Intelligent Mango Classification and Pricing System Using Optimized CNN Architectures

Wardah Haya

Department of Artificial Intelligence, Faculty of Computer Science & Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan.

[u2022622@giki.edu.pk](mailto:u2022622@giki.edu.pk)

*Abstract*— The way the mango varieties are classified and priced in Pakistani markets is prone to distortion, leading to distorted trade and loss of purchasing power. It describes a novel machine learning method by using the EfficientBNet2, a convolutional neural network, which optimizes the classification of mangoes and predicts their market price in real time. Based on a dataset of local mango photos and government-regulated price data, the proposed method matched up with more than 96% classification accuracy and was able to correctly recognize mango varieties. For accessibility and user-friendliness, an easy-to-use Flutter mobile app was created that used the predictive properties of the EfficientBNet2 model. It offers users access to real-time prices and product specifications which enables them to make the most of their purchases. Through the mango mislabeling study, this research can help improve market transparency and stability and reveals how deep learning and mobile technologies can revolutionize solving long-standing market problems.

*Index Terms* — Deep Learning, EfficientBNet2, TensorFlow, Mango Classification, Price prediction, Flutter, Consumer transparency.

# INTRODUCTION

One of the most popular fruits in Pakistani agriculture and cuisine is mangoes. But the local mango market is continually plagued by label misinformation and pricing manipulation. High-end products like Anwar Ratol and Chaunsa are substituted with lower-cost ones like Fajri or Sindhri, depriving customers of profits and undermining consumer confidence. Prosellers often take advantage of consumer ignorance and exploit the price inequalities and market inefficiencies. This problem doesn’t only concern local markets, as cheaper mangoes are mislabelled and exported as high-quality varieties, leading to transnational scams.

This paper attempts to overcome these issues by using powerful machine learning. Deep learning, in particular convolutional neural networks (CNNs), have become famous for their good performance on image classification, object recognition and segmentation. AI applications have been successful in many fields, ranging from agriculture to medical services, linguistics to economics. Using these improvements, in this work we used a CNN model named EfficientNet-B2 that was trained on the ImageNet dataset and optimised for a locally sampled dataset of Pakistani mango varieties. This helps in classifying mangoes accurately and lays the groundwork for combating market abuse.

A unique component of this study is combining the classification model with real-time government-provided pricing information. The system is embedded in a full-stack mobile app with the Flutter framework, and users can easily identify mango types and check current market prices. Photo-scanning a mango allows users to instantly see its variety and determine its state-regulated price per kilogram. This functionality gives consumers the tools to make the right buying choices, and prevents them from being overcharged because of label confusion or swindling by vendors.

Previous research on mango classification has been largely focused on feature data, or assessing mango properties by using CSV files. These techniques work well for researchers, but not for non-technical users. In addition, the earlier papers focused only on foreign mango varieties, which omitted the study of local species in Pakistan. While tabular data models are relatively easier to use, they do not provide easy-to-use solutions for the average user. This work fills this void by using computer vision to train models for local mangoes to ensure mango classification accuracy and accessibility to the general population.

Not only does this study fill an important consumer protection gap, it also outlines a new way of using AI to increase market transparency. The effort will enhance consumer confidence, prevent fraud, and encourage the fair trade of mangoes by offering an easy-to-use mobile app that integrates seamless mango category and real-time price data.

# METHODOLOGY

This paper used a labeled dataset consisting of more than 2,000 images of Pakistani mangoes spread across eight classes (Angar Ratol, Chaunsa, Sindhri). Images were taken between 20 days after the hardening phase and one week before the harvest phase, which meant that the dataset contained a range of visual traits. We partitioned the data into 3 subsets, based on the Sklearn library: 80% training, 10% validation, and 10% testing. They rigorously preprocessed the data to maintain diversity and quality of training data, which accounted for the excellent classification performance of the model.

At the heart of this study was the design of a deep learning model with the EfficientNet-B2 framework implemented using the Keras library. Pre-trained on the ImageNet dataset, efficientNet-B2 was tailored for the Pakistani mango dataset in order to give it a higher generality. Training involved optimisation with two algorithms, Adam and Adamax. The hyperparameters (initial learning rate 0.001, batch size 30), and momentum value 0.99 were carefully set up to make the model run optimally. L1 and L2 regularizations also applied to activity, kernel, and bias components (0.006 and 0.016, respectively) in order to avoid overfitting and for robust training. The training was 14 epochs and the target accuracy was 0.9, which was good enough to keep the model running.

