Shahid Bahonar University of Kerman
Agricultural Biotechnology Journal
Iranian Biotechnology Society
Print ISSN: 2228-6705 Online ISSN: 2228-6500

Automated soybean leaf disease detection using convolutional neural network and machine learning

Mate Sanjay

*Corresponding author, Government Polytechnic Daman, India. E-mail: matesanju@gmail.com

Suryawanshi Renuka

School of Computer Science and Technology, Dr.Vishwanath Karad World Peace University, Pune, India. E-mail: renukasurya1117@gmail.com

Thalor Meenakshi

AISSMS Institute of Information Technology, Pune, India. E-mail: meenakshi.thalor@aissmsioit.org

Tidake Ankita

Ajeenkya DY Patil School of Engineering, Pune, India. E-mail: ankita.tidake@gmail.com

Abstract

Objective

Agriculture stays a central driver of economic growth and food security in many developing countries, with soybean serving as one of the most important cash crops worldwide. Owing to its rich nutritional value, soybean is broadly consumed in human diets and livestock feed, and its use as a milk alternative is increasing because of its developed shelf life after processing. Nevertheless, soybean production is often threatened by a broad range of foliar diseases that vary based on soil conditions, climatic factors, and cropping patterns. These diseases create substantial losses in both yield and quality, underscoring the need for effective, automated, and accurate diagnosis techniques. This investigation aims to design and evaluate a soybean leaf disease diagnosis framework by comparing traditional machine learning methods with deep learning approaches.

Materials and methods

A curated dataset of 1,500 soybean leaf images covering multiple disease categories was utilized, divided into an 80:20 ratio for training and testing. Prior to model training, the images underwent preprocessing steps like resizing, grayscale conversion, histogram equalization, and segmentation to improve feature extraction. Two approaches were implemented: a support vector machine (SVM) with handcrafted features representing traditional machine learning, and a custom convolutional neural network (CNN) representing deep learning. The CNN was designed with three convolutional layers, max-pooling, and fully connected layers, incorporating data augmentation, dropout regularization, and adaptive learning rate tuning. Hyperparameter optimization was carried out to further raise efficiency.

Results

The proposed CNN achieved a validation accuracy of 100%, meaningfully outperforming the SVM classifier, which reached only 44.44%. Although minor indications of overfitting were observed, the CNN maintained robust generalization capability across varied leaf samples, demonstrating its reliability in practical applications.

Conclusions

The outcomes clearly establish the superiority of deep learning over conventional machine learning methods for plant disease recognition. Future study will centralize on employing transfer learning with pre-trained networks, as well as developing real-time diagnosis systems applying mobile applications integrated with IoT sensors. The proposed framework supplies a scalable and reliable pathway toward precision agriculture, proposing substantial benefits for sustainable crop management and improved productivity.

Keywords: convolutional neural network (CNN), image preprocessing, machine learning, precision agriculture, soybean leaf disease

Paper Type: Research Paper.

Citation: Mate, S., Suryawanshi, R., Thalor, M., & Tidake, A. (2025). Automated soybean leaf disease detection using convolutional neural network and machine learning. *Agricultural Biotechnology Journal,* 17(4), 151-168.

Agricultural Biotechnology Journal, 17(4), 151-168. DOI: 10.22103/jab.2025.25153.1694

Received: June 29, 2025. Received in revised form: September 10, 2025.

Accepted: September 11, 2025. Published online: November 30, 2025.

Publisher: Shahid Bahonar University of Kerman & Iranian Biotechnology Society.

Introduction

Soybean is one of the most important crops worldwide, serving as a major source of protein and oil. However, different diseases, containing bacterial pustule, downy mildew, and frogeye leaf spot, meaningfully reduce both yield and quality. Traditional disease diagnosis methods rely on manual inspection, which is subjective, time-consuming, and impractical for large-scale monitoring (Singh & Shukla, 2021; Yu et al., 2022). The rapid development of high-throughput and high-efficiency technologies has greatly expanded data generation in biology and biotechnology (Mohammadabadi et al., 2024a). Such data, derived from molecules containing metabolites, proteins, RNA, and DNA, supply critical discernments into the structure, task, and dynamics of living systems (Mohammadabadi et al., 2024b). Functional genomics, for example, aims to describe the task and interactions of these biomolecules to explain phenotypic variation. Similarly, in breeding programs, genetic improvement can be accelerated by accurate identification of superior animals as parents of the next generation (Mohammadabadi et al., 2024). Artificial neural networks have been proposed as powerful alternatives to traditional regression methods, capable of handling nonlinear, complex, and noisy biological data (Mohammadabadi et al., 2024b). Given that “omics” datasets are often too large and complex for visual or purely statistical analysis, machine intelligence and artificial intelligence have emerged as indispensable tools (Ghotbaldini et al., 2019). In agriculture, deep learning has demonstrated remarkable success in automated plant disease diagnosis, enabling rapid and precise classification of infected leaves (Singh & Shukla, 2021; Yu et al., 2022; Mohammadabadi et al., 2024). Numerous investigations have employed convolutional neural networks (CNNs) for soybean disease classification, with architectures like VGG16, ResNet50, and MobileNetV2 achieving high accuracy (Ma et al., 2017; Zhang, Wu, & Chen, 2021; Jain & Sharma, 2024). Transfer learning has further increased efficiency, exclusively for small, labeled datasets (Wu, Zhang, & Meng, 2019; He et al., 2016; Lin, Chen, & Yan, 2013). Hybrid models combining CNN feature extraction with support vector machine (SVM) classifiers have also been developed to improve robustness (Jain & Sharma, 2020; Mahobiya, Iyer, & Kolhe, 2023; Ray & Ray, 2021). In addition, hyperspectral imaging and UAV-based remote sensing have been investigated for large-scale field-level disease monitoring (Tetila et al., 2019; Zhang, Shao, & Shen, 2020; Thapliyal et al., 2024). In spite of these advancements, persistent challenges include dataset imbalance, environmental variability, and computational efficiency (Ji, Liu, & Wu, 2021). Earlier investigations demonstrated that traditional machine learning approaches like SVMs and Random Forest classifiers achieve only moderate efficiency because they rely heavily on handcrafted features like Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT). For instance, SVM classifiers achieved accuracies as low as 44.44%, underscoring the need for more robust approaches (Singh & Shukla, 2021). By contrast, CNNs automate feature extraction directly from raw image data, substantially improving classification accuracy (Wu, Zhang, & Meng, 2019). Recent investigations further highlight the benefits of transfer learning, progressive CNN architectures, and hybrid models in overcoming the limitations of handcrafted feature engineering (Jain &

