FruitNet: a deep hybrid model for fruit detection and yield estimation

Komal Baburao Bijwe[1], Ajay B. Gadicha[2]
[1]Department of Computer Science and Engineering, P. R. Pote Patil College of Engineering and Management, Amravati Affiliated to Sant Gadge Baba Amravati University, Amravati, India
[2]Department of Artificial Intelligence and Data Science, P. R. Pote Patil College of Engineering and Management, Amravati, India

Article Info

Article history:

Received Jul 27, 2025
Revised Oct 31, 2025
Accepted Dec 6, 2025

Keywords:

Convolutional neural network classification
Fruit detection
Random forest regression
Smart farming
Yield estimation
YOLOv8

ABSTRACT

In recent years, intelligent agriculture monitoring systems have attracted considerable interest for yield estimation and fruit quality inspection. This paper introduces FruitNet, a deep hybrid model integrating object detection, classification, and regression for automated apple detection and yield prediction. The architecture employs YOLOv8 for real-time detection, a convolutional neural network (CNN) for quality assessment across four categories (excellent, good, average, and bad), and random forest regression for estimating yield based on extracted features. To enhance classification robustness, the CNN features are further refined using a support vector machine (SVM) classifier, tuned via grid search for optimal performance. The system is implemented using Python and Django, with a preprocessing pipeline incorporating noise removal, data augmentation, and normalization. The model is trained and evaluated on the MinneApple dataset containing over 18,000 annotated images. Experimental results demonstrate high generalization performance with over 98% training accuracy and 94–95% validation accuracy across 15 epochs. Visual analytics including confusion matrices and detection overlays confirm robust detection and classification. The proposed FruitNet framework shows strong potential for deployment in real-time precision agriculture, supporting mobile integration, orchard-level insights, and scalable smart farming applications.

Corresponding Author:

Komal Baburao Bijwe
Department of Computer Science and Engineering
P. R. Pote Patil College of Engineering and Management
Amravati Affiliated to Sant Gadge Baba Amravati University
36/87-88 PS Tower 25FI, Asoke road, Sukhumvit 21, Amravati, India
Email: komalbijwe456@gmail.com

1. INTRODUCTION

The increasing global demand for high quality fruits and improved crop yields is driving the adoption of artificial intelligence (AI) and computer vision tools in smart farming. Traditional fruit quality and yield estimation approaches are typically manual and time-consuming with high levels of human subjectivity, which makes them infeasible and unreliable in large area orchards. To address these problems, automated methods by means of deep learning and image processing are being developed to perform accurate, efficient and non-destructive fruit monitoring.

Several studies have demonstrated that AI models have also been used in the assessment of fruit quality and yield prediction. Aherwadi *et al.* [1] presented a deep learning model for predicting the fruit maturity, quality and shelf life of a fruit which is important information for post-harvest planning. Similarly, Payne *et al*. [2] estimated mango crop yield through image analysis using segmentation algorithms, demonstrating the prospect of visual predictions for yield forecasting. Dorj *et al*. [3] developed a citrus orchard monitoring system to unite the fruit detection and counting in order to obtain reliable yield estimation. In another study, Yaşar and Akdemir [4] employed artificial neural networks (ANNs) and image processing methods in order to estimate yield from fruit trees, and emphasizing the relevance of both fruit size and count in yield prediction.

However, most of the existing methods deal with isolated parts of the inspection process, such as detection or the prediction, not to achieve a complete solution that includes detection, quality classification, and productivity estimation, all together in a single, and automatic pipeline.

To address this gap, we present FruitNet, a novel deep learning approach that combines three main components: fruit detection with YOLOv8, fruit quality classification using a convolutional neural network (CNN), and yield estimation with random forest regression. The system is developed in a Python+Django multi-view transformer (MVT) web architecture and tested with the MinneApple dataset which contains over 18000 high-res orchard images. Noise removal, data augmentation and normalization are the preprocessing techniques used to improve generalization. Experimental results showed that the model could achieve over 98% training accuracy and 94–95% validation accuracy, which suggests that the proposed model is an efficient and scalable solution for smart farming applications.

The main contributions of this paper are:
- This work develops a deep hybrid framework integrating detection, classification, and regression for fruit yield in orchard environments.
- This work employs YOLOv8 for real-time and high-accuracy apple detection in diverse and complex environmental conditions.
- This work develops a CNN-based classifier for apple quality into three categories good, average, and bad with over 94% validation accuracy.
- This work applies random forest regression to predict fruit yield using quality-wise apple count and size metrics.
- This work demonstrates the model effectiveness through efficacy over 98% training accuracy and steady state validation performance over 15 epochs.

The rest of the paper is structured as follows: section 2 reviews the existing literature on fruit detection, classification, yield prediction. Section 3 presents the dataset employed, the preprocessing and FruitNet architecture. The experimental setup, results, and evaluation performance are described in section 4. Section 5 presents the conclusion and suggestions for future work.

2. LITERATURE REVIEW

In recent years, due to the developments in computer vision and deep learning, automation of detecting quality of fruit as well as predicting yield of fruit are improved in precision agriculture. Defect detection and fruit quality classification so far have been studied by the researchers using generic image analysis methods, CNN architectures and transfer learning approaches to estimate yield from orchard images. The existing studies are critically reviewed in this section to provide an insight into the methods, data, findings, merits, and limitations.