We further refined the EfficientNet-B2 code to make it faster on the new dataset. We created a batch normalization layer to help smooth out and accelerate the training, and two dense layers to train complex relationships on the dataset. We introduced a dropout layer to reduce overfitting by randomly disabling neurons during training. In layers intermediate and final we used the rectified linear unit (ReLU) for activation functions and in layer final we used the softmax function to compute probability distributions for all the eight mango classes. Continual data switching between epochs further enhanced generalization and stability of the model to a validity accuracy of 98%

This final model was implemented as a full-stack mobile application built on top of Flutter, with Python Flask as the backend. The app lets you snap a photo of a mango, assign it to one of the eight pre-set categories, and get the price on that day. This fusion of deep learning with user-friendly interface bridges the distance between technology and consumer use, allowing for accurate mango classification and giving consumers power to make informed buying decisions.

With this careful approach, the research obtained a robust and trustworthy classification model that provided real-life solutions to the problem of mango mislabeling and market price irregularities in Pakistan.

# LITERATURE REVIEW

Mango classification has fascinated scientists who want to create automated tools to identify mangoes. Other research has been done using machine vision, artificial intelligence and image processing to overcome the problem of mango classification.

Thinh et al. (2019) [22] a machine vision and AI-based mango classification algorithm. This work used CNNs to classify the mango images with impressive accuracy, and demonstrated the power of deep learning in fruit classification. The paper pointed out that image datasets were essential for preparing powerful models for accurate classification.

Thong et al. (2019) [23] expanded the use of CNNs by combining image processing with artificial intelligence for mango classification. Their work also showed that image features were efficient in improving classification performance, making CNNs an extremely reliable source of agricultural tech knowledge.

In another important study, Nandi et al. (2014) [24] developed a machine vision-based method to automatically sort and grade mangoes. In contrast to the studies above, this study used a Gaussian Mixture Model to characterize mangoes by maturity and size. This work provided useful information about how machine learning algorithms are used in grading and quality control in the agriculture industry.

Although these experiments point to improvements in mango classification, these studies tend to use mostly foreign mango datasets and models trained for the task at hand, i.e. sorting, grading, or classification based on visual features. However, there’s a gap in the use of regionally appropriate datasets, particularly for Pakistani mango varieties. This disconnect highlights the need for better research that not only tackles classification but adds other features (price estimation) that are specific to the local market conditions.

This paper attempts to overcome this by building a mango classification algorithm trained on a Pakistani mango dataset. Using cutting edge deep learning techniques and a hands-on mobile app, this research not only improves classification accuracy, it also gives consumers the knowledge to make good buying choices.

|  |
| --- |
| Fig. 1. A sample of the dataset used alongwith sample images for each class of mangoes present in the dataset. |

# RESULTS

The model trained on a locally selected dataset of Pakistani mangoes was practically error-free and it has proven to be a powerful and accurate mango classification tool. The model calculates a softmax function to return the most likely class as classification output.

*Model Training and Loss Reduction*

In training the loss value gradually decreases as the number of training epochs increases, indicating that the model learnt all the required patterns and features to correctly classify. The loss went down from 8 to 4 during the training process as illustrated by the loss vs. epochs plot (Figure 3), which had a negative slope during training.

*Accuracy Trends and Overfitting Prevention*

Our accuracy vs epochs graph displayed a positive gradient, which was consistent with continuous model training gains. Yet accuracy gains plateaued after a given number of epochs, which meant that overfitting was at risk. In order to alleviate this, training was stopped as much as possible so that the model would remain simple and generalizable to invisible data.

*Evaluation Metrics and Model Robustness*

It used generic measurements like Accuracy, Precision, Recall, and F1 Score to gauge the performance of the model. Both the test and validation sets showed accuracy close to perfect which further demonstrated the validity and strength of the mango classification model. A confusion matrix (Figure 4) revealed very good classification scores for each of the eight existing mango types.

*Real-Time Implementation*

In order to make the model easier to use, it was connected to a real-time backend server that pulls daily mangoes price data from the Ministry of Food and Agriculture’s website. It will not only categorize mangoes but will provide you with real-time price details so you can make informed purchase decisions.

*Continuous Improvement and Augmentation*

The app features an incremental improvement methodology. Each time a user scans a mango image, it gets recorded on the server, providing you with opportunities to periodically retrain the model using both old and new data. Additionally, image augmentation can be applied to add more datasets when data is limited, thus maintaining long-term increases in model performance.