Sharma, 2024; Mahobiya et al., 2023). Data augmentation, dropout regularization, and adaptive learning rate tuning have also been applied to mitigate overfitting and raise generalization across diverse leaf images (Ma et al., 2017; Wu et al., 2019). Remote sensing technologies like UAV-based hyperspectral imaging expand the scope to large-scale monitoring (Tetila et al., 2019), while emerging approaches like federated learning supply chances for collaborative model training without data sharing (Thapliyal et al., 2024). This investigation shows an optimized CNN-based soybean disease classification model that integrates data augmentation, dropout regularization, adaptive learning rate adjustment, and hyperparameter tuning. The proposed framework is systematically compared with traditional machine learning and other deep learning approaches to validate its effectiveness. By addressing key challenges like overfitting, data scarcity, and classification of visually similar diseases, this study contributes to developing a practical, scalable, and reliable solution for real-world deployment in precision agriculture.

Materials and methods

Dataset details: The dataset applied in this investigation consists of 1,500 processed images of soybean leaves gathered applying mobile devices. It includes both healthy and diseased leaf samples across multiple classes. The dataset is publicly available and can be applied to train machine learning models without additional preprocessing. It also serves as a benchmark for evaluating predictive models for plant disease diagnosis. Two prior investigations are exclusively relevant, as they supply detailed descriptions of datasets for soybean leaf disease identification and classification, with emphasis on conformation, acquisition methods, preprocessing steps, and applicability.

Dataset from the hypergraph cell membrane computing network study (Huang et al., 2024): This investigation employed a curated dataset containing 1,010 soybean leaf images, evenly divided between healthy (505 images) and diseased (505 images) samples. The images were captured in real agricultural environments near Huainan, China, applying a high-resolution Fuji X-A5 camera, ensuring that the dataset reflected natural variations in disease symptoms and environmental situations. Prior to model input, the images were preprocessed by resizing them to 227 × 227 pixels and applying normalization for uniformity. Feature extraction was carried out through three parallel modules: pyramid convolution (to capture multi-scale details), standard convolution (to extract local features like edges and textures), and a U-shaped network (to combine shallow and deep features). A dynamic attention mechanism was further integrated to emphasize the most informative features, effectively merging outputs from the three extraction pathways. The dataset was divided into training, validation, and testing subsets to enable robust model evaluation. Given the relatively modest dataset size, the model architecture was specifically designed to maximize feature learning and improve generalization. Outcomes demonstrated that the model achieved an accuracy of 98%, outperforming different established deep learning architectures and showing strong potential for real-world application in soybean disease diagnosis.

Table 1. Comparison of numerous investigations on soybean leaf disease diagnosis, emphasizing methodology, algorithms, and outcomes

Title	Year	Methodology	Algorithm	Achievement
Classification and Recognition of Soybean Leaf Diseases in Madhya Pradesh	2023	Image Processing and Deep Learning	CNN	Developed a system to identify Angular Leaf Spot, Bacterial Blight, and Soybean Rust with high accuracy.
Classification System for Soybean Leaf Disease Based on Deep Learning	2020	Deep Learning on Real Crop Images	CNN	Proposed a classification system that achieved meaningful accuracy on multi-leaf real crop images.
Federated Learning with CNNs for Soybean Leaf Disease Classification	2024	Federated Learning with CNNs	CNN	Developed a federated learning-based classification model with strong accuracy on real crop images.
Automatic Recognition of Soybean Leaf Diseases Using UAV Images and Deep Convolutional Neural Networks	2019	UAV Imaging and Deep Learning	Deep CNN	Successfully recognized diseases from UAV-captured images with high precision.
Soybean Leaf Disease Classification Using Machine Learning Techniques	2020	Machine Learning with Feature Extraction	Random Forest, SVM	Compared numerous classifiers, with Random Forest achieving the highest accuracy.
Deep Learning-Based Soybean Leaf Disease Classification Using Image Augmentation	2020	Deep Learning with Augmented Data	CNN	Improved model efficiency by applying image augmentation techniques.

Dataset from the India soybean leaf study (Kotwal et al., 2024): This study introduced a larger dataset of 3,363 soybean leaf images gathered primarily in Maharashtra, India. Images were acquired over different crop cycles applying smartphone cameras in natural outdoor situations, capturing a broad range of health and disease variations representative of real field environments. The dataset is organized into six categories: healthy leaves, vein necrosis, dry leaf, Septoria brown spot, root images, and bacterial leaf blight. Class distributions varied from relatively small (root images) to different hundred (leaf disease categories), reflecting the natural prevalence of these situations. All images were resized to 300 × 300 pixels, and data augmentation techniques, like rotation, zoom, and brightness adjustment, were applied to raise dataset diversity and reduce overfitting through training. The dataset was split into training and testing subsets applying an 80:20 ratio to ensure fair efficiency evaluation. Importantly, it is openly accessible to researchers and supplies a practical benchmark for developing and validating machine learning and deep learning models for soybean leaf disease classification.

