. A deep learning model, like CNN, is the best for image classification because of its high accuracy level. That is why the CNN model is most suitable for detecting diseases from the fruit image datasets.

Around the world, fresh fruits contributed approximately 622.80 US dollars in revenue in 2022 [4]. Moreover, according to an article [5], every year around 45% of all fruits and vegetables are wasted due to their freshness. Since fruit and vegetable waste results in a significant loss to the economy, we have suggested a model in this research that will identify the freshness of fruits. Using our model we can easily determine which fruits we need to eat before they become stale. We have used MobileNetV2 and InceptionV3 models as these models are pre-trained on huge datasets. Besides, we have combined two datasets to boost our model which will help to classify fruit freshness more accurately than other models. Moreover, MobileNetV2 performs well on mobile devices so it will help people easily detect fruit freshness using their mobile devices.

In the next section, Section 2 of the paper, we have described previously done works in detecting fruit diseases. In Section 3, we have discussed the dataset of our research purpose. After that, the methodologies and steps of our research work are briefly described in Section 4. Next, the result and analysis part of this research work is shown in Section 5. Finally, we conclude our paper by mentioning future works in Section 6.

II. RELATED WORKS

Fruit disease detection and classification task has been done previously by many researchers using the help of advanced technologies. Malathy et al. suggested a method to identify fruit diseases using image processing techniques in their research paper [6]. In their article, they used CNN, a deep learning model, to identify fruit diseases. For their research, they have gathered a fruit image dataset from Kaggle that includes three different fruit diseases, including Sooty Blotch, Bitter Rot, and Powdery Mildew. Their collected images are in .JPG format. At first, they performed preprocessing in the image dataset by rotating some images by 90 degrees and resizing the images by 250 x 250 pixels in size. After that, they performed image segmentation and feature extraction. Finally, the CNN model is used to detect diseases from fruits. The applied method achieved around 97% accuracy in fruit disease detection.

A recent research work [7] presented a model to detect defects from the surface of Mangosteen fruit by using the Convolutional Neural Network. In this research, the image dataset is collected manually by researchers by checking the shape of the fruits. Based on the shapes, the researchers classified images into two categories: fine and defect. In the image preprocessing step, they resized the images in 512 x 512 pixels to clearly show the surface of the fruit. The image datasets are passed into the CNN model as an input and detect defects from the surface of Mangosteen. In this paper, 120 images are used to test the dataset where 30 images were from the defect category and 90 images were from the fine category. They have used 4-fold cross-validation to improve the accuracy of their classification. Around 97.5% accuracy is achieved to detect the defects by using their proposed algorithm.

Another proposed paper for fruit disease identification [8] used K-means clustering and multi-class SVM model to detect diseases from image datasets. The authors identified three different apple diseases: apple rot, apple blotch, and apple scab in their proposed paper. The dataset used in this paper consists of 391 apple fruit images. The dataset contains fresh apple and defected apple images where defected apple images are further classified into three different classes: Apple rot, Apple blotch, and Apple scab images. Dubey et al. followed three different steps in their research. At first, by using the K-means clustering method, they performed defect segmentation. After that, they used LBP, CLBP, CCV, and a histogram for extracting characteristics from the image datasets. Finally, a multi-class SVM is used to train and classify the apple images into different disease categories. Up to 93% classification accuracy is attained using the suggested approach.

Mia et al. in their research work [9], proposed an approach to identify cucumber disease using ML algorithms and transfer learning models. The dataset used in this research work consists of 525 cucumber images with six different classes of cucumber diseases and one class of healthy cucumber images. As the dataset is too small, the researchers performed various data augmentation techniques like shifting, rotating, scaling, cropping, etc. After augmentation, the dataset consists of 4200 images. They split the data into an 80:20 ratio for training and testing. The training dataset is further partitioned for training and validation. After that, they separately implement ML algorithms and transfer learning models to detect cucumber diseases. Among different ML algorithms and transfer learning models, Random forest and MobileNet achieved the highest accuracy of 89.93% and 93.23% respectively.