The successful implementation of this deep learning algorithm into a real-life app has enabled users to accurately identify mangoes and gain access to additional features like price fetching. The application, with the assistance of a powerful and general classifier, illustrates how advanced ML can be used to solve real-world problems. Figures 3 and 4 depict the training graphs and confusion matrix, and Figure 5 shows the sample predictions and model confidence. Table I summarizes the evaluation parameters.

# TABLE I METRICS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-score | Support |
| Anwar Ratol | 0.98 | 0.98 | 0.98 | 22 |
| Chaunsa (Black) | 0.97 | 0.97 | 0.97 | 16 |
| Chaunsa (Summer Bahisht) | 0.96 | 0.96 | 0.96 | 28 |
| Chaunsa (White) | 0.96 | 0.96 | 0.96 | 12 |
| Dosehri | 0.97 | 0.97 | 0.97 | 23 |
| Fajri | 0.97 | 0.97 | 0.97 | 19 |
| Langra | 0.97 | 0.97 | 0.97 | 27 |
| Sindhri | 0.96 | 0.96 | 0.96 | 13 |
| Accuracy |  | 0.97 | |  |
| Macro avg | 0.97 | 0.97 | 0.97 | 160 |
| Weighted avg | 0.97 | 0.97 | 0.97 | 160 |

# DISCUSSION

They have made positive progress on mango classification using a deep learning model. However, there are certain drawbacks that have to be overcome to further improve the performance of the model and to expand its scope. This is primarily due to the limited size of the dataset and the number of mango classes currently being implemented in the model. To overcome this, additional classes can be added by adding new folders to the dataset, where each class has as many images as the current one.

This balancing process will preserve the model's accuracy and robustness. Additionally, training with various batch sizes will let you fine-tune the model to find the best configuration to get the best accuracy. An extra layer between them might also allow the model to differentiate between closely similar mango classes to improve classification performance and usability.

When the dataset increases in size, the model size will grow automatically, making it difficult to deploy on low-resource hardware. The trained Keras model was then converted to TFLite, decreasing the size from under 50MB to around 30MB in order to alleviate this. Further optimizations made the model smaller to 9 MB, which is ideal for use on edge machines with minimal storage overhead. Furthermore, images are uploaded on a server instead of on a client which further reduces the storage requirements for clients. This dual strategy model reduction, and server-side image processing allows the system to run very efficiently on low-end hardware without compromising on performance.

A green apple with white squares

Description automatically generated

Fig. 2. A sample image from the test dataset displayed

While the model was able to cope well with single mango image, bulk scanning situations with multiple mangoes per image presented challenges. This restriction is due to the training architecture currently used for single mango images. The future effort to overcome this is to scale up the dataset to include more than one mangoed image. This will give the model more power for bulk scans and better accuracy in these situations.

|  |  |
| --- | --- |
| Hyper Parameters and Configurations | |
| Epochs | 12 |
| Learning rate | 0.001 |
| Mini batch size | 30 |
| Optimizer | Adam, Adamax |
| Momentum | 0.99 |
| Learning Factor | 0.5 |
| *L*1 Regularization | Activity, Bias, 0.06 |
| *L*2 Regularization | Kernel, 0.016 |
| Samples in training set | 1280 |
| Samples in validation set | 160 |
| Samples in test set | 160 |

# TABLE II

THE TABLE SHOWS THE CONFIGURATION OF HYPER PARAMTERS

USED IN THE STUDY.

Future plans involve to expand the app’s reach beyond Pakistan through datasets that feature foreign mango varieties. This will not only make the app more general but also allow for better classification performance on a wider range of mango species.