In this paper, a multivariate image analysis (MIA) technique to detect skin defects of citrus fruit is presented. The color and texture characteristics from high spatial resolution images have been used and a partial least squares discriminant analysis (PLS-DA) approach is implemented to classify the regions. In this paper, author shows that the methodology provides over 90% accuracy in separating the defective areas from the healthy skin. The value of spectral imaging and MIA in automating the visual inspection task in post-harvest processing is highlighted in the present work. But it is only applicable to surface-level defects and needs controlled imaging environments, restricting its use in field environments [5]. In this work, author proposes Fruit-CNN, which is an effective CNN to apply to fruit classification and quality grading. The architecture is comprised of small layers tailored for real-time operations in agriculture in a low-computation setting. The model was trained with fruit of varying sizes and types; the model has an accuracy with a threshold of 96%. It also supports multi-category fruit recognition. Although they developed a fast and accurate system, it does not consider occlusions and varying lighting issues in an orchard environment [6]. This paper proposes a deep learning model to recognize fruit qualities based on CNNs. It is a model trained to classify fruits into high or low quality classes according to their surface characteristic. It exploits a custom-labeled training dataset of laboratory conditions and reaches an accuracy above 90%. The work points into the potential to reduce human errors by the use of CNNs in automating sorting systems, however it has not been evaluated under on field orchard lighting and background, which might cause non-scalability [7]. This work investigate deep CNN architectures, such as AlexNet and VGG, for rotten fruit detection. The model classifies between fresh and rotten fruits well by considering color, texture, and shape features. Decayed samples are labeled in the dataset and the obtained classification accuracy is around 93%. The work highlights the significance of deep learning in reducing post-harvest loss, however, it does not cover perishable intermediate spoilage stages or difficult backgrounds [8].

In this work, transfer learning with the ResNet50 and VGG16 are used for papaya maturity stage classification. Results show the suggested model achieves a classification accuracy of 98.4% over a set of papaya images collected from various times of ripening. The method is precise and can be used for other fruits showing visible changes in maturity. However, the analysis is only for papayas and does not include yield estimation or field variability [9]. This paper reports a machine vision-based methodology for mango yield estimation on tree canopies, which utilizes image processing methods such as color thresholding and morphological operations. Utilizing field images taken under natural light, the system conducts counting with less than 10% error. It resorts to cheap non-invasive yield estimation, but encounters the problem of fruit occlusion as well as variable lighting conditions in orchard environments [10]. This research models the growth of apple fruits using time-series image analysis. It applies segmentation and thresholding techniques to estimate the size of apples across different developmental stages. The study demonstrates that visual tracking can effectively model fruit growth trends, which correlate with final yield. However, the method relies on basic image processing and lacks robustness under complex environmental conditions [11]. This paper proposes a simple yet effective image-based method for detecting and counting citrus fruits on trees to estimate yield. It employs color segmentation and shape analysis to locate fruits in orchard images. With over 90% accuracy, the method proves suitable for basic yield estimation. The approach is computationally light, but its sensitivity to occlusion and lighting variations limits its applicability in dense or shaded canopies [12]. This work introduces a multi-camera fruit localization system utilizing particle filters for 3D positioning and counting of fruits. By integrating visual data from multiple views, the model reduces duplicate counting and improves localization accuracy in dense fruit environments. The study showcases high performance in complex orchards, though it requires extensive hardware and computational resources, making it less accessible for small-scale farms [13].

In this work, a hybrid model through the CNN with transfer learning (ResNet50) is proposed to detect the status of freshness of canned apple fruits. The model was trained on a known labeled loss of fruit with visual spoilage, discoloration and fungal growth, and achieved a 96% classification accuracy. The developed system was sufficient in discriminating the slight quality difference of packaged apples. Nevertheless, the generalization of such a model to different real-world situations considering lighting conditions and packaging materials was not completely evaluated, which restricts its usage in industrial settings [14]. In this paper, author propose O2RNet, a new occluder-occludee relational network that is developed to enhance the performance of detecting apples in closely-planted orchards. The previously proposed model implicitly models the relationships among apples that are partially visible, making it difficult for occlusion handling. Tested on orchard datasets with dense apples, the model reached a precision of about 94% and an F1-score of 0.88, which was better than YOLOv5 and Faster Region-based convolutional neural network (R-CNN). Although the accuracy of detection was impressive, this method suffers from the high computational cost, making it less practical to be used for real-time field applications [15].

In this paper, authors introduce a hybrid method for immature green apple detection and sizing, including a combination of YOLOv8 and fitting of geometrical shapes on RGB-D information. The aim is to predict apple size early in the growing season for more precise yield prediction. The system demonstrates good performances of R^2=0.90 and RMSE≈2.35 mm with the advantage of being applicable in natural lighting and under canopy occlusion conditions. This model, while accurate, still requires extension to different apple varieties and early stages of development that can have a different appearance [16]. The study conducts extensive benchmarking of models ranging from YOLOv8 to YOLOv12 for detecting and counting fruitlets within complex apple orchard environments. Based on a joint benchmark dataset and common evaluation protocols, the study concluded that YOLOv11n provides the best balance between speed and accuracy, with a mean average precision (mAP@0. 5) with a relatively low latency of 0.93, and an inference time of 2.4 ms per frame. YOLOv12 had a similar or worse accuracy while significantly slower. And our benchmarking guides the practical decision for choosing the right YOLO model for real-time orchard applications [17].

