International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Sunflower Seed Oil Yield Prediction using Machine Learning

M G Srinivasa[1*], Amrutha D N[2], Prajwal K P[3], Prakruthi P[4], Rachana Prabhu[5]
1 Professor, 2,3,4,5 Students
Department of Electronics and Communication Engineering
Maharaja Institute of Technology Thandavapura, Nanjangud, Mysore, India

Abstract: Sunflower (Helianthus annuus L.) is a highly cultivated oilseed crop, appreciated for its high-quality edible oil. Manual rating of seed yield potential is time consuming and subjective in nature. This study proposes a deep learning-based sunflower seed oil yield prediction system using MobileNetV2. In this study, the model was trained to classify seeds into high- and low-yield categories based on visual features of seeds, including texture, shape, and color. The framework achieved an accuracy rate of 92.3% and was deployed in a Gradio web interface for real-time prediction. This system enhances low-cost, efficient and scalable tools in precision agriculture and seed quality assessment [1].

Keywords: Sunflower Seeds, Deep Learning, MobileNetV2, Image Classification, Precision Agriculture

1. INTRODUCTION

Sunflower (Helianthus annuus L.) is one of the most important oilseed crops, valued for its edible oil and nutritional benefits. The oil yield of sunflower seeds depends on several visible traits such as seed size, color uniformity, maturity, and the specific genetic variety. Traditionally, farmers and workers assess these qualities manually, but this process is slow, subjective, and becomes unreliable when dealing with large quantities of seeds [2][3].

With the rise of precision agriculture, automated methods using computer vision and deep learning are becoming increasingly useful [4]. Convolutional Neural Networks (CNNs) can automatically learn important features from seed images, eliminating the need for manual measurements. Among these, MobileNetV2 stands out as a lightweight and efficient model, making it suitable for real-time applications on devices with limited processing power [5].

In this project, a MobileNetV2-based model is developed to classify sunflower seeds into high-yield and low-yield categories using only visual features [6]. To make the system easy to use in real-world agricultural settings, a Gradio interface is also integrated for simple and real-time interaction. Overall, this approach aims to offer a fast, accurate, and affordable solution for evaluating seed quality [7].

2. PROBLEM STATEMENT AND OBJECTIVE

2.1 Problem Statement

Sunflower seed oil yield is usually judged through manual visual inspection, but this method is often subjective, inconsistent, and too slow for large-scale evaluation. While laboratory tests can give accurate results, they are expensive and not suitable for everyday use in farming [8]. Therefore, there is a clear need for an automated, reliable, and affordable way to classify sunflower seeds based on their oil-yield potential. This study addresses that need by using a deep learning-based image classification model [9].

2.2 Objective

The main goal of this study is to develop an automated, image-based system that can estimate the oil yield potential of sunflower seeds using deep learning. The work focuses on using the MobileNetV2 architecture to automatically extract visual features from seed images and classify them as either high-yield or low-yield. The model is thoroughly evaluated to ensure that it is accurate, consistent, and suitable for real-time applications. A user-friendly Gradio interface is also developed to make the system easy to use in practical settings. Overall, this approach aims to offer a cost-effective and scalable solution [11][12].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

3. METHODOLOGY

The proposed system is designed to classify sunflower seeds into high-yield and low-yield categories based on visual characteristics using a deep learning model. The methodology consists of four main stages: dataset preparation, image preprocessing, model development and training, and deployment through a user interface [14].

A collection of images of sunflower seeds was curated, covering seeds with either high or low yield potential. The images were taken under the same lighting and background conditions in order to control unwanted variance and noise. Before feeding this image data to the model, the dataset was resized to the input size required by MobileNetV2, and pixel values were normalized (using zero mean and unit standard deviation). The dataset was also diversified using data augmentation to reduce redundancy while enriching the data through rotation, flipping, and zooming [15]. MobileNetV2 was selected for this work due to its light weight and effectiveness in extracting features from images efficiently and accurately.

The model was updated with a custom softmax layer to accommodate two-class (binary) classification, on top of the pre-trained MobileNetV2 backbone. The model was trained against the labeled image dataset, and hyper parameters such as learning rate, number of epochs, and batch size were selected and tuned. Throughout training, accuracy and loss values were monitored for both the training data and a separate validation dataset to assess generalization to unseen data. The performance of the model was evaluated using training accuracy, validation accuracy, and confusion matrix analysis [15].