Preprocessing steps: Preprocessing is essential for improving the quality of input data and enhancing classification accuracy. The subsequent steps were applied:

- Image resizing: All images were resized to 224 × 224 pixels to ensure uniformity for deep learning models.

- **Grayscale conversion:** Applied only for classical feature extraction methods to reduce computational complexity.

- **Image enhancement:** Histogram equalization was applied to improve contrast, and edge diagnosis techniques were applied to highlight diseased regions.

- **Thresholding and segmentation:** Methods like Otsu's thresholding and K-means clustering were employed to segment affected areas of the leaf.

- Data augmentation: Random rotations, flips, and brightness adjustments were applied to improve dataset diversity and reduce overfitting through model training.

Feature extraction: Extracting meaningful features is crucial for distinguishing between healthy and diseased leaves. Figures 1–3 show the key stages of image processing and feature extraction for soybean leaf disease diagnosis.

Figure 1. Diseased soybean leaf exhibiting vein necrosis symptoms

The three stages of image processing are described below:

- Original image: The original image (left in Figure 2) displays the soybean leaf in its natural color. Yellow and brown discolorations characteristic of vein necrosis are clearly visible, serving as a reference for disease identification.

- Grayscale image: The grayscale image (center in Figure 2) converts the original color image to shades of gray. This simplifies the image by removing color information while preserving shape, texture, and vein patterns. Grayscale images are commonly applied in computer vision tasks to reduce computational complexity while retaining essential structural details.

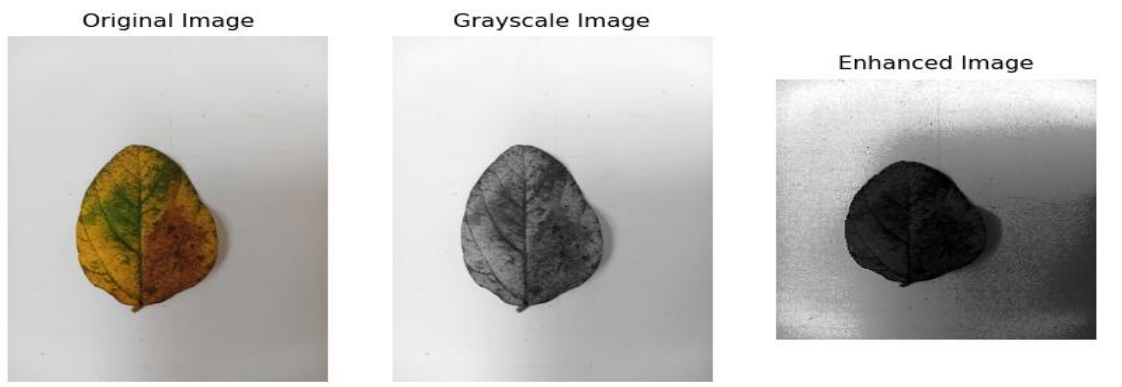

Figure 2. Soybean leaf image processing stages: Original, Grayscale, and Increased views

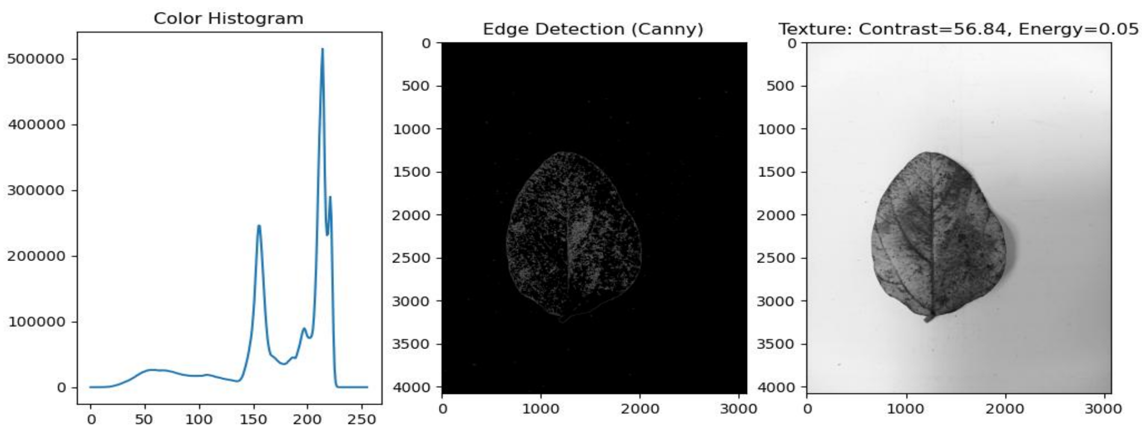

Figure 3. Feature extraction techniques for soybean leaf image analysis, containing color, edge, and texture features

- Enhanced image: The increased image (right in Figure 2) applies techniques like histogram equalization or contrast adjustment to improve feature visibility. This step accentuates the contrast between healthy and affected areas, making disease symptoms, containing vein patterns and lesions, more distinguishable for both human observers and machine learning algorithms. These preprocessing and feature extraction steps ensure that the dataset is well-prepared for training machine learning and deep learning models for accurate soybean leaf disease classification.

Color histogram, edge detection, and texture analysis:

- Color histogram: The color histogram (left in Figure 3) represents the distribution of color intensities within the leaf image. Peaks in the histogram illustrate often occurring color values. This supplies worth information about dominant hues and color variations in the leaf, which can be related to different disease symptoms, like yellowing created by vein necrosis.

- Edge detection (Canny): The edge detection image (center in Figure 3) highlights boundaries and edges of the leaf, containing the outline, veins, and lesion areas. Canny edge detection is an effective method for revealing structural details, which assists algorithms in recognizing lesion boundaries, vein patterns, and other features critical for differentiating healthy from diseased tissue.

