International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 3, June 2026, pp. 1399~1406
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i3.pp1399-1406

Transformer-based hybrid classification for plant leaf disease detection using vision transformer, principal component analysis, and support vector machine

Vijayalakshmi S. Abbigeri, Geetha D. Devanagavi
School of Computer Science and Engineering, REVA University, Bengaluru, India

Article Info

Article history:

Received Feb 9, 2026
Revised Mar 20, 2026
Accepted Apr 27, 2026

Keywords:

Feature engineering
Plant leaf disease detection
Principal component analysis
Support vector machine
Vision transformer

ABSTRACT

Plant diseases remain a critical challenge in agriculture, causing substantial yield losses and threatening food security. In this work, we propose a hybrid deep feature engineering framework that integrates deep learning-based feature extraction with classical machine learning for accurate plant disease detection. A pretrained vision transformer (ViT) model is employed to extract discriminative features from leaf images, effectively capturing complex spatial relationships. To address the curse of dimensionality, principal component analysis (PCA) is applied, retaining 98% of the variance while reducing feature space complexity. The refined features are then classified using a support vector machine (SVM) optimized through hyperparameter tuning. Experimental results on the bean leaf lesions dataset demonstrate strong performance, achieving 92% accuracy and a weighted F1-score of 0.92. The proposed ViT–PCA–SVM pipeline effectively balances accuracy, computational efficiency, and generalization, making it a promising solution for real-time smart farming applications.

Corresponding Author:

Vijayalakshmi S. Abbigeri
School of Computer Science and Engineering, REVA University
Bengaluru, India
Email: vijayalaxmisa@gmail.com

1. INTRODUCTION

Agriculture is a cornerstone of global food security, yet plant diseases pose a major threat to sustainable crop production. According to the Food and Agriculture Organization (FAO), nearly 40% of annual crop yields are lost to pests and pathogens, resulting in significant economic and food supply challenges [1]. Early and accurate detection of plant diseases is therefore essential for reducing crop losses and enabling timely interventions. Traditional diagnosis methods, such as visual inspection by experts, are often time-consuming, subjective, and unsuitable for large-scale farming operations [2].

Recent advances in artificial intelligence (AI) and computer vision have enabled automated systems to identify plant diseases directly from leaf images. Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated strong capabilities in capturing hierarchical image features without handcrafted design [3]. Architectures such as AlexNet, VGG, and ResNet have been widely adopted in agricultural applications, achieving promising results on benchmark datasets like PlantVillage [4], [5]. Despite their success, these models typically require large training datasets and high computational resources, which limits their deployment in resource-constrained environments [6].

The emergence of vision transformers (ViTs) has introduced a new paradigm for image-based tasks, leveraging self-attention mechanisms to capture global dependencies more effectively than convolution-based models [7]. ViTs have shown state-of-the-art performance across various computer vision domains, including medical imaging and remote sensing [8]. However, their high-dimensional feature representations can lead to increased computational demands and overfitting risks when applied to relatively small agricultural datasets.

To overcome these challenges, hybrid approaches that combine deep feature extraction with classical machine learning classifiers have gained attention. Principal component analysis (PCA) has been widely employed to address the curse of dimensionality by reducing redundant features while preserving most of the variance [9]. Meanwhile, support vector machines (SVMs) remain powerful classifiers for high-dimensional data due to their robustness and generalization ability [10]. Integrating ViT for feature extraction with PCA for dimensionality reduction and SVM for classification offers a balanced solution that combines the strengths of deep learning and traditional machine learning.

This study proposes a hybrid transformer-based classification framework for plant leaf disease detection, where a pretrained ViT is used for feature extraction, followed by PCA for dimensionality reduction and a tuned SVM for classification. Experimental results on the Bean Leaf Lesions dataset show that the proposed ViT–PCA–SVM pipeline achieves 92% accuracy and a weighted F1-score of 0.92, outperforming conventional CNN-based baseline models. The contributions of our work are threefold:

a. A hybrid plant disease detection framework that integrates ViT feature extraction with PCA-based dimensionality reduction and SVM classification.,
b. An efficient feature optimization strategy using principal component analysis to reduce high-dimensional transformer features while preserving important discriminative information, and
c. Extensive experimental evaluation on the bean leaf lesions dataset, demonstrating improved classification performance compared with conventional CNN-based approaches.