3.1 Dataset Preparation

Sunflower seed images were collected to represent both high-yield and low-yield categories, ensuring that the dataset covered the necessary visual differences between the two groups. The images were taken under controlled lighting and a uniform background to avoid shadows, reflections, or distractions that could affect the model's learning process. Each seed sample was carefully positioned to maintain consistency in angle and distance from the camera. This controlled setup helped produce clean, high-quality images that allowed the deep learning model to focus directly on important visual features such as seed size, texture, shape, and color uniformity [16]. Collecting the data in this systematic way improved the reliability of the dataset and supported better model performance.

3.2 Image Processing

All images were resized to a standard resolution that matches the input requirements of the MobileNetV2 model, ensuring consistency during training. Preprocessing also involved normalizing pixel values so that the model could learn more efficiently and avoid bias toward brighter or darker images [17]. To make the dataset more diverse and improve the model's ability to generalize, several augmentation techniques—such as rotation, zooming, flipping, and slight shifts—were applied. These techniques help simulate real-world variations and reduce the chances of over fitting, allowing the model to perform better on new, unseen images [17].

3.3 Model Training using MobileNetV2

MobileNetV2 is a lightweight deep learning model designed for fast and efficient image classification. It uses depth-wise separable convolutions to extract useful features while keeping computation low, making it ideal for real-time applications [18].

3.4 System Deployment

After achieving satisfactory accuracy, the trained model was integrated into a Gradio interface to enable real-time prediction. Users can upload seed images and instantly receive yield classification results, making the system practical for agricultural usage [18].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

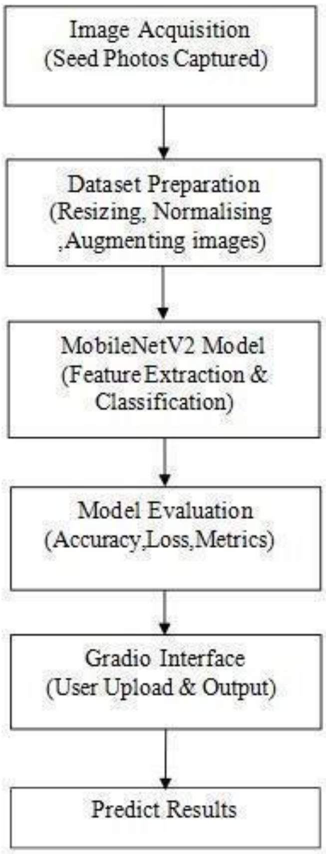

Fig -1: Block diagram of proposed classification system

The block diagram illustrates the complete workflow of the sunflower seed oil yield prediction system using deep learning. The process starts with image acquisition, where high-quality images of sunflower seeds are captured under controlled lighting and background conditions to minimize noise and variability. These images are then subjected to dataset preparation, which includes resizing the images to match the input dimensions of the model, normalizing pixel values for efficient learning, and applying data augmentation techniques such as rotation, flipping, and zooming to increase dataset diversity and prevent overfitting [18]. The prepared images are fed into the MobileNetV2 model, which acts as both a feature extractor and a classifier by automatically learning discriminative visual features like seed texture, shape, surface patterns, and color characteristics, and classifying the seeds into high-yield or low-yield categories. The performance of the trained model is assessed using evaluation metrics such as accuracy, loss, precision, recall, and confusion matrix analysis to verify its reliability and consistency. To enable real-world usability, the trained model is deployed through a Gradio-based web interface that allows users to upload sunflower seed images and receive real-time classification results without requiring technical expertise. Finally, the system generates prediction results indicating the oil yield potential of the sunflower seeds, offering a fast, accurate, and user-friendly solution for automated seed quality assessment in precision agriculture [19].

Table -1: Performance metrics of sunflower seeds oil yield

Class	Precision	Recall	F1-Score	Accuracy (%)
High-Yield Seeds	0.93	0.94	0.93	93.1
Low-Yield Seeds	0.91	0.90	0.90	91.5
Average / Overall	0.92	0.92	0.92	92.3

Table 1 presents the performance evaluation of the MobileNetV2-based sunflower seed oil yield classification model. The model achieved an overall accuracy of 92.3%, with balanced precision, recall, and F1-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

score for both high-yield and low-yield seed classes, indicating reliable and consistent classification performance.

Fig -2: (I) Low-yield seed image (II) High-yield seed image

High-yield sunflower seeds are typically larger and more uniform, while low-yield seeds are smaller and irregular, and these visual differences are used by the system to classify the seeds accurately.

4. SOFTWARE REQUIREMENTS

The proposed sunflower seed oil yield prediction system is developed using deep learning and computer vision techniques. Hence, appropriate software tools and frameworks are required to support image preprocessing, model training, evaluation, and deployment. The software requirements for this project are listed and described below.