- Texture Analysis: The texture image (right in Figure 3) displays the leaf after processing for texture features, with quantitative metrics like contrast (56.84) and energy (0.05). Texture analysis captures subtle variations in the leaf surface, like roughness, smoothness, or regular patterns, applying statistical methods. These properties are highly informative for classifying plant diseases that manifest as alters in leaf texture. Together, these feature extraction techniques supply a wide characterization of leaf images, enabling machine learning models to accurately identify and classify soybean diseases based on color, shape, and texture cues. These preprocessing and feature extraction stages are essential for improving the accuracy of automated soybean disease diagnosis.

Feature extraction methods: This investigation employed a combination of handcrafted and deep learning-based feature extraction techniques:

- Texture features: Extracted applying Gray Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) to capture leaf surface patterns.

- Color features: Histogram-based analysis was applied to quantify variations in leaf color.

- Edge detection: Canny and Sobel operators were applied to highlight affected regions.

- Deep CNN features: Pre-trained models, like VGG16, were utilized for automatic feature learning and extraction from leaf images.

Classification Approaches: Two primary classification strategies were evaluated: Machine Learning (ML) and Deep Learning (DL):

- Machine Learning: Handcrafted features were applied to train traditional classifiers, containing Support Vector Machine (SVM) and Random Forest. The ML approach achieved a maximum accuracy of 44.44%, restricted by feature dependencies and the inherent constraints of handcrafted feature extraction.

- Deep Learning: A custom Convolutional Neural Network (CNN) was designed for automatic feature learning. The dataset was split into 80% training and 20% validation subsets. The CNN architecture consisted of convolutional layers with ReLU activation, max-pooling layers, fully connected layers, and dropout regularization. The model achieved 100% validation accuracy after 10 epochs, meaningfully outperforming the traditional ML approach.

Hyperparameter optimization: To raise model efficiency, different hyperparameters were tuned:

- Learning rate: An adaptive learning rate plan was employed to optimize convergence.

- Batch size: Experiments were managed with batch sizes of 16, 32, and 64 to identify the optimal value.

- Dropout regularization: Dropout layers were applied to prevent overfitting.

- Loss function: The cross-entropy loss task was applied and optimized applying the Adam optimizer. These optimization strategies contributed to improved model generalization and high classification accuracy for soybean leaf disease diagnosis.

Results and discussion

The deep learning model meaningfully outperformed traditional machine learning models. The proposed CNN-based classifier achieved the highest accuracy of 100% with minimal training loss and robust generalization. Properly tuned hyperparameters and an augmented dataset contributed to this efficiency. The training loss was reduced nearly to zero, and validation loss approached an optimal threshold. The model demonstrated strong robustness against variations in leaf images due to data augmentation.

Classification efficiency: Table 2 shows the precision, recall, F1-score, and support for each class. In contrast, traditional machine learning approaches achieved a maximum accuracy of 44.4%, emphasizing the limitations of handcrafted feature-based methods. The feature extraction process applied to soybean leaf images, showed in the previous figures, contributed to the superior efficiency of the CNN model.

Comparison of Deep Learning Models: Different CNN architectures exhibit unique characteristics in terms of accuracy, precision, recall, fault tolerance, and computational requirements. Error handling was addressed by applying appropriate class weights. Table 3 summarizes the efficiency comparison between the proposed CNN model and other commonly applied architectures, containing MobileNetV2, ResNet50, and EfficientNetB0.

Table 2. Classification efficiency metrics for soybean disease detection model

Class	Precision	Recall	F1-Score	Support
0	0.61	0.34	0.43	59
1	0.50	0.04	0.07	27
2	0.69	0.18	0.29	49
3	0.42	0.73	0.53	55
4	0.38	0.77	0.51	44
Macro Avg	0.52	0.41	0.37	234
Weighted Avg	0.52	0.44	0.40	234

Table 3. Comparison of Deep Learning Models for soybean disease classification

Model	Accuracy	Precision	Recall	Training Time	Inference Time	Model Size
Proposed CNN	100%	1.0	1.0	41 s/epoch	0.005 s	11 M params
MobileNetV2	~95%	High	High	10 s/epoch	0.002 s	2 M params
ResNet50	98%	High	High	35 s/epoch	0.008 s	25 M params
EfficientNetB0	99%	Very High	Very High	30 s/epoch	0.003 s	5 M params

Model efficiency analysis: Figure 4 shows the training and validation accuracy of the proposed CNN model across epochs. The model achieved 100% training accuracy, denoting strong learning capability. However, such high accuracy may sometimes illustrate potential overfitting, and care should be taken to ensure generalization to unseen data.

Overall, the proposed CNN model demonstrates superior efficiency in soybean disease diagnosis compared to both traditional machine learning approaches and other popular CNN architectures. High accuracy, precision, and recall, combined with effective training and inference times, make this model highly suitable for practical agricultural applications. The outcomes demonstrate that the proposed CNN model achieves superior classification accuracy (100%) compared to traditional machine learning methods and other deep learning architectures, containing MobileNetV2, ResNet50, and EfficientNetB0. These outcomes are in line with prior investigations reporting CNNs' effectiveness in automatically capturing complex features in plant disease images, outperforming classical approaches that rely on handcrafted features. In contrast, models like MobileNetV2 propose faster training and inference times with lower computational requirements but at a slight cost in accuracy, making them suitable for resource-constrained deployment scenarios. EfficientNetB0 and ResNet50 supply strong classification efficiency;

however, they request greater computational resources, emphasizing the trade-off between accuracy and efficiency.