In a research paper [10], different deep learning models like R-CNN, AlexNet, ResNet, VGG-16, and various transfer learning models are used to grade the quality of mango fruit. The dataset used in this paper is collected from the AICUP2020 dataset [1], which contains 6,400 mango images of different qualities. Before training the dataset, they performed various steps like data prepossessing, background removal, and data augmentation. The input images are resized to 224x224 pixels. To increase the amount of data, they performed rotation, flip, zoom in/out, and brightness control in the dataset. The authors used various CNN models like AlexNet, VGG-16, and ResNet to train and test the dataset. From their research, it is found that VGG-16 performs the best while grading the quality of Mangoes with an accuracy score of around 83.6%.

Kumar et al. in their research article [11], proposed a novel model based on a deep learning technique to detect and classify fresh and rotten fruits. In their research, they have used a dataset from Kaggle which contains 8,400 images of apples, bananas, and oranges. The dataset is labeled into 6 classes where 3 classes are for fresh fruits and the remaining 3 classes are for rotten fruits. For training purposes, 90% of the dataset is used and 10% is used for testing their model. Before training the model, they performed some preprocessing steps like resizing and labeling. After that, they applied the CNN model to extract features from the dataset and then performed classification. Their model achieved around 97.14% accuracy is classifying fresh and damaged fruits from the image dataset.

An article [12] by Awate et al. introduced an approach to fruit disease detection using color and texture analysis. In the paper, they applied the ANN model to identify and classify diseases. The authors used two different image databases in their paper to train their models. After that, they applied K-means clustering for image augmentation, and four feature vectors: morphology, color, the structure of the hole, and the texture of the fruits are used in feature extraction. Finally, they used the ANN model to match the patterns and classify diseases.

III. DATASET

For our research, at first, we collected two different fruit image datasets from Kaggle [13],[14]. After that, we combined the two datasets and prepared a new dataset to increase the amount of data. The new dataset contains 18,096 fruit and vegetable images of apples, bananas, oranges, tomatoes, bitter gourd and capsicums which were captured using a mobile phone camera. The images of the dataset are grouped into 2 different categories: fresh and rotten. Like fresh-apple, fresh-banana, fresh-orange, fresh-tomato, fresh-capsicum, fresh-bitter gourd, rotten-apple, rotten-banana, rotten-orange, rotten-capsicum, rotten-bitter gourd and rotten tomato. All the images are in .png format. Among 18,096 image datasets, 14,802 images are used to train our model and 3,294 images are used to test our model. The number of images of all categories of fruits and vegetables in training dataset is shown in Table I. All the images of the dataset are 512 x 512 pixels in size.

In below Figure 1 and Figure 2 we have shown some images from our dataset.

[1]https://www.kaggle.com/datasets/dowbatw/aicup2020

Fig. 1. Sample of Dataset (a) Fresh Apple (b) Rotten Apple

Fig. 2. Sample of Dataset (a) Fresh Orange (b) Rotten Orange

TABLE I
TRAINING SET IMAGE COUNT

Fruit Name	Number of Images
Apple	4035
Banana	3805
Orange	3061
Tomato	1731
Capsicum	1620
Bitter Gourd	550

IV. METHODOLOGY

We have used transfer learning models like MobileNetV2 in our research to detect fruit freshness. Before training and testing our model, we performed some pre-processing steps in our dataset. The workflow diagram of our research is shown in below Figure 3.

A. *Data Pre-processing*

As we have combined two different datasets to increase the amount of data, we created a .csv file to label all the images of the dataset. In data pre-processing step, we resized the images to 224 x 224 pixels in size because deep learning models train faster on this size images, also the models can work more efficiently with lesser resolution. We also decoded the image to the 'uint8' tensor in pre-processing step.