4.1 Operating System

Windows, Linux, or macOS operating systems may be used for accessing the development environment. Since the project is executed primarily in a cloud-based environment, no specialized local OS configuration is needed. A standard system capable of running a web browser is sufficient.

4.2 Google Colab (Cloud-Based Development Environment)

Google Colab is used as the primary platform for implementing and training the deep learning model. It provides free GPU support, allowing computationally intensive model training without requiring high-end local hardware. Colab also supports real-time collaboration, dataset handling, visualization, and notebook execution.

4.3 Python Programming Language (Version 3.7 or Above)

Python serves as the core development language due to its extensive machine learning ecosystem, readability, and rapid prototyping capability. Python supports numerous libraries required for data manipulation, neural network modeling, and image processing.

4.4 Tensor Flow and Keras Framework

The MobileNetV2 deep learning model used in this project is implemented using the Tensor Flow/Keras framework. Tensor Flow provides GPU acceleration and optimized deep learning operations. Keras offers simplified APIs for model building, training, evaluation, and fine-tuning. These frameworks enable efficient implementation of transfer learning for sunflower seed classification.

4.5 Open CV (Open Source Computer Vision Library)

Open CV is used for image preprocessing tasks such as resizing, normalization, and augmentation. These steps enhance the dataset quality and improve the model's generalization capability. OpenCV ensures consistent input preparation for the MobileNetV2 architecture.

4.6 NumPy and Pandas Libraries

NumPy is used for handling matrices and numerical operations involved in image arrays and model predictions. Pandas assists in dataset organization, label management, and structured data manipulation. Both libraries play a crucial role in efficient data handling during preprocessing and analysis.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

4.7 Gradio Interface Framework

Gradio is used to create an easy and interactive web interface for real-time predictions. With this interface, users can upload sunflower seed images and instantly get the yield classification results. This makes the final model simple to use and accessible for farmers, researchers, and field workers—even if they have no programming knowledge.

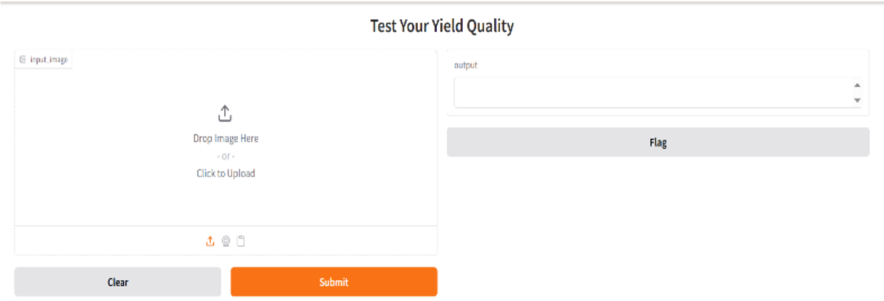

Fig -3: User interface to capture or upload the seed image

This interface represents the Gradio-based web application developed for sunflower seed oil yield prediction. It provides a simple and interactive platform where users can upload sunflower seed images using the upload section on the left. After uploading, the user can submit the input to the system using the Submit button, which sends the image to the trained deep learning model for processing [20]. The predicted output, indicating the seed yield category, is displayed in the output panel on the right. The interface also includes a Clear option to reset inputs, making the system user-friendly, efficient, and suitable for real-time agricultural applications without requiring technical expertise [21].

5. RESULTS AND DISCUSSION

5.1 Model Performance and Analysis

The proposed sunflower seed oil yield classification system achieved an overall accuracy of 92.3%, demonstrating strong capability in distinguishing between high-yield and low-yield sunflower seeds. The precision, recall, and F1-score values indicate balanced performance across both classes, showing that the model reliably learns discriminative visual features from seed images without bias toward a specific category.

5.2 Confusion Matrix and Classification Behavior

The confusion matrix analysis shows that most sunflower seed samples were correctly classified, with only a small number of misclassifications between high-yield and low-yield categories. This indicates effective separation of seed classes and confirms that the model performs consistently. Minor misclassifications can be attributed to visual similarities between certain seeds, such as overlapping size or texture characteristics.

Table -2: Confusion matrix for sunflower seeds classification

Actual / Predicted	**High-Yield Seeds**	**Low-Yield Seeds**
High-Yield Seeds	56 (True Positive)	4 (False Negative)
Low-Yield Seeds	5 (False Positive)	55 (True Negative)

Accuracy Calculation

The overall classification accuracy is calculated as:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$= (56 + 55) / 120 = 92.3\%$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Table 2 shows the confusion matrix obtained from the test dataset. The model correctly classified 56 high-yield and 55 low-yield sunflower seed samples, resulting in an overall accuracy of 92.3%. The low number of misclassifications indicates strong discrimination capability between the two yield categories.