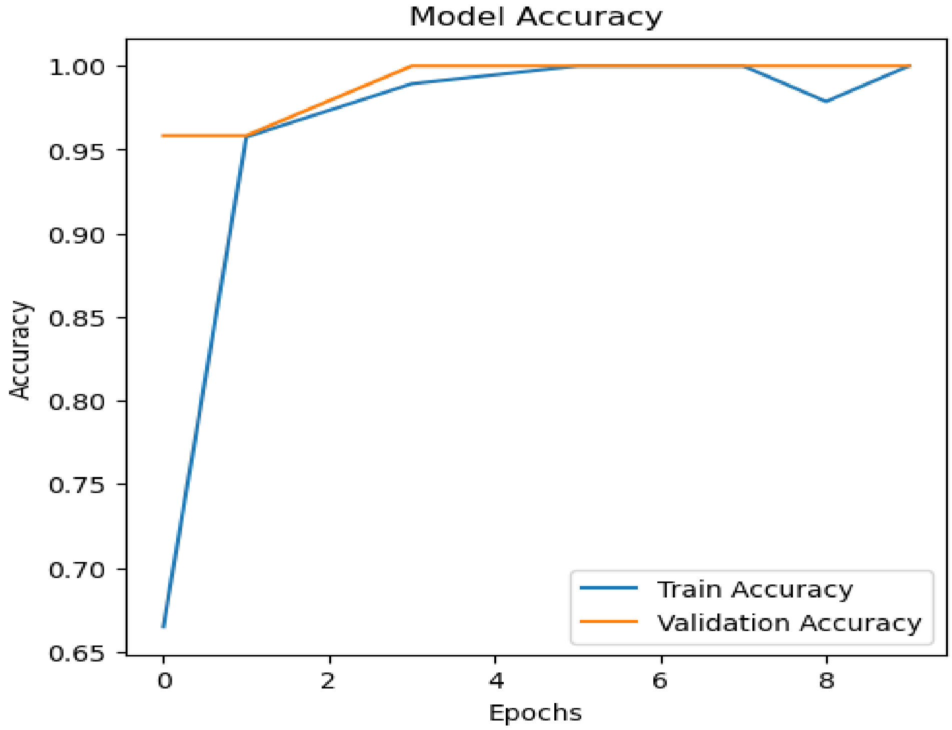

Figure 4. Training and validation accuracy of the proposed CNN model across epochs

The proposed model's innovative multi-branch feature extraction and dynamic attention mechanisms contribute to improved robustness and generalization, exclusively on restricted datasets, a common challenge noted in related study. These architectural features position our method favorably compared to similar approaches, which have announced classification accuracies ranging from 90% to 99%. In spite of the excellent efficiency, indications of overfitting propose that additional strategies, like increased data augmentation or transfer learning, could further strengthen model generalization. Future work should also attend optimizing model size and inference speed to facilitate practical, real-time deployment in field situations. Overall, these outcomes reinforce and develop the growing body of evidence supporting the effectiveness of deep learning for plant disease diagnosis, exclusively in identifying complex and subtle symptom patterns in soybean leaves. Figure 5 shows the training and validation loss over the course of model training. The loss curves supply discernment into how well the model fits the training data and generalizes to the validation set, with convergence denoting effective learning and minimal overfitting.

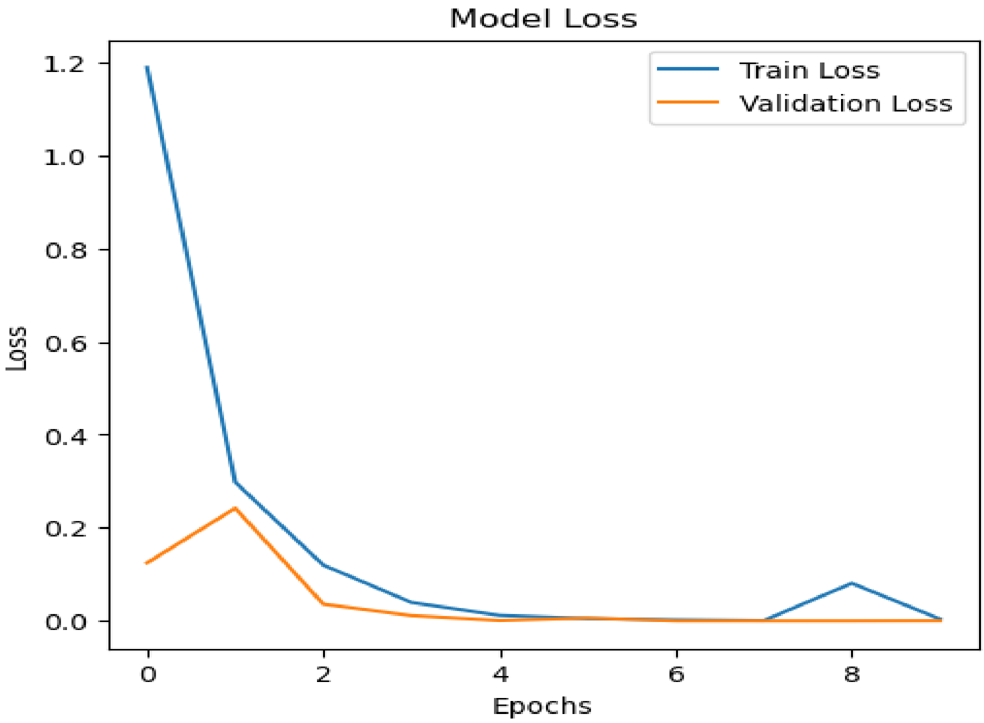

Figure 5. Training and validation loss of the proposed CNN model across epochs

The accuracy plot (top) shows that both training and validation accuracy rapidly improve, reaching near-perfect scores within a few epochs, and then stay stable. This illustrates that the model quickly and effectively learns the underlying patterns in the data.