A graph of a graph of a graph

Description automatically generated with medium confidence

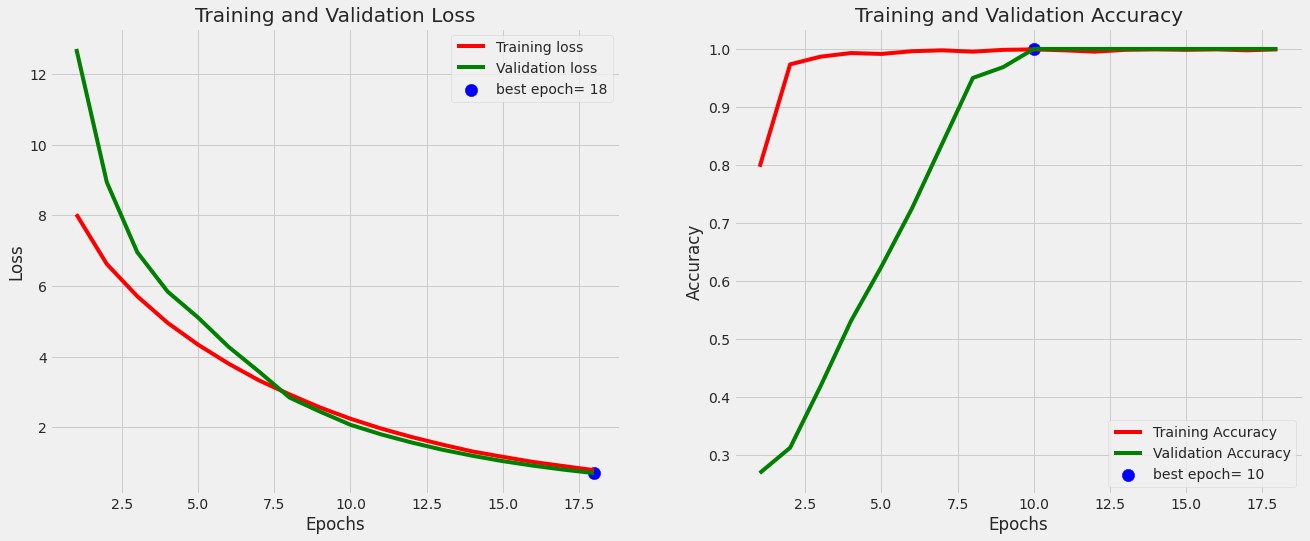


Fig. 3. Training and Validation Accuracy and Loss graphs

These enhancements will allow it to serve a larger user base, which makes the application easier to use and enables its worldwide use. Further, subsequent releases of the app will aim to provide accessibility by offering localized languages and text-to-speech or speech-to-text support. In order to fulfill these goals, data collection will be based on user submitted scans to allow for a collaborative growth of the dataset. This continuous data accumulation will not only improve the accuracy of the app but will also make it more useful and useful to a wide variety of users.