5.3 Practical Significance and System Effectiveness

The use of the MobileNetV2 architecture, combined with image preprocessing and data augmentation, ensures efficient feature extraction and low computational complexity. This makes the system suitable for real-time deployment through a Gradio-based interface. Overall, the proposed approach offers a fast, accurate, and user-friendly solution for automated sunflower seed oil yield assessment, providing practical benefits for farmers and the wider agriculture sector.

6. CONCLUSION

The project Sunflower Seed Oil Yield Prediction Using Deep Learning successfully demonstrates the application of computer vision and neural network models in agricultural quality assessment. By using the MobileNetV2 architecture, the system efficiently classifies sunflower seeds into high-yield and low-yield categories based on visual characteristics such as texture, color, and surface uniformity. The use of Google Colab for model training and Gradio for front-end deployment enables easy accessibility and real-time predictions without the need for high-end hardware. The achieved results indicate that deep learning-based image classification can serve as a reliable and fast alternative to traditional manual inspection methods. Overall, the proposed system offers a practical, cost-effective, and user-friendly solution to support farmers, researchers, and agricultural industries in making informed decisions regarding seed selection and yield optimization.

7. FUTURE ENHANCEMENTS

Although the system works well under controlled conditions, there is still room to improve and expand it. One way to make the model more robust is to increase the size of the dataset by adding images taken under different lighting and background settings. This would help the model perform better in real-world conditions. Currently, the system only supports binary classification, but future versions could include multi-class classification to identify medium-yield or defective seeds for more detailed grading.

The project can also be extended by integrating IoT-based seed scanning devices or developing a mobile app, making the solution easier to use directly in the field. Another useful improvement would be to move beyond classification and use regression techniques to estimate the exact oil yield of each seed, giving more precise and quantitative results. Deploying the system on cloud platforms could further increase its scalability and allow farmers from different regions to access the tool easily.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Department of Electronics and Communication Engineering, Maharaja Institute of Technology, Thandavapura, for providing the necessary facilities and support to carry out this research work. We extend our heartfelt thanks to our guide and faculty members for their valuable guidance, continuous encouragement, and technical support throughout the project. We also acknowledge the use of open-source tools and platforms such as Python, Tensor Flow, Google Colab, and Gradio, which greatly contributed to the successful implementation of this work.

REFERENCES

[1] M. Sawarkar et al., "Seed quality evaluation using digital image processing techniques in MATLAB," Int. J. Adv. Res. Comput. Sci., vol. 8, no. 5, 2017.

[2] S. Cvejić, V. Mladenov, and A. Popović, "Application of machine learning algorithms for sunflower oil yield prediction from hybrid data," Comput. Electron. Agric., vol. 212, 2023.

[3] X. Liu et al., "Non-destructive classification of sunflower seeds using NIR spectroscopy and 1D CNN," Infrared Physics & Technology, vol. 128, 2023.

[4] Y. Huang, W. Li, H. Zhang, and C. Zhang, "Hyperspectral imaging for non-destructive assessment of sunflower seed vigor and moisture," Comput. Electron. Agric., vol. 209, 2023.

[5] M. Rashad et al., "A comprehensive review of modern classification techniques for machine learning applications," J. Theor. Appl. Inf. Technol., 2020.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

