BMC Plant Biology

RESEARCH

Open Access

Potato plant disease detection: leveraging hybrid deep learning models

Jackson Herbert Sinamenye[1*], Ayan Chatterjee[2] and Raju Shrestha[1]

Abstract

Agriculture, a crucial sector for global economic development and sustainable food production, faces significant challenges in detecting and managing crop diseases. These diseases can greatly impact yield and productivity, making early and accurate detection vital, especially in staple crops like potatoes. Traditional manual methods, as well as some existing machine learning and deep learning techniques, often lack accuracy and generalizability due to factors such as variability in real-world conditions. This study proposes a novel approach to improve potato plant disease detection and identification using a hybrid deep-learning model, EfficientNetV2B3+ViT. This model combines the strengths of a Convolutional Neural Network - EfficientNetV2B3 and a Vision Transformer (ViT). It has been trained on a diverse potato leaf image dataset, the "Potato Leaf Disease Dataset", which reflects real-world agricultural conditions. The proposed model achieved an accuracy of 85.06%, representing an 11.43% improvement over the results of the previous study. These results highlight the effectiveness of the hybrid model in complex agricultural settings and its potential to improve potato plant disease detection and identification.

Keywords Potato plant disease, Diverse dataset, Hybrid model, Deep learning, EfficientNet, Convolutional Neural Network (CNN), Vision Transformer (ViT)

Introduction

Agriculture is a fundamental part of human society, providing food and significantly contributing to the global economy [1–3]. It is particularly crucial in developing countries, where it is often the major source of income and employment [4, 5]. However, it faces numerous challenges, including climate change, soil degradation, and diseases, which threaten its productivity and sustainability [1, 2, 6]. Among various crops, potatoes hold significant importance as a staple food for millions of people around the globe. However, they are susceptible to a wide array of diseases caused by various pathogens, including bacteria, viruses, and fungi [7]. These diseases can destroy entire fields, leading to substantial financial losses for farmers and wider economic and food supply challenges, potentially worsening food insecurity and affecting the livelihoods of those in agriculture [8].

Traditional methods for disease diagnosis heavily rely on human observation and manual tests, which are prone to errors and inconsistencies. Moreover, they can be time-consuming and impractical for large-scale operations [9–11]. The potential for human error is high given the complex nature of plant diseases and hardly distinguishable differences in their symptoms. For example, early blight and late blight, though caused by different pathogens, present similar symptoms in their early stages, making them difficult to distinguish without expert knowledge or laboratory testing [5, 12]. Furthermore, achieving consistent and accurate diagnoses across vast agricultural landscapes is a significant challenge given the variability in disease manifestations and environmental conditions. Factors such as temperature,

*Correspondence:
Jackson Herbert Sinamenye
jacksonherberts@gmail.com
[1] Department of Computer Science, Oslo Metropolitan University (OsloMet), Oslo, Norway
[2] Department of Digital Technology, STIFTELSEN NILU, Kjeller, Norway

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

humidity, and soil composition can influence the development and appearance of diseases, adding another layer of complexity to disease identification [9–11]. The limitations of traditional methods for disease diagnosis and classification underscore the need for more advanced, scalable, and accurate solutions. This is where artificial intelligence (AI), particularly advanced deep learning techniques, becomes highly relevant. Deep learning models can automatically learn hidden features from large datasets, recognize complex patterns, and make accurate predictions, making them well-suited for disease detection and identification. However, effectiveness of these models is contingent upon the quality and diversity of the training data. Recent studies have demonstrated the potential of AI in agriculture, particularly in plant disease classification using computer vision. Deep learning models, specifically Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), have shown significant promise in this area [4, 5, 13–15]. Recent studies and reviews further highlight the significant advancements in the application of deep learning for plant disease management [16]. However, most of these studies have been limited to a narrow range of diseases and have utilized datasets collected under conditions which do not fully represent the variability and complexity or real-world agricultural settings.

The primary objective of this research is to enhance the prediction accuracy of potato diseases using plant image-based analysis, which contributes to improved management of potato crop health and advances the application of AI in the agricultural domain. The main contributions of this study are –

- Design and development of a novel and effective hybrid deep learning model, EfficientNetV2B3+ViT, that significantly enhances the accuracy of potato plant disease detection and identification.
- Use of a diverse dataset, which covers a wide range of diseases and provides a novel and challenging context for training and testing the proposed model.
- Establishment of new benchmarks for improved potato plant disease detection on a diverse image dataset through rigorous testing and optimization.

Unlike previous approaches that primarily focus on either CNNs or Vision Transformers independently, our hybrid model specifically addresses the challenges of detecting potato diseases in diverse field conditions. Our EfficientNetV2B3+ViT model differs from existing hybrid approaches in two key aspects; (1) We employ a unique feature fusion strategy that preserves both local texture features from CNN and global contextual information from ViT, rather than using a sequential architecture; and (2) We validate our approach on a challenging real-world dataset that contains diverse illumination conditions, backgrounds, and disease manifestations, unlike the controlled environments used in most previous studies. Recent works such as those by Dai et al., who developed PPLC-Net [17] and DFN-PSAN [18], focus primarily on attention mechanisms with CNN architectures, while our approach leverages the complementary strengths of fundamentally different architectural paradigms. Similarly, while Pal and Kumar [19] explored segmentation-based approaches, and Dai et al. [20] investigated cross-modal fusion, our method focuses specifically on the challenges of potato plant disease detection under variable field conditions through a novel CNN-Transformer fusion framework.

The remainder of this paper is structured as follows. "Related works" section provides a comprehensive review of existing literature concerning the utilization of AI in agriculture, focusing specifically on potato plant disease detection and identification. We examine the evolution of methodologies from traditional machine learning approaches to deep learning models and more recently, advancements in vision transformers and hybrid models. "Materials and methods" section describes the materials and methods used in this research including model selection, dataset used, data preprocessing, model training and evaluation. "The proposed model" section offers a detailed explanation of our proposed hybrid model, EfficientNetV2B3+ViT. "Implementation and experimentation" section outlines the implementation and experimentation. "Results" section presents the results, while "Discussion" section provides a discussion. Finally, "Conclusion" section summarizes the findings and proposes avenues for future research.

Related works

Recent advancements in AI have demonstrated considerable potential in the field of plant disease image classification. Researchers have utilized a spectrum of models ranging from conventional machine learning approaches to modern deep learning architectures and combinations of CNNs and vision transformers. These models have proven effective in accurately categorizing various plant diseases using image datasets. For instance, in study [21], researchers applied various traditional machine learning models, including Naive Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF), to classify diseases in maize crops based on image data whereby RF achieved the highest accuracy of 79.23% as compared to the rest of the classification techniques.

As the research progressed, deep learning models, particularly CNNs, gained significant prominence. For instance, a study in [22] applied a CNN-based method for plant disease detection. A simulation analysis conducted on sample images, evaluating time complexity and the area of the infected region using image processing techniques, achieved an accuracy of 88.8%. Further advancements in the field have involved the application of pre-trained models such as VGG, ResNet, and DenseNet for plant disease image classification. KP and Anitha [23] utilized these pre-trained models and achieved the highest accuracy of 98.27% with the DenseNet model, on the "Plant Village" image dataset for grape plant diseases. The success of the DenseNet model was attributed to its ability to identify features across a complex spectrum and its robust performance irrespective of dataset size. Similarly, Hong et al. [24] implemented different CNN architectures including ResNet50, Xception, MobileNet, ShuffleNet, and Densenet121_Xception for tomato leaf disease image classification. They utilized transfer learning to reduce the training time and computational costs. Among the tested architectures, Densenet_Xception achieved the highest recognition accuracy of 97.10%. Moreover, Habiba and Islam [25] applied the VGG16 deep CNN classifier to recognize unhealthy tomato plants and classify specific diseases. Using the "Plant Village" image dataset, the study achieved a classification performance of about 95.5% top-1 accuracy and 99% top-2 accuracy. These studies collectively highlight the effectiveness of deep learning models, particularly CNN pre-trained models in classifying plant diseases setting a strong precedent for further research in this area.

Recently, there has been a growing interest in the application of deep learning for plant disease classification across various crops. The classification of plant diseases has emerged as a significant research area in the agricultural industry, with deep learning and image-processing techniques proving particularly effective [26]. Convolutional Neural Networks (CNNs) have been tremendously helpful in the diagnosis and classification of plant diseases [16], enabling faster, more accurate, and less human-dependent diagnosis [27]. For example, recent work has shown high accuracy in detecting diseases in tomato [16], grape [28], apple [27], and sugarcane leaves [26]. These studies highlight the growing importance and success of deep learning applications in managing plant health across diverse agricultural contexts.

Building on these CNN-based approaches, ViTs have also shown promise for plant disease image classification. For instance, Borhani et al. [29] proposed a lightweight deep learning approach based on ViT for real-time automated plant disease classification. Similarly, Parez et al. [30] introduced a fine-tuned technique, called GreenViT, for detecting plant infections and diseases based on ViTs. The results obtained were 100%, 98% and 99% on Plant Village, Data Repository of Leaf images and merged dataset respectively, outperforming state-of-the-art CNN models for detecting plant diseases demonstrating the potential of ViTs in this field. In another study, Thai et al. [31] introduced FormerLeaf, a transformer-based leaf disease detection model along with two optimization methods to enhance the model performance. The study led to a reduction of the model size by 28%, decrease in training and inference time by 10%, and outperformed the state-of-the-art models in most cases.

In addition to ViTs, researchers have explored hybrid models that combine ViTs with other deep-learning models, such as CNNs. Thakur et al. [32] proposed a hybrid model that combines a ViT with a CNN but is still regarded as lightweight, with only 0.85 million trainable parameters, making it suitable for IoT-based agriculture systems. The model achieved an accuracy of 98.86% on the Plant Village dataset which comprised 38 plant categories, and 89.24% on Embrapa which comprised 93 classes. Similarly, Li et al. [33] proposed a computationally efficient deep learning architecture based on the mobile vision transformer for real-time detection of plant diseases. The model was designed to be highly accurate and low-cost, making it suitable for deployment on mobile devices with limited computational power. It was tested on three datasets namely; wheat, coffee and rice achieving an accuracy of 93.6%, 85.4% and 93.1% respectively. In a different study, Yu et al. [34] suggested a new deep learning-based framework based on inception convolution and vision transformer, called ICVT, which could automatically find plant diseases. The ICVT architecture not only models local spatial information but also focuses on the learning of high-level information for plant disease identification. According to the reported results, the network achieved an average accuracy of 77.54%, 86.89%, 99.94%, and 99.22%, on the PlantDoc, AI2018, PlantVillage, and ibean respectively. Furthermore, Thakur et al. [35] proposed a vision transformer convolutional neural network model, called "PlantXViT", for plant disease identification. The model combined the capabilities of traditional CNN with a vision transformer to efficiently identify a large number of plant diseases for several crops including apple, maize, rice, and plants in Plant Village and Embrapa datasets.

In the context of potato farming, several studies have explored the use of AI models for identifying diseases. A common theme across these studies is the use of the Plant Village dataset, which contains images of potato leaves affected by early and late blight plus images of healthy leaves. For instance, Iqbal and Talukder [36] proposed an image processing and machine

learning-based system for identifying and classifying potato leaf diseases, achieving an accuracy of 97% with the Random Forest classifier. Building on this, Arshad et al. [14] developed a hybrid deep learning model, PLDPNet, which combined deep features from VGG19 and Inception-V3 and used vision transformers for final prediction. The model achieved an overall accuracy of 98.66% and F1-score of 96.33%. However, this study acknowledged the limitations of the available labelled potato diseases dataset, which restricted their ability to train the model comprehensively. Shaheed et al. [37] developed an advanced system, called EfficientRMT-Net, which integrated ViT and ResNet50 architectures. The model achieved an accuracy of 97.65% on a general image dataset and 99.12% on specialized potato leaf images from the Plant Village dataset. The study [38] used pre-trained models including InceptionV3, VGG16, and VGG19 for feature extraction to detect leaf diseases. Among these models, VGG19 in combination with logistic regression achieved the highest classification accuracy of 97.8% on the test dataset. Furthermore, Mahum et al. [13] proposed an improved deep-learning model, the DenseNet201, to classify leaf diseases into five classes. The model, trained on the PlantVillage dataset and additional data that was manually gathered, achieved an accuracy of 97.2% on the testing set. Lastly, Chakraborty et al. [39] explored deep learning models for the automated recognition of late and early blight diseases. The VGG16 model outperformed the others (VGG19, MobileNet, and ResNet50) with an initial accuracy of 92.69%. After fine-tuning, the accuracy improved to 97.89%.

Table 1 consolidates and summarizes the major studies, outlining the models used, applications, datasets employed, and performance results.

Despite significant advancements in the field of plant disease detection and identification using deep learning models, the scope of these studies has often been limited due to the constraints of the available image datasets. The recent release of a "Potato Leaf Disease" image dataset by Shabrina et al. [40], which presents a more complex and realistic representation of agricultural conditions represents a significant step forward. Their preliminary study on this dataset using various deep learning models has shown promising results; however, there is room for improvement. This research seeks to improve the performance of deep-learning models for potato plant disease detection and identification using this diverse dataset. By leveraging advanced techniques, we have designed, developed, and validated a hybrid deep-learning model that can detect and identify diverse images of potato diseases as captured under real-world agricultural conditions more accurately.

Materials and methods

This section details the materials and methods employed in this study, which includes the selection of models, datasets, data pre-processing, training, and evaluation of the proposed model.

Model selection and design

Deep learning models, including CNNs and ViTs, have demonstrated significant effectiveness across various plant disease detection tasks, as extensively documented in recent systematic reviews [41]. To address the detection and identification of potato diseases, we employ advanced deep learning techniques, specifically focusing on feature extraction and fine-tuning methods. We propose a hybrid model designed to leverage the strengths of both CNN and ViT architectures. The choice of EfficientNetV2B3 and ViT as the backbone models in our hybrid approach was based on several key considerations.

EfficientNetV2B3

EfficientNetV2B3 is a highly efficient CNN architecture that has demonstrated strong performance on a wide range of image classification tasks [42]. Its lightweight design and excellent accuracy-to-parameter ratio make it well suited for resource constrained settings, such as mobile devices or edge computing. Furthermore, previous studies have shown the effectiveness of EfficientNet models for plant disease classification, particularly on the PlantVillage dataset [43]. These factors, along with its superior performance compared to other CNN models on a similar task [40], led us to select EfficientNetV2B3 as the CNN component of our hybrid model. The architecture of EfficientNetV2B3, like other EfficientNetV2 models, is based on CNNs [44], renowned for their capability to effectively capture local features in images. This is accomplished through convolutional layers that apply a series of filters to the input, enabling the model to learn a rich hierarchy of features at varying levels of abstraction. These features range from simple patterns such as edges and textures in the lower layers to more complex, high-level object parts in the deeper layers. Additionally, EfficientNetV2B3 incorporates several design optimizations that enhance its efficiency and performance. These include the use of depthwise separable convolutions [45], which reduce the computational complexity of the model and skip connections [46], which improve the flow of gradients during training and enable the model to learn more complex functions.

ViT

ViT, on the other hand, is a recently proposed transformer-based architecture that has achieved

Table 1 Summary of related studies. The accuracy values in the last column correspond to the datasets listed in the study and in the order they are given in the previous column

Ref	Models	Application	Dataset	Performance (accuracy %)
[21]	NB, DT, KNN, SVM, RF	Disease classification in maize	PlantVillage dataset	79.23 % (Random Forest)
[22]	CNN-based method	Plant disease detection	PlantVillage	88.80%
[23]	Pre-trained models (VGG, ResNet, DenseNet)	Plant disease classification	Plant Village	98.27% (DenseNet)
[24]	ResNet50, Xception, MobileNet, ShuffleNet, Densenet121_Xception	Tomato leaf disease classification	PlantVillage	97.10% (Densenet_Xception)
[25]	VGG16	Tomato plant disease classification	Plant Village	95.50%
[26]	EfficientNet-b0 through EfficientNet-b7, EfficientNetv2-small, EfficientNetv2-medium, EfficientNetv2-large, ResNetv2-50, and InceptionV4	Disease detection in sugarcane leaves	Sugarcane Leaf Dataset	EfficientNet-b6 (93.39%)
[28]	14 CNN and 17 vision transformer models	Classification of grape leaves and diagnosis of grape diseases	PlantVillage and Grapevine datasets	CNN + ViT (Swinv2-Base) (100%)
[27]	ResNet50, InceptionV4, Xception, DenseNet121, EfficientNetV2_m, and VGG13	Classification of apple diseases (on leaves)	PlantVillage dataset	EfficientNetV2_m (100%)
[16]	Res2 Next50, Res2 Net50 d, VGG16, and DenseNet121	Detecting diseases in tomato leaves	Small dataset with 13,875 tomato images	Res2 Next50 (99.85%)
[29]	ViT, hybrid of CNN and ViT	Real-time automated plant disease classification	Wheat Rust, Rice Leaf Disease and Plant Village	Balance between accuracy and prediction speed
[30]	ViT (GreenViT)	Plant disease detection	Plant Village, Data Repository of Leaf Images and a merged dataset	100.00%, 98.00% and 99.00% respectively
[31]	ViT (FormerLeaf)	Cassava leaf disease detection	Cassava leaf disease dataset	Reduce model size by 28.00% and decrease inference speed by 10.00%
[32]	Hybrid model (ViT + CNN)	Plant disease detection	Plant Village and Embrapa	Accuracy of 98.86% and 89.24% respectively
[33]	ViT (PMVT)	Real-time detection of plant diseases	wheat, coffee, and rice	93.60%, 85.40% and 93.10% respectively
[34]	Inception Convolutional ViT	Automatic plant disease identification	PlantDoc, AI2018, PlantVillage, ibean	77.54%, 86.89%, 99.94%, and 99.22%
[35]	ViT enabled CNN (PlantXViT)	Plant disease identification	Apple, Embrapa, Maize, PlantVillage, and Rice	93.55%, 89.24%, 92.59%, 98.86%, and 98.33%
[36]	Image processing and machine learning-based system	Potato leaf disease identification and classification	PlantVillage	97.00% (Random Forest)
[14]	PLDPNet (VGG19 + Inception-V3 + ViT)	Potato leaf disease classification	Plant Village	98.66% accuracy, 96.33% F1-score
[37]	EfficientRMT-Net (ViT + ResNet50)	Potato leaf disease classification	Plant Village (General, specialized)	97.65% (general), 99.12%(specialized)
[38]	InceptionV3, VGG16 and VGG19	Potato leaf disease detection	Plant Village	97.80% (VGG19 + logistic regression)
[13]	DenseNet201	Potato leaf disease classification	Plant Village, additional data	97.20%
[39]	VGG 16, VGG 19, MobileNet and ResNet50	Late and early blight diseases recognition in potato crops	Plant Village	97.89% (VGG 16 after fine-tuning)

state-of-the-art results on various vision tasks [47]. Unlike CNNs, ViT can capture long-range dependencies and global context in images, which may be particularly relevant for plant disease detection, where symptoms can manifest across different parts of the leaf or plant. The self-attention mechanism in ViT allows it to focus on the most informative regions of the image, potentially enhancing its discriminative power. Considering the size of our dataset, we opted for the ViT base model as it offers a better balance between efficiency and performance compared to other more complex ViT variants [48].

In the ViT base model, an image is initially segmented into a grid of non-overlapping patches, each measuring 16x16 pixels. Each patch is then linearly transformed into a one-dimensional vector by flattening and multiplying by a learnable matrix. This sequence of vectors is subsequently fed into a standard transformer encoder. The transformer encoder captures the global context of the image by modeling long-range dependencies between these patches. This is facilitated by the self-attention mechanism of the transformer, which enables each patch to attend to all other patches in a weighted manner. The weights are determined by the similarity between patches, allowing the model to emphasize the most relevant parts of the image. The output of the transformer encoder is a sequence of vectors, each representing a patch of the image enriched with contextual information from all other patches. The vector corresponding to the first patch, or a special "classification" token added to the sequence, is then processed through a linear layer to yield the final classification output.

By combining EfficientNetV2B3 and ViT in a hybrid model, we seek to leverage their complementary strengths. EfficientNetV2B3 serves as a feature extractor, effectively capturing local features in images, while ViT captures the global context by modeling long-range dependencies between patches. The extracted features from both models are concatenated and used for the final classification task. This hybrid approach allows us to leverage the strengths of both architectures, potentially leading to improved disease detection and identification performance. The lightweight nature of EfficientNetV2B3 helps to balance the computational cost of the ViT module, making the hybrid model more practical for real-world deployment.

Datasets

This study utilizes a recently published Potato Leaf Disease dataset [40], comprising potato leaf images captured under authentic farming conditions in Central Java, Indonesia. This dataset represents a significant advancement over previous datasets, particularly the Plant Village dataset [43], due to its realistic and intricate depiction of typical field conditions. The increased complexity is attributed to the diverse disease types, variations in backgrounds, and differences in image orientations and distances. Figure 1 provides sample images in the two datasets. Figure 1a shows a sample image from the Potato Leaf Disease dataset, showing the complexity and variability of real-world farming conditions, while Fig. 1b shows a leaf image from the Plant Village dataset, which typically has more uniform and less varied image features.

The Potato Leaf Disease dataset consists of images of potato leaves affected by six different types of pathogens, such as bacteria, fungi, nematodes, pests, phytophthora, and viruses, along with a healthy class, totalling seven classes. These classes were labelled by plant disease experts from the Department of Plant Protection, Faculty of Agriculture, Universitas Gadjah Mada [40]. The images were captured using various smartphone cameras, resulting in a wider variation in the dataset, including diverse backgrounds and varying directions and distances of the images. The images are in.jpg format with a size of 1500 x 1500 pixels, and the dataset includes a total of 3,076 images. The distribution of images per class in the dataset is shown in Table 2.

Sample images from each class in the Potato Leaf Disease dataset are shown in Fig. 2. Each column represents a different disease class showcasing the variety of diseases and conditions present in the dataset. Each class

(a) (b)

Fig. 1 Sample of healthy images from (**a**) the Potato Leaf disease dataset and (**b**) the Plant Village dataset

Table 2 Distribution of images per disease class in the Potato Leaf Disease dataset

	Disease	Number of images	Distribution (%)
1	Bacteria	569	18.5
2	Fungi	748	24.3
3	Nematode	68	2.2
4	Pest	611	19.9
5	Phytopthora	347	11.3
6	Virus	532	17.3
7	Healthy	201	6.5

exhibits unique symptoms and patterns on the leaves. For example, leaves infected by viruses show reduced leaf size and crinkling, mild mottling or mosaic, and necrosis. Leaves infected by phytophthora can be observed as dark brown to black lesions which can enlarge into circular and necrotic patches if left untreated. Leaves infected by nematodes appear yellowish, with symptoms similar to those of water and nutrient deficiencies. Leaves infected by fungi show circular patterns manifested along leaf edges and/or slightly sunken leaf spots with yellow borders and concentric ring appearance, and/or yellow leaves with powdery patches. Leaves infected by bacteria wilt without turning yellow or necrotic. Leaves affected by pests show tissue distortion and/or holes and/or dotted leaves with silver colour or chlorotic material and/or with mined route on leaves. Healthy leaves exhibit a uniform green colour in all parts of the leaves and perfect leaf shape without imperfections.

Dataset analysis and characteristics

The Potato Leaf Disease dataset presents several challenges that impact model development and evaluation. First, the class distribution is notably imbalanced with the Nematode class (2.2%) and Healthy class (6.5%) being significantly underrepresented compared to Fungi (24.3%) and Bacteria (18.5%). This imbalance could potentially bias the model toward majority classes, a challenge we address through careful validation strategies and performance metrics that account for class imbalance (macro-averaging).

Second, the dataset reflects real-world agricultural conditions, capturing several sources of variability:

- **Background variations**: Images contain diverse backgrounds including soil, other plants, and agricultural implements.
- **Illumination differences**: Varying lighting conditions from direct sunlight to shade.
- **Perspective and scale variations**: Images captured from different angles and distances.

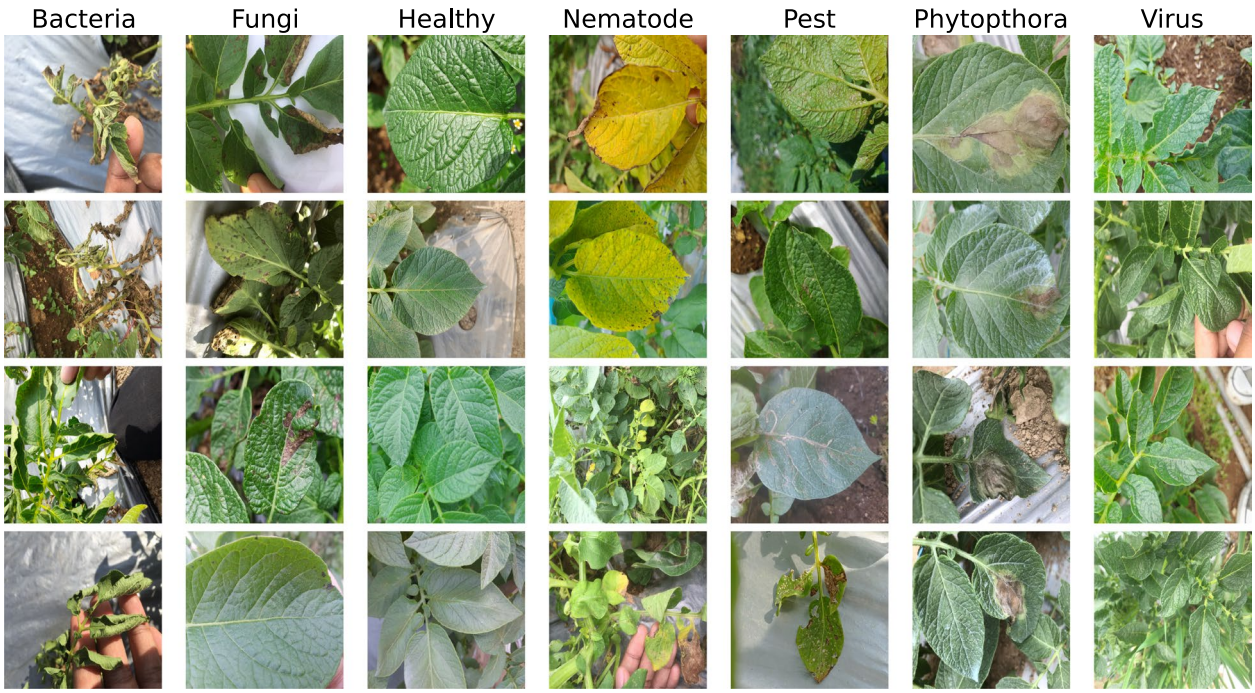

Fig. 2 Sample images from each class in the Potato Leaf Disease dataset. Each column represents a disease class, showcasing the variety of diseases and conditions present in the dataset

- **Disease progression stages**: Different manifestations of the same disease depending on severity and stage.

These characteristics make the dataset particularly challenging compared to controlled datasets like Plant Village but also more representative of the conditions in which the model would be deployed in practice. The real-world variability in the dataset serves as both a challenge and an opportunity to develop models with greater robustness and generalizability to field conditions. To assess the model's ability to generalize across different datasets, we conducted additional testing on the Plant Village dataset, which provides a more controlled set of images with standardized backgrounds and lighting conditions. This cross-dataset evaluation helps to evaluate the model's adaptability and potential for transfer learning across different agricultural contexts.

Data pre-processing and splitting

The dataset used was subjected to several preprocessing steps. Initially, all images were resized to a standard dimension of 256x256 pixels to ensure uniform input to the model. Subsequently, the images were augmented using a variety of transformations to enrich the diversity of the training data and enhance the model's ability to generalize. Specifically, images were randomly cropped to a size of 224x224 pixels with an adjusted scale for more intense zooming (scale of 0.95 to 1.05) and a ratio of 0.75 to 1.33. Images were also randomly flipped both horizontally and vertically, and rotated within a range of −40° to 40° following recommendations by Arshad et al. [14]. The saturation of the images was randomly adjusted with a factor of 0.8 and the hue with a factor of 0.021. Additionally, the images were randomly translated (shifted) horizontally and vertically by a factor of 0.13 and the scale was randomly adjusted by a factor of 0.8 to 1.2 for an additional zooming effect. Finally, the images were normalized using the mean and standard deviation of the ImageNet dataset ([0.485, 0.456, 0.406], [0.229, 0.224, 0.225]), aligning with common practice when using pretrained models and ensuring that the input values fell within a range that the model was designed to work with. After preprocessing, the dataset was split into training, validation, and test sets. The dataset was split into training (90%) and test (10%) sets, with the training set further divided into final training and validation sets using the same 9:1 ratio.

Training, tuning, and evaluation

During the training phase, the proposed model learns to make predictions based on the provided training dataset to enable the model to adjust its internal parameters to predict as accurately as possible. In the tuning phase, certain hyperparameters of the model are adjusted to optimize its performance. These hyperparameters include learning rate, batch size, and dropout rate. The tuning is performed using the validation dataset to ensure the model is not overly fitted to the training data.

The performance of the model is rigorously evaluated using a suite of metrics to provide a comprehensive assessment of its capability to correctly classify potato plant diseases. These metrics encompass Accuracy, Precision, Recall (Sensitivity), F1 Score, and Matthews Correlation Coefficient (MCC) [49–51]. Test Accuracy offers a direct measure of performance by calculating the ratio of correct predictions to the total number of input samples. Precision, defined as the ratio of correctly predicted positive observations to the total predicted positives, evaluates the model's effectiveness in minimizing false-positive predictions. Recall, or Sensitivity, measures the model's ability to accurately identify positive instances by calculating the ratio of correctly predicted positive observations to all actual positives. The F1 Score, which is the harmonic mean of Precision and Recall, provides a balanced measure that considers both false positives and false negatives. The MCC is a robust metric that evaluates the quality of classifications by incorporating true and false positives and negatives, making it particularly suitable for datasets with imbalanced class distributions, as is the case in this study. For Precision, Recall, and F1 Score, 'macro' averaging is employed. This approach calculates the metric independently for each class and then averages the results, ensuring equal treatment of all classes regardless of their size, which is crucial for handling imbalanced datasets [52]. The formulas for these metrics are given below.

$$\begin{aligned}
\text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\
\text{Precision} &= \frac{TP}{TP + FP} \\
\text{Recall} &= \frac{TP}{TP + FN} \\
\text{F1 Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
\text{MCC} &= \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
\end{aligned}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives respectively.

The proposed model

The proposed hybrid model intergrates CNN and Transformer architectures through a parallel feature extraction and fusion approach. Figure 3 illustrates the complete architecture which can be divided into three main components/modules which include the following:

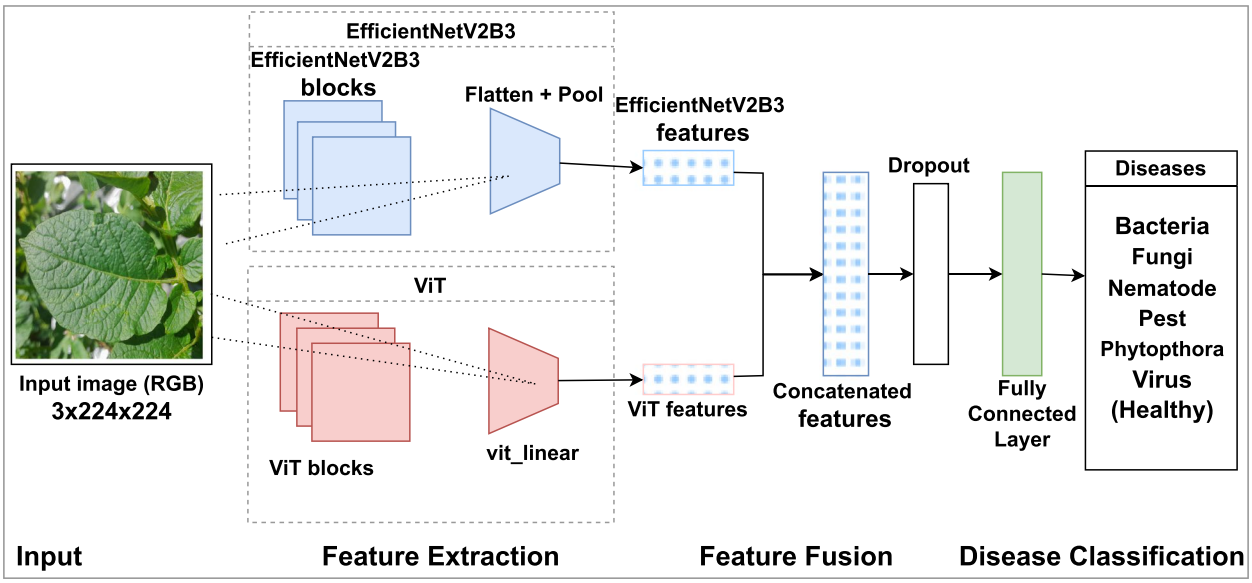

Fig. 3 Architecture of the proposed model, EfficientNetV2B3+ViT

- **Feature Extraction**: Consists of parallel paths through the EfficientNetV2B3 and ViT models each extracting complementary feature representations from the input image.
- **Feature Fusion**: Combines the feature vectors from both models through concatenation, followed by a dropout layer to enhance generalization.
- **Classification**: A fully connected layer with softmax activation that maps the fused features to class probabilities for the seven disease categories.

The input RGB images (3×224×224) are simultaneously processed through both feature extraction paths. The EfficientNetV2B3 path extracts local textural and structural features while the ViT path captures global contextual information and long-range dependencies. This parallel processing strategy ensures that both local and global information are preserved throughout the network.

The classifier head of the EfficientNetV2B3 is removed, transforming the model into a dedicated feature extraction module. The parameters of the model are frozen, preventing updates during the training process. This strategy maintains the model's lightweight and computational efficiency. The output from the CNN model is subsequently flattened and globally average pooled, which reduces the dimensionality of the output and provides a compact feature representation, focusing on local features of the image data. Similarly, an additional linear layer is incorporated into the ViT model to further reduce the dimensionality of its output. The parameters of the ViT model are frozen, ensuring that it remains lightweight and focused on feature extraction. The ViT model is designed to capture the global context of the image data.

The outputs from both the CNN and ViT models are then concatenated. By combining the local features captured by the CNN model with the global context captured by the ViT model, the proposed model is able to create a comprehensive feature representation of the input data. A dropout layer is applied to the concatenated features which helps mitigate overfitting during training time. Following this is a fully connected layer that serves as a classifier. This layer employs a softmax activation function to output class probabilities, with the class having the highest probability selected as the model's prediction. Leveraging the strengths of CNN and Transformer architectures through a hybrid combination and the innovative feature concatenation and dropout layers enhances the model's robustness and generalization capability. The rationale behind our hybrid architecture design stems from the complementary nature of CNNs and Vision Transformers. EfficientNetV2B3 was specifically selected for its balanced accuracy-to-parameter ratio and optimized mobile-focused design, making it ideal for potential field deployment in agricultural settings. The CNN component excels at capturing local textural features crucial for identifying specific disease patterns like spots, lesions, and discoloration at the leaf surface level. Meanwhile, the ViT component was chosen for its ability to model long-range dependencies between image regions, allowing the model to understand the global context of

disease spread across the entire leaf. This is particularly important for diseases that manifest with patterns that span large areas of the leaf or have characteristic spatial distributions. The concatenation of features from both models, rather than using a sequential architecture, enables the preservation of both feature types without one dominating the other.

The dropout layer (with rate 0.2) between the concatenated features and the final classifier serves two critical purposes: (1) preventing overfitting to training data, especially important given the limited size of the dataset relative to the model's capacity, and (2) ensuring robust feature integration by randomly dropping connections during training, which forces the model to learn redundant representations and not rely too heavily on either the CNN or Transformer features alone. This balanced integration approach is crucial for handling the diverse manifestations of potato plant diseases in real-world agricultural environments.

Implementation and experimentation

The proposed hybrid framework is implemented using PyTorch and trained on a high-performance computing environment, eX3 at Simula Research Laboratory, using Nvidia Volta V100 GPUs. The model training spans 70 epochs, utilizing the cross-entropy loss function with an initial learning rate of 0.0001 and a batch size of 64. These parameters were selected based on preliminary experiments, balancing training speed and model performance. The cross-entropy loss function is apt for classification tasks, measuring the dissimilarity between predicted and true distributions. The chosen learning rate is sufficiently small to ensure convergence yet large enough for efficient learning, while the batch size of 64 aligns with the memory constraints of the GPUs used. The proposed model employs the Adam optimizer, which integrates the advantages of stochastic gradient descent with momentum and root mean square propagation. Therefore, it is computationally efficient and consumes less memory [49]. To enhance model generalization, a learning rate (α) scheduler is incorporated. This scheduler reduces the α by a factor of 0.5 if no improvement in validation loss is observed over 5 consecutive epochs. Combined with a dropout rate of 0.2, this approach helped prevent overfitting. After training and tuning the hybrid model, further experimentation involved an ablation study with the individual models; the EfficientNetV2B3 and ViT models. This was done to understand the contributions of each model within the hybrid framework. Both models were evaluated individually under the same training settings as the primary study. Additionally, we have compared the performance of the proposed model with a previous study [40] on the Potato Leaf Disease dataset and tested the proposed model on the Plant Village dataset.

Results

Table 3 shows the results of the evaluation of the proposed hybrid model, EfficientNetV2B3+ViT on the diverse Potato Leaf Disease dataset. The results show the model achieving a test accuracy of 85.06%, a precision of 82.86% a recall of 85.29%, an F1 score of 83.77%, and an MCC of 0.82. The table also shows the results from an ablation study. The pre-trained ViT model, adapted for this task achieved a test accuracy of 77.92%, a precision of 76.56%, a recall of 72.03%, an F1 score of 73.44% and an MCC of 0.73. The EfficientNetV2B3 model, also adapted for this task, slightly outperformed the ViT model with a test accuracy of 79.55%, a precision of 78.02%, a recall of 80.97% an F1 score of 79.01%, and an MCC of 0.75. This ablation study shows that both individual models contribute significantly to the overall performance of the hybrid model.

The MCC scores further validate the model's performance, providing a balanced measure that accounts for all four quadrants of the confusion matrix, which is particularly informative given the class imbalance and potential confusions observed. The performance of the EfficientNetV2B3+ViT was compared with that of a previous study [40] on the Potato Leaf Disease dataset. Our model outperformed the previous study's EfficientNetV2B3 across all metrics with a notable 11.43% improvement in test accuracy (see Table 3). Similarly, the EfficientNetV2B3 used in our ablation study also outperformed the same model from the previous study.

Table 3 Performance comparison of ViT, EfficientNetV2B3, EfficientNetV2B3 [40] and the hybrid (EfficientNetV2B3+ViT) models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	MCC
ViT	77.92	76.56	72.03	73.44	0.73
EfficientNetV2B3	79.55	78.02	80.97	79.01	0.75
EfficientNetV2B3+ViT	**85.06**	**82.86**	**85.29**	**83.77**	**0.82**
EfficientNetV2B3 [40]	73.63	74.28	73.63	73.02	-
EfficientNetV2B3+ViT (Plant Village Dataset)	98.15	98.73	93.41	95.74	0.97

The performance metrics of the EfficientNetV2B3+ViT model on the Plant Village dataset are also given in Table 3. The model achieved an accuracy of 98.15% on this dataset which is 13.09% higher than its performance on the Potato Leaf Disease dataset. Comparative performance of the individual models and the proposed hybrid model based on various metrics are graphically shown in Fig. 4.

Detailed analysis of the results reveals variations in the performance of the model across different disease classes (see Table 4). The 'Bacteria' class achieved the highest accuracy (93.42%) and MCC (0.91), suggesting that the model is highly effective at identifying bacterial diseases based on the available symptoms. In contrast, the 'Pest' class had the lowest accuracy (60.66%) and MCC (0.59), indicating significant challenges in accurately detecting pest-related damage. This may be attributed to the potential overlap in symptoms between pest damage and other foliar diseases, particularly those caused by fungi. As shown in the confusion matrix (see Fig. 5), 20% of 'Pest' instances were misclassified as 'Fungi', while 10% of 'Fungi' instances were misclassified as 'Pest'. These findings underscore the need for further research to develop more discriminative features or incorporate additional data sources to improve pest detection accuracy.

To further validate our model's performance, we compare our results with recent state-of-the-art approaches. Although these models were not originally tested on the exact same dataset, we can draw meaningful comparisons based on their reported performance on similar agricultural disease detection tasks. Recent work by Dai et al. [17] with their PPLC-Net achieved excellent results on controlled datasets (>99% accuracy) using attention mechanisms and dilated convolutions. Although direct comparison is not possible, the principles of their attention mechanism could be incorporated into our framework to potentially enhance performance further. Dai et al. [20]'s cross-modal fusion approach (ITF-WPI) achieved 97.98% accuracy for pest recognition by incorporating both image and text data. Although our approach focuses solely on image data, the superior performance of our hybrid model (85.06%) compared to single-modality approaches on our challenging dataset suggests that our feature fusion strategy effectively captures complementary information. Pal and Kumar [19]'s AgriDet framework addresses occlusion through segmentation before classification. Although effective for controlled environments, our approach demonstrates robustness to background variability without requiring explicit segmentation, making it more suitable for direct field deployment.

The robustness of our model is further validated by its consistent performance in different disease classes, with particularly strong results for the Bacteria class (93.42%) and reasonable performance even for the challenging Pest class (60.66%). This class-wise performance variability provides insight into areas for future improvement through targeted data augmentation or model refinement.

Discussion

The results highlight the potential of the proposed hybrid model, EfficientNetV2B3+ViT, in detection and identification of potato leaf diseases. The high accuracy,

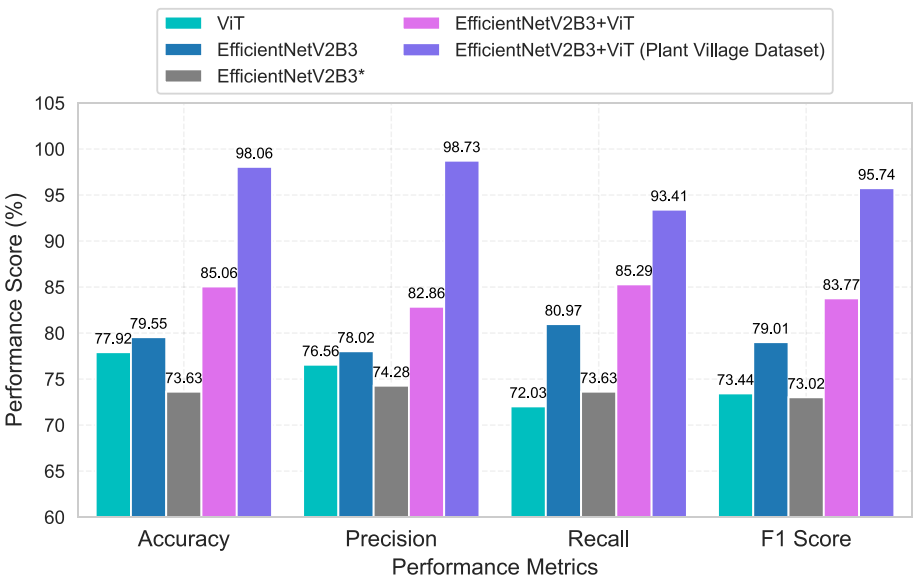

Fig. 4 Comparative performance of ViT, EfficientNetV2B3, EfficientNetV2B3* [40] and the Hybrid Model on Various Metrics

Table 4 Class-wise performance metrics of the proposed hybrid model. The highest accuracy and MCC scores are highlighted in bold and the lowest scores are underlined

Class	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	MCC
Bacteria	**93.42**	93.42	93.42	93.42	**0.91**
Fungi	84.42	73.86	84.42	78.79	0.71
Nematode	83.33	83.33	83.33	83.33	0.83
Pest	<u>60.66</u>	72.55	60.66	66.07	<u>0.59</u>
Phytopthora	73.91	65.38	73.91	69.39	0.67
Virus	82.98	84.78	82.98	83.87	0.81
Healthy	72.22	86.67	72.22	78.79	0.78

precision and recall of the model indicate its reliability in disease identification and its capacity to detect true positive instances. However, the model's performance varies across different classes. One of the main limitations of this study is the challenge posed by the 'Pest' class, which exhibited significant confusion with the 'Fungi' class. As evident from the confusion matrix in Fig. 5, 20% of 'Pest' instances were misclassified as 'Fungi', while 10% of 'Fungi' instances were misclassified as 'Pest'. This confusion probably arises from the overlap in features between these classes, such as similar leaf damage patterns as described in "Materials and methods" section. This highlights the difficulty in distinguishing between biotic stresses based solely on foliar symptoms and the need for more nuanced features or additional data modalities to accurately differentiate between these classes. Given the observed confusion between the 'Pest' and 'Fungi' classes, future work could investigate tailored approaches or feature engineering techniques specifically designed to improve the differentiation of these visually similar disease symptoms. The confusion matrix also reveals that 17% of 'Nematode' instances were misclassified as 'Fungi'. Future studies could further refine the dataset, specifically for nematode symptoms, to explore potential variations and inconsistencies in symptoms between different stress factors.

Another limitation is the potential variability in the model performance when applied to different datasets or crops. Although the proposed model achieved strong results on the potato dataset used in this study, its generalization to other contexts remains to be validated. It is important to recognize that the specific characteristics of the training data, such as the range of diseases and imaging conditions, may impact the model's performance in new settings. Despite these limitations, our study demonstrates the promise of hybrid deep learning approaches for the detection of plant disease and highlights the importance of using diverse real-world datasets for training and evaluation. By acknowledging and addressing

Fig. 5 Confusion matrix of the prediction distribution of the proposed EfficientNetV2B3+ViT model

these challenges, we can continue to advance the development of reliable, scalable, and practical AI solutions for the management of agricultural diseases.

The ablation study has demonstrated both the EfficientNetV2B3 and ViT models contributing significantly to the performance of the hybrid model. Notably, the combination of these models in the hybrid setup leads to improved performance, illustrating the effectiveness of the chosen hybrid approach. Compared to the previous study [40], our hybrid model outperforms the EfficientNetV2B3 model across all metrics. This improvement highlights the effectiveness of the hybrid approach, which combines the strengths of both ViT and EfficientNetV2B3. This combination allows the hybrid model to capture a wider range of patterns in the data, contributing to its superior performance. Furthermore, the superior performance of EfficientNetV2B3 in our ablation study compared to the previous study, can be attributed to a combination of factors: the use of learning rate scheduling during training, the careful application of regularization techniques such as dropout, and the incorporation of varied data augmentations which may not have been used in the previous study. These measures collectively enhanced the model's robustness and generalizability.

While our investigation yields promising results, it also uncovers some limitations. The complexity and variability of the dataset make achieving near-perfect accuracy a challenge. Additionally, our study does not currently provide insights into the explainability of the models, a critical aspect for fostering trust and interpretability in such deep learning applications. Future research could address these limitations by exploring alternative architectures, hybrid models, and integrating explainable AI techniques to gain better insights into the models' prediction processes. Future research could also explore the use of synthetic image generation processes to balance the dataset, which may further improve overall detection accuracy, particularly for underrepresented classes.

Conclusion

The study demonstrates the potential of the EfficientNetV2B3+ViT hybrid model for improving the accuracy of potato plant disease detection and identification. The proposed model achieved an accuracy of 85.06% on a diverse dataset that reflects real-world agricultural conditions, outperforming the individual CNN and ViT models. However, the model performance varied across different disease classes, with the 'Pest' class posing particular challenges due to potential feature overlap with the 'Fungi' class. Further testing on additional datasets and crops is necessary to establish the broader applicability and generalizability of the proposed approach.

Future research should focus on refining the model architecture, incorporating explainable AI techniques to provide insight into the decision-making process, and investigating the transferability of the approach to other crops and diseases. Furthermore, integrating complementary data modalities and developing user-friendly interfaces and workflows could enhance the model's performance and facilitate its integration into real-world agricultural practices.

Despite limitations, this study represents a significant step towards leveraging advanced AI techniques to improve disease diagnosis in complex agricultural settings. By addressing the challenges identified and building on the foundation laid by this research, future work can contribute to the development of more accurate, reliable, and impactful AI solutions for plant disease detection.

Acknowledgements
The research presented in this paper has benefited from the high performance computing infrastructure at Simula Research Laboratory, namely Experimental Infrastructure for Exploration of Exascale Computing (eX3), which is financially supported by the Research Council of Norway.

Authors' contributions
J.H.S, A.C and R.S formulated the concept and designed the methodology. J.H.S collected data, and performed experiments and data analysis. A.C and R.S provided supervision and additional guidance and suggestions relevant to the experiments and analysis. J.H.S initially drafted the paper and all the authors reviewed and edited the paper.

Data availability
This research used two publicly available datasets: Plant Village Dataset and Potato Leaf Disease Dataset.

Code availability
The code is available at https://github.com/HJacksons/potato-efficientViT

Declarations

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 9 July 2024 Accepted: 5 May 2025
Published online: 16 May 2025

References

1. Attri I, Awasthi LK, Sharma TP. Machine learning in agriculture: A review of crop management applications. Multimed Tools Appl. 2023;83(5):12875–915. https://doi.org/10.1007/s11042-023-16105-2.
2. Sharma A, Jain A, Gupta P, Chowdary V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. IEEE Access. 2021;9:4843–73. https://doi.org/10.1109/ACCESS.2020.3048415.
3. Pawlak K, Kołodziejczak M. The Role of Agriculture in Ensuring Food Security in Developing Countries: Considerations in the Context of the

Problem of Sustainable Food Production. Sustainability. 2020;12(13). https://doi.org/10.3390/su12135488.
4. Ayoub Shaikh T, Rasool T, Rasheed Lone F. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. Comput Electron Agric. 2022;198:107119. https://doi.org/10.1016/j.compag.2022.107119.
5. Wani JA, Sharma S, Muzamil M, Ahmed S, Sharma S, Singh S. Machine learning and deep learning based computational techniques in Automatic Agricultural Diseases Detection: Methodologies, applications, and challenges. Arch Comput Methods Eng. 2021;29(1):641–77. https://doi.org/10.1007/s11831-021-09588-5.
6. Chakraborty SK, Chandel NS, Jat D, Tiwari MK, Rajwade YA, Subeesh A. Deep learning approaches and interventions for futuristic engineering in Agriculture. Neural Comput & Applic. 2022;34(23):20539–73. https://doi.org/10.1007/s00521-022-07744-x.
7. Ayaz M, Li CH, Ali Q, Zhao W, Chi YK, Shafiq M, et al. Bacterial and Fungal Biocontrol Agents for Plant Disease Protection: Journey from Lab to Field, Current Status, Challenges, and Global Perspectives. Molecules. 2023;28(18). https://doi.org/10.3390/molecules28186735.
8. Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, et al. The persistent threat of emerging plant disease pandemics to Global Food Security. Proc Natl Acad Sci. 2021;118(23). https://doi.org/10.1073/pnas.2022239118.
9. Kuswidiyanto LW, Noh HH, Han X. Plant Disease Diagnosis Using Deep Learning Based on Aerial Hyperspectral Images: A Review. Remote Sens. 2022;14(23). https://doi.org/10.3390/rs14236031.
10. Orchi H, Sadik M, Khaldoun M. On Using Artificial Intelligence and the Internet of Things for Crop Disease Detection: A Contemporary Survey. Agriculture. 2022;12(1). https://doi.org/10.3390/agriculture12010009.
11. Abdullah HM, Mohana NT, Khan BM, Ahmed SM, Hossain M, Islam KS, et al. Present and future scopes and challenges of plant pest and disease monitoring: Remote sensing, image processing, and artificial intelligence perspectives. Remote Sens Appl Soc Environ. 2023;32:100996. https://doi.org/10.1016/j.rsase.2023.100996.
12. Arshaghi A, Ashourian M, Ghabeli L. Potato diseases detection and classification using Deep Learning Methods. Multimed Tools Appl. 2022;82(4):5725–42. https://doi.org/10.1007/s11042-022-13390-1.
13. Mahum R, Munir H, Mughal ZUN, Awais M, Khan FS, Saqlain M, et al. A novel framework for potato leaf disease detection using an efficient deep learning model. Hum Ecol Risk Assess Int J. 2023;29(2):303–26. https://doi.org/10.1080/10807039.2022.2064814.
14. Arshad F, Mateen M, Hayat S, Wardah M, Al-Huda Z, Gu YH, et al. PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction. Alex Eng J. 2023;78:406–18. https://doi.org/10.1016/j.aej.2023.07.076.
15. Thakur PS, Khanna P, Sheorey T, Ojha A. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Syst Appl. 2022;208:118117. https://doi.org/10.1016/j.eswa.2022.118117.
16. Kunduracioglu I. Utilizing ResNet Architectures for Identification of Tomato Diseases. J Intell Decis Mak Inform Sci. 2024;1:104–119. https://doi.org/10.59543/jidmis.v1i.11949.
17. Dai G, Fan J, Tian Z, Wang C. PPLC-Net: Neural network-based plant disease identification model supported by weather data augmentation and multi-level attention mechanism. J King Saud Univ Comput Inform Sci. 2023;35(5):101555. https://doi.org/10.1016/j.jksuci.2023.101555.
18. Dai G, Tian Z, Fan J, Sunil CK, Dewi C. DFN-PSAN: Multi-level deep information feature fusion extraction network for interpretable plant disease classification. Comput Electron Agric. 2024;216:108481. https://doi.org/10.1016/j.compag.2023.108481.
19. Pal A, Kumar V. AgriDet: Plant Leaf Disease severity classification using agriculture detection framework. Eng Appl Artif Intell. 2023;119:105754. https://doi.org/10.1016/j.engappai.2022.105754.
20. Dai G, Fan J, Dewi C. ITF-WPI: Image and text based cross-modal feature fusion model for wolfberry pest recognition. Comput Electron Agric. 2023;212:108129. https://doi.org/10.1016/j.compag.2023.108129.
21. Panigrahi KP, Das H, Sahoo AK, Moharana SC. Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms. In: Das H, Pattnaik PK, Rautaray SS, Li KC, editors. Progress in Computing, Analytics and Networking. Singapore: Springer Singapore; 2020. pp. 659–669. https://doi.org/10.1007/978-981-15-2414-1_66.
22. Shrestha G, Deepsikha, Das M, Dey N. Plant Disease Detection Using CNN. In: 2020 IEEE Applied Signal Processing Conference (ASPCON). 2020. pp. 109–113. https://doi.org/10.1109/ASPCON49795.2020.9276722.
23. KP A, Anitha J. Plant disease classification using deep learning. In: 2021 3rd International Conference on Signal Processing and Communication (ICPSC). 2021. pp. 407–411. https://doi.org/10.1109/ICSPC51351.2021.9451696.
24. Hong H, Lin J, Huang F. Tomato Disease Detection and Classification by Deep Learning. In: 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). 2020. pp. 25–29. https://doi.org/10.1109/ICBAIE49996.2020.00012.
25. Habiba SU, Islam MK. Tomato Plant Diseases Classification Using Deep Learning Based Classifier From Leaves Images. In: 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD). 2021. pp. 82–86. https://doi.org/10.1109/ICICT4SD50815.2021.9396883.
26. Kunduracıoğlu I, Paçal I. Deep Learning-Based Disease Detection in Sugarcane Leaves: Evaluating EfficientNet Models. J Oper Intell. 2024;2(1):321–235. https://doi.org/10.31181/jopi21202423.
27. Kunduracioglu I. CNN Models Approaches for Robust Classification of Apple Diseases. Comput Decis Mak Int J. 2024;1:235–251. https://doi.org/10.59543/comdem.v1i.10957.
28. Kunduracioglu I, Pacal I. Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. J Plant Dis Protect. 2024;131(3):1061–80. https://doi.org/10.1007/s41348-024-00896-z.
29. Borhani Y, Khoramdel J, Najafi E. A deep learning based approach for automated plant disease classification using vision transformer [HTML]. Sci Rep. 2022. https://doi.org/10.1038/s41598-022-15163-0.
30. Parez S, Dilshad N, Alghamdi NS, Alanazi TM, Lee JW. Visual Intelligence in Precision Agriculture: Exploring Plant Disease Detection via Efficient Vision Transformers. Sensors. 2023;23(15). https://doi.org/10.3390/s23156949.
31. Thai HT, Le KH, Nguyen NLT. FormerLeaf: An efficient vision transformer for Cassava Leaf Disease detection. Comput Electron Agric. 2023;204:107518. https://doi.org/10.1016/j.compag.2022.107518.
32. Thakur PS, Chaturvedi S, Khanna P, Sheorey T, Ojha A. Vision transformer meets convolutional neural network for plant disease classification. Ecol Inform. 2023;77:102245. https://doi.org/10.1016/j.ecoinf.2023.102245.
33. Li G, Wang Y, Zhao Q, Yuan P, Chang B. PMVT: a lightweight vision transformer for plant disease identification on mobile devices. Front Plant Sci. 2023;14. https://doi.org/10.3389/fpls.2023.1256773.
34. Yu S, Xie L, Huang Q. Inception convolutional vision transformers for plant disease identification. Internet Things. 2023;21:100650. https://doi.org/10.1016/j.iot.2022.100650.
35. Thakur PS, Khanna P, Sheorey T, Ojha A. Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT. 2022. https://doi.org/10.48550/arXiv.2207.07919.
36. Iqbal MA, Talukder KH. Detection of Potato Disease Using Image Segmentation and Machine Learning. In: 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET). 2020. pp. 43–47. https://doi.org/10.1109/WiSPNET48689.2020.9198563.
37. Shaheed K, Qureshi I, Abbas F, Jabbar S, Abbas Q, Ahmad H, et al. EfficientRMT-Net—An Efficient ResNet-50 and Vision Transformers Approach for Classifying Potato Plant Leaf Diseases. Sensors. 2023;23(23). https://doi.org/10.3390/s23239516.
38. Tiwari D, Ashish M, Gangwar N, Sharma A, Patel S, Bhardwaj S. Potato Leaf Diseases Detection Using Deep Learning. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS); 2020. pp. 461–466. https://doi.org/10.1109/ICICCS48265.2020.9121067.
39. Chakraborty KK, Mukherjee R, Chakroborty C, Bora K. Automated recognition of optical image based potato leaf blight diseases using deep learning. Physiol Mol Plant Pathol. 2022;117:101781. https://doi.org/10.1016/j.pmpp.2021.101781.
40. Shabrina NH, Indarti S, Maharani R, Kristiyanti DA, Irmawati, Prastomo N, et al. A novel dataset of potato leaf disease in uncontrolled environment. Data Brief. 2024;52:109955. https://doi.org/10.1016/j.dib.2023.109955.

41. Pacal I, Kunduracioglu I, Alma MH, Deveci M, Kadry S, Nedoma J, et al. A systematic review of deep learning techniques for plant diseases. Artif Intell Rev. 2024;57(11). https://doi.org/10.1007/s10462-024-10944-7.
42. Tan M, Le QV. EfficientNetV2: Smaller Models and Faster Training. 2021. https://doi.org/10.48550/arXiv.2104.00298.
43. Hughes DP, Salathe M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. 2016. arXiv:1511.08060.
44. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–323. https://doi.org/10.1109/5.726791.
45. Chollet F. Xception: Deep Learning With Depthwise Separable Convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. https://doi.org/10.1109/CVPR.2017.195.
46. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
47. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 2021. arXiv:2010.11929.
48. Scabini L, Sacilotti A, Zielinski KM, Ribas LC, Baets BD, Bruno OM. A Comparative Survey of Vision Transformers for Feature Extraction in Texture Analysis. 2024. arXiv:2406.06136.
49. Chatterjee A, Prinz A, Riegler MA, Meena YK. An automatic and personalized recommendation modelling in activity eCoaching with deep learning and ontology. Sci Rep. 2023;13(1):10182. https://doi.org/10.1038/s41598-023-37233-7.
50. Chatterjee A, Pahari N, Prinz A, Riegler M. AI and semantic ontology for personalized activity eCoaching in healthy lifestyle recommendations: a meta-heuristic approach. BMC Med Inform Decis Mak. 2023;23(1):278.
51. Chatterjee A, Pahari N, Prinz A, Riegler M. Machine learning and ontology in eCoaching for personalized activity level monitoring and recommendation generation. Sci Rep. 2022;12(1):19825. https://doi.org/10.1038/s41598-023-30029-9.
52. Grandini M, Bagli E, Visani G. Metrics for Multi-Class Classification: an Overview. 2020. https://doi.org/10.48550/arXiv.2008.05756.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.