Fruit Quality and Defect Image Classification with Conditionalgan Data Augmentation

B.Gayathri[1], D.Bhavith Sai [2], A. Prudhvi Raj[3], Dr.A,Theja[4]
[1,2,3] *UG Scholars, Department of Electronics and communication Engineering, TKR College of Engineering and Technology,*
[4] *Associate Professor, Department of Electronics and Communication Engineering, TKR College of Engineering and Technology,*

***Abstract* -The quality and safety of fruits play a vital role in the agricultural industry, influencing both consumer satisfaction and market value. However, traditional image classification techniques face challenges such as limited and imbalanced datasets, as well as high variability in fruit appearance due to differences in size, shape, color, and surface defects. This project proposes an innovative fruit quality classification and defect detection system using Conditional Generative Adversarial Networks (GANs) for data augmentation. By generating realistic synthetic images representing various fruit conditions, the system enhances the training process of classification models, improving their accuracy and robustness. The integration of both real and augmented data helps the model effectively recognize subtle quality differences, making it more resilient to diverse real-world conditions. This automated approach supports large-scale fruit sorting, reduces reliance on manual inspection, and promotes more sustainable agricultural practices by minimizing food waste and ensuring consistent quality.**

Keywords: **Fruit Quality Classification, Defect Detection, Conditional GANs, Data Augmentation, Automated Sorting, Agricultural Technology, Synthetic Image Generation, Computer Vision, Deep Learning, Smart Farming**

I. INTRODUCTION

The quality and safety of fruits are essential factors in the agricultural industry, directly influencing both consumer satisfaction and the economic value of produce. Consumers expect fresh, visually appealing, and defect-free fruits, while producers and retailers rely on maintaining high-quality standards to secure market demand. Ensuring consistent fruit quality across large-scale operations is challenging due to natural variations in fruit appearance caused by differences in size, shape, color, and surface imperfections. These variations make manual inspection labor-intensive, error-prone, and inefficient, especially in high-volume agricultural supply chains.

Conventional image categorization techniques have been used to automate fruit quality inspection ,However, these methods frequently fail. Because of the limited availability of labeled data, imbalanced datasets, and the wide range of visual differences between healthy and defective fruits. Inconsistent lighting, environmental factors, and complex surface textures further hinder the performance of conventional classification algorithms. These limitations result in lower accuracy and reduced reliability, ultimately impacting both operational efficiency and product quality in the supply chain.

This study suggests using Conditional Generative Adversarial Networks (GANs) for data augmentation in order to overcome these difficulties., enabling the creation of realistic synthetic images that capture diverse fruit conditions, including various defects. By combining real and synthetic data, the classification model can be trained on a more comprehensive and balanced dataset, allowing it to better distinguish between healthy and defective fruits. This approach enhances the accuracy, robustness, and generalizability of the classification system, even under variable real-world conditions.

II. METHODOLOGY

A. Existing Methodologies

In the modern fruit processing industry, manual inspection and simple image analysis techniques are the mainstays of traditional methods for identifying faulty fruits. These techniques frequently classify fruit quality using basic image processing algorithms, which can be laborious and inaccurate. Standard machine learning models may be used by some current systems, but they usually have trouble with fruit appearance variations and need a lot of labeled data to train.

B. Proposed Methodologies

The proposed system aims to make fruit quality checking smarter, faster, and more reliable by addressing a key problem in current AI training: data imbalance. Farmers and companies often have plenty of images of perfect fruits but very few of damaged ones. This lack of variety makes it difficult for AI models to accurately identify subtle defects like bruises or mold, especially in real-world settings where speed and precision are critical.

To solve this, the system uses Conditional GANs (Generative Adversarial Networks) to generate realistic synthetic images of fruits with different types of defects. These GANs can be guided to produce specific fruit conditions, allowing for a balanced and diverse training dataset. By creating fake but highly believable images of spoiled fruits, we give the AI more chances to learn what “bad” looks like—even in rare or early-stage defects.

Once the dataset includes both real and synthetic images, a Convolutional Neural Network (CNN) is trained to classify fruit quality. This model becomes better at spotting even small or hidden imperfections, improving its accuracy and reliability in sorting. The CNN doesn't just memorize examples; it learns deeper patterns that help it perform well across various real-world conditions like lighting changes or fruit variations.

Overall, this combined approach leads to a robust fruit sorting system that can automatically and accurately separate good fruits from bad ones. It saves time, reduces labor costs, cuts down on waste, and ensures better quality produce reaches consumers. By blending real and AI-generated data, the system becomes more efficient and dependable ready to meet the demands of modern agriculture and high-speed fruit processing environments.