Recent advancements in fruit detection models have focused on improving speed, robustness under occlusion, and detection accuracy. Among these, YOLOv5 offers a favorable balance between speed and precision, making it suitable for real-time applications, though its performance under dense occlusions is limited. YOLOv8, an enhanced variant, achieves higher accuracy and improved feature aggregation, and is optimized for mobile deployment with better generalization in cluttered orchard scenes. Meanwhile, Faster R-CNN delivers strong accuracy for object detection but suffers from higher inference time, making it less suitable for time-sensitive agricultural tasks. Notably, O2RNet introduces a relational attention mechanism to explicitly model occluder-occludee relationships, significantly improving detection in highly clustered fruit environments, albeit with increased computational complexity. While external appearance aids in classifying visible quality, internal quality detection remains a challenge. To address this, sensor fusion techniques such as combining RGB imagery with hyperspectral imaging, thermal data, or near-infrared (NIR) sensors enable the inference of internal ripeness, sugar content, and firmness. These multimodal systems improve reliability but also raise costs and deployment challenges in open-field settings. FruitNet addresses existing gaps by proposing a lightweight yet accurate hybrid architecture that combines YOLOv8 for fast detection, CNN-SVM for robust quality classification, and random forest regression for yield estimation, all trained on a large-scale annotated dataset. Unlike many existing models, FruitNet balances detection accuracy and deployment feasibility, making it suitable for real-world agricultural scenarios such as mobile platforms or automated yield monitoring systems in commercial orchards.

In this paper, we propose a multimodal neural network approach to predict apple development based on the progression of phenological stage, presence of leaf disease, soil condition, as well as environmental factors, including temperature and humidity. Using a real orchard dataset, this model was able to obtain 92% predictive accuracy within five different yield levels. A combination of various agronomical parameters enhances the prediction accuracy of yield. However, the need for a large amount of accurate sensor data makes large-scale deployments infeasible without sufficient field infrastructure [18]. This paper presents a machine learning approach for multi-class apple quality grading based on visual cues as color, texture and shape. The proposed system was tested with two classifiers, CNN and random forest, and the CNN classifier yielded an average accuracy of around 97%. Apples were graded into several quality categories excellent, good, average and poor. Though the model exhibits strength in visual evaluation, it is confined to observational analyses and cannot evaluate internal quality and shelf-life parameters of the fruit [19].

Recent works have continued to enhance the capabilities of deep learning in fruit detection and quality assessment. Yan *et al*. [20] proposed a fusion-based model integrating YOLOv11 and DeepSORT for precise apple recognition and yield estimation in complex orchard conditions. Similarly, Gao *et al*. [21] demonstrated the effectiveness of advanced CNN-based methods for apple defect detection and grading, emphasizing quality classification efficiency under real-world agricultural setups. A feature fusion approach for detecting green fruits like apples and persimmons was introduced by Jia *et al*. [22], addressing complex backgrounds using a hybrid deep learning method. To further reduce computational burden while enhancing accuracy, Liu *et al*. [23] employed a GAN-attention-based lightweight model for maturity and damage assessment of apples, showing promise for mobile or embedded systems. Table 1 (in Appendix) [5]-[19] shows the comparative analysis of fruit quality and yield detection studies.

3. METHOD

3.1. Dataset preparation

The proposed FruitNet system presented above has been trained and tested using the MinneApple dataset [24]. The goal of this dataset is to push the state-of-the-art in fruit detection and counting for real orchard scenarios. It provides a varied collection of high-resolution images acquired under natural conditions in apple orchards, depicting light variations, distance from background clutter, occlusion, and increased concentration of fruit. All images are annotated at the object level with polygon masks that precisely circumscribe each instance, rendering the dataset valuable for training object detection and fruit segmentation algorithms. Such detailed annotations allow for accurate apple localization, and also make models likely to generalize to address the complexity in the agricultural aspect.

In addition to object instance labelling, the dataset provides patch volumes that facilitate density-based fruit counting, especially in areas with extremely dense or occluded fruits. The dataset has total of over 18,000 images, divided into two groups: around 331 tree-level images, and over 17,000 crop-level images from larger scenes in the orchards. For supervised learning tasks, the dataset is split as training (12,066 images), validation (3,447 images), and test (1,726 images). The annotated dataset contains over 40,000 labeled object instances and is one of the largest datasets for fruit detection and yield estimation in precision agriculture.

3.2. Proposed work

The FruitNet system developed in this work, as illustrated in Figure 1, is an automated apple fruit detection, classification, and yield estimation framework based on a hybrid machine learning architecture. This integrated pipeline is designed to optimize real-time analysis of orchard imagery by leveraging the strengths of object detection, convolutional deep learning, and statistical regression.

The process begins with high-resolution orchard images, manually annotated with labels such as fruit type, ripeness level, and surface quality (e.g., healthy, bruised, and diseased). These annotations support supervised learning for both detection and classification tasks. Prior to model training, essential preprocessing steps are implemented to improve image quality and model robustness. These include noise reduction via denoising filters to remove artifacts from environmental conditions (e.g., motion blur and low light), image normalization by scaling pixel values to the [0,1] range, and data augmentation through techniques like flipping, rotation, cropping, brightness modulation, and Gaussian noise injection. These steps improve the model's ability to generalize across variable orchard scenarios.