B. *TensorFlow Hub*

TensorFlow hub is an open repository of pretrained reusable machine learning models. In this hub there a repository name 'TFHub.dev'[2] provides many pre-trained models like VGG-19, InceptionV3, BERT, ResNet50, etc. These pretrained models are used in different classification, segmentation , detection, prediction, creation, etc works. The hub is divided into different problem domains like text, image, video, audio. By downloading, installing and importing tensorflow_hub, we can easily use these models in our research works. TensorFlow hub helps us to achieve results more accurately as the models available in it are already pre-trained using a huge number of dataset. In our research work, we used the MobileNetV2 and InceptionV3 model of TensorFlow hub to detect and classify fruit freshness from image dataset.

[2]https://tfhub.dev/

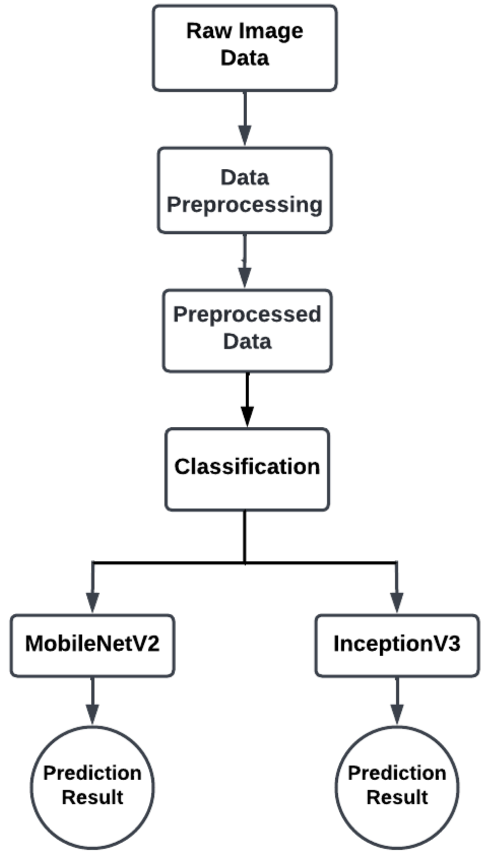

Fig. 3. Workflow Diagram

C. *MobileNetV2*

MobileNetV2 is a transfer learning model that works well with mobile devices. In our research work we have used MobileNetV2 to classify fresh and stale fruit images from the image dataset. We have used the default model with 12,024 trainable parameters and with 5,432,713 non-trainable parameters. The value of the input shape is (224, 224, 3). As mobileNetV2 is pretrained on the large ImageNet dataset, it can classify images more accurately than other models. Also, as it works well with mobile devices we have used this mobile in our research.

D. InceptionV3

A well known convolutional neural network, Inception v3 is used to help with object detection and image analysis. It is also a pretrained model on ImageNet dataset. In this research, we have used the default InceptionV3 model. As this model reduces training time, and it shows greater accuracy in ImageNet dataset, we have chosen this model for classifying fruit images.

E. Model Training Parameters

We utilized the same parameters for both models in our research. A batch size of 32 was used to train the models. Additionally, for training models, at first we trained our models over 100 epochs but due to Stop function there arises overfitting and then training stops. So, to prevent this we trained our models around 10 epochs. For training and testing purposes, our dataset is divided 70:20:10 portion. So, for training purpose, we have used 70% of our dataset, for validation we have used 20% and for testing purpose, we have used 10% of our dataset.

V. Result and Analysis

Both the pre-trained models scored quite well in classifying fresh and state fruits. From the below Table II, we can see that, MobileNetV2 scored around 98% accuracy in predicting fresh and stale fruits and vegetables. It achieved around 97% precision, 92% Recall and around 91% F1 score. On the other hand, InceptionV3 scored around 90% accuracy from the test data. We can say that, among the two proposed models, MobileNetV2 scored better than InceptionV3 with an accuracy score of 98%.