The loss plot (bottom) shows a steep decrease in both training and validation loss, approaching zero after different epochs. This behavior proposes that the model fits the data well, with minimal classification errors. The close alignment of the training and validation curves further illustrates that there is no meaningful overfitting, demonstrating strong generalization efficiency.

Conclusions: The diagnosis and classification of soybean leaf diseases is a critical task in precision agriculture, as it directly affects crop yield and food security. Conventional manual inspection methods are time-consuming, prone to errors, and need expert intervention. Recent advancements in machine learning (ML) and deep learning (DL) have enabled automated solutions that improve disease identification with higher accuracy and efficiency. Soybean is a major agricultural crop, and timely disease diagnosis is essential to minimize yield loss. Traditional methods, however, fail to supply scalable, rapid, and reliable solutions. Although previous study has investigated numerous ML and DL techniques, challenges like insufficient datasets, poor generalization, and restricted real-time implementation stay. Addressing these limitations needs robust models capable of high accuracy and computational efficiency. In this investigation, a dataset of 1,500 processed soybean leaf images was utilized. Image preprocessing included resizing, grayscale conversion, histogram equalization, and segmentation. Feature extraction techniques, like Haralick texture, color histograms, and edge detection, were applied. Traditional classifiers, containing Support Vector Machine (SVM) and Random Forest, were trained on these extracted features, achieving a maximum accuracy of 44.44%, emphasizing the limitations of classical ML approaches. A Convolutional Neural Network (CNN) was implemented for automatic feature learning. After hyperparameter tuning and data augmentation, the CNN model achieved 100% classification accuracy with minimal validation loss (0.0001). Training time was optimized applying GPU acceleration, reducing processing delays. The model demonstrated robustness across multiple test cases, affirming its applicability in real-world scenarios. Future study will centralize on deploying the trained model in real-time field situations via mobile applications. Expanding the dataset to include diverse environmental conditions and integrating Internet of Things (IoT) sensors for real-time disease monitoring will further raise applicability and scalability. Additionally, comparative investigations involving hybrid models that combine ML, DL, and transformer-based architectures will be investigated to push classification accuracy beyond current benchmarks. Overall, this investigation demonstrates the meaningful potential of deep learning for high-accuracy, automated soybean leaf disease detection and supplies a foundation for practical precision agriculture applications.

Author contributions

Sanjay Mate: Classification approach, Renuka Amit Mane: Dataset equalization and preprocessing, Meenakshi Thalor: Feature extraction, Ankita Harshad Tidake: Hyperparameter optimization.

Data availability statement

The dataset applied in this investigation is publicly available. Original data: Jameer Kotwal, Ramgopal Kashyapa, Mohd. Shafi Pathan. *Data in Brief*, 53 (2024), 110216. Published by Elsevier Inc., 2352-3409/© 2024. This is an open-access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). The India soybean leaf dataset can be accessed at: https://data.mendeley.com/datasets/bshkvgbzpt/1 DOI: 10.1016/j.dib.2024.110216

Acknowledgment

The authors would like to express their gratitude to J. Kotwal, Mohd. Shafi Pathan, and R. Kashyapa for their efforts in collecting the soybean dataset from Tal: Barshi, Dist: Solapur, Maharashtra, India (PIN: 413406).

Ethical considerations

This article, *"Automated Soybean Leaf Disease Detection Using Convolutional Neural Networks and Machine Learning"*, represents the authors' original work and has not been published elsewhere.

Funding

This study was managed without financial support from any institution or funding agency.

Conflict of interest

The authors declare no competing interests.

References

Cheng, X., Zhang, Y., Chen, Y., Wu, Y., & Yue, Y. (2017). Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture, 141*, 351–356. https://doi.org/10.1016/j.compag.2017.08.005

Dixit, S., Kumar, A., Haripriya, A., Bohre, K., & Srinivasan, K. (2023, April). Classification and recognition of soybean leaf diseases in Madhya Pradesh and Chhattisgarh using deep learning methods. In *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)* (pp. 1–6). IEEE. https://doi.org/10.1109/PCEMS58491.2023.10136030

Fang, W., Yue, L., & Dandan, C. (2020, November). Classification system study of soybean leaf disease based on deep learning. In *2020 International Conference on Internet of Things and Intelligent Applications (ITIA)* (pp. 1–5). IEEE. https://doi.org/10.1109/ITIA50152.2020.9312252

Ghotbaldini, H., Mohammadabadi, M. R., Nezamabadi-pour, H., Babenko, O., & Amiri Ghanatsaman, Z. (2019). Predicting breeding value of body weight at 6-month age using artificial neural networks in Kermani sheep breed. *Acta Scientiarum. Animal Sciences, 41*, Article e45282. https://doi.org/10.4025/actascianimsci.v41i1.45282

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Herrera, P. J., Dorado, J., & Ribeiro, Á. (2014). A novel approach for weed type classification based on shape descriptors and a fuzzy decision-making method. *Sensors, 14*(8), 15304–15324. https://doi.org/10.3390/s140815304

Huang, Y., Song, H., Han, T., Xu, S., Wang, Z., Liu, Q., & Wang, X. (2024). A hypergraph cell membrane computing network model for soybean disease identification. *Scientific Reports, 14*(1), Article 29637. https://doi.org/10.1038/s41598-024-81325-x

Jain, E., & Sharma, P. (2024, November). Deep learning-based soybean leaf disease classification using VGG16 and data augmentation techniques. In *2024 International Conference on Cybernation and Computation (CYBERCOM)* (pp. 710–714). IEEE. https://doi.org/10.1109/CYBERCOM63683.2024.10803140

Ji, M., Liu, P., & Wu, Q. (2021). Feasibility of hybrid PSO-ANN model for identifying soybean diseases. *International Journal of Cognitive Informatics and Natural Intelligence, 15*(4), 1–16. https://doi.org/10.4018/IJCINI.290328