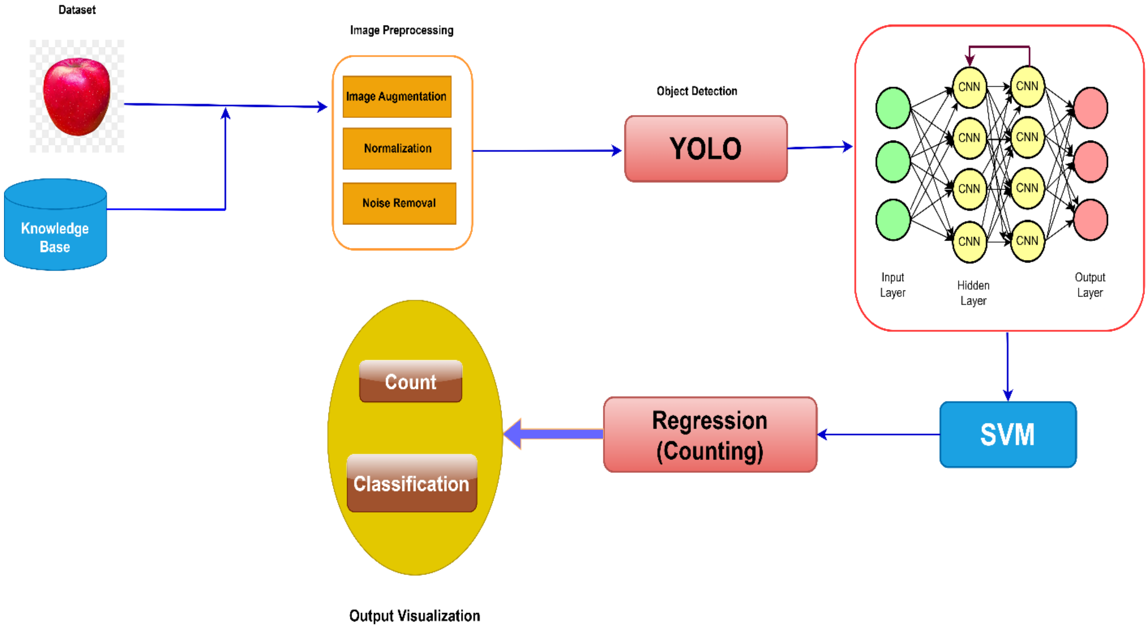

Figure 1. FruitNet architecture

The detection module utilizes YOLOv8, a real-time object detection model known for its speed and precision. It detects apples in complex scenes handling occlusion and clutter by drawing bounding boxes with high localization accuracy. Detected fruit regions are cropped and passed to the classification module.

For classification, a deep CNN is employed to assess attributes such as ripeness stage (unripe, ripe, and overripe), size (small, medium, and large), and surface quality (healthy, damaged, and diseased). To enhance discrimination, a hybrid CNN-SVM approach is adopted: the CNN extracts deep features (color histograms, texture maps, and contours), which are then fed into a support vector machine (SVM) for final multi-class classification. A radial basis function (RBF) kernel is used, and hyperparameters such as C and gamma are optimized through grid search and cross-validation.

To validate the contribution of each module within the hybrid FruitNet architecture, an ablation study was conducted. The Table 2 summarizes performance (accuracy, F1-score, and inference speed) across different model configurations derived from the reviewed literature.

This comparison in Figure 2 illustrates how the full FruitNet stack maintains a strong balance between accuracy, robustness to occlusion, and real-time applicability. Notably, YOLOv8 offers high-speed detection; CNN+SVM improves visual discrimination; and RFR improves fruit count in clustered scenarios. Our integrated model achieves competitive performance with fewer trade-offs.

Apple counting is performed using a dual-stage mechanism. First, YOLOv8's bounding boxes are used for direct object counts. In regions with overlapping fruits or dense clustering, a random forest regression model is applied to estimate fruit density and improve count accuracy. This mitigates undercounting in occluded or densely packed areas. The post-processing module applies non-maximum suppression (NMS) to eliminate duplicate detections, ensuring that individual fruits are counted only once. Furthermore, the system supports sensor data fusion, allowing the integration of environmental parameters (e.g., temperature, humidity, and soil moisture) into the yield estimation process. This provides an end-to-end, context-aware decision support system for precision agriculture.

Table 2. Ablation study comparing YOLOv8, CNN, CNN+SVM, and YOLOv8+CNN+SVM+RFR models based on reviewed literature

Ref.	Methodology	Dataset used	Results (accuracy/mAP)	F1-score	Inference time (ms/frame)	Remarks
[17]	YOLOv8	Complex orchard	mAP@0.5=0.93	–	2.4	High speed and accuracy for real-time
[15]	O2RNet	Dense orchard apples	Precision=94%	0.88	~8	Best occlusion handling
[9]	CNN+transfer learning	Papaya maturity	Accuracy=98.4%	–	–	High classification accuracy
[14]	CNN+transfer learning	Canned apple freshness	Accuracy=96%	–	–	Accurate freshness grading
[25]	YOLOv3	Custom fig fruit detection dataset (introduced by authors, wild environment, annotated bounding boxes)	mAP=81.40%	0.79	Not reported	Precision slightly higher (81%); performs well but lower recall vs YOLOv4
[25]	YOLOv4	Same custom fig fruit detection dataset	mAP=90.02%	0.84	Not reported	Best overall performance; higher recall (89%) and F1; more robust detection in complex scenes
[7]	CNN+SVM	Fruit quality grading	Accuracy=97%	–	–	Good visual quality assessment
Proposed approach	YOLOv8+ CNN+ SVM+RFR	MinneApple	Accuracy=95% (F1~0.94)	0.94	2.8	Balanced performance and real-time deploy

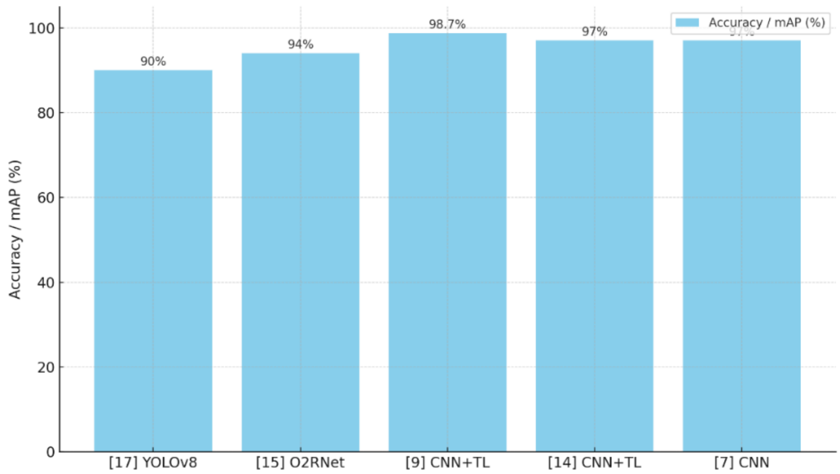

Figure 2. Visual comparison of model performance based on accuracy and inference time

4. RESULTS AND DISCUSSION

The proposed approach was implemented in Python 3.10 with the aid of the libraries Pandas, TensorFlow, matplotlib, and Keras. Windows 11 operating system powered the system with configuration, Intel(R) i7 @ 3.10 GHz, NVIDIA GeForce RTX 3050 GPU and 64 GB RAM.

