Alexandria Engineering Journal 78 (2023) 406–418

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Original Article

PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction

Fizzah Arshad [a,1], Muhammad Mateen [a,b,*], Shaukat Hayat [a], Maryam Wardah [a], Zaid Al-Huda [c], Yeong Hyeon Gu [d,*], Mugahed A. Al-antari [d,*,1]

[a] Department of Computer Science, Air University Multan Campus, Multan 61000, Pakistan
[b] School of Electronic and Information Engineering, Soochow University, Suzhou, China
[c] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, Sichuan, China
[d] Department of Artificial Intelligence, College of Software & Convergence Technology, Daeyang AI Center, Sejong University, Seoul 05006, Korea

ARTICLE INFO

Keywords:
Potato leaf disease
Plant leaf disease prediction
Hybrid prediction AI mode
Deep learning
Feature fusing and concatenation
Vision Transformer (ViT)

ABSTRACT

Agricultural productivity plays a vital role in global economic development and growth. When crops are affected by diseases, it adversely impacts a nation's economic resources and agricultural output. Early detection of crop diseases can minimize losses for farmers and enhance production. In this study, we propose a new hybrid deep learning model, PLDPNet, designed to automatically predict potato leaf diseases. The PLDPNet framework encompasses image collection, pre-processing, segmentation, feature extraction and fusion, and classification. We employ an ensemble approach by combining deep features from two well-established models (VGG19 and Inception-V3) to generate more powerful features. The hybrid approach leverages the concept of vision transformers for final prediction. To train and evaluate PLDPNet, we utilize the public potato leaf dataset: early blight, late blight, and healthy leaves. Utilizing the strength of segmentation and fusion feature, the proposed approach achieves an overall accuracy of 98.66%, and F1-score of 96.33%. A comprehensive validation study is conducted using Apple (4 classes) and tomato (10 classes) datasets achieving impressive accuracies of 96.42% and 94.25%, respectively. These experimental findings confirm that the proposed hybrid framework provides more effective and accurate detection and prediction of potato crop diseases, making it a promising candidate for practical applications.

1. Introduction

Agriculture serves as the foundation of every civilization. In the past, it held a position of dominance, but over time, its productivity has declined due to various factors such as political, social, environmental, and climate conditions [1]. Consequently, it has become the second largest sector in Pakistan. Currently, agriculture contributes more than 21% to the GDP and employs 45% of the workforce. A significant portion of the country's population, approximately 63%, resides in rural areas and relies directly or indirectly on this sector for their livelihood. The agriculture sector exhibits strong interconnections with the overall economy, despite being overlooked in statistical data [2]. Pakistan's extensive agricultural land is dedicated to cultivating various crops, forming the backbone of the country's economy. Among these crops, potato holds significant importance as the fourth most widely grown and produced crop in Pakistan. The majority, approximately 95%, of potato cultivation takes place in Punjab, specifically in regions like Sahiwal, Okara, Jhang, and Kasur. The remaining production is distributed across KPK, Sindh, and Baluchistan. Recent statistics indicate that Pakistan produces approximately 4.0 million tons of potatoes, cultivated across a total area of 170,300 ha [3]. The consumption of potatoes in Pakistan has been steadily increasing, with the annual per capita intake rising from 10 kg a decade ago to 15 kg at present. Globally, potatoes rank as the third most widely produced and consumed crop, trailing behind rice

* Corresponding authors at: Department of Artificial Intelligence, College of Software & Convergence Technology, Daeyang AI Center, Sejong University, Seoul 05006, Korea (Yeong Hyeon Gu and Mugahed A. Al-antari); School of Electronic and Information Engineering, Soochow University, Suzhou, China (Muhammad Mateen).

E-mail addresses: fizzaa995@gmail.com (F. Arshad), muhammadmateen@suda.edu.cn (M. Mateen), shaukat.hayat@aumc.edu.pk (S. Hayat), Maryam.Wardah@aumc.edu.pk (M. Wardah), zaidalhuda@swjtu.edu.cn (Z. Al-Huda), yhgu@sejong.ac.kr (Y.H. Gu), en.mualshz@sejong.ac.kr (M.A. Al-antari).

[1] Authors equally contributed to this work.

https://doi.org/10.1016/j.aej.2023.07.076
Received 16 February 2023; Received in revised form 23 June 2023; Accepted 28 July 2023
Available online 1 August 2023
1110-0168/

and wheat, with billions of people worldwide consuming it in various forms [4]. Additionally, potatoes serve as a vital food source and can be locally grown in different regions across the globe. Notably, China stands as the leading producer of potatoes, with an annual production of approximately 90,321,442 tons, followed by India, Ukraine, and Ireland, producing 48,529,000 tons, 22,503,970 tons, and 273,000 tons respectively, according to a recent study [5].

Potatoes play a vital role in maintaining good health, as they aid in preventing heart diseases and are rich in fiber. They possess high levels of antioxidants that contribute to combating diseases such as cholesterol and imbalanced sugar levels [6]. It is important to note that potato crops thrive best in loose, loamy, and sandy loam soils enriched with organic matter, as alkaline and saline soils are unsuitable for cultivation [7]. The occurrence of plant diseases holds significant importance as they directly impact the quantity and quality of plants. These diseases can be attributed to factors such as fungi, bacteria, viruses, and more. Early identification and detection of these diseases enable the implementation of preventive measures, minimizing both production and economic losses. In the past, disease classification and detection heavily relied on visual observation by experts [8]. However, this approach often proves impractical due to limited expert availability in remote areas and the time-consuming nature of the process. As a solution, the emergence of tools like image analysis has proven to be an effective approach for early disease detection in plants. Diseases typically manifest specific symptoms on leaves, making it feasible to employ various deep learning algorithms and image processing techniques for disease identification and detection [9]. Both approaches contribute to addressing food security concerns and enhancing productivity levels. In recent years, agriculture has continued to be a major component of our economy [10]. As technology and research have advanced, it has become apparent that various diseases affect potato crops, leading to reduced yields and significant tuber losses. This issue has become a cause for concern within the agriculture industry. Nonlinear dynamics encompasses important concepts such as bifurcation, chaos, and delay systems. These concepts are crucial for studying complex systems that exhibit behaviors beyond the predictability of linear mathematical models. The dynamic behavior of neural networks has consistently captivated the scientific and technological communities [11]. Delays, which arise from the time it takes for signals to travel between neurons, can significantly influence the dynamic behavior of neural networks. Extensive research has demonstrated that delays have a profound impact on network behavior. Introducing delays can induce instability and give rise to complex phenomena such as Hopf bifurcation, Hopf-Hopf bifurcation, pitchfork bifurcation, chaotic attractors, and more [12,13]. Therefore, through extensive research, it has been determined that numerous diseases pose a threat to potato crops. Consequently, the primary aim and objective of the proposed research endeavor are to accurately detect the presence of blight on the crop and promptly identify the specific diseases afflicting it.

Crop diseases are detrimental alterations to key functions of plants caused by bacteria, fungi, pests, or viruses. These diseases have a profound impact on agricultural production and pose a significant challenge to the economy and environment. As per existing assessments, 60% of the global population relies on agriculture as their primary food source. To meet the needs of a growing population, the Food and Agriculture Organization (FAO) emphasizes the necessity to increase the global food supply by 70% [14]. Plant diseases account for approximately 25% of crop losses, which translates to a substantial amount of food that could feed around 600 million people. Among these diseases, plant leaf diseases play a significant role in reducing agricultural productivity. The potato crop holds immense importance both domestically and globally, serving as a staple for local consumption and export. Consequently, addressing plant diseases becomes crucial for our country's agricultural sector and economy. By effectively tackling this challenge, we can enhance potato quality, reduce the prevalence of diseased crops, and achieve higher yields. This approach not only strengthens the economy but also contributes to the advancement of agriculture. To address this issue, we have developed a hybrid deep learning framework that combines various deep learning models to extract deep features using a fusion strategy. The main contributions of this work are summarized as follows;

- A novel deep learning framework is developed to achieve efficient and precise disease prediction, encompassing a hybrid approach that combines ensemble deep learning CNN for deep feature extraction and transformer for classification.
- To enhance the accuracy of disease prediction in the proposed PLDPNet framework, automatic segmentation of potato leaf diseases is carried out using the adopted U-Net architecture.
- A comprehensive evaluation ablation study is carried out to assess the reliability of the proposed AI framework and the individual impact of each component within the framework.

The rest of this paper is organized as follows. A review of relevant literature to this study is presented in section 2. The details of the proposed framework are presented in section 3. The results of the experimental study are reported and discussed in section 4. Finally, section 5 presents the most finding conclusions from this study.