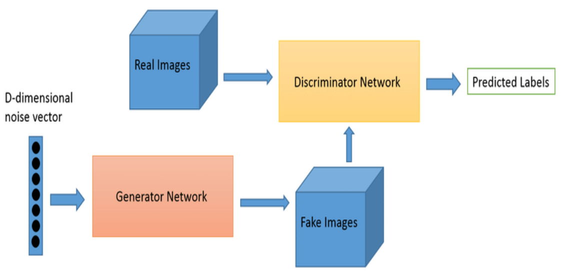

Fig: 1. Methodology Diagram

III. IMPLEMENTED DESIGN

Block Diagram:

This block diagram showcases about the data representation steps that need to be followed in the given represented procedure for the data augmentation.

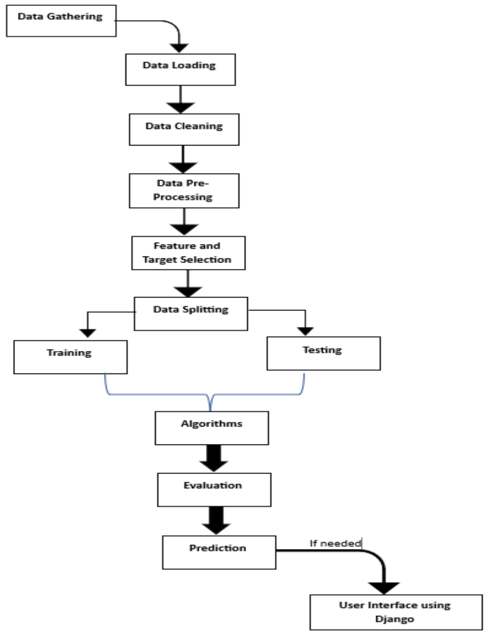

Fig: 2 .Block Diagram

The process begins with data gathering, where raw data is collected from various sources. This data is then loaded into the working environment for analysis. Following this, data cleaning is performed to handle missing values, remove duplicates, and correct inconsistencies. The cleaned data undergoes pre-processing, which may include normalization, encoding, and other transformations to make it suitable for modeling. Next, feature and target selection is carried out to identify the most relevant input variables (features) and the outcome variable (target) for prediction. The dataset is then split into training and testing sets to enable model training and evaluation. The training set is used to train machine learning models using selected algorithms, while the testing set helps assess the model’s performance. The results are evaluated to determine the model’s accuracy and reliability. Once a satisfactory model is achieved, it can be used for prediction on new data. If needed, the model is deployed through a user interface built with Django, allowing users to interact with the system seamlessly.

System Architecture:

System architecture is a conceptual framework that explains the composition and actions of multiple components and subsystems.

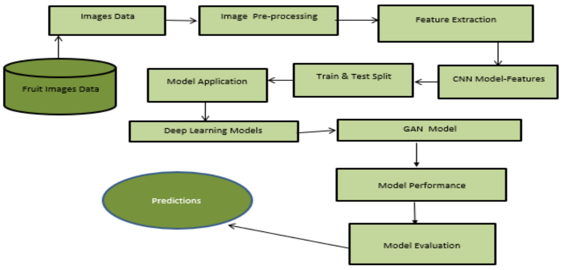

Fig: 3. System Architecture Diagram

Activity Diagram:
An activity diagram is a type of flowchart that shows how a system or process operates. It demonstrates the flow of control from one activity to another.

Purpose:
•To model the system's dynamic component
•To visualize business processes, system behaviors, or user workflows.
•Used heavily in software engineering, especially with UML (Unified Modeling Language).

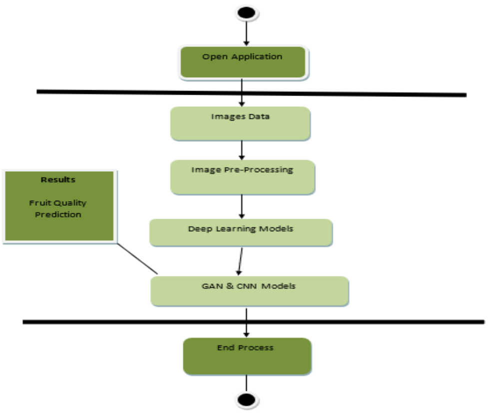

Fig: 4. Activity Diagram

An Activity Diagram shows the step-by-step flow of a system's operations. It demonstrates how a system's workflow progresses from one action to the next. To simulate the dynamic behavior of a system. To visualize business processes, system behaviors, or user workflows Commonly used in software engineering, especially with UML (Unified Modeling Language):