[6] Anonymous, "Deep learning-based classification of sunflower seed images," Unpublished summary, 2018–2021.
[7] M. Hernández et al., "Genetic variability in sunflower seed oil content: machine learning-based predictive analysis," Industrial Crops and Products, 2022.
[8] R. Ortega et al., "Oil content estimation in sunflower seeds using NIR spectroscopy and PLS regression," J. Sci. Food Agric., 2020.
[9] D. Radanović et al., "Machine learning models for predicting sunflower hybrid performance and oil yield," Agronomy, 2021.
[10] A. Tubić et al., "Analysis of sunflower seed quality using computer vision and texture features," Helia, 2019.
[11] M. Khaki and L. Wang, "Crop yield prediction using deep neural networks," Frontiers in Plant Science, 2019.
[12] S. Jeong et al., "Machine learning-based crop yield prediction using remote sensing," Remote Sensing, 2022.
[13] K. Jain et al., "Agricultural yield prediction using ML techniques," Int. J. Adv. Comput. Sci., 2021.
[14] M. Chlingaryan et al., "Machine learning approaches for crop yield prediction and nitrogen status estimation: A review," Comput. Electron. Agric., 2018.
[15] Y. Sun and H. Wang, "Deep learning for multispectral yield prediction," Agricultural Systems, 2020.
[16] A. Sagan et al., "Predicting plant oil content using ML models: A general framework," Plants, 2021.
[17] N. Li et al., "Yield estimation of oil crops using UAV imagery and machine learning," Precision Agriculture, 2022.
[18] Seed Classification, Quality Assessment & Computer Vision (Applicable).
[19] N. Zhang et al., "Computer vision-based seed classification and quality detection: A review," Sensors, 2021.
[20] N. J. Slaughter et al., "Image-based quality estimation of seeds using ML," Biosystems Engineering, 2020.
[21] R. Tetila et al., "Convolutional neural networks for seed classification in agriculture," Comput. Electron. Agric., 2020.
[22] A. K. Singh et al., "Assessment of seed traits using deep learning and image analysis," Information Processing in Agriculture, 2022.
[23] J. Han et al., "Deep CNN-based seed morphology analysis for quality and germination prediction," Computers and Electronics in Agriculture, 2023.
[24] L. Feng et al., "Deep learning for hyperspectral image-based crop and seed quality analysis," Remote Sensing, 2019.
[25] L. Zhao et al., "Estimation of oil content in oilseeds using hyperspectral imaging," Sensors, 2020.
[26] K. Yamada et al., "Non-destructive seed classification based on NIR spectroscopy and ML," Journal of Near Infrared Spectroscopy, 2021.
[27] R. B. Walker et al., "Machine learning for chemical composition prediction from hyperspectral data," ISPRS Journal of Photogrammetry, 2021.
[28] P. Mahesh et al., "Seed moisture & oil content detection using NIR and ML," Postharvest Biology and Technology, 2022.
[29] J. Kamilaris and F. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Comput. Electron. Agric., 2018.
[30] N. Saleem et al., "Applications of machine learning in smart farming: A review," Information Processing in Agriculture, 2022.
[31] P. Singh et al., "AI and ML techniques for precision agriculture and crop analysis," Sustainable Computing, 2021.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

AUTHORS PROFILE

M G Srinivasa: M.Tech in VLSI and Embedded Systems from SJCE Mysore; Ph.D. in the field of Wireless Body Area Network (WBAN) for real-time wearable physiological parameter monitoring and algorithms to classify health status, from Visvesvaraya Technological University, Belgaum, under the guidance of Dr. P S Pandian, Scientist, G, LRDE, DRDO. Presently working as Professor in the Department of Electronics and Communication Engineering, Maharaja Institute of Technology, Thandavapura, with 19 years of teaching and 8 years of industrial experience. Field of interest includes Analog Electronics, Signals & Systems, Linear Integrated Circuits, Antenna and Propagation, Control Systems, Digital Signal Processing, and Embedded System Design. Actively involved in IETE professional body activities, including workshops, seminars, and national-level model and paper presentation contests. Instrumental in signing MOUs with software and hardware companies for the benefit of students in securing internship training and placements.

Amrutha D N: Final-year Electronics and Communication Engineering student at Maharaja Institute of Technology, Thandavapura. A result-oriented student with strong fundamentals in VLSI design, digital design, and embedded systems, along with knowledge of IT domains such as programming and problem-solving. Has hands-on experience using software tools such as Xilinx, MATLAB, and Cadence for design, simulation, and implementation of academic and project-based work. Committed to continuous learning and adapting to new technologies, with the goal of securing a challenging role in the VLSI industry or IT sector.

Prajwal K P: Final-year Electronics and Communication Engineering student at Maharaja Institute of Technology, Thandavapura. Possesses solid knowledge of embedded systems, IoT, and basic VLSI concepts. Skilled in programming using C and Python with practical exposure to Arduino and sensor interfacing. Has foundational understanding of machine learning concepts using Python libraries. Skilled in handling responsibilities both independently and as part of a team.

Prakruthi P: Final-year student of Electronics and Communication Engineering at Maharaja Institute of Technology, Thandavapura, with a strong foundation in circuit design, embedded systems, and VLSI design. Eager to apply academic knowledge to real-world projects and contribute to innovative technological solutions. Strong analytical, problem-solving, and logical thinking abilities. Skilled in programming using C and Python with practical exposure to Arduino and sensor interfacing.

Rachana Prabhu: Pursuing final year at Maharaja Institute of Technology, Thandavapura with the specialization of Electronics and Communication Engineering. An enthusiastic fresher with strong fundamentals in VLSI, digital design, and embedded systems, along with knowledge of Python and machine learning, and hands-on experience through basic deep learning-based projects. Also used OpenCV and NumPy for image processing and Gradio to create a simple interface for real-time predictions.