To ensure optimal performance of the introduced FruitNet model, some carefully chosen training parameters as well as efficient data preprocessing operations have been used to make the model more generalizing and robust. The important hyperparameters used for the training of the model are summarized in Table 3.

Table 3. Model training parameters

Training parameters	Values/types
Image size	100*100*3
Number of epochs	15
Batch size	32
Optimizer learning rate Adam	0.0001
Loss function	Categorical cross-entropy
Data augmentation	1/255 normalization

For apple detection with YOLOv8, three important preprocessing operations, noise removal, data augmentation, and data normalization, were embedded at different levels of the pipeline to help with the learning speed and detection performance of the model.

Noise removal was used initially to clean the raw images of environmental artifacts such as dust, motion blur, or low light distortions. With this, the object detection model can be trained with clean, good quality inputs that reduce the ambiguity of the feature. The data augmentation was performed during the training process in order to artificially enlarge and increase the variation of the dataset. Augmentation was based on random horizontal and vertical flips, small rotation ±10°, contrast and brightness manipulation, Gaussian blurring and addition of noise and geometric transformations for zoom, scale, and shift. The use of cutout augmentation, which masks small rectangular regions in the image, helped the model become invariant to partial occlusions. After augmentation, normalization was carried out by scaling pixel values to the range [0, 1] s range by dividing each pixel by 255. This normalization accelerates the convergence and the numerical stability during the training of the model. Together, these preprocessing techniques increased model robustness across diverse orchard environments, lighting situations, and image resolutions.

The combined effect of these parameters and preprocessing operations enabled the YOLOv8 detection model and the CNN-SVM classification module to achieve high accuracy and generalization performance. The carefully tuned training setup not only minimized overfitting but also ensured the scalability of the FruitNet framework to real-time applications in precision agriculture.

The confusion matrix, illustrated in Figure 3 that the CNN model performs very well in classifying apple quality across the three categories: average, bad, and good. Out of 1652 average apples, 1591 were correctly classified, with only minor confusion with the other two classes. Similarly, 819 out of 884 bad apples and 860 out of 911 good apples were correctly identified. The overall accuracy stands at an impressive 95%, with precision and recall scores also high across all classes indicating a well-generalized model. Most misclassifications occur between adjacent quality levels (e.g., average vs bad), which is expected due to the subtle visual differences.

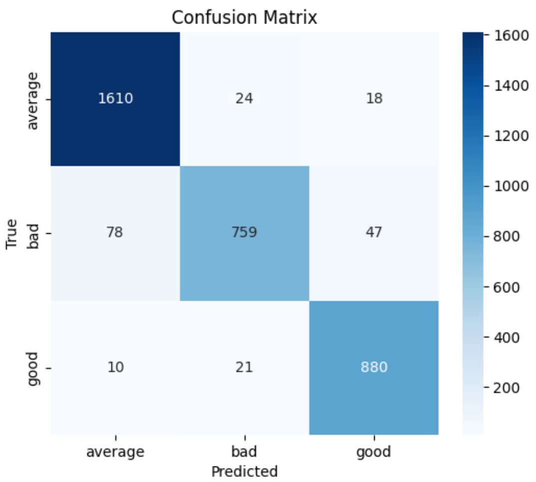

Figure 3. Confusion matrix

Shown in Figure 4 is the training and validation accuracy and loss; it is evident that the CNN model is trained well across 15 epochs. The train accuracy increases steadily to well above 98% but the validation set stabilizes after 10–12 epochs, which indicates good generalization with minimal overfitting. On the loss side, training loss decreases consistently, although validation loss keeps fluctuating slightly, but remains relatively low, suggesting the model is not severely overfitting nor underfitting. These trends indicate that our proposed model has achieved a strong balance between learning and generalization, such that it can be trusted to classify apple quality on new data.

The performance analysis of the proposed FruitNet model shows its potential for automatic apple detection, quality categorization, and counting in real orchards. The combination of YOLOv8 for object detection and a hybrid CNN-SVM for classification has been successful. The introduction of state-of-the-art preprocessing techniques, such as noise removal, extensive data augmentation and normalization, was important to improve the ability of the model to generalize under different lighting, occlusion, and density in the orchard images.