IV. RESULT

In this project, the classification of fruit quality and defect types was significantly enhanced through the use of Conditional Generative Adversarial Networks (cGANs) for data augmentation. By generating diverse and realistic synthetic images conditioned on specific fruit quality labels, the training dataset was effectively expanded, addressing issues of class imbalance and limited data. Convolutional Neural Networks (CNNs) were then employed to perform classification tasks, achieving improved accuracy and robustness compared to models trained on original datasets alone. Evaluation metrics such as precision, recall, F1-score, and overall accuracy demonstrated that the use of cGAN-generated images contributed to more reliable and consistent classification outcomes, particularly in underrepresented defect categories. This approach proved beneficial for real-world applications where data diversity and model generalization are critical.

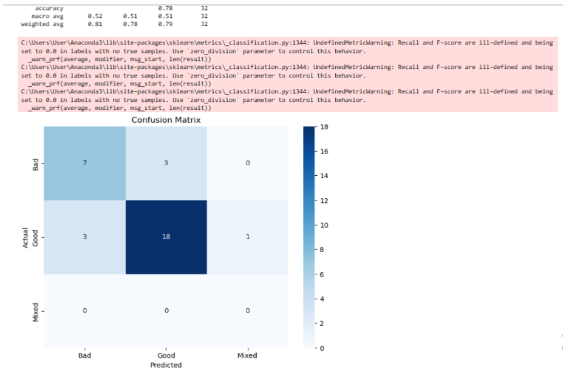

Fig:5.Confusion Matrix

The above figure showcases about the confusion matrix that tells about the trained model to the bit size ratio of the given image data unit.Training and Validation Accuracy plot for Fruit Quality Classification is shown bellow

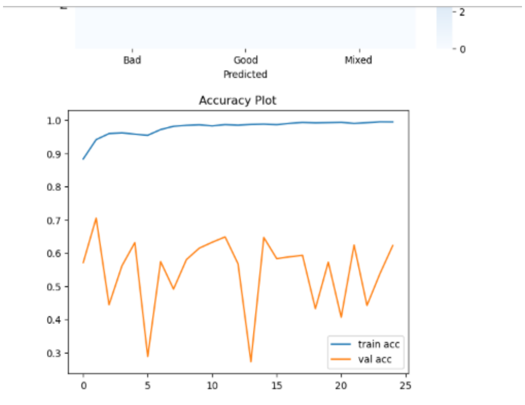

Fig:6. Accuracy Plot

Training Accuracy (blue line): Increases rapidly and remains very high (close to 1.0), indicating the model fits the training data well.
Validation Accuracy (orange line): Fluctuates significantly and stays relatively low (mostly

between 0.3 and 0.7), showing instability and poor generalization.

The model is likely overfitting—it performs very well on training data but fails to generalize to unseen validation data.

Good fruit quality prediction result with confidence score:

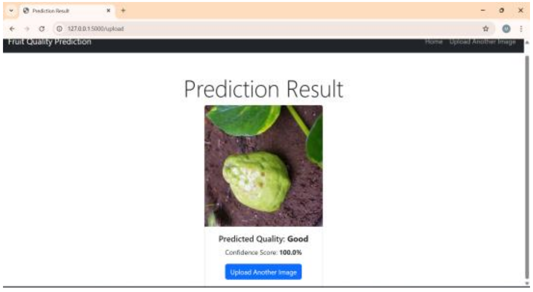

Fig:7. Predicted output as Good

This above image shows the quality of the fruit is type-good and defined the confidence score as 100%.So that it is good to consume

Bad fruit quality prediction result with confidence score:

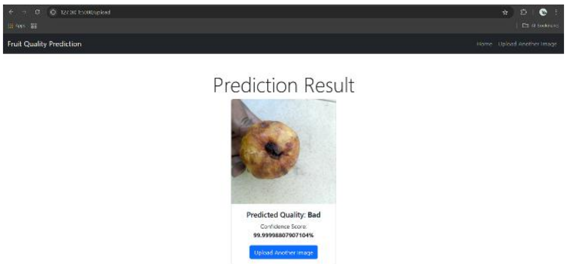

Fig:8. Predicted output as Bad

This above image shows the quality of the fruit is type-bad and defined the confidence score as 99.9%.So that it is not good to consume