Kotwal, J., Kashyap, R., & Pathan, M. S. (2024). An India soybean dataset for identification and classification of diseases using computer-vision algorithms. *Data in Brief, 53*, Article 110216. https://doi.org/10.1016/j.dib.2024.110216

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90. https://doi.org/10.1145/3065386

Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv*. https://doi.org/10.48550/arXiv.1312.4400

Ma, J., Du, K., Zheng, F., Zhang, L., & Sun, Z. (2018). Disease recognition system for greenhouse cucumbers based on deep convolutional neural network. *Transactions of the Chinese Society of Agricultural Engineering, 34*(12), 186–192. https://doi.org/10.11975/j.issn.1002-6819.2018.12.022

Ma, Z., Wang, Y., Zhang, T., Wang, H., Jia, Y., Gao, R., & Su, Z. (2022). Maize leaf disease identification using deep transfer convolutional neural networks. *International Journal of Agricultural and Biological Engineering, 15*(5), 187–195. https://doi.org/10.25165/j.ijabe.20221505.6658

Mahobiya, C., Iyer, S., & Kolhe, S. (2023). Hybrid features based classification of insect and leaf disease of soybean plants using random forest classifier. *Multimedia Tools and Applications, 82*(2), 1231–1251. https://doi.org/10.1007/s11042-022-13376-x

Mohammadabadi, M., Akhtarpoor, A., Khezri, A., Babenko, O., Stavetska, R. V., Tytarenko, I., Ievstafiieva, Y., Buchkovska, V., Slynko, V., & Afanasenko, V. (2024). The role and diverse applications of machine learning in genetics, breeding, and biotechnology of livestock and poultry. *Journal of Agricultural Biotechnology, 16*(4), 413–442. https://doi.org/10.22103/jab.2025.24662.1644

Mohammadabadi, M., Kheyrodin, H., Afanasenko, V., Babenko, O., Klopenko, N., Kalashnyk, O., Ievstafiieva, Y., & Buchkovska, V. (2024). The role of artificial intelligence in genomics. *Journal of Agricultural Biotechnology, 16*(2), 195–279. https://doi.org/10.22103/jab.2024.23558.1575

Ray, A., & Ray, H. (2021). Performance analysis of machine learning classifiers on different healthcare datasets. In A. K. Manocha, S. Jain, M. Singh, & T. Paul (Eds.), *Emerging technologies in data mining and information security: Proceedings of IEMIS 2020, Volume 2* (pp. 99–111). Springer Nature Singapore. https://doi.org/10.1007/978-981-33-4367-2_11

Singh, B. P., & Shukla, S. (2021). Classification and recognition of soybean leaf disease detection using convolutional neural network (CNN). In *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1134–1138). IEEE. https://doi.org/10.1109/ICECA52323.2021.9676017

Srilakshmi, A., & Geetha, K. (2023). A novel framework for soybean leaves disease detection using DIM-U-Net and LSTM. *Multimedia Tools and Applications, 82*(18), 28323–28343. https://doi.org/10.1007/s11042-023-14775-6

Sun, H., Xu, H., Liu, B., He, D., He, J., Zhang, H., & Geng, N. (2021). MEAN-SSD: A novel real-time detector for apple leaf diseases using improved light-weight convolutional neural networks. *Computers and Electronics in Agriculture, 189*, Article 106379. https://doi.org/10.1016/j.compag.2021.106379

Tetila, E. C., Machado, B. B., Menezes, G. K., Oliveira, A. D. S., Alvarez, M., Amorim, W. P., Belete, N. A. S., Silva, G. G., & Pistori, H. (2020). Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters, 17*(5), 903–907. https://doi.org/10.1109/LGRS.2019.2932385

Thapliyal, N., Thapliyal, S., Kukreja, V., & Mehta, S. (2024, March). Disruptive tech in agriculture: Federated learning CNNs for soybean leaf disease classification. In *2024 3rd International Conference for Innovation in Technology (INOCON)* (pp. 1–6). IEEE. https://doi.org/10.1109/INOCON60754.2024.10511434

Wang, X., Pan, T., Qu, J., Sun, Y., Miao, L., Zhao, Z., Li, X., Liu, Y., & Zhu, R. (2023). Diagnosis of soybean bacterial blight progress stage based on deep learning in the context of data-deficient. *Computers and Electronics in Agriculture, 212*, Article 108170. https://doi.org/10.1016/j.compag.2023.108170

Wu, Q., Zhang, K., & Meng, J. (2019). Identification of soybean leaf diseases via deep learning. *Journal of the Institution of Engineers (India): Series A, 100*(4), 659–666.

Yu, M., Ma, X., Guan, H., & Zhang, T. (2023). A diagnosis model of soybean leaf diseases based on improved residual neural network. *Chemometrics and Intelligent Laboratory Systems, 237*, Article 104824. https://doi.org/10.1016/j.chemolab.2023.104824

Yu, M., Ma, X., Guan, H., Liu, M., & Zhang, T. (2022). A recognition method of soybean leaf diseases based on an improved deep learning model. *Frontiers in Plant Science, 13*, Article 878834. https://doi.org/10.3389/fpls.2022.878834

Zhang, D., Shao, J., Li, X., & Shen, H. T. (2021). Remote sensing image super-resolution via mixed high-order attention network. *IEEE Transactions on Geoscience and Remote Sensing, 59*(6), 5183–5196. https://doi.org/10.1109/TGRS.2020.3009918

Zhang, K., Wu, Q., Liu, A., & Meng, X. (2018). Can deep learning identify tomato leaf disease? *Advances in Multimedia, 2018*, Article 6710865. https://doi.org/10.1155/2018/6710865

انجمن بیوتکنولوژی ایران
مجله بیوتکنولوژی کشاورزی
دانشگاه شهید باهنر کرمان
شاپا چاپی: ۶۷۰۵-۲۲۲۸ شاپا الکترونیکی: ۶۵۰۰-۲۲۲۸