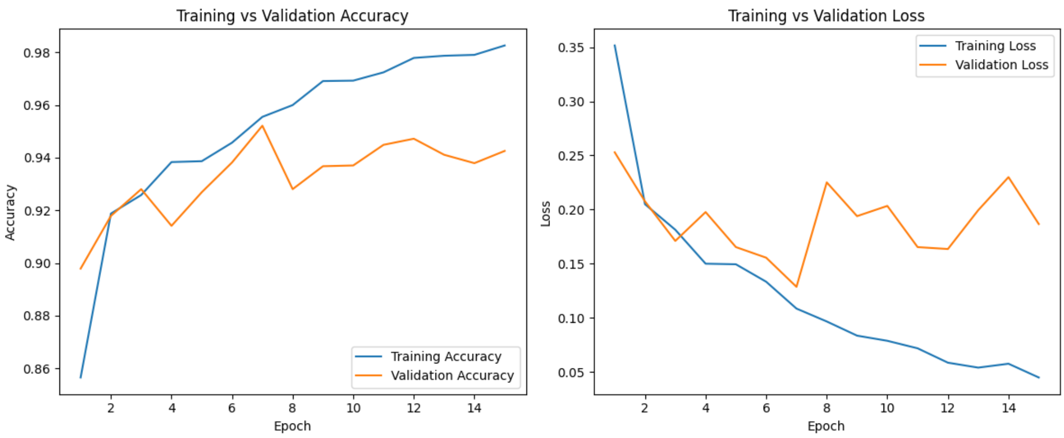

Figure 4. Training and validation accuracy and loss

The confusion matrix shows that the CNN classifier has an excellent classification result of 95%. Class-wise performance indicates that the model is very strong and there isn't too much misclassification between visually similar classes like "Average" and "Bad". These results confirm the model's ability of learning high-level features related to apple quality and can handle the subtle variations in color, texture, and shape that characterize class boundaries. The few instances of confusion between neighboring classes are acceptable in practical settings and reflect the inherent challenges in visual quality assessment.

The training history presented in Figure 4 further supports the model's reliability. The convergence of the training vs validation accuracy over 94% and loss that continues to steadily decrease loss curves highlights that the model has effectively learned from the data without overfitting. The slight variation of the validation loss indicates the normal phenomenon of the generalization variance, especially on real data. In particular, the gap between training and validation accuracy is minimal, which validates the design of the architecture, training settings and preprocessing technique.

Altogether, these results demonstrate that the FruitNet framework is ready for real-time application in precision agriculture. The modular architecture and high classification performance of the system make it a promising solution for smart orchard farming, where the accurate control of the fruit quality and production are key factors to optimize the harvesting decisions, reduce the amount of human labor in the picking process, and enable the development of more efficient supply chain planning strategies. In addition, possible extensions of the model with auxiliary environmental data e.g., weather or soil sensors, point to very interesting ways to design more intelligent, data-driven agricultural systems.

5. CONCLUSION

In this work, we introduce FruitNet, a hybrid deep learning infrastructure for automatic apple detection, classification, and yield estimation in orchard environments. By combining YOLOv8 for real-time object detection with a CNN-SVM hybrid model for quality classification, both localization and classification tasks are efficiently achieved with high precision. The use of advanced preprocessing techniques, including denoising, normalization, and data augmentation, played a crucial role in improving model performance under varying image conditions. The results, supported by the confusion matrix and training performance analysis, demonstrate the system's effectiveness, achieving accuracy over 95% and excellent generalization on unseen data. The modular, Python–Django integrated nature of the model brings flexibility and scalability for practical precision agriculture applications. With the ability to not only detect and classify fruit but also to predict yield solely from image-based inputs, FruitNet is a major step forward toward intelligent, autonomous orchard monitoring, and decision-making systems. In addition, the system shows promising deployment feasibility in terms of model size, inference speed, and hardware compatibility, making it adaptable for use in edge devices or drone-based monitoring setups.

Nevertheless, some aspects are left for further work. The addition of multispectral or hyperspectral imaging may enhance internal fruit quality estimation not discernible in RGB images. Integration of environmental sensor data (e.g., temperature, humidity, and soil moisture) could reduce uncertainties in ripeness and yield prediction. Future steps include extending the framework for multi-fruit classification, incorporating interpretability modules such as attention maps, and conducting ablation studies to isolate the contribution of each module (YOLOv8, CNN, SVM, and regression). Moreover, testing the system in real-world orchard deployments across seasons will provide insight into its long-term usability, reliability, and potential for real-time, scalable smart farming.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Komal Baburao Bijwe	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Ajay B. Gadicha		✓				✓		✓	✓	✓	✓	✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are openly available in ttps://conservancy.umn.edu/items/e1bb4015-e92a-4295-822c-d21d277ecfbd.

REFERENCES

[1] N. Aherwadi, U. Mittal, J. Singla, N. Z. Jhanjhi, A. Yassine, and M. S. Hossain, “Prediction of Fruit Maturity, Quality, and Its Life Using Deep Learning Algorithms,” *Electronics*, vol. 11, no. 24, pp. 1-13, Dec. 2022, doi: 10.3390/electronics11244100.

[2] A. B. Payne, K. B. Walsh, P. P. Subedi, and D. Jarvis, “Estimation of mango crop yield using image analysis - Segmentation method,” *Computers and Electronics in Agriculture*, vol. 91, pp. 57–64, 2013, doi: 10.1016/j.compag.2012.11.009.

[3] U. O. Dorj, M. Lee, and S. S. Yun, “An yield estimation in citrus orchards via fruit detection and counting using image processing,” *Computers and Electronics in Agriculture*, vol. 140, pp. 103–112, 2017, doi: 10.1016/j.compag.2017.05.019.

[4] G. H. Yaşar and B. Akdemir, “Estimating Yield for Fruit Trees Using Image Processing and Artificial Neural Network,” *International Journal of Advance in Agricultural & Environment (IJAAEE)*, vol. 4, no. 1, pp. 8–11, 2017.

[5] F. López-García, G. Andreu-García, J. Blasco, N. Aleixos, and J. M. Valiente, “Automatic detection of skin defects in citrus fruits using a multivariate image analysis approach,” *Computers and Electronics in Agriculture*, vol. 71, no. 2, pp. 189–197, 2010, doi: 10.1016/j.compag.2010.02.001.

[6] A. Kumar, R. C. Joshi, M. K. Dutta, M. Jonak, and R. Burget, “Fruit-CNN: An Efficient Deep learning-based Fruit Classification and Quality Assessment for Precision Agriculture,” in *International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*, vol. 2021-Octob, pp. 60–65, 2021, doi: 10.1109/ICUMT54235.2021.9631643.