TABLE II
Accuracy score

Model	Accuracy	Precision	Recall	F1-score
MobileNetV2	0.98	0.97	0.92	0.91
InceptionV3	0.90	0.89	0.89	0.92

The accuracy comparison of previously approached models and our proposed model is given in the below Table III.

TABLE III
Comparison of Accuracy

Reference	Model	Classification Accuracy
[6]	CNN	97%
[8]	SVM	93%
[9]	RF	89.93%
[10]	VGG - 16	83.6%
Our Approach	**MobileNetV2**	**98%**

In the Epoch Vs Accuracy 4, the white line indicates the training data and the orange line indicates the validation data. From the graph we can see that when the number of epoch increases, the value of training accuracy increases. Similarly, then the value of epoch increases, the validation accuracy also increases.

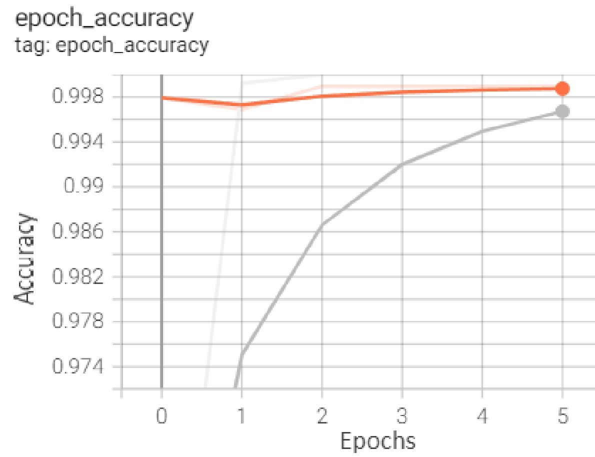

Fig. 4. Epoch vs Accuracy for training and validation(MobileNetV2)

In case of Epoch Vs Loss graph 5, we can see that when the value of epoch increases, training loss decreases. On the other hand, if the epoch number increases, at first the value of validation loss decreases and after sometime it remains the same.

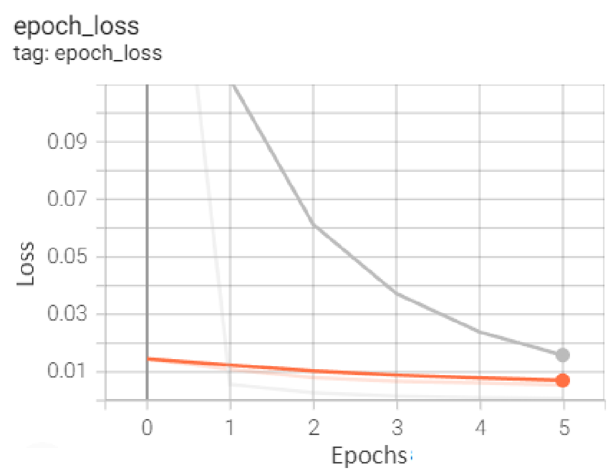

Fig. 5. Epoch vs Loss for training and validation(MobileNetV2)

From the confusion matrix of Figure 6, we can see the performance of our classification task.

VI. Conclusion

Food waste, food-borne illnesses, and economic loss due to wastage of fruits and vegetables could be greatly reduced by automated vision-based systems which can detect freshness of