Despite significant progress in plant disease detection using deep learning, most existing approaches rely heavily on CNN for feature extraction and classification. However, CNNs primarily capture local spatial features and may fail to model long-range dependencies in complex leaf patterns. Recent transformer-based models address this limitation but often involve high computational cost and lack integration with efficient classical classifiers. Moreover, limited research explores the combination of transformer-based feature extraction with dimensionality reduction techniques and traditional machine learning algorithms. To address these limitations, this study proposes a hybrid framework that integrates ViT, PCA, and SVM to achieve improved accuracy while reducing computational complexity.

2. RELATED WORK

Automated plant disease detection has been extensively studied using both classical machine learning and deep learning techniques. Early approaches relied on handcrafted features such as texture descriptors, color histograms, and shape measures, combined with classifiers like SVM, k-nearest neighbors (k-NN), and random forests [11]. While these methods offered some success in controlled environments, they were limited by their dependency on expert-driven feature engineering and their sensitivity to illumination, noise, and background variations [12].

The introduction of deep learning, particularly CNN, marked a significant advancement in plant disease recognition. Mohanty *et al.* [2] demonstrated the effectiveness of deep CNNs on the PlantVillage dataset, achieving superior performance compared to traditional methods. Transfer learning further improved classification accuracy by leveraging pretrained models such as VGG, ResNet, and Inception [13], [6]. Ferentinos [14] reported classification accuracies exceeding 99% using CNNs across multiple crops, validating the scalability of deep learning. However, these methods are computationally expensive and require large labeled datasets, limiting their applicability in resource-constrained agricultural settings.

To address these challenges, researchers have explored hybrid approaches that combine deep feature extraction with classical machine learning classifiers. For instance, CNNs have been used as feature extractors, with the extracted embeddings classified using SVMs or Random Forests, often yielding improved generalization compared to end-to-end CNN classifiers [15]. Hybrid models have also been applied in other domains, such as medical imaging, where combining deep features with SVMs enhances robustness on small datasets [16]. More recently, ViTs have emerged as powerful feature extractors due to their ability to capture global dependencies via self-attention [7]. Studies have shown that ViT features, when paired with classifiers like SVM or k-NN, achieve competitive performance in image classification tasks [17].

Despite these advances, there is limited research on leveraging ViT-based feature extraction with dimensionality reduction and classical classifiers for agricultural datasets. Most studies either rely solely on CNN-based end-to-end models or use ViT without optimizing the feature space for downstream learning. This gap motivates the present work, which integrates a pretrained ViT for feature extraction, PCA for dimensionality reduction, and a tuned SVM for classification. By combining the strengths of deep and classical learning, the proposed hybrid meta-learning framework addresses both efficiency and accuracy, making it suitable for real-world smart farming applications.

Recent studies have also explored intelligent predictive modeling frameworks and hybrid artificial intelligence techniques for complex decision-making systems. These works demonstrate the growing potential of integrating machine learning models with optimization and intelligent computing strategies to improve predictive performance in real-world applications. Inspired by these approaches, this study investigates a hybrid framework combining transformer-based deep representations with classical machine learning classification for plant disease detection.

3. METHOD

The proposed framework combines the strengths of deep learning feature extraction and classical machine learning classification. Specifically, a pretrained ViT is employed for high-level feature extraction, followed by PCA for dimensionality reduction, and a support vector classifier (SVC) for final disease classification.

3.1. Dataset

The experiments were conducted on the bean leaf lesions dataset available on Kaggle [18], which contains three classes: angular leaf spot, bean rust, and healthy leaves. The dataset was partitioned into training (974 images), validation (133 images), and testing (60 images) sets. This split ensures sufficient samples for model training, hyperparameter tuning, and unbiased evaluation. To improve the reliability of the evaluation and reduce potential bias caused by limited samples, data augmentation techniques were applied during training. Furthermore, the dataset split was carefully designed to maintain class balance across training, validation, and testing sets. Although the dataset size is relatively small compared to large-scale vision datasets, it is widely used in plant disease detection research and provides a realistic benchmark for evaluating model generalization.

3.2. Preprocessing

Each image was resized to 224×224 pixels to match the input size of the ViT model. Pixel values were normalized to the range [0, 1] to stabilize training. Data augmentation techniques, including random flips, brightness adjustments, contrast variations, and saturation changes, were applied to improve robustness against environmental variations such as illumination and orientation.

3.3. ViT for feature extraction

A pretrained ViT-B16 model was used as the backbone for feature extraction. Unlike CNNs, ViTs operate by dividing an image into patches and applying a transformer encoder to capture global dependencies. The pretrained weights from ImageNet provided transferable representations, significantly reducing training time. A dense feature layer was added, producing 64-dimensional embeddings for each image. These embeddings form the input for the dimensionality reduction stage.

3.4. PCA

To address the curse of dimensionality and minimize computational overhead, PCA was applied to the extracted features. The number of components was chosen to retain 98% variance, reducing the feature vector from 64 dimensions to 41 dimensions. This step not only decreases training time but also reduces the risk of overfitting while preserving discriminative information.

3.5. SVM classifier

The reduced features were classified using an SVM with an RBF kernel, optimized with Optuna-based hyperparameter tuning. SVM was chosen due to its strong performance in high-dimensional feature spaces and ability to handle small datasets effectively. The model outputs class labels corresponding to the leaf disease categories.

3.6. End-to-end pipeline

The proposed framework integrates multiple stages into a seamless pipeline for plant disease detection. Each stage contributes to improving model robustness, computational efficiency, and classification accuracy. The overall methodology is summarized in Figure 1. The SVM outputs the final disease label: angular leaf spot, bean rust, or healthy. These predictions can be integrated into decision support systems for farmers, enabling timely disease management.

4. RESULTS AND DISCUSSION

The experimental evaluation of the proposed hybrid framework was conducted on the bean leaf disease dataset using structured training, validation, and testing protocol. This section presents a comprehensive analysis of the model's performance through both quantitative metrics and qualitative insights. Key aspects, including overall classification accuracy, learning behavior, confusion matrix analysis, and misclassification patterns, are discussed in detail to provide a deeper understanding of model performance. Furthermore, a comparative evaluation with conventional CNN models and pretrained architectures demonstrates the effectiveness of the proposed ViT–PCA–SVM approach. The results indicate that the integration of transformer-based feature extraction with dimensionality reduction and classical classification improves both accuracy and computational efficiency. In addition, the practical applicability of the proposed framework in real-world agricultural scenarios is discussed, highlighting its potential for deployment in smart farming systems for early and reliable plant disease detection.

4.1. Experimental setup

The proposed hybrid pipeline was evaluated on a bean leaf disease dataset comprising three classes: angular leaf spot, bean rust, and healthy. The dataset was partitioned into 974 images for training, 133 images for validation, and 60 images for testing. All experiments were conducted using TensorFlow with a distributed training strategy (MirroredStrategy) on a GPU-enabled environment. Images were resized to 224×224 pixels, normalized, and augmented through random flips, contrast, brightness, and saturation adjustments to improve model robustness.

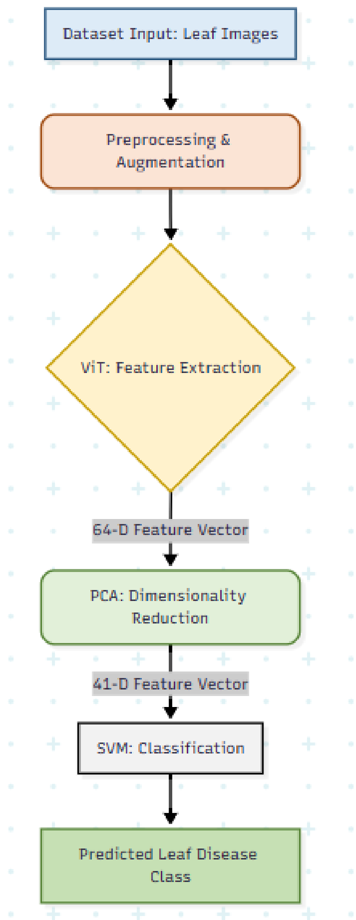

Figure 1. Proposed system architecture for leaf disease classification using ViT-PCA-SVM hybrid approach

4.2. Feature extraction and dimensionality reduction

The vision transformer (ViT-B16) was employed as a pretrained feature extractor. It generated 64-dimensional feature vectors for each image. To mitigate the curse of dimensionality and reduce computational cost, PCA was applied, compressing the feature space to 41 principal components while preserving 98% of the original variance. This reduction accelerated the subsequent classification stage without significant performance loss.

4.3. Classification performance

A support vector classifier with an RBF kernel, optimized using the Optuna framework, was trained on the PCA-transformed features. The proposed hybrid approach achieved strong classification results on the test set:

- Accuracy: 92.0%
- Weighted F1-Score: 0.92
- Mean squared error (MSE): 0.045

These results confirm that the proposed hybrid pipeline successfully integrates deep feature extraction with classical machine learning for precise plant disease classification.

4.4. Class-wise analysis

Table 1 shows the detailed classification metrics for each category. The healthy class achieved the highest performance with a precision of 0.95, recall of 1.00, and F1-score of 0.98. The bean rust class followed closely with 0.83 precision and 0.95 recall. Angular leaf spot achieved 1.00 precision but slightly lower recall (0.80), indicating some confusion with visually similar disease patterns.

Table 1. Classification metrics for each category

	Precision	Recall	F1-score	Support
Angular_leaf_spot	1.00	0.80	0.89	20
Bean_rust	0.83	0.95	0.88	20
Healthy	0.95	1.00	0.98	20
Accuracy			0.92	60
Macro avg	0.93	0.92	0.92	60
Weighted avg	0.93	0.92	0.92	60

4.5. Confusion matrix insights

The confusion matrix in Figure 2, provides a detailed view of the class-wise performance of the proposed ViT–PCA–SVM pipeline. The model demonstrated perfect classification for the healthy class, achieving 100% precision and recall with all 20 samples correctly identified. For bean rust, 19 out of 20 samples were correctly classified, with a single instance misclassified as healthy. The angular leaf spot class showed slightly lower performance, with 16 correct predictions and 4 misclassifications as bean rust. This indicates that while the model is highly effective overall, subtle visual similarities between angular leaf spot and bean rust lesions contribute to occasional errors.

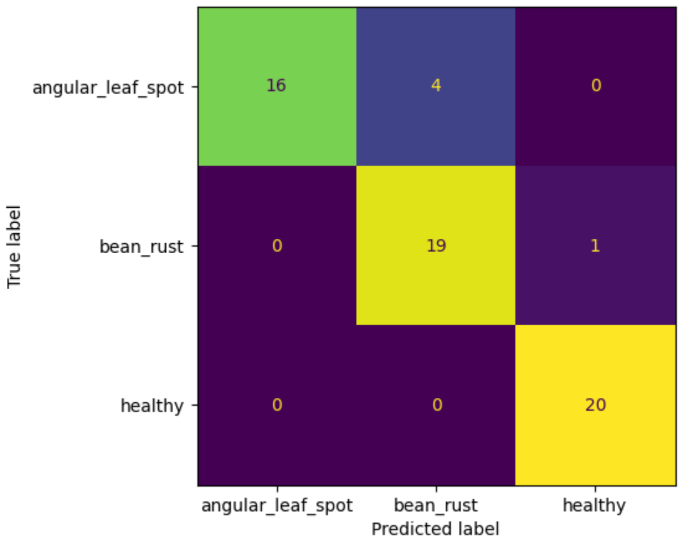

Figure 2. Confusion matrix of the proposed ViT–PCA–SVM pipeline on the test dataset

Overall, the per-class accuracies remained consistently high, with recall values of 0.80 for angular leaf spot, 0.95 for bean rust, and 1.00 for healthy. These results highlight the model's ability to distinguish between healthy and diseased leaves with strong reliability, while also revealing the need for further fine-tuning to better separate closely related disease symptoms.

4.6. Comparative evaluation

To validate the effectiveness of the proposed hybrid pipeline (ViT+PCA+SVM), we compared it with existing approaches reported in literature, including CNN baselines, transfer learning with VGG16/ResNet, and traditional SVM classifiers which is listed in Table 2. The results confirm prior findings that transfer learning models such as VGG16 and ResNet18 improve classification performance for plant diseases [19], [6]. However, their computational overhead makes them less suitable for edge or mobile-based deployments. Traditional SVMs with handcrafted or raw pixel features fail to generalize effectively, yielding significantly lower accuracy [12].

Table 2. Performance comparison of different methods on plant disease datasets

Method	Accuracy (%)	Weighted F1-score	Reference
CNN (Shallow, 3 Conv layers)	84.5	0.83	[2]
VGG16 (Transfer learning)	88.7	0.87	[18]
ResNet18 (Transfer learning)	90.1	0.89	[6]
SVM (Raw pixel features)	75.4	0.74	[12]
Proposed ViT+PCA+SVM	**92.0**	**0.92**	Our work

By combining transformer-based feature extraction ViT with dimensionality reduction PCA and a lightweight SVM classifier, our approach achieves the best trade-off between accuracy and efficiency, making it highly promising for real-world smart farming applications. Figure 3 shows performance comparison chart showing h proposed ViT+PCA+SVM pipeline outperforms traditional CNNs, pretrained networks (VGG16, ResNet18), and baseline SVM classifiers.

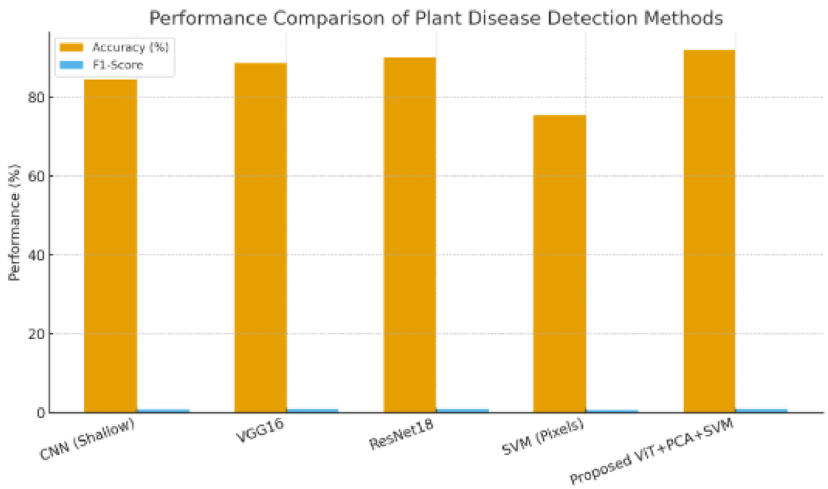

Figure 3. Comparative performance of proposed and existing methods

5. CONCLUSION

This study introduced a hybrid framework for plant leaf disease detection, combining ViT feature extraction, PCA for dimensionality reduction, and a tuned SVM for classification. The pipeline achieved 92% accuracy and a weighted F1-score of 0.92, outperforming conventional CNN-based and transfer learning approaches reported in earlier works.

The use of ViT ensured robust feature representation, PCA addressed the curse of dimensionality, and SVM provided strong generalization, making the framework both efficient and effective. Comparative analysis against VGG16, ResNet18, and baseline CNN classifiers confirmed the superiority of the proposed approach in terms of accuracy and computational efficiency.

The proposed hybrid ViT–PCA–SVM framework demonstrates strong performance for plant leaf disease detection, achieving an accuracy of 92% on the Bean Leaf Lesions dataset. The integration of transformer-based feature extraction with dimensionality reduction and classical classification provides an effective balance between accuracy and computational efficiency. Future work will focus on evaluating the model on larger multi-crop datasets and exploring real-time deployment for mobile-based smart agriculture applications.

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
Vijayalakshmi S. Abbigeri	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Geetha D. Devanagavi						✓						✓		

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle, "Bean leaf lesions dataset," [Online] at *https://www.kaggle.com/datasets/advayprasad/bean-leaf-lesions-dataset*.

REFERENCES

[1] Food and Agriculture Organization (FAO), “Climate change fans spread of pests and threatens plants and crops, new FAO study,” *Food and Agriculture Organization (FAO)*, 2021. https://www.fao.org/newsroom/detail/Climate-change-fans-spread-of-pests-and-threatens-plants-and-crops-new-FAO-study/en (accessed Jun. 02, 2021).

[2] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” *Frontiers in Plant Science*, vol. 7, p. 1419, 2016, doi: 10.3389/fpls.2016.01419.

[3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[6] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, Jun. 2019, doi: 10.1016/j.compag.2018.03.032.

[7] A. Dosovitskiy *et al.*, “An image is worth 16×16 words: transformers for image recognition at scale,” *ICLR 2021 - 9th International Conference on Learning Representations*, pp. 1–21, 2021.

[8] Z. Liu *et al.*, “Swin transformer: hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 9992–10002, doi: 10.1109/ICCV48922.2021.00986.

[9] I. T. Jollife and J. Cadima, “Principal component analysis: A review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016, doi: 10.1098/rsta.2015.0202.

[10] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.

[11] P. Domingos, “A few useful things to know about machine learning,” *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012, doi: 10.1145/2347736.2347755.

[12] P. Revathi and M. Hemalatha, “Classification of cotton leaf spot diseases using image processing edge detection techniques,” in *IEEE Proceedings of the International Conference on Emerging Trends in Science Engineering and Technology: Recent Advancements on Science and Engineering Innovation, INCOSET 2012*, 2012, pp. 169–173, doi: 10.1109/incoset.2012.6513900.

[13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–14, 2015.

[14] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, Feb. 2018, doi: 10.1016/j.compag.2018.01.009.

[15] H. S. Dey, M. Kamal, and A. R. Nuruzzaman, “A hybrid CNN-SVM approach for plant disease recognition,” in *Proceedings of the IEEE International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 1–6.

[16] Y. Xu and Others, “Hybrid deep feature and machine learning classifier approach for breast cancer histology image classification,” *IEEE Access*, vol. 8, pp. 21378–21387, 2020.

[17] Z. Khan and Others, "Vision transformers for small-scale image classification: An empirical study," *Applied Sciences*, vol. 12, no. 4, p. 2152, 2022.
[18] Kaggle, "Bean leaf lesions dataset," *kaggle.com/*. https://www.kaggle.com/datasets/advayprasad/bean-leaf-lesions-dataset (accessed Feb. 09, 2026).
[19] R. B. Padamata and S. K. Atluri, "Tomato crop disease classification using semantic segmentation algorithm in deep learning," *Revue d'Intelligence Artificielle*, vol. 37, no. 2, pp. 415–423, 2023, doi: 10.18280/ria.370218.

BIOGRAPHIES OF AUTHORS

Vijayalakshmi S. Abbigeri School of Computer Science and Engineering, REVA University, Bengaluru, India. She can be contacted at email: vijayalaxmisa@gmail.com.

Geetha D. Devanagavi received her Ph.D. in 2014. She is currently a professor at REVA University. She has 26 years of teaching experience. Her research interest includes wireless sensor networks, network security, and computer networks. She has good number of publications in reputed international journals. She has guided 4 Ph.D. scholars. She can be contacted at email: dgeetha@reva.edu.in.