تشخیص خودکار بیماری‌های برگ سویا با استفاده از شبکه عصبی کانولوشنی و یادگیری ماشین

ماته سانجی

نویسنده مسئول، پلی‌تکنیک دولتی دامان، هند. ایمیل: matesanju@gmail.com

سوریوانشی رنوکا

دانشکده علوم و فناوری رایانه، دانشگاه صلح جهانی دکتر ویشوانات کاراد، پونه، هند. ایمیل: renukasurya1117@gmail.com

تالور میناکاشی

مؤسسه فناوری اطلاعات AISSMS، پونه، هند. ایمیل: meenakshi.thalor@aissmsioit.org

تیداکه آنکیتا

دانشکده مهندسی اجینکیا دی.وای. پاتیل، پونه، هند. ایمیل: ankita.tidake@gmail.com

تاریخ دریافت: ۱۴۰۴/۰۴/۰۸ تاریخ دریافت فایل اصلاح شده نهایی: ۱۴۰۴/۰۶/۱۹ تاریخ پذیرش: ۱۴۰۴/۰۶/۲۰

چکیده

هدف: کشاورزی همچنان یکی از عوامل اصلی رشد اقتصادی و امنیت غذایی در بسیاری از کشورهای در حال توسعه به شمار می‌رود و سویا یکی از مهم‌ترین محصولات نقدی در جهان محسوب می‌شود. به دلیل ارزش غذایی بالای آن، سویا به طور گسترده در رژیم‌های غذایی انسانی و خوراک دام مصرف می‌شود و استفاده از آن به عنوان جایگزین شیر نیز به دلیل ماندگاری بیشتر پس از فرآوری رو به افزایش است. با این حال، تولید سویا اغلب در معرض طیف وسیعی از بیماری‌های برگی قرار دارد که بسته به شرایط خاک، عوامل اقلیمی و الگوهای کشت متفاوت هستند. این بیماری‌ها باعث کاهش چشمگیر در عملکرد و کیفیت محصول می‌شوند و ضرورت استفاده از روش‌های تشخیص کارآمد، خودکار و دقیق را برجسته می‌سازند. هدف این مطالعه طراحی و ارزیابی یک چارچوب تشخیص بیماری برگ سویا از طریق مقایسه روش‌های سنتی یادگیری ماشین با رویکردهای یادگیری عمیق است.

مواد و روش‌ها: یک مجموعه‌داده انتخاب‌شده شامل ۱۵۰۰ تصویر برگ سویا در دسته‌های مختلف بیماری به کار گرفته شد که به نسبت ۸۰:۲۰ برای آموزش و آزمون تقسیم گردید. پیش از آموزش مدل‌ها، تصاویر تحت پردازش‌های اولیه نظیر تغییر اندازه، تبدیل به طیف خاکستری، همسان‌سازی هیستوگرام و بخش‌بندی قرار گرفتند تا استخراج ویژگی بهبود یابد. دو رویکرد پیاده‌سازی

شد: ماشین بردار پشتیبان (SVM) با ویژگی‌های استخراج‌شده به صورت دستی که نمایانگر یادگیری ماشین سنتی است، و یک شبکه عصبی کانولوشنی (CNN) سفارشی که نمایانگر یادگیری عمیق است. شبکه CNN با سه لایه کانولوشنی، لایه‌های بیشینه‌برداری (max-pooling) و لایه‌های کاملاً متصل طراحی شد و از تکنیک‌های افزایش داده، منظم‌سازی Dropout و تنظیم تطبیقی نرخ یادگیری استفاده گردید. همچنین بهینه‌سازی ابرپارامترها برای ارتقای عملکرد انجام شد.

نتایج: شبکه CNN پیشنهادی به دقت اعتبارسنجی ۱۰۰٪ دست یافت که به طور قابل توجهی بهتر از طبقه‌بند SVM با دقت ۴۴/۴۴٪ بود. اگرچه نشانه‌های جزئی بیش‌برازش مشاهده شد، اما CNN توانست قابلیت تعمیم قوی خود را در میان نمونه‌های متنوع برگ حفظ کند و قابلیت اطمینان آن در کاربردهای عملی را نشان دهد.

نتیجه‌گیری: یافته‌ها برتری آشکار یادگیری عمیق نسبت به روش‌های متداول یادگیری ماشین در شناسایی بیماری‌های گیاهی را تأیید می‌کنند. پژوهش‌های آینده بر استفاده از یادگیری انتقالی با شبکه‌های از پیش آموزش‌دیده و توسعه سامانه‌های تشخیص بلادرنگ مبتنی بر برنامه‌های تلفن همراه همراه با حسگرهای اینترنت اشیا متمرکز خواهد بود. چارچوب پیشنهادی مسیری مقیاس‌پذیر و قابل اعتماد برای کشاورزی دقیق فراهم می‌کند و مزایای قابل توجهی برای مدیریت پایدار محصولات و بهبود بهره‌وری در پی خواهد داشت.

واژگان کلیدی: بیماری برگ سویا، پیش‌پردازش تصویر، شبکه عصبی کانولوشنی (CNN)، یادگیری ماشین، کشاورزی دقیق

نوع مقاله: پژوهشی

استناد: ماته سانجی، سوریوانشی رنوکا، تالور مینکاشی، تیداکه آنکیتا (۱۴۰۴). تشخیص خودکار بیماری‌های برگ سویا با استفاده از شبکه عصبی کانولوشنی و یادگیری ماشین. *مجله بیوتکنولوژی کشاورزی*، ۱۷(۴)، ۱۵۱–۱۶۸.

Publisher: Shahid Bahonar University of Kerman & Iranian Biotechnology Society.