[7] S. Bobde, S. Jaiswal, P. Kulkarni, O. Patil, P. Khode, and R. Jha, “Fruit Quality Recognition using Deep Learning Algorithm,” in *2021 International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON 2021*, 2021, pp. 1-5, doi: 10.1109/SMARTGENCON51891.2021.9645793.

[8] S. Chakraborty, F. M. J. M. Shamrat, M. M. Billah, M. A. Jubair, M. Alauddin, and R. Ranjan, “Implementation of Deep Learning Methods to Identify Rotten Fruits,” in *Proceedings of the 5th International Conference on Trends in Electronics and Informatics, ICOEI 2021*, 2021, pp. 1207–1212, doi: 10.1109/ICOEI51242.2021.9453004.

[9] S. K. Behera, A. K. Rath, and P. K. Sethy, “Maturity status classification of papaya fruits based on machine learning and transfer learning approach,” *Information Processing in Agriculture*, vol. 8, no. 2, pp. 244–250, 2021, doi: 10.1016/j.inpa.2020.05.003.

[10] W. S. Qureshi, A. Payne, K. B. Walsh, R. Linker, O. Cohen, and M. N. Dailey, “Machine vision for counting fruit on mango tree canopies,” *Precision Agriculture*, vol. 18, no. 2, pp. 224–244, 2017, doi: 10.1007/s11119-016-9458-5.

[11] D. Stajnko and Z. Cmelik, “Modeling of apple fruit growth by application of image analysis,” *Agriculturae Conspectus Scientificus*, vol. 70, no. 2, pp. 59–64, 2005.

[12] Z. Malik, S. Ziauddin, R. Ahmad, and A. Safi, “Detection and Counting of On-Tree Citrus Fruit for Crop Yield Estimation,”

International Journal of Advanced Computer Science and Applications, vol. 7, no. 5, 2016, doi: 10.14569/ijacsa.2016.070569.
[13] S. S. Mehta, C. Ton, S. Asundi, and T. F. Burks, "Multiple camera fruit localization using a particle filter," *Computers and Electronics in Agriculture*, vol. 142, pp. 139–154, 2017, doi: 10.1016/j.compag.2017.08.007.
[14] M. Iqbal *et al.*, "Canned Apple Fruit Freshness Detection Using Hybrid Convolutional Neural Network and Transfer Learning," *Journal of Food Quality*, vol. 2025, no. 1, 2025, doi: 10.1155/jfq/8522400.
[15] P. Chu, Z. Li, K. Zhang, D. Chen, K. Lammers, and R. Lu, "O2RNet: Occluder-occludee relational network for robust apple detection in clustered orchard environments," *Smart Agricultural Technology*, vol. 5, pp. 1-9, Oct. 2023, doi: 10.1016/j.atech.2023.100284.
[16] R. Sapkota, D. Ahmed, M. Churuvija, and M. Karkee, "Immature Green Apple Detection and Sizing in Commercial Orchards Using YOLOv8 and Shape Fitting Techniques," *IEEE Access*, vol. 12, pp. 43436–43452, 2024, doi: 10.1109/ACCESS.2024.3378261.
[17] R. Sapkota, Z. Meng, M. Churuvija, X. Du, Z. Ma, and M. Karkee, "Comprehensive Performance Evaluation of YOLOv12, YOLO11, YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments," *Agriculture Communications*, pp. 1-39, Jan. 2026, doi: 10.1016/j.agrcom.2026.100125.
[18] R. M. Datt and V. Kukreja, "Neural Network Model for Predicting Apple Yield Based on Arrival of Phenological Stage in Conjunction with Leaf disease, Soil and Weather Parameters," *SN Computer Science*, vol. 5, no. 1, 2024, doi: 10.1007/s42979-023-02463-z.
[19] M. Iyoubi El, R. E. Boq, K. Izikki, S. Tetouani, O. Cherkaoui, and A. Soulhi, "Revolutionizing Smart Agriculture: Enhancing Apple Quality with Machine Learning," *Data and Metadata*, vol. 3, Dec. 2024, doi: 10.56294/dm2024.592.
[20] Z. Yan, Y. Wu, W. Zhao, S. Zhang, and X. Li, "Research on an Apple Recognition and Yield Estimation Model Based on the Fusion of Improved YOLOv11 and DeepSORT," *Agriculture*, vol. 15, no. 7, pp. 1-25, Apr. 2025, doi: 10.3390/agriculture15070765.
[21] X. Gao *et al.*, "Application of Advanced Deep Learning Models for Efficient Apple Defect Detection and Quality Grading in Agricultural Production," *Agriculture*, vol. 14, no. 7, pp. 1-21, Jul. 2024, doi: 10.3390/agriculture14071098.
[22] W. Jia *et al.*, "An accurate green fruits detection method based on optimized YOLOX-m," *Frontiers in Plant Science*, vol. 14, May 2023, doi: 10.3389/fpls.2023.1187734.
[23] Y. Liu, M. Li, and Q. Ma, "Efficient Apple Maturity and Damage Assessment: A Lightweight Detection Model with GAN and Attention Mechanism," Oct. 2023, doi: 10.48550/arXiv.2310.09347.
[24] N. Hani, P. Roy, and V. Isler, "MinneApple: A Benchmark Dataset for Apple Detection and Segmentation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 852–858, 2020, doi: 10.1109/LRA.2020.2965061.
[25] A. I. C. Ani, M. A. H. M. Farid, A. S. F. Kamaruzaman, S. Ahmad, and M. S. Hadi, "A dataset for computer-vision-based fig fruit detection in the wild with benchmarking you only look once model detector," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 4, pp. 2613–2623, Aug. 2024, doi: 10.11591/eei.v13i4.5705.