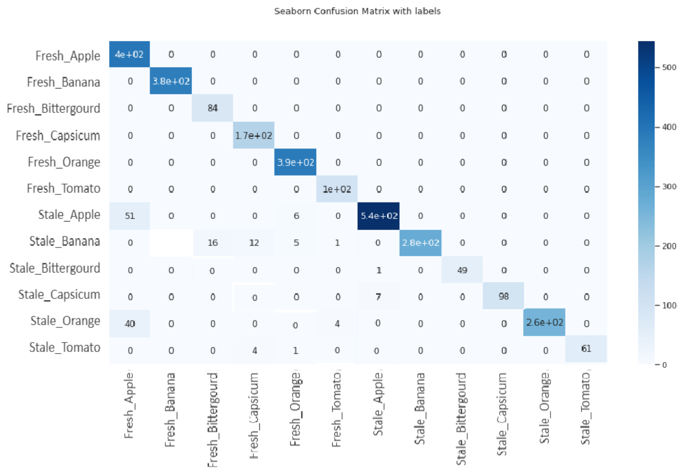

Fig. 6. Confusion Matrix

fruits and vegetables. In our research work we have implemented an advanced model to detect and classify freshness of different fruits and vegetables. For accurately detecting and classifying the freshness of different fruits and vegetables, we have used two pre-trained models like MobileNetV2 and InceptionV3 in our research. Based on the color and texture of the images, the classification task has been performed. Among the two proposed models, MobileNetV2 scored the highest. It achieved around 98% accuracy to detect freshness of fruits and vegetables. As we have used MobileNetV2, in the future we will test our model using real-time images. Moreover, in order to improve the effectiveness of our system, we will increase our dataset by collecting additional fruit and vegetable images, such as mango, papaya, guava, cucumber, carrot, etc.

References

[1] *Overview*. [Online]. Available: https://www.worldbank.org/en/topic/agriculture/overview.

[2] Admin, *Top 10 agricultural producing countries in the world*, 2021. [Online]. Available: https://www.tractorjunction.com/blog/top-10-agricultural-producing-countries-in-the-world/.

[3] *Bangladesh gdp annual growth rate2022 data - 2023 forecast - 1994-2021 historical*. [Online]. Available: https://tradingeconomics.com/bangladesh/gdp-growth-annual.

[4] *Produced but never eaten: A visual guide to food waste*. [Online]. Available: https://www.theguardian.com/environment/ng-interactive/2015/aug/12/produced-but-never-eaten-a-visual-guide-to-food-waste.

[5] *Fresh fruits - worldwide: Statista market forecast*. [Online]. Available: https://www.statista.com/outlook/cmo/food/fruits-nuts/fresh-fruits/worldwide.

[6] S. Malathy, R. Karthiga, K. Swetha, and G. Preethi, “Disease detection in fruits using image processing,” in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, IEEE, 2021, pp. 747–752.

[7] L. M. Azizah, S. F. Umayah, S. Riyadi, C. Damarjati, and N. A. Utama, “Deep learning implementation using convolutional neural network in mangosteen surface defect detection,” in *2017 7th IEEE international conference on control system, computing and engineering (ICCSCE)*, IEEE, 2017, pp. 242–246.

[8] S. R. Dubey and A. S. Jalal, “Adapted approach for fruit disease identification using images,” in *Image processing: Concepts, methodologies, tools, and applications*, IGI Global, 2013, pp. 1395–1409.

[9] M. J. Mia, S. K. Maria, S. S. Taki, and A. A. Biswas, “Cucumber disease recognition using machine learning and transfer learning,” *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3432–3443, 2021.

[10] S.-L. Wu, H.-Y. Tung, and Y.-L. Hsu, “Deep learning for automatic quality grading of mangoes: Methods and insights,” in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2020, pp. 446–453.

[11] T. B. Kumar, D. Prashar, G. Vaidya, V. Kumar, S. Kumar, and 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, 2022.

[12] A. Awate, D. Deshmankar, G. Amrutkar, U. Bagul, and S. Sonavane, “Fruit disease detection using color, texture analysis and ann,” in *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, IEEE, 2015, pp. 970–975.

[13] R. R. Potdar, *Fresh and stale images of fruits and vegetables*, 2021. [Online]. Available: https://www.kaggle.com/datasets/raghavrpotdar/fresh-and-stale-images-of-fruits-and-vegetables.

[14] S. R. Kalluri, *Fruits fresh and rotten for classification*, 2018. [Online]. Available: https://www.kaggle.com/datasets/sriramr/fruits-fresh-and-rotten-for-classification/metadata.