And it is in bad condition.Thus gives the system function type as bad fruit

Mixed fruit quality prediction result with confidence score

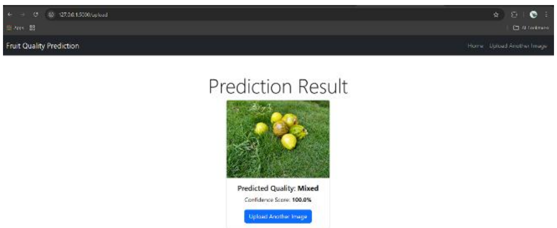

Fig:9. Predicted output as mixed

This above image shows the quality of the fruit is type-Mixed and defined the confidence score as 100%.So that it is not sure to tell that which fruite is good or bad and to avoide any mis judumental issue while goindg through various fruite types. Which have different shape,colour and sizes,

V. CONCLUSION

This project presents an advanced fruit quality classification and defect detection system using Adversarial Networks with Conditional Generative Functions(GANs) for data augmentation. By generating realistic synthetic images, the system addresses challenges related to limited and imbalanced datasets, improving the accuracy and robustness of fruit classification models. The system achieves an accuracy of 78%, demonstrating its effectiveness in identifying fruit quality and defects. The integration of both real and augmented data enhances the system's ability to identify subtle quality variations, making it more reliable across different fruit types and conditions. This automation reduces dependency on manual inspection, ensuring faster and more consistent quality control in large-scale agricultural operations. Ultimately, the proposed system contributes to improving product quality, reducing food waste, and supporting more sustainable farming practices.

B. *Future and scope*

Future work can focus on extending the system to classify a wider range of fruits and detect additional quality attributes such as internal defects or chemical residues. Integrating multi-sensor data, including hyperspectral imaging, could further enhance defect detection accuracy. The system can also be enhanced using advanced explainable AI techniques to provide clear insights into why a fruit was classified as defective. Additionally, deploying this system on mobile or edge devices could enable real-time quality assessment directly at farms or packaging centers. Continuous updates to the dataset and model retraining will ensure adaptability to evolving agricultural standards and changing environmental conditions.

REFERENCES

[1] K. Deshmukh, R. Kasture, S. Bhoite, S. L. Tade and S. A. Vaishnav, "Performance Analysis of Fruit Quality Detection Using Computer Vision and Object Detection," *2023 7th International Conference On Computing, Communication,*

Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-8, doi: 10.1109/ICCUBEA58933.2023.10392090.

[2] B. Chaudhari, D. Patil, H. Patil, R. Patil and J. Gawali, "Fruit Defect Inspection System Using Image Processing and IoT Framework," *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, 2023, pp. 1-6, doi: 10.1109/ESCI56872.2023.10099913.

[3] 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," *2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, Brno, Czech Republic, 2021, pp. 60-65, doi: 10.1109/ICUMT54235.2021.9631643.

[4] S. S. Hari, M. Sivakumar, P. Renuga, S. karthikeyan and S. Suriya, "Detection of Plant Disease by Leaf Image Using Convolutional Neural Network," *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, Vellore, India, 2019, pp. 1-5, doi: 10.1109/ViTECoN.2019.8899748.

[5] Goodfellow, I., et al. (2014).They invented GANs (Generative Adversarial Networks), which help computers create fake but realistic images, like fruits with defects.

[6] Mirza, M., & Osindero, S. (2014).These researchers improved GANs by adding labels (like "apple with bruise") so the model can create specific types of images — that's Conditional GANs (cGANs).

[7] Shorten, C., & Khoshgoftaar, T. M. (2019).They wrote a big review about different ways to make your image dataset bigger using tricks like rotations, flips, and GANs.

[8] LeCun, Y., Bengio, Y., & Hinton, G. (2015).They explained deep learning basics — how models like CNNs actually "learn" from images.

[9] Odena, A., Olah, C., & Shlens, J. (2017).They developed techniques (like Auxiliary Classifier GANs) that make generated images even sharper and more reliable for training.

[10] Radford, A., Metz, L., & Chintala, S. (2016).These guys introduced Deep Convolutional GANs (DCGANs), which are better at making realistic images, especially useful for fruits and other objects.

[11] Krizhevsky, A., Sutskever, I., & Hinton, G. (2012).They created AlexNet, one of the first CNNs that made a big splash in image classification tasks.

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016).They built ResNet, a powerful CNN model that solves some big training problems — it's great for classifying tricky fruit images.

[13] Wang, Z., She, Q., & Ward, T. E. (2020).They wrote about how GANs are used to create synthetic data across different industries, including agriculture.

[14] Frid-Adar, M., et al. (2018).They showed how GANs can create synthetic medical images for better training — the same idea applies to fruits: make fake examples to train better models.

[15] Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017).They invented CycleGANs, where you can transform images from one type to another — like healthy fruit to bruised fruit, without needing pairs.