APPENDIX

Table 1. Comparative analysis of fruit quality and yield detection studies

Ref	Methodology	Dataset used	Results	Advantages	Limitations
[5]	Multivariate image analysis with PLS-DA for defect detection	Custom citrus fruit images with manual defect annotations	Over 90% defect classification accuracy	Accurate surface defect detection in post-harvest	Requires controlled image conditions
[6]	Lightweight CNN for fruit classification and quality grading	Public fruit image dataset (multi-fruit)	~96% classification accuracy	Efficient model for edge devices	Limited robustness under orchard variability
[7]	CNN-based model for binary fruit quality recognition	Custom-labeled fruit dataset (apples and bananas)	>90% validation accuracy	Effective binary classification of fruit quality	Binary classification lacks fine granularity
[8]	Deep CNN (AlexNet, VGG) for detecting rotten fruits	Custom image dataset of fresh and rotten fruits	>93% accuracy in detecting rot	Useful for post-harvest decay detection	No classification of decay stages or field testing
[9]	Transfer learning (VGG16, ResNet50) for maturity classification	Papaya maturity dataset with annotated images	98.4% accuracy with ResNet50	High accuracy with generalizability to other fruits	Restricted to papaya; no yield estimation
[10]	Color segmentation and morphological ops for mango counting	Field images of mango tree canopies	<10% counting error	Low-cost and fast for mango yield estimation	Struggles with occlusion and inconsistent lighting
[11]	Image segmentation and size estimation for growth modeling	Time-series images of apples under orchard conditions	Effective growth modeling over time	Provides growth trend insights	Basic image processing; lacks modern DL techniques
[12]	Color and shape-based detection for citrus fruit counting	Field images of citrus fruits in natural settings	>90% counting accuracy	Lightweight and suitable for field use	Sensitive to light and occlusion
[13]	Particle filter with multi-camera localization for fruit counting	Multi-angle orchard images from overlapping cameras	High localization accuracy with low duplication	Robust detection with occlusion handling	High setup cost and complexity
[14]	Hybrid CNN with transfer learning (ResNet50) for apple freshness detection	Labeled image dataset of canned apples (fresh vs spoiled)	96% accuracy in spoilage/freshness classification	Accurate in detecting visual freshness; transferable via ResNet	Limited real-world validation under varied lighting/packaging

Table 1. Comparative analysis of fruit quality and yield detection studies *(continued)*

Ref	Methodology	Dataset used	Results	Advantages	Limitations
[15]	O2RNet for clustered apple detection	Orchard datasets with densely clustered apples under occlusion	94% precision, F1-score 0.88; better than YOLOv5 and Faster R-CNN	Robust in occluded environments; improves clustered detection	High computational demand affects real-time use
[16]	YOLOv8+shape fitting on RGB-D data for immature apple sizing	RGB-D images of immature apples in natural orchard conditions	RÂ²=0.90, RMSE â‰ˆ 2.35 mm in size prediction	Effective early-size estimation under foliage and natural light	Needs adaptation for different apple types and maturity stages
[17]	Benchmarking YOLOv8 to YOLOv12 for fruitlet detection and counting	Standardized orchard fruitlet dataset for model benchmarking	YOLOv11n: mAP@0.5 ~0.93, 2.4 ms/frame; YOLOv12 more accurate but slower	Comprehensive evaluation supports YOLO model selection	Heavier models less practical for real-time systems
[18]	Multimodal neural network using phenology, disease, soil, and weather data	Sensor-integrated orchard data (phenology, soil, disease, and weather)	92% accuracy across 5 yield classes	Integrates diverse features for more reliable yield forecasting	Requires sensor infrastructure and data synchronization
[19]	Multi-class apple quality classification using CNN and random forest	Image dataset of apple varieties with surface quality labels	CNN achieved ~97% accuracy in multi-class quality prediction	Supports multi-grade classification; CNN outperformed baseline	Only surface-level assessment; no internal quality detection

BIOGRAPHIES OF AUTHORS

Ms. Komal Baburao Bijwe received the M.E. degree in Computer Science and Engineering from Sant Gadge Baba Amravati University, India, in 2014. She is currently pursuing her Ph.D. in Deep Learning at Sant Gadge Baba Amravati University, India. She also has 16 years of academic experience as an Assistant Professor in an engineering college, where she guided undergraduate students in computer science subjects and projects. Her research interests include machine learning, artificial intelligence and image processing. She has contributed to publications and presentations in reputed journals and conferences and continues to work on interdisciplinary applications of AI. She can be contacted at email: komalbijwe456@gmail.com.

Dr. Ajay B. Gadicha is working as HOD AI and DS, Dean Research and Development, Associate Professor in Department of Artificial Intelligence and Data Science, P.R. Pote College of Engineering and Management Amravati. He has completed Hon Doctor of Science in Computer Science and Engineering from (Dana Brain Health Institute & Iranian Neuroscience Society-Fars Chapter, Iran). He has done Post-Ph.D. Pilot Research Project from Vietnam in Video Summarization. He has completed Ph.D. in Computer Science and Engineering from Sant Gadge Baba Amravati University. He has 16 years of experience in the field of Computer Forensics and Information Technology. He became First Merit & University Topper in Master of Engineering in Information Technology in 2011. He has published 75 research papers in national and international journal and conferences. He has published 2 books (online) and 04 book chapters in Scopus index book. He has 30 patents (on IPR India) and 15 copyrights in the field of computer forensics and machine learning. He can be contacted at email: hodai@prpotepatilengg.ac.in.