2. Related works

Various techniques, including image processing, machine learning, and deep learning algorithms, have been employed to track and identify plant diseases, resulting in significant advancements. To identify diseases on potato leaves, a segmentation approach utilizing K-means clustering has been applied to extract features such as area, color, and texture from image samples. Neural network algorithms are then used for disease classification and identification [15,16]. Another approach involves image segmentation combined with feature extraction, such as mean, standard deviation, and contrast, followed by neural network algorithms for disease classification in cotton and tomato crops. In 2017, Islam et al. introduced the application of image segmentation on the plantVillage dataset, specifically focusing on potato leaves, and employed multiclass support vector machine for classification [17]. Additionally, Barrea et al. proposed the LeafNet architecture, based on convolutional neural networks (CNNs), and tested it on freely available datasets such as Flavia, LeafSnap, and foliage [18]. Furthermore, Sladojevic at al. developed a deep convolutional neural network model to classify 13 different types of diseases [19]. A method was devised to classify and identify fungal diseases in the grapes dataset. This approach utilizes support vector machine and K-means clustering for feature extraction and classification [20]. Additionally, Singh et al. presented an algorithm that combines image segmentation with soft computing techniques for leaf disease detection [21]. The color co-occurrence method is employed for feature extraction, and an SVM classifier is used for disease classification. David et al. introduced a Few-Shot learning approach for disease classification, where the Inception V3 network is fine-tuned to learn general plant characteristics [22]. A Siamese network and triplet loss, along with a baseline CNN, are utilized in this approach. Tiwari et al. proposed an identification algorithm that incorporates the Chan-Vase algorithm for segmentation. The classification is performed using a regression neural network, and the training dataset is processed with the RPN algorithm [23]. Neural networks are employed to locate and retrieve defected leaves. The algorithm achieves an accuracy of 83.57%. Subsequently, Sujatha et al. presented a model that employs transfer learning with pre-trained models such as VGG19, VGG16, and Inception-V3 [6]. The classifiers used include SVM, KNN, and Neural Networks. After evaluating their performance, it is determined that VGG19, in conjunction with logistic regression, exhibits relatively good results. Researchers have also explored disease prediction and classification using a hybrid approach combining convolutional neural networks and auto-encoders. The encoding part of the auto-

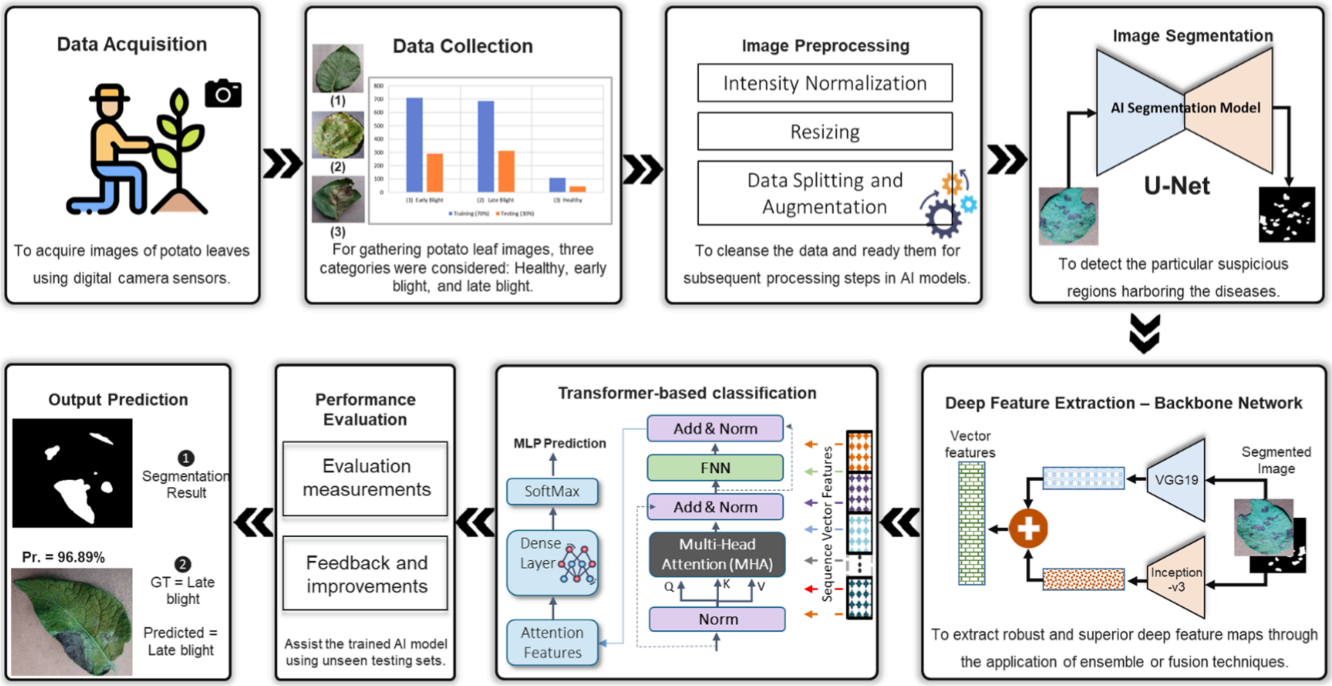


Fig. 1. Overview of the proposed PLDPNet AI framework providing an end-to-end perspective for predicting potato leaf diseases.

encoders is utilized to extract useful features. This work has been conducted on three crops: potato, tomato, and maize. Another system is proposed in [24] for the identification and classification of guava leaf diseases using D-CNN, along with pooling and fully connected layers. Softmax is employed as an activation function.

The researchers in [25] aimed to introduce a disease detection system using a deep learning algorithm. They began by acquiring and preprocessing the images, followed by implementing the CNN structure. Shima Ramesh and colleagues [26] proposed the identification of plant diseases using machine learning algorithms. They employed Random Forest (RF) classification and a feature extraction algorithm to classify diseased and healthy images. Subsequently, Lee et al. presented a system for potato disease detection based on the CNN network and image processing techniques, utilizing the PlantVillage dataset [27]. The image underwent noise removal through Gaussian filtering, followed by desired area extraction, color conversion, and implementation of CNN with ReLU activation function. The Adam optimizer was utilized, and resource and convolution layers were optimized for reduced usage without compromising accuracy. Additionally, Samer proposed a CNN-based system for identifying and classifying potato leaf diseases using the PlantVillage dataset, with 1700 images for training and 300 images for testing [14]. In [28], Gajjar et al. proposed a model that involved preprocessing image enhancement, feature extraction, and the use of a CNN model, which yielded favorable results compared to SVM. Talukder et al. introduced a model that incorporated image pre-processing, image segmentation, and the RF classifier, achieving superior results compared to SVM [17]. Similarly, Ashok et al. proposed a method for identifying and classifying tomato leaf diseases using the CNN algorithm and preprocessing techniques [29]. Finally, in recent research proposed by Lee et al. [30], a model utilizing the CNN algorithm, Adam optimizer, and cross-entropy analysis was developed, along with a training image database, leading to promising outcomes.

Venkataramana et al. employed the CNN and SVM models for classification, achieving a remarkable accuracy of 99.4% [31]. Joe et al. [32] proposed a RCNN model to predict the potato leaf diseases and its valuated based on precision and recall, yielding 98.1% and 81.9% respectively. Paul et al. proposed an enhanced approach for the Crow Search Algorithm (CSA) by incorporating dynamic awareness probability and implementing a Lévy flight movement to improve the exploration and exploitation phases of the original CSA [33]. This study addresses the challenge of rescheduling train timetables during disastrous situations to minimize journey time delay. They propose a Bat Algorithm (BA) approach that dynamically updates the timetable, considering multiple tracks and platforms. The proposed approach outperforms existing methods in terms of efficiency and optimization in this field was presented in [34]. Anim et al. proposed a model that utilizes ResNet50 and Saliency maps by applying data augmentation and hyperparameter tuning, the model achieves a precision of 99% and a recall of 99.3% [35]. Another work proposes the use of the RCNN and CBAM models, achieving an accuracy of 99% [36]. In 2022, Rabbia Mahum et al. proposed a model for the detection of potato leaf diseases. They employed the DesNet architecture with an additional transition layer, DesNet201, for classification. The model utilized the PlantVillage dataset and achieved an accuracy of 97.2% [9]. In 2023, Trishita et al. presented a new method that incorporates data augmentation, data partitioning, and data shuffling [37]. They applied three deep learning algorithms, SVM, VGG16, and CNN, which yielded accuracies of 87%, 92%, and 98% respectively. In the same year, Alok et al. proposed a model that utilizes a hierarchical deep learning convolutional neural network for disease detection and classification. They employed the intuitionist Fuzzy local binary pattern for feature extraction. This method achieved an accuracy of 95.77% [38]. Also in 2023, Olushola Olawuyi and Serestina Viriri proposed a model that employs ResNet50 for the detection and classification of plant diseases using the ImageNet dataset. This model achieved a precision of 77% and an accuracy of 98.0% [39]. In recent years, agriculture has remained a significant sector of our economy. Technological advancements and research have revealed the presence of various diseases in potato crops, leading to reduced yields and significant tuber losses. This issue poses a significant concern in the agricultural industry. Consequently, extensive research has identified numerous diseases affecting potato crops. The primary objective of this proposed research is to accurately detect the presence of blight in the crop and promptly identify and classify the specific diseases with high accuracy. To achieve this, a novel deep learning framework has been developed, combining the strengths of deep learning and machine learning approaches for classification. This hybrid approach

Table 1
Leaf Potato disease dataset distribution per class: Early Blight, Late Blight, and Healthy leaves.

Disease Class	Training (70%)	Testing (30%)	Total
(1) Early Blight	710	290	1,000
(2) Late Blight	687	313	1,000
(3) Healthy	109	43	152
Total	1,506	646	2,152

outperforms single-model approaches in terms of feature extraction. Moreover, the proposed framework includes image pre-processing and segmentation as essential stages to enhance disease prediction accuracy. These stages form the core foundation of the PLDPNet framework. The integration of this model can significantly contribute to improving potato crop productivity in the agriculture sector.

3. Material and methods

The PLDPNet framework proposes an end-to-end approach for predicting potato leaf diseases, consisting of two main sequential stages: segmentation and classification. To achieve segmentation, we adopt and employ the U-Net deep learning model, known for its ability to effectively leverage image structure and perform high-resolution segmentation. The model follows a "U" shaped architecture, with a contracting path for capturing context and an expanding path for precise localization. In the classification stage, we combine deep features extracted from two well-established deep learning CNN models (i.e., VGG19 and Inception-V3) based on the segmented regions of interest (ROIs) obtained from the segmentation step. By employing a fusion strategy, we enhance knowledge on diverse diseases and improve prediction performance. In general, the AI framework encompasses data acquisition and collection, image pre-processing, automatic segmentation using U-Net, ensemble deep feature extraction and fusion, and classification task using the transformer concept. Fig. 1 provides an overview of the proposed deep learning framework for predicting potato leaf diseases.