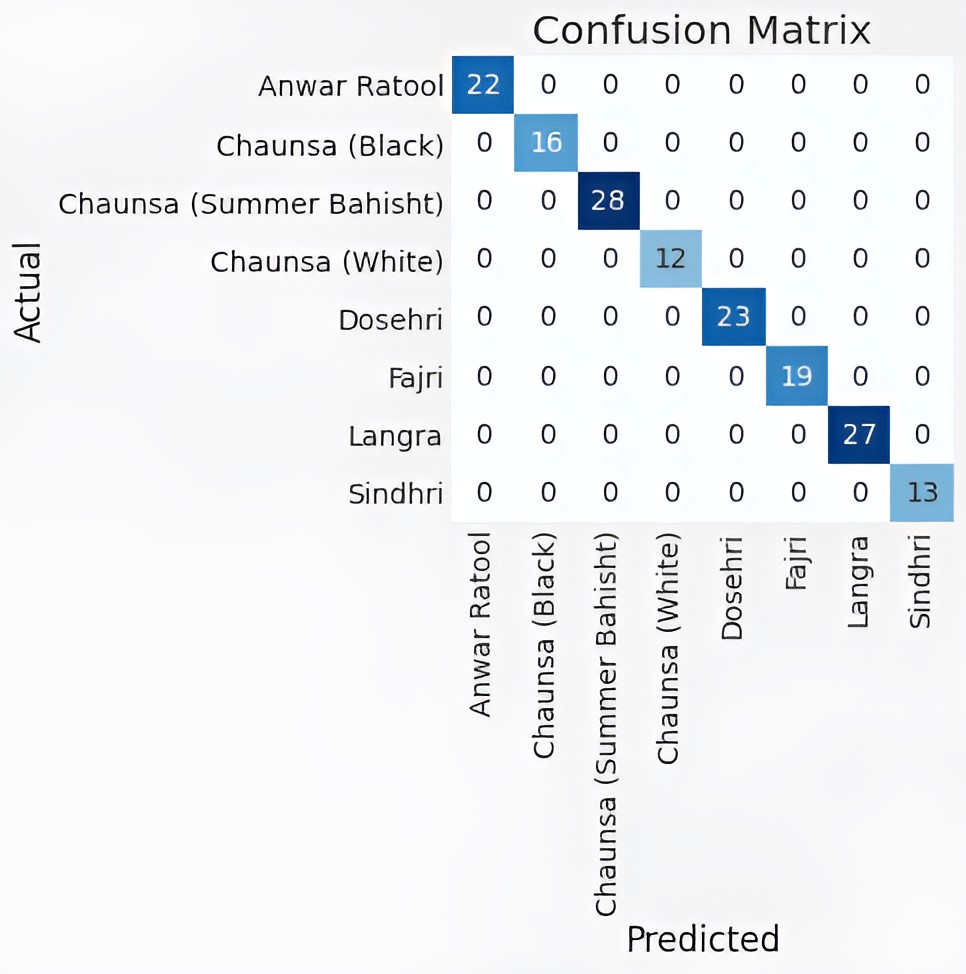


Fig. 4. The confusion matrix of all of the classes available.



Fig. 6. Results of our model.

In a nutshell, while the existing model has a strong performance and real-world usage, future enhancements to dataset size, model tuning, bulk scanning, and worldwide implementation will greatly increase its efficiency. These changes will make the app a multipurpose, trusted, and world-converging mango classification and price assistance tool.

# CONCLUSION

Ultimately, this study successfully cracks the mango mislabeling and exploitation problem with an accurate deep learning model. The model’s close to-perfect classification accuracy helps customers make informed purchases and ensures they get the correct mango type. Providing access to real-time market prices via an easy-to-use mobile app, the study helps consumers compare prices on the go and avoid paying more than necessary. The simple user interface of the app enables the app to be used by many users, and makes it accessible to the general public. Seeing the future plans to expand in more mango varieties and spread across the globe, this research represents an important first step towards giving more power to the consumers and more transparency in the mango market.

# REFERENCES

1. H. Javed, U. Arshad, Z. ul Abideen, and R. H. Ali, “A new rhythm in a1: Convolutional neural networks for music genre classification,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 222–227.
2. A. B. Siddique, M. A. Bakar, R. H. Ali, U. Arshad, N. Ali, Z. U. Abideen, T. A. Khan, A. Z. Ijaz, and M. Imad, “Studying the effects of feature selection approaches on machine learning techniques for mushroom classification problem,” in *2023 International Conference on IT and Industrial Technologies (ICIT)*. IEEE, 2023, pp. 1–6.
3. T. Ahmed, A. Maaz, D. Mahmood, Z. ul Abideen, U. Arshad, and R. H. Ali, “The yolov8 edge: Harnessing custom datasets for superior realtime detection,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 38–43.
4. I. Mueed, U. Arshad, and R. H. Ali, “Revolutionizing campus exploration with gikilens: A deep learning-powered object detection app,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 315–320.
5. U. Arshad, H. H. Ellahie, Z. ul Abideen, and R. H. Ali, “An enhanced genetic algorithm framework for efficient solutions to capacitated vehicle routing problems,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 141–146.
6. H. Ashfaq, U. Arshad, R. H. Ali, Z. ul Abideen, M. H. Shah, T. A. Khan, A. Z. Ijaz, N. Ali, and A. B. Siddique, “Mitigating crop losses: Ai-enabled disease detection in tomato plants,” in *2023 International Conference on Frontiers of Information Technology (FIT)*. IEEE, 2023, pp. 190–195.
7. M. H. Shah, M. A. Bakar, R. H. Ali, Z. U. Abideen, U. Arshad, A. Z. Ijaz, N. Ali, M. Imad, and S. Nabi, “Investigating novel machine learning based intrusion detection models for nsl-kdd data sets,” in *2023 International Conference on IT and Industrial Technologies (ICIT)*. IEEE, 2023, pp. 1–6.
8. A. Haider, A. B. Siddique, R. H. Ali, M. Imad, A. Z. Ijaz, U. Arshad, N. Ali, M. Saleem, and N. Shahzadi, “Detecting cyberbullying using machine learning approaches,” in *2023 International Conference on IT and Industrial Technologies (ICIT)*. IEEE, 2023, pp. 1–6.
9. A. Shabbir, R. H. Ali, M. Z. Shabbir, Z. U. Abideen, T. A. Khan, A. Z. Ijaz, N. Ali, M. Imad, and M. A. Bakar, “Genetic algorithm-based feature selection for accurate breast cancer classification,” in *2023 International Conference on IT and Industrial Technologies (ICIT)*. IEEE, 2023, pp. 1–6.
10. M. Shehzadi, U. Arshad, Z. Abideen, R. H. Ali, A. A. Khan, and A. Z. Ijaz, “Identifying covid-19 through x-ray and ct scan images using machine learning,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 56–61.
11. A. Shabbir, R. H. Ali, M. Z. Shabbir, Z. U. Abideen, A. Z. Ijaz, N. Ali, M. H. Shah, S. Nabi, and K. Perveen, “Stock price forecasting using hidden markov models,” in *2023 International Conference on IT and Industrial Technologies (ICIT)*. IEEE, 2023, pp. 1–6.
12. M. Iftikhar, R. H. Ali, M. Saleem, U. Arshad, A. Z. Ijaz, N. Ali, M. Imad, M. A. Bakar, and A. Aftab, “Exploring the potential of hmms in linguistics for part of speech (pos) tagging,” in *2023 International Conference on IT and Industrial Technologies (ICIT)*. IEEE, 2023, pp. 1–6.
13. H. B. Khalid, A. B. Siddique, and R. H. Ali, “Custom hidden markov models for effective part-of-speech tagging,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 33–37.
14. A. Mashhood, Z. ul Abideen, U. Arshad, R. H. Ali, A. A. Khan, and B. Khan, “Innovative poverty estimation through machine learning approaches,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 154–158.
15. M. Iftikhar, R. H. Ali, M. Saleem, N. Shahzadi, U. Arshad, T. A. Khan, A. Z. Ijaz, N. Ali, and M. Imad, “Optimizing airline networks: A comparative analysis of graph-based techniques,” in *2023 International Conference on IT and Industrial Technologies (ICIT)*. IEEE, 2023, pp. 1–6.
16. O. K. Majeed, Z. ul Abideen, U. Arshad, R. H. Ali, A. Habib, and R. Mustafa, “Adaptivecloset: Reinforcement learning in personalized clothing recommendations,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 305–309.
17. M. H. Ishaq, R. Mustafa, U. Arshad, Z. ul Abideen, R. H. Ali, and A. Habib, “Deciphering faces: Enhancing emotion detection with machine learning techniques,” in *2023 18th International Conference on Emerging Technologies (ICET)*. IEEE, 2023, pp. 310–314.
18. D. Worasawate, P. Sakunasinha, and S. Chiangga, “Automatic classification of the ripeness stage of mango fruit using a machine learning approach,” *AgriEngineering*, vol. 4, no. 1, pp. 32–47, 2022.
19. P. Borianne, J. Sarron, F. Borne, and E. Faye, “Deep mango cultivars: cultivar detection by classification method with maximum misidentification rate estimation,” *Precision Agriculture*, vol. 24, no. 4, pp. 1619– 1637, 2023.
20. V. Gautam, R. K. Ranjan, P. Dahiya, and A. Kumar, “Esdnn: A novel ensembled stack deep neural network for mango leaf disease classification and detection,” *Multimedia Tools and Applications*, vol. 83, no. 4, pp. 10989–11015, 2024.
21. B. Jayanthi and L. S. Kumar, “Comparison of convolutional neural networks architectures for mango leaf classification,” *International Journal of Computational Vision and Robotics*, vol. 14, no. 1, pp. 84–98, 2024.
22. N. T. Thinh, N. D. Thong, H. T. Cong, and N. T. T. Phong, “Mango classification system based on machine vision and artificial intelligence,” in *2019 7th International Conference on Control, Mechatronics and Automation (ICCMA)*. IEEE, 2019, pp. 475–482.
23. N. D. Thong, N. T. Thinh, and H. T. Cong, “Mango classification system uses image processing technology and artificial intelligence,” in *2019 International Conference on System Science and Engineering (ICSSE)*. IEEE, 2019, pp. 45–52.
24. C. Nandi, B. Tudu, and C. Koley, “Machine vision based techniques for automatic mango fruit sorting and grading based on maturity level and size,” *Sensing technology: current status and future trends II*, pp. 27–46, 2014.
25. Directorate Food KPK, “Price list detail,” *Journal of Botanical Applications*, 2023.
26. M. Grinberg, “Flask web development with python tutorial,” [https://blog. miguelgrinberg.com/post/the-flask-mega-tutorial-part-i-hello-world,](https://blog.miguelgrinberg.com/post/the-flask-mega-tutorial-part-i-hello-world) 2018, accessed: May 11, 2023.
27. R. Iqbal, H. M. Hakim, and A. Hakim, “Mango variety and grading dataset,” *Mendeley Data*, vol. V1, 2021.