3.1. Dataset

To achieve the objectives of this study, we utilize the benchmark potato leaf dataset extracted from the publicly available PlantVillage dataset [40]. This dataset is collected by a collaboration between EPFL University of Switzerland and Penn State University of America. The dataset comprises both healthy and diseased leaves, aiming to develop

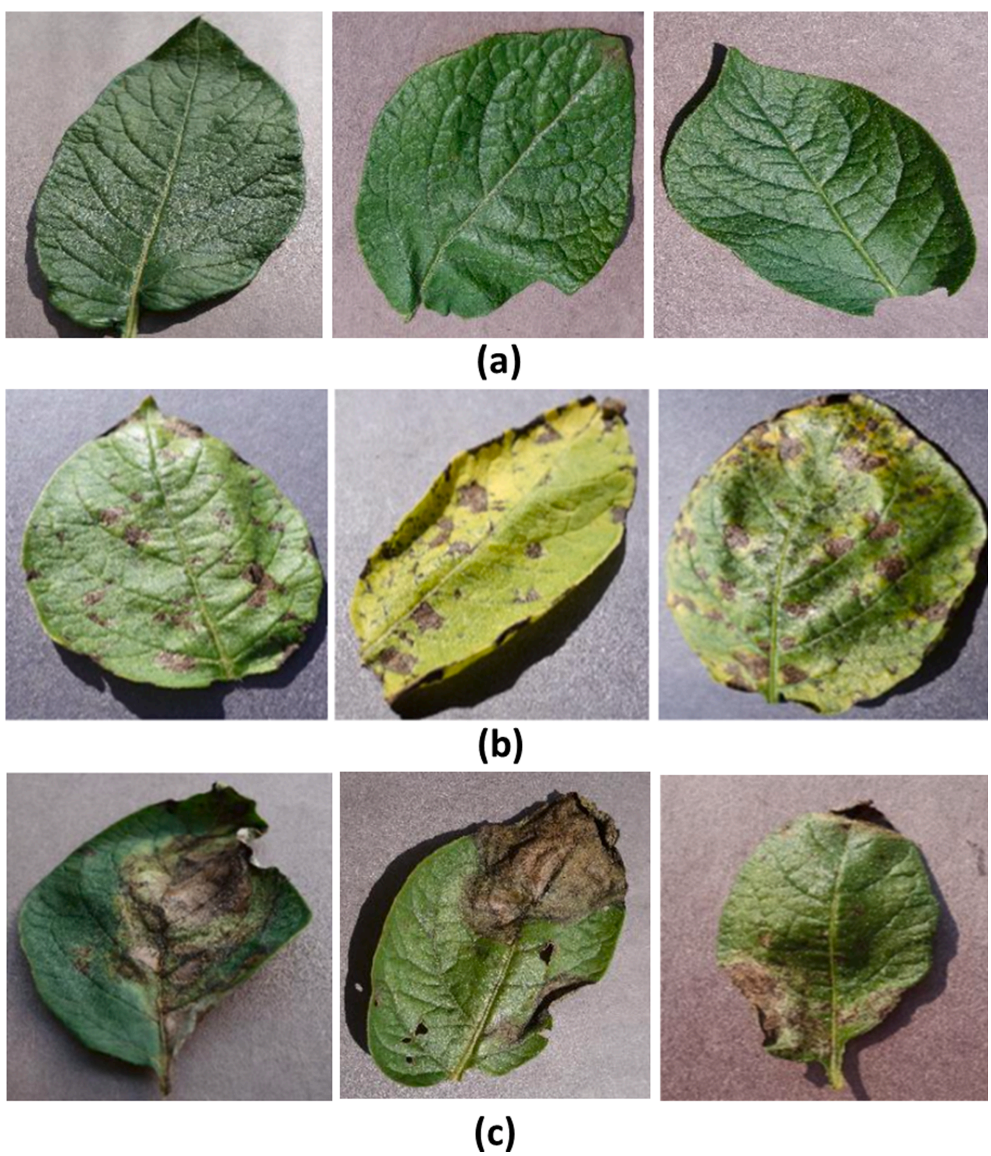


Fig. 2. Examples of potato leaves in different health conditions: (a) Healthy, (b) Early blight, and (c) Late blight.

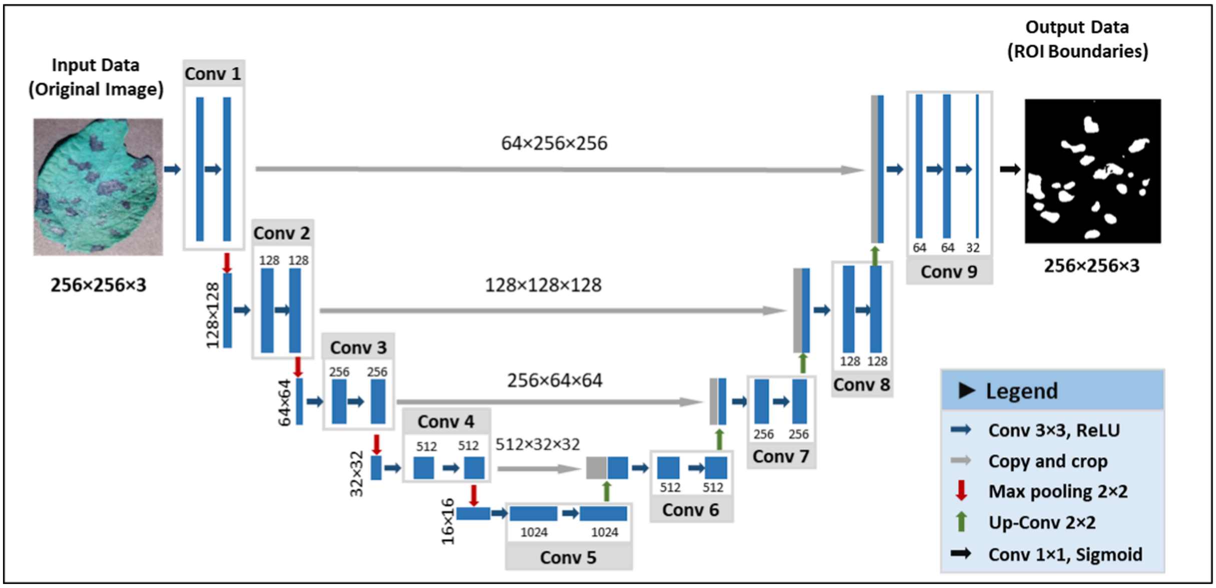


Fig. 3. The adopted U-Net segmentation model employed for segmenting the plant leaf diseases.

machine learning models capable of accurately identifying plant species and detecting diseases. Our research focuses specifically on potato leaf diseases, incorporating AI technology for multi-classification prediction. Internal expert annotators have provided segmentation labels (masks) to denote disease boundaries on the leaves, as well as classification labels for different diseases (classes). The experimental study of the proposed framework employs a randomly divided dataset, with 70% allocated for training and 30% for testing. Additionally, a validation set is derived from the training set, representing 10% of the data. Table 1 presents the distribution of images per class used in the experimental study.

Fig. 2 depicts some instances of potato leaf for healthy and disease classes (i.e., early blight and late blight). Early blight is a fungal disease that affects potato crops. Infected potato leaves exhibit circular lesions in shades of grayish white and dark brown. This fungal infection can impact tubers, leaves, and stems, leading to reduced tuber size, storability, yield, and crop marketability. On the other hand, potato late blight is another fungal disease that can infect both potato tubers and foliage at any stage of crop development. This fungus thrives in cool and moist conditions. Temperature ranges typically between 50 to mid-90s Fahrenheit promote disease progression in the field. If late blight infects potato leaves, circular patterns manifest along the leaf edges. The lesions rapidly expand, forming large, black, and dark brown areas, often with a greasy appearance.

3.2. Pre-processing

Image preprocessing techniques are utilized to effectively format the images for training machine learning models [41,42]. These techniques improve image quality and aid AI models in optimizing their trainable parameters [43–46]. To address the inherent variability in disease classification, preprocessing steps are applied to all images before any splitting. This includes intensity normalization, resizing, and annotation (for segmentation and classification tasks) to enhance disease prediction performance. Normalization adjusts pixel values to a fixed range of [0, 255], ensuring fair parameter training and overall prediction performance.

Furthermore, potato leaf images are resized to a standardized dimension of 256×256 pixels using bi-cubic interpolation [43]. AI models typically expect input images to have a specific size to ensure that all images are standardized, promoting consistent processing by the model. Additionally, resizing images to a smaller dimension reduces the computational requirements of the AI model. Smaller images necessitate fewer computational resources, resulting in faster and more efficient training and inference processes. After pre-processing and splitting the images into training and testing sets, data augmentation is exclusively performed on the training set to increase its size. This is crucial to meet the AI model's requirement for a large dataset, enhancing learning and optimizing trainable parameters. Consistent with previous research works [47–50], augmentation techniques incorporating photometric and geometric distortions are applied to expand the training image dataset. For photometric distortion, we empirically adjust the hue, saturation, and value of the images by 0.021, 0.8, and 0.43, respectively. Regarding geometric distortion, random scaling of 0.95, translation of 0.13, and rotation in the range of −0.40 to 0.40 rad are applied. Additionally, we employ the Mosaic and MixUp augmentation methods with probabilities of 1 and 0.1, respectively.

3.3. Potato leaf image automatic segmentation via U-Net

Image segmentation involves dividing an image into smaller segments representing different classes or objects within the image [51]. This technique allows for the analysis of specific regions, such as a leaf, facilitating accurate disease identification and classification [43,52]. By segmenting the image, we can isolate the relevant areas for analysis and avoid potential inaccuracies or false detections that may arise from analyzing the entire image. Moreover, segmentation helps remove background noise and irrelevant information, thereby enhancing the accuracy of disease detection systems. Recent literature has introduced various segmentation models based on convolutional networks [43,52], vision transformers (ViT) [53], and attention mechanisms. One prominent model is the U-Net, originally developed for biomedical image segmentation to identify infected areas rather than simply classifying the presence of infection [54]. The U-Net architecture comprises an Encoder and a Decoder. The Encoder encodes and generates precise high-level feature maps using 2D convolution and pooling layers, while the Decoder reconstructs the feature maps to match the original input image dimensions (spatial resolution). With its proven track record in image segmentation across computer vision and medical research domains, we have chosen the U-Net model for the potato leaf image segmentation task in our proposed framework. The primary goal of the U-Net in our framework is to predict the areas affected by blight. U-Net typically employs image-masks pairs for training and optimizing the loss function. It utilizes 3×3 convolutions with Rectified Linear Unit (ReLU) activation, complemented by batch normalization to enhance network

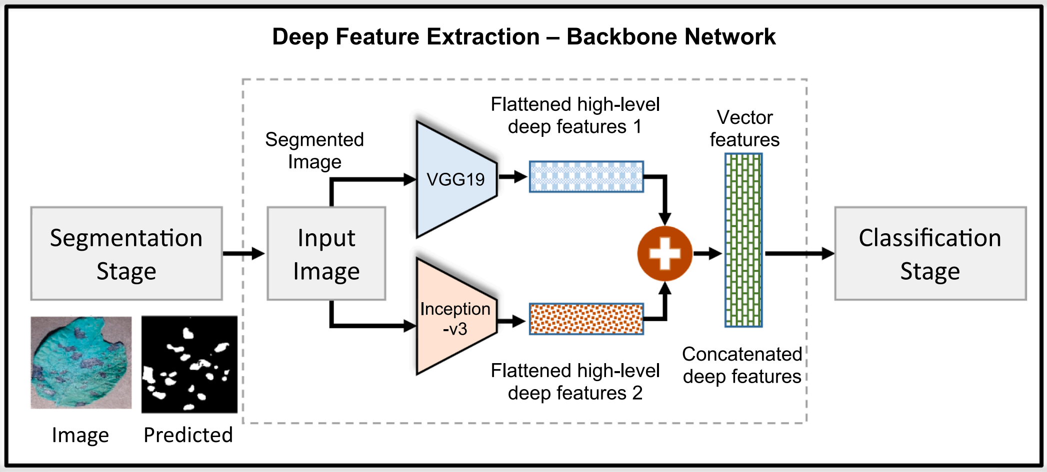


Fig. 4. High-level deep feature extraction using the feature concatenation strategy.

stability and performance. Max pool layers are employed to reduce the spatial resolution of feature maps during training when working with large input image sizes, effectively reducing the number of trainable parameters. As shown in Fig. 3, The U-Net model consists of an encoder, bridge, and decoder. The encoder gradually decreases image size while increasing depth. Whereas, the decoder increases image size while decreasing depth. In the U-Net model, a Sigmoid activation function is utilized to produce a binary mask as the output. The architecture captures image context through the contracting path and combines it with fine-grained details using feature map concatenation in the expanding path to achieve segmentation. Each blue box corresponds to a multi-channel feature map, with the channel count indicated at the top. The gray boxes represent copied feature maps, and the arrows depict various operations.

3.4. Deep feature extraction and fusion

As depicted in Fig. 4, feature concatenation is employed as an ensemble learning strategy to extract robust high-level deep features. The proposed model's backbone network is constructed based on VGG19 and Inception-V3. By concatenating features from multiple models or diverse sources, the ensemble model can access a larger feature space [41,42,45,46,55]. Each individual model within the ensemble captures different aspects or representations of the input data. Combining these features enables the ensemble model to possess greater capacity, potentially learning more complex relationships and patterns. Moreover, the feature fusion strategy is generally more resilient to outliers, noise, and data variations. Aggregating features from multiple models helps mitigate the impact of individual model errors or biases, enhancing robustness when working with real-world data that may contain anomalies or uncertainties. Furthermore, the feature ensemble strategy, through the combination of diverse models or features, often yields superior generalization compared to individual models. Each model in the ensemble contributes its strengths and weaknesses, and their predictions are combined via feature concatenation to mitigate biases or errors [45]. This approach results in improved overall performance, particularly when handling complex or noisy datasets. In essence, utilizing different models or feature sources captures complementary information about the same input data. By combining these features through concatenation, the ensemble model leverages diverse perspectives, leading to a more comprehensive representation of underlying patterns and an enhanced understanding of the data.

3.5. The proposed hybrid AI-based classification

$$Attention(Q, K, V) = Softmax\left(\frac{Qk^t}{\sqrt{d_k}}\right)v \tag{1}$$

As shown in Fig. 1, the construction of the hybrid AI-based framework for potato leaf prediction involves leveraging the benefits of ensemble transfer learning and incorporating the Vision Transformer (ViT) with multiple heads for classification. The ensemble model's backbone combines high-level features from VGG19 and Inception-V3, followed by average pooling to convert the 3D features into a 2D space. Data reshaping and resizing ensure compatibility for merging the feature vectors into a single space vector called Vector features. This latent vector serves as input for the classification stage using ViT. Indeed, the ViT is chosen for its precise object detection based on valuable, derived features, while self-attention features contribute to excellent performance and reduced reliance on vision-specific biases [41,42]. Self-attention is the core component of ViT, assigning varying weights to determine the importance of each input data point in an encoder-decoder configuration. Unlike CNN models focusing on adjacent pixels within the receptive field, they struggle with distant pixel relationships. To overcome this limitation, the attention mechanism emerges as a new approach, identifying informative parts of input images, discarding redundancy, and minimizing false-negative results. For this work, we adopt and retrain the ViT based on the derived feature vector from the backbone ensemble network. The proposed transformer encoder includes a multi-head self-attention (MHA) layer, normalization layers, classification heads with two dense layers, and a Softmax regression function for final potato leaf disease classification. Fig. 1 illustrates the linear concatenation of 16×16 2D patches into 1D vectors for both AI models, passing through the transformer encoder with multi-head self-attention (MSA) and MLP blocks. To establish the relationship between each patch and all other patches within a single input sequence, the MHA employs the dot product form of attention, defined as:

$$\mathrm{MultiHead}(\mathrm{Q},\ \mathrm{K},\ \mathrm{V}) = \mathrm{Concat}(\mathrm{head}_1,\ \cdots,\ \mathrm{head}_h)\ W^o$$

where

$$\mathrm{headi} = \mathrm{Attention}(\mathrm{QW}_i^Q,\ \mathrm{KW}_i^K,\ \mathrm{VW}_i^V) \tag{2}$$

Here, Q, V, and K represent the query vector, value dimensional vector, and key vector, respectively. The term d_k denotes the variance of the product Qk^t, which has a zero-mean. Additionally, the product is normalized by dividing it by the $\sqrt{d_k}$. The SoftMax function converts the scaled dot product into an attention scores. Such mechanism forms the core of ViT module, providing parallel attention to comprehend the overall content of the input potato leaf images. Through multi-head attention, the model can attend to input from multiple representation subspaces and different positions simultaneously. Multi-head attention linearly extends the queries, keys, and values h times using various learned linear projections. The calculation can be expressed as:

where the projections indices are $W_i^Q \in R^{d_{model}\ x\ d_k}$, $W_i^K \in R^{d_{model}\ x\ d_k}$, $W_i^V \in R^{d_{model}\ x\ d_v}$ and $W^o \in R^{hd_v\ x\ d_{model}}$. The MLP block consists of a non-linear layer with Gaussian error linear unit (GeLU) activation, comprising 1,024 neurons, as well as normalization and dropout layers with a 50% dropout rate.

3.6. Experimental settings

3.6.1. Segmentation settings

For the purpose of automatically identifying the area of blight on potato leaves, we employ the U-Net model for image segmentation. The training of the proposed model is conducted using a 100 epochs and minimum batch size of 32, which is chosen to accommodate GPU memory limitations. Additionally, all potato leaf images are resized to a standardized dimension of 256×256 pixels. Furthermore, a normalization process is applied to adjust the image pixel values to a fixed range of [0, 255]. This ensures fair training of the model's parameters and enhances the overall performance of disease prediction. Moreover, Adam optimizer is used for training the model. A binary cross-entropy loss function is used for training and validating the segmentation U-Net model.

3.6.2. Classification settings

During training, the multi-scale training strategy is employed to facilitate prediction learning across various resolutions of input images. In the inference stage, predictions are generated for unseen datasets that were previously provided. In this proposed work, inference is performed to determine whether the leaf disease is classified as late blight or early blight. Additionally, a mini-batch size of 8 and a training duration of 100 epochs are utilized to train and validate the proposed AI prediction framework.

3.7. Environmental execution

To execute the proposed PLDPNet framework, the experimental studies are conducted on a PC featuring the following specifications: an Intel(R) Core(TM) i7-10700KF CPU @ 3.80 GHz, 32.0 GB of RAM, six CPUs, and one NVIDIA GeForce RTX 3060 GPU.

3.8. Evaluation metrics

To assess the proposed framework, the primary evaluation metrics of accuracy, precision, recall, and F1-score are employed. Accuracy provides an overall measure of the model's ability to correctly predict cases from each class. Precision evaluates the model's capability to avoid labeling a negative instance as positive, calculated as the ratio of true positives to the sum of true positives and false positives. Recall gauges the model's ability to identify positive instances, represented by the ratio of true positives to the sum of true positives and false negatives. Finally, Lastly, the F1-score is a weighted harmonic mean of recall and precision, indicating the percentage of correct positive predictions. All these metrics assess the performance of both the segmentation and classification stages. In the case of segmentation, these measures are computed at a pixel-wise level, whereas for classification evaluation, they are computed at the image-level. In addition, the mean intersection over union (mIoU) index is used to evaluate only the segmentation performance. The mathematical definitions of these metrics are derived from the multi-classification confusion matrix, which are presented as follows,

Table 2
Segmentation evaluation weighted metrics of AI models that used to segment the potato leaf diseases (i.e., early blights and late blights). The normal cases are not involved for the segmentation process.

Segmentation AI Model	Evaluation Metrics (%)				
	Accuracy	Precision	Recall	F1-score	mIoU
U-Net	96.80	97.01	96.02	96.56	94.06
U^2-Net	95.20	95.05	93.25	93.50	91.15
FCN	89.18	90.72	89.60	88.54	89.87
SegNet	84.60	84.65	85.65	87.77	85.25
DeepLabv3+	83.03	83.65	84.95	87.09	84.96

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + +\text{FP} + \text{FN}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{F1} - \text{score} = 2 \times \frac{(\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})}, \tag{6}$$

$$\text{mIoU} = \frac{|\text{X} \cap \text{Y}|}{|\text{X} \cup \text{Y}|} \tag{7}$$

where *TP*, *FP*, *FN* and *TN* indicate the true positive, false positive, false negative, and true negative values in the confusion matrix, respectively. The mIoU index is specifically used for segmentation evaluations, quantifying the similarity between the ground-truth (GT) labeled mask and the corresponding segmented or predicted mask. The quantities $|X|$ and $|Y|$ represent the number of pixels in the predicted and labeled binary masks, respectively.

4. Experimental results and discussion

4.1. Segmentation results

For the segmentation task, we employ well-acceptable deep learning models including U-Net [54], U^2-Net [51], fully convolutional network (FCN) [52], SegNet [43,52], and DeepLabv3 + [51]. This allows for a direct comparison using the same input dataset, training conditions, settings, and execution environment. The purpose of incorporating the pre-segmentation stage in our proposed PLDPNet is to focus on regions suspected of having diseases and reduce bias towards healthy pixels that may exist between unhealthy ones. This targeted approach highlights deep features that can be extracted solely from those unhealthy regions. However, in the ablation study, we compare the final classification results with and without the image segmentation step to ensure the significance of the segmentation pre-processing in achieving improved classification outcomes. Prior to segmentation, the images are resized to fit the required dimensions of each segmentation model, while the training environment settings remain consistent. The segmentation stage is specifically used to segment the abnormal instances, such as potato leaves with early and late blight diseases, while the normal instances are directly passed to the feature extraction stage without any segmentation. By incorporating the normal class, we enhance the classification performance and establish a comprehensive AI framework capable of simultaneously predicting both normal and abnormal cases.

Table 2 shows the segmentation evaluation results of U-Net, U^2-Net, FCN, SegNet, and DeepLabv3 +. It is clear show that the segmentation results of U-Net outperform other models.

Considering the results obtained from the experimental evaluation,

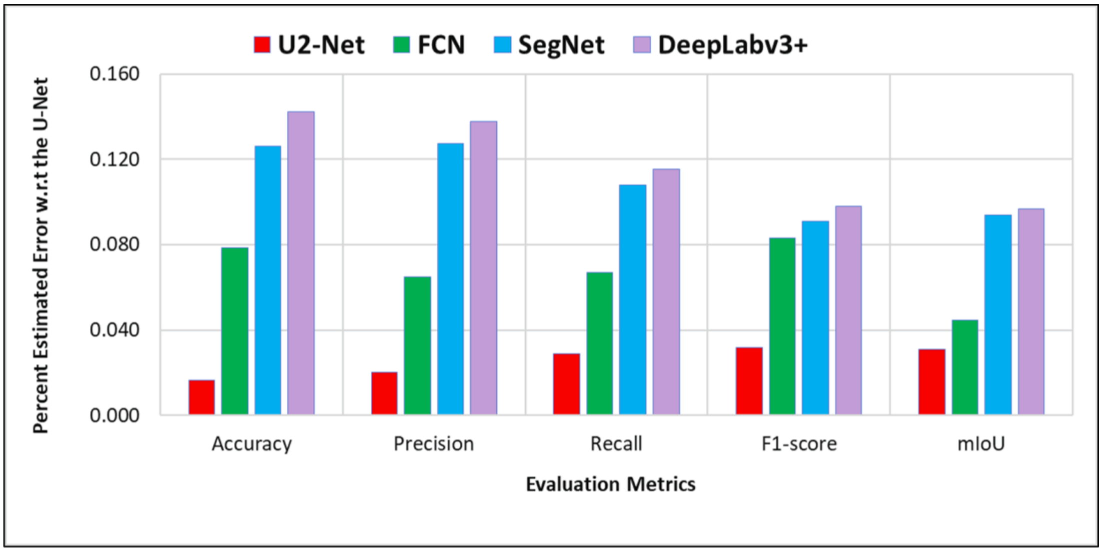


Fig. 5. Percent estimated error with respect to the best AI segmentation U-Net model.

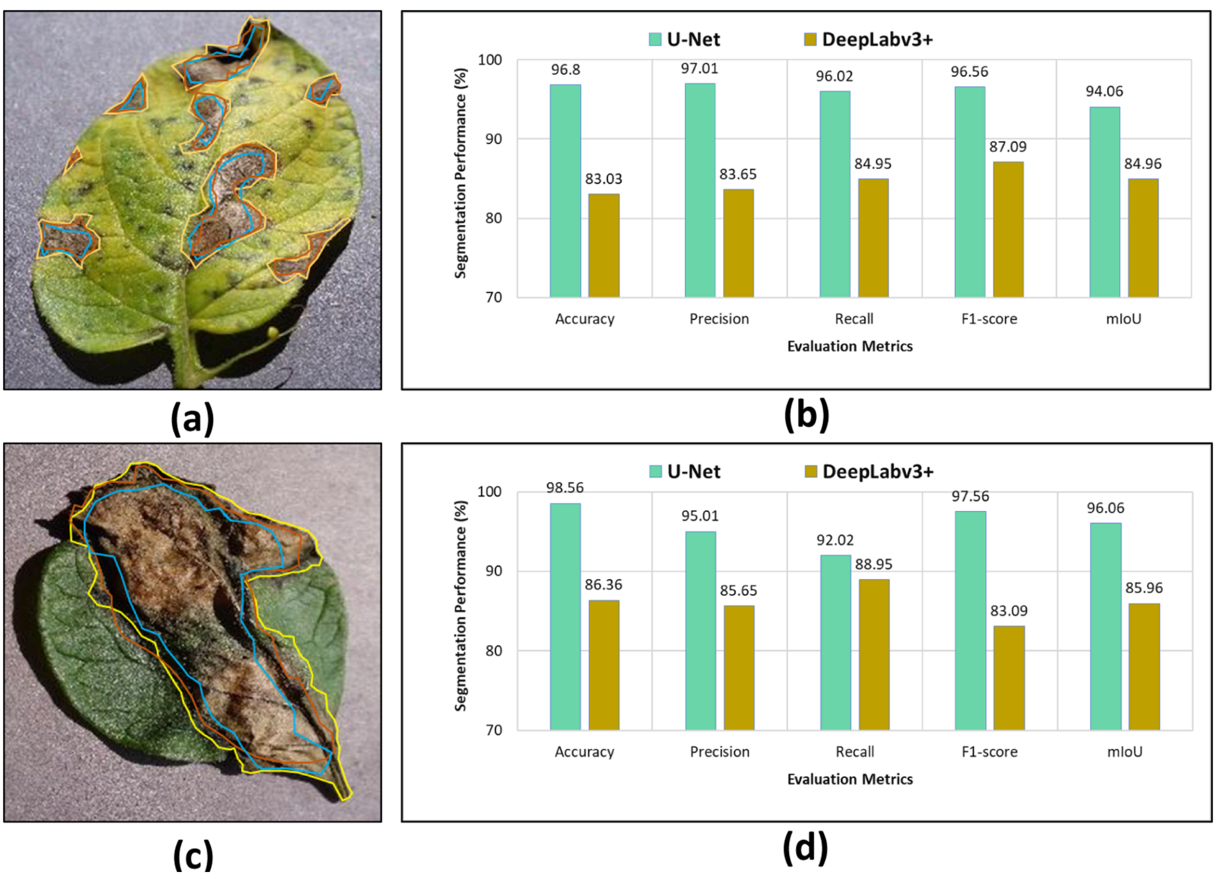


Fig. 6. Examples of visual segmentation results for comparison between the best and worst segmentation models employed in this study. (a) and (c) depict input potato leaf images corresponding to early and late blight classes, respectively. The disease regions are enclosed by boundary shapes, denoting the ground truth segmentation label (Yellow), regions predicted by U-Net (Orange), and regions predicted by DeepLabv3+ (Blue). (b) and (d) illustrate the evaluation metrics per image presented in (a) and (c) for both U-Net and DeepLabv3+, respectively.

the U-Net model demonstrated superior performance compared to other AI segmentation models. It achieved an overall weighted accuracy of 96.80%, surpassing, U^2-Net, FCN, SegNet, and DeepLabv3 + by 1.6%, 7.62%, 12.2%, and 13.77%, respectively. Taking into account all the evaluation metrics, Fig. 5 displays the percentage of estimated error in comparison to the best-performing U-Net segmentation model. It is evident that the U^2-Net demonstrates the closest performance to the U-Net, whereas DeepLabv3 yields the poorest segmentation results. The variation in segmentation performance can be attributed to differences in the deep structure of the AI models, the number of trainable parameters, or the inherent characteristics of the data. Addressing this challenge, AI models strive to generate unique solutions for diverse data

Table 3
Classification prediction evaluation results per class for potato leaf diseases prediction of the proposed PLDPNet. Only unseen testing images are used for the evaluation task.

Segmentation AI Model	Evaluation Metrics (%)			
	Accuracy	Precision	Recall	F1-score
Healthy	98.92	91.0	93.0	92.0
Early Blight	98.76	99.0	98.0	99.0
Late Bright	98.30	98.0	98.0	98.0
Average	**98.66**	**96.0**	**96.33**	**96.33**

types. To summarize the segmentation step, we choose to utilize the superior U-Net model to construct the proposed end-to-end deep learning framework, PLDPNet. Additionally, Fig. 6 illustrates visual examples of the superior and inferior segmentation outcomes obtained through the utilization of U-Net and DeepLabV3 + models for both early and late blight classes. For the sake of visibility and clarity, we have omitted the results of other models. The evaluation results for each image demonstrate the superior performance achieved when employing the U-Net model.

4.2. Classification results

4.2.1. Classification results of the proposed PLDPNet

To assess the classification stage, the segmented images are directly inputted into the feature extraction process to obtain high-level deep features. To achieve this, two popular deep AI models, namely VGG19 and Inception-v3, are utilized. The ViT with multiple head prediction is utilized for the ultimate prediction of potato leaf diseases. The classification evaluation outcomes for each class are consolidated in Table 3.

As indicated in Table 3, the proposed model achieves the highest classification results for potato leaf diseases, with an average of 98.66%, 96.0%, 96.33%, and 96.33% for overall classification accuracy, precision, recall, and F1-score across all classes. The lowest results are observed for the healthy class, which can be attributed to the limited number of images in that category. Notably, the F1-score serves as the most reliable measure for evaluating the overall performance, considering its ability to address class imbalances. However, in this study, weighted evaluation metrics are utilized to demonstrate the contribution rate of each class. All evaluation metrics are calculated using the multi-classification confusion matrix depicted in Fig. 7**(a)**. Fig. 7**(b)** showcases the confusion matrix when the segmentation module is excluded from the PLDPNet pipeline.

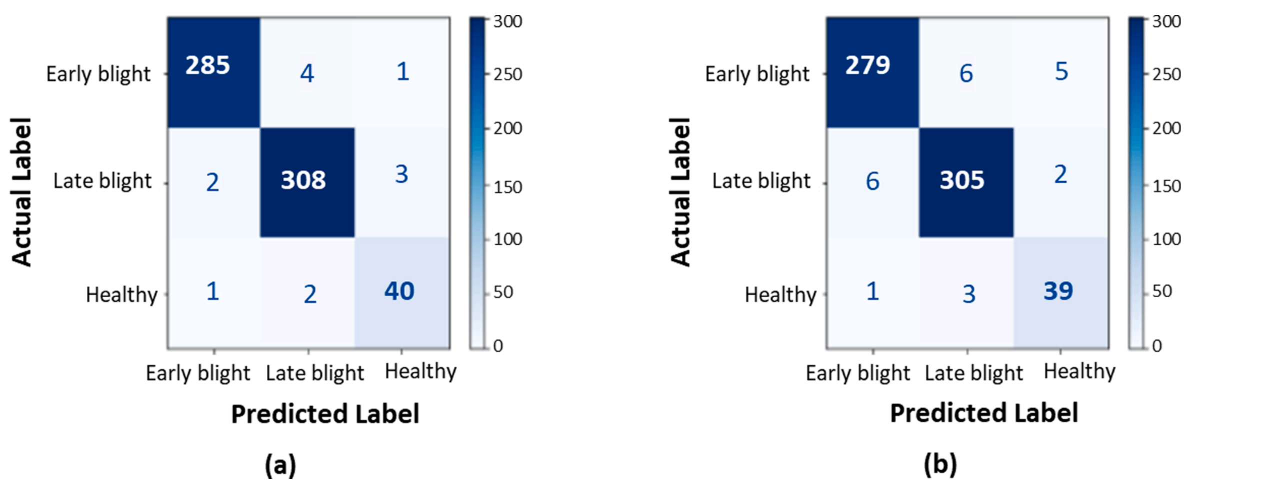


Fig. 7. Classification evaluation performance of the proposed PLDPNet represented by confusion matrices: (a) with segmentation, and (b) without segmentation.

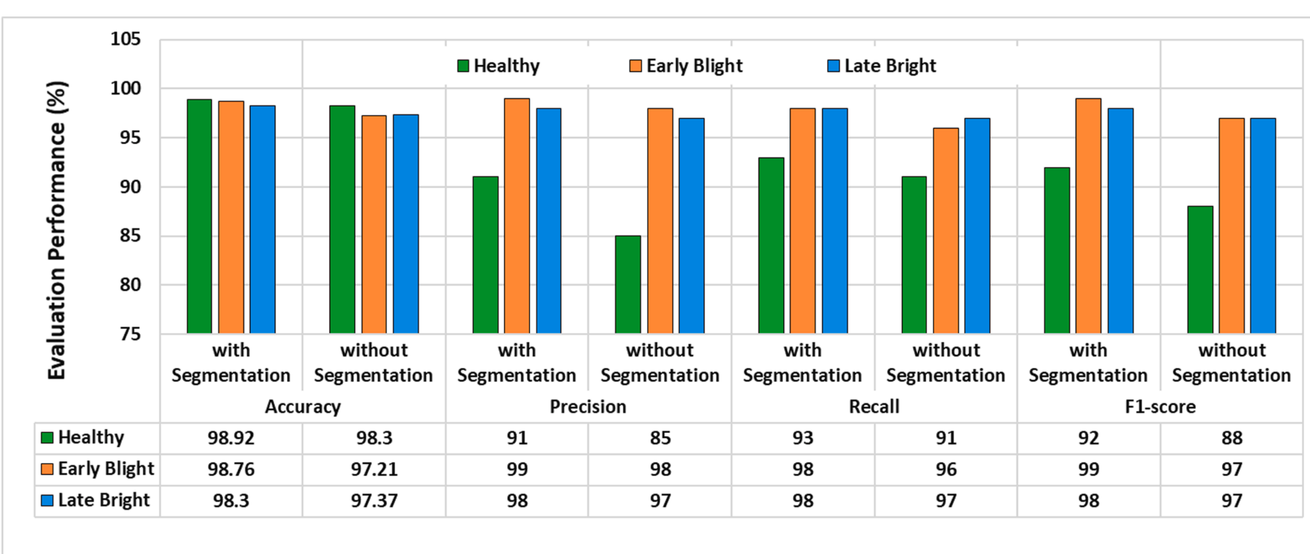


Fig. 8. Comparison Classification performance of the proposed PLDPNet with and without segmentation module.

Table 4
Classification outcomes for each individual transfer learning model employed in selecting the backbone network. The segmentation and ViT-based classification models are employed in these experiments.

Segmentation AI Model	Evaluation Metrics (%)			
	Accuracy	Precision	Recall	F1-score
VGG19	95.87	88.33	93.0	90.33
Inception-V3	95.87	87.67	91.0	89.0
DenseNet201	95.05	86.0	90.0	88.0
Xception	94.43	83.67	88.67	86.33
ResNet50	94.32	83.33	90.33	86.0
Proposed Ensemble Model (VGG19 + Inception-V3)	**98.66**	**96.0**	**96.33**	**96.33**

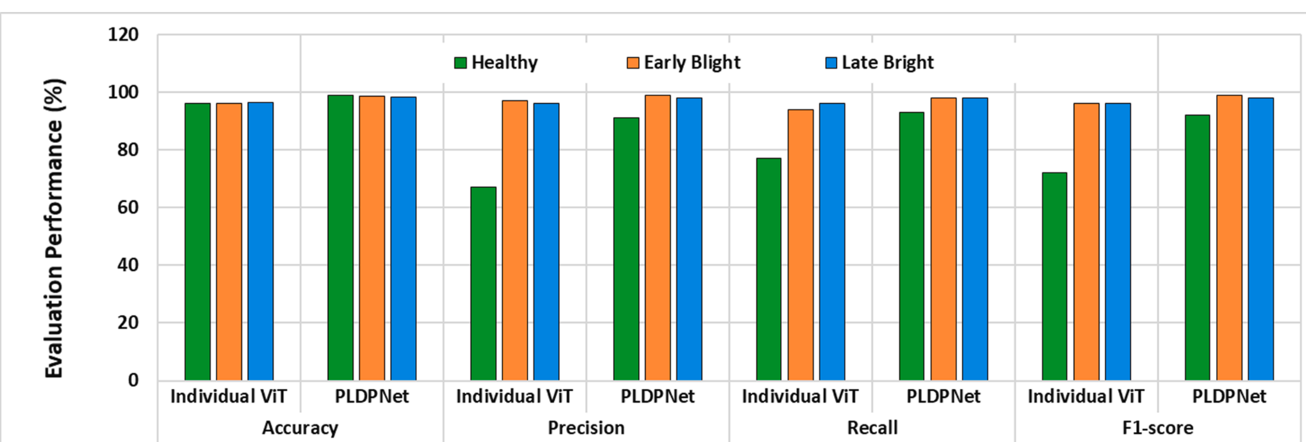
4.2.2. Ablation study

In this section, we explore the impact of different modules and settings in the proposed framework to demonstrate their contribution to the overall PLDPNet framework. This investigation is crucial to assess the significance of each module within the entire framework. Each experimental study presents the evaluation results when excluding at least one module, providing insights into the overall prediction performance. We use the same training settings and environment to train all models for conducting the ablation study.

4.2.2.1. The impact of the segmentation module. The evaluation results of our proposed PLDPNet framework with and without the segmentation module are compared in Fig. 8. The results highlight the significant contribution of the segmentation process, leading to improvements in overall accuracy by 0.62%, 1.55%, and 0.93% for the healthy, early blight, and late blight classes, respectively. Similarly, the improvement rates are clearly recorded for healthy, early blight, and late blight classes in terms of 4.0%, 2.0%, and 1.0%, respectively.

4.2.2.2. The impact of the ensemble fused feature extraction. We evaluate the final classification results by considering different options for the backbone network, which generates deep features in our proposed framework. The AI models are fine-tuned and evaluated using the same training settings, training environment, and same dataset distribution for training, testing, and evaluation. Table 4 presents the comparative outcomes obtained by utilizing five distinct transfer learning models as the backbone network. Given the superior classification performance of VGG19 and Inception-V3 networks, we select them as the feature extractors for our PLDPNet using an ensemble or fusion strategy. Both segmentation and ViT-based classification models are employed in these experiments. Results clearly demonstrate that the ensemble module outperforms individual ones. Compared to the superior VGG19 model, the PLDPNet shows improvements of 2.97%, 7.67%, 3.33%, and 6.03% in terms of accuracy, precision, recall, and F1-score respectively.

4.2.2.3. The proposed PLDPNet against the individual ViT. As a classification comparison study with respect to various scenarios for our proposed PLDPNet framework, we use the fusion transformer to learn the global information directly from the input images without segmentation and deep feature extraction modules. This is check and compare the capability of the individual ViT for disease prediction. In this scenario, the preprocessed images undergo a division into a grid of equally sized patches. Each patch is considered a token, similar to a word token in NLP. These patches are then flattened to form a sequence of 1D vectors, known as latent vectors. Unlike CNNs, Transformers do not inherently encode positional information. To address this, ViT incorporates position information into the input sequence by adding position embeddings to each patch vector. This embedding indicates the spatial location of the patch within the image. The sequence of patch vectors, along with their respective position embeddings, is passed through an initial linear projection layer to convert the patch vectors into a higher-dimensional representation, enabling the model to capture more complex patterns. Next, the encoded sequence undergoes processing by a stack of Transformer encoder layers. Each encoder layer comprises a multi-head self-attention mechanism and a feed-forward neural network. The self-attention mechanism enables the model to focus on different parts of the image and capture relationships between patches. The feed-forward network applies non-linear transformations to the patch representations. Following the Transformer encoder layers, a classification head is appended to the encoded sequence. Typically, this head includes fully connected layers with a softmax activation at the top to perform the classification task. We compare the evaluation prediction results in terms of all evaluation metrics and present it in Fig. 9. To conduct this study, the training settings and datasets are utilized in a similar manner to ensure a fair comparison. The results shown in Fig. 9 approve the capability of the PLDPNet in improving the overall classification performance by 2.48%, 9.33%, 7.33%, 8.33% and in terms of accuracy,



	Accuracy: Individual ViT	Accuracy: PLDPNet	Precision: Individual ViT	Precision: PLDPNet	Recall: Individual ViT	Recall: PLDPNet	F1-score: Individual ViT	F1-score: PLDPNet
Healthy	95.98	98.92	67	91	77	93	72	92
Early Blight	96.13	98.76	97	99	94	98	96	99
Late Bright	96.44	98.3	96	98	96	98	96	98

Fig. 9. Comparison Classification performance of the proposed PLDPNet against the individual ViT using the same training and testing settings. This study is conducted for the same dataset of training, validation, and testing.

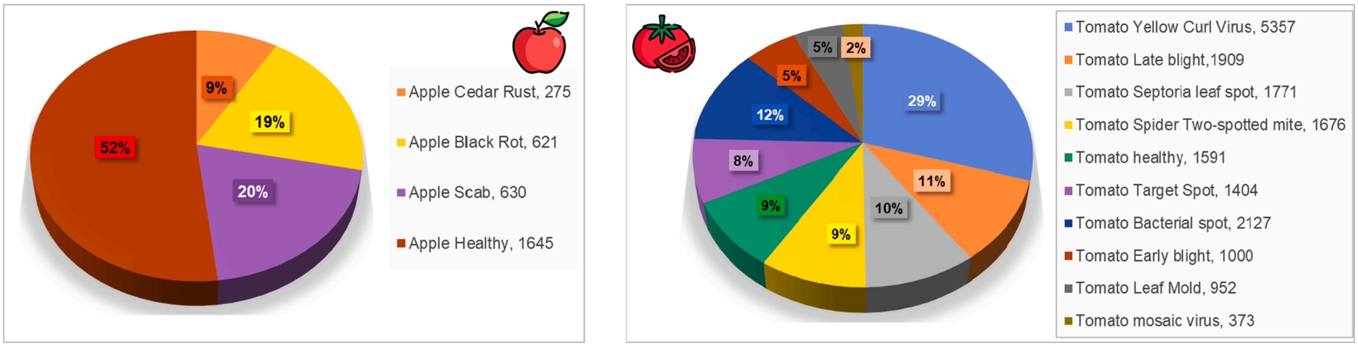


Fig. 10. Validation dataset of (a) apple (4 classes) and (b) Tomato (10 classes). The class name and image count are overlaid accordingly. These datasets are used to separately validate the proposed PLDPNet on the same execution training environment and training settings.

Table 5
Validation classification results of the PLDPNet using potato, apple, and tomato leaf diseases datasets. The apple and Tomato datasets are used as a testing sets.

Dataset	Evaluation Metrics (%)			
	Accuracy	Precision	Recall	F1-score
Potato leaf diseases (3 classes)	97.63	93.33	94.67	94.0
Apple leaf diseases (4 classes)	96.42	92.77	93.26	92.78
Tomato leaf diseases (10 classes)	94.25	90.08	63.82	93.89

precision, recall, and F1-score.

4.2.2.4. Validate the proposed PLDPNet for other plant diseases. To assess the effectiveness of the proposed PLDPNet in predicting leaf diseases in other plants such as tomatoes and apples, we perform an ablation study using the PlantVillage tomato and apple leaf disease datasets. The tomato dataset comprises ten classes with a total of 18,160 images, while the apple dataset consists of four classes with 3,171 images. Fig. 10 shows the apple and Tomato data distribution indicating the number of images for each class. We evaluate the performance of the PLDPNet framework on each dataset separately to assess its capability in predicting diseases in different plants. The neurons of classification head are addopted slightly according to the number of classes. Since the binary segmentation masks for disease localization are not available in these datasets (i.e., tomato and apple), we conduct these experiments without the segmentation module (U-Net). The average evaluation results of the PLDPNet using each dataset individually are compared in Table 5.

4.2.3. Comparison results with other state of the art methodologies

Indeed, the direct comparison among multiple AI models using the same training settings, execution environment, and training/testing dataset is the best approach. We have done such comparison in section 4.2.2.2. On the other hand, the indirect comparison is also acceptable if we consider the final prediction results regardless the training environments or data split. In this section, we compare the evaluation results of the proposed framework indirectly with other state-of-the-art published methods. Table 6 summarizes the indirect comparison in terms of all evaluation metrics even for different plant datasets.

4.3. Limitations and future work

The limited availability of labeled plant diseases limits our ability to train the proposed model comprehensively, encompassing all modules: segmentation, feature extraction, and classification. To assess the effectiveness of the proposed AI-based model in predicting plant diseases, even those unseen during training, we conducted an ablation study on tomato and apple datasets. Incorporating the segmentation module posed a challenge but yielded improved prediction results. However, training each module proved time-consuming due to numerous ablation studies. In the future, we aim to continue labeling datasets for the remaining classes and reassess the model's performance. Additionally, employing explainable AI could be advantageous, as it would provide visual heat maps that uncover the inner workings of the black box AI model, highlighting the specific areas it focuses on to generate reliable predictions. The objective for our future endeavors is to merge various and diverse datasets, enabling the model to predict a wide range of diseases across various plant species.

5. Conclusion

Early plant prediction plays a critical role in saving crops and enhancing global quality of life. This study introduces a novel AI-based framework for predicting potato leaf diseases in a multi-classification scenario: healthy, early blight, and late blight. Additionally, we evaluate the execution and prediction performance of the PLDPNet

Table 6
Indirect comparison evaluation results of the proposed PLDPNet framework against the other existing AI models in the literature.

Reference, (Year)	Target Plant	Dataset	AI Model	Evaluation Metrics (%)			
				Accuracy	Precision	Recall	F1-score
Oppenheim et al., (2017) [56]	Potato	PlantVillage	VGG16	96.0	–	–	–
Kurmi et al., (2022) [1]	Potato	PlantVillage	CNN	95.35	–	–	–
Wang et al., (2017) [57]	Apple	PlantVillage	VGG16	90.40	–	–	–
Mishra et al., (2020) [5]	Corn	PlantVillage	DCNN	88.46	–	–	–
Kurmi et al., (2022) [1]	Tomato	PlantVillage	CNN	92.60	–	–	–
Sujatha et al., (2021) [6]	Citrus	Private	Inception-v3	89.0	89.20	–	89.0
			VGG-16	89.50	89.60	–	89.50
			VGG-19	87.40	87.70	–	87.40
Subetha et al., (2021) [58]	Apple	Kaggle	VGG-19	87.70	–	–	–
Ours, (2023)	Potato (End-to-End)	PlantVillage	The proposed PLDPNet	98.66	96.0	96.33	96.33
	Apple (Validation without Segmentation)			96.42	92.77	93.26	92.78
	Tomato (Validation without Segmentation)			94.25	90.08	63.82	93.89

framework using apple and tomato datasets. Fortunately, our proposed model demonstrates remarkable prediction results by incorporating segmentation, ensemble feature extraction, and classification through the latest ViT with its multiple heads. We believe this model holds great promise as a candidate for further testing with different datasets, offering support to the financial sector in protecting livelihoods and providing smart service solutions to farmers as well.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2022-0-00739, MLOps Platform for Machine learning pipeline automation). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2022-00166402).

References

[1] Y. Kurmi, P. Saxena, B.S. Kirar, S. Gangwar, V. Chaurasia, A. Goel, Deep CNN model for crops' diseases detection using leaf images, Multidim. Syst. Sign. Process. 33 (2022) 981–1000.
[2] S. A. Raza, Y. Ali, and F. Mehboob, "Role of agriculture in economic growth of Pakistan," 2012.
[3] PotatoPro. (2023, [Accessed: June, 2032]). *PotatoPro - Everything you need to know about potatoes*. Available: https://www.potatopro.com.
[4] T. Shami, S. Adil, S. Hassan, M. Bashir, Spatial market integration and price transmission in major potato markets of Punjab, Pakistan, Indian J. Sci. Technol. 13 (2020) 2328–2335.
[5] S. Mishra, R. Sachan, D. Rajpal, Deep convolutional neural network based detection system for real-time corn plant disease recognition, Procedia Comput. Sci. 167 (2020) 2003–2010.
[6] R. Sujatha, J.M. Chatterjee, N. Jhanjhi, S.N. Brohi, Performance of deep learning vs machine learning in plant leaf disease detection, Microprocess. Microsyst. 80 (2021), 103615.
[7] N.E.M. Khalifa, M.H.N. Taha, L.M. Abou El-Maged, A.E. Hassanien, Artificial intelligence in potato leaf disease classification: a deep learning approach, Machine Learn. Big Data Analytics Paradigms: Analysis, Applications and Challenges (2021) 63–79.
[8] H. Afzaal, A.A. Farooque, A.W. Schumann, N. Hussain, A. McKenzie-Gopsill, T. Esau, et al., Detection of a potato disease (early blight) using artificial intelligence, Remote Sens. (Basel) 13 (2021) 411.
[9] R. Mahum, H. Munir, Z.-U.-N. Mughal, M. Awais, F. Sher Khan, M. Saqlain, et al., "A novel framework for potato leaf disease detection using an efficient deep learning model," *Human and Ecological Risk Assessment: an*, Int. J. (2022) 1–24.
[10] C.C. Bonik, F. Akter, M.H. Rashid, A. Sattar, "A convolutional neural network based potato leaf diseases detection using sequential model," in, Int. Conference for Adv. Technol. (ICONAT) 2023 (2023) 1–6.
[11] J. Rashid, I. Khan, G. Ali, S.H. Almotiri, M.A. AlGhamdi, K. Masood, Multi-level deep learning model for potato leaf disease recognition, Electronics 10 (2021) 2064.
[12] C. Xu, Z. Liu, P. Li, J. Yan, L. Yao, Bifurcation mechanism for fractional-order three-triangle multi-delayed neural networks, Neural Process. Lett. (2022) 1–27.
[13] C. Xu, D. Mu, Z. Liu, Y. Pang, M. Liao, C. Aouiti, New insight into bifurcation of fractional-order 4D neural networks incorporating two different time delays, Commun. Nonlinear Sci. Numer. Simul. 118 (2023), 107043.
[14] S. I. Mohamed, "Potato Leaf Disease Diagnosis and Detection System Based on Convolution Neural Network," 2020.
[15] G. Athanikar, P. Badar, Potato leaf diseases detection and classification system, Int. J. Comput. Sci. Mob. Comput. 5 (2016) 76–88.
[16] C. U. Kumari, S. J. Prasad, and G. Mounika, "Leaf disease detection: feature extraction with K-means clustering and classification with ANN," in *2019 3rd international conference on computing methodologies and communication (ICCMC)*, 2019, pp. 1095-1098.
[17] M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," in *2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE)*, 2017, pp. 1-4.
[18] P. Barré, B.C. Stöver, K.F. Müller, V. Steinhage, LeafNet: a computer vision system for automatic plant species identification, Eco. Inform. 40 (2017) 50–56.
[19] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep neural networks based recognition of plant diseases by leaf image classification, Comput. Intell. Neurosci. 2016 (2016).
[20] G. Li, Z. Ma, and H. Wang, "Image recognition of grape downy mildew and grape powdery mildew based on support vector machine," in *Computer and Computing Technologies in Agriculture V: 5th IFIP TC 5/SIG 5.1 Conference, CCTA 2011, Beijing, China, October 29-31, 2011, Proceedings, Part III 5*, 2012, pp. 151-162.
[21] V. Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques, Information processing in Agric. 4 (2017) 41–49.
[22] D. Argüeso, A. Picon, U. Irusta, A. Medela, M.G. San-Emeterio, A. Bereciartua, et al., Few-Shot Learning approach for plant disease classification using images taken in the field, Comput. Electron. Agric. 175 (2020), 105542.
[23] D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel, and S. Bhardwaj, "Potato leaf diseases detection using deep learning," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 461-466.
[24] M. R. Howlader, U. Habiba, R. H. Faisal, and M. M. Rahman, "Automatic recognition of guava leaf diseases using deep convolution neural network," in *2019 international conference on electrical, computer and communication engineering (ECCE)*, 2019, pp. 1-5.
[25] M.A. Jasim, J.M. Al-Tuwaijari, "Plant leaf diseases detection and classification using image processing and deep learning techniques," in, Int. Conference on Comput. Sci. Software Eng. (CSASE) (2020) 259–265.
[26] S. Ramesh, R. Hebbar, M. Niveditha, R. Pooja, N. Shashank, and P. Vinod, "Plant disease detection using machine learning," in *2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C)*, 2018, pp. 41-45.
[27] T.-Y. Lee, J.-Y. Yu, Y.-C. Chang, and J.-M. Yang, "Health detection for potato leaf with convolutional neural network," in *2020 Indo–Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)*, 2020, pp. 289-293.
[28] R. Gajjar, N. Gajjar, V.J. Thakor, N.P. Patel, S. Ruparelia, Real-time detection and identification of plant leaf diseases using convolutional neural networks on an embedded platform, Vis. Comput. (2021) 1–16.
[29] H. Nagamani, H. Sarojadevi, Tomato leaf disease detection using deep learning techniques, Int. J. Adv. Comput. Sci. Appl. 13 (2022).
[30] T.-Y. Lee, I.-A. Lin, J.-Y. Yu, J.-M. Yang, Y.-C. Chang, High efficiency disease detection for potato leaf with convolutional neural network, SN Comput. Sci. 2 (2021) 297.
[31] A. Venkataramana K.S. Kumar N. Suganthi R. Rajeswari Prediction of Brinjal Plant Disease Using Support Vector Machine and Convolutional Neural Network Algorithm Based on Deep Learning Journal of Mobile Multimedia 2022 pp. 771–788-771–788.
[32] J. Johnson, G. Sharma, S. Srinivasan, S. K. Masakapalli, S. Sharma, J. Sharma, *et al.*, "Enhanced field-based detection of potato blight in complex backgrounds using deep learning," *Plant Phenomics,* vol. 2021, 2021.
[33] K. Paul, P. Sinha, S. Mobayen, F.F. El-Sousy, A. Fekih, A novel improved crow search algorithm to alleviate congestion in power system transmission lines, Energy Rep. 8 (2022) 11456–11465.
[34] P. Dalapati and K. Paul, "Optimal scheduling for delay management in railway network using hybrid bat algorithm," in *Intelligent Computing in Control and Communication: Proceeding of the First International Conference on Intelligent Computing in Control and Communication (ICCC 2020)*, 2021, pp. 91-103.
[35] A. O. Anim-Ayeko, C. Schillaci, and A. Lipani, "Automatic Blight Disease Detection in Potato (Solanum tuberosum L.) and Tomato (Solanum lycopersicum, L. 1753) Plants using Deep Learning," *Smart Agricultural Technology,* p. 100178, 2023.
[36] M. Wang, B. Fu, J. Fan, Y. Wang, L. Zhang, C. Xia, Sweet potato leaf detection in a natural scene based on faster R-CNN with a visual attention mechanism and DIoU-NMS, Eco. Inform. 73 (2023), 101931.
[37] T. Acharjee, S. Das, and S. Majumder, "Potato Leaf Diseases Detection Using Deep Learning," *International Journal of Digital Technologies,* vol. 2, 2023.
[38] A. Kumar, V.K. Patel, Classification and identification of disease in potato leaf using hierarchical based deep learning convolutional neural network, Multimed. Tools Appl. (2023) 1–27.
[39] O. Olawuyi and S. Viriri, "Plant Diseases Detection and Classification Using Deep Transfer Learning," in *Pan-African Artificial Intelligence and Smart Systems Conference*, 2022, pp. 270-288.
[40] A. Al-Dabbagh. (2021, February). *PlantVillage Dataset*. Available: https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset.
[41] C.C. Ukwuoma, Z. Qin, M.B.B. Heyat, F. Akhtar, O. Bamisile, A.Y. Muad, et al., A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images, J. Adv. Res. (2022).
[42] C.C. Ukwuoma, D. Cai, M.B.B. Heyat, O. Bamisile, H. Adun, Z. Al-Huda, et al., Deep Learning framework for rapid and accurate respiratory COVID-19 prediction using chest X-ray images, J. King Saud University-Computer and Information Sciences (2023), 101596.
[43] M.A. Al-Antari, M.A. Al-Masni, M.T. Choi, S.M. Han, T.S. Kim, A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification, Int. J. Med. Inf. 117 (Sep 2018) 44–54.
[44] M.A. Al-Antari, T.-S. Kim, Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms, Comput. Methods Programs Biomed. (2020).
[45] A.M. Al-Hejri, R.M. Al-Tam, M. Fazea, A.H. Sable, S. Lee, M.A. Al-Antari, ETECADx: ensemble self-attention transformer encoder for breast cancer diagnosis using full-field digital X-ray breast images, Diagnostics 13 (2022) 89.
[46] R.M. Al-Tam, A.M. Al-Hejri, S.M. Narangale, N.A. Samee, N.F. Mahmoud, M.A. Al-Masni, et al., A hybrid workflow of residual convolutional transformer encoder for breast cancer classification using digital X-ray mammograms, Biomedicines 10 (2022) 2971.

[47] M.A. Al-antari, C.-H. Hua, J. Bang, S. Lee, Fast deep learning computer-aided diagnosis of COVID-19 based on digital chest x-ray images, Appl. Intell. 51 (2021) 2890–2907.

[48] A.M. Farhan, S. Yang, A.Q. Al-Malahi, M.A. Al-antari, MCLSG: Multi-modal classification of lung disease and severity grading framework using consolidated feature engineering mechanisms, Biomed. Signal Process. Control 85 (2023), 104916.

[49] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2778-2788.

[50] M. AlElaiwi, M.A. Al-antari, H.F. Ahmad, A. Azhar, B. Almarri, J. Hussain, VPP: visual pollution prediction framework based on a deep active learning approach using public road images, Mathematics 11 (2022) 186.

[51] Z. Al-Huda, B. Peng, R.N.A. Algburi, M.A. Al-antari, A.-J. Rabea, D. Zhai, A hybrid deep learning pavement crack semantic segmentation, Eng. Appl. Artif. Intel. 122 (2023), 106142.

[52] M.A. Al-Masni, M.A. Al-Antari, M.T. Choi, S.M. Han, T.S. Kim, Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks, Comput. Methods Programs Biomed. 162 (Aug 2018) 221–231.

[53] Y. Gulzar, S.A. Khan, Skin lesion segmentation based on vision transformers and convolutional neural networks—a comparative study, Appl. Sci. 12 (2022) 5990.

[54] O. Ronneberger, P. Fischer, T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, munich, germany, october 5–9, 2015*, Proceedings, Part III 18 (2015) 234–241.

[55] C.C. Ukwuoma, Z. Qin, V.K. Agbesi, B.M. Cobbinah, S.B. Yussif, H.S. Abubakar, et al., Dual_Pachi: Attention-based dual path framework with intermediate second order-pooling for Covid-19 detection from chest X-ray images, Comput. Biol. Med. 151 (2022), 106324.

[56] D. Oppenheim, G. Shani, Potato disease classification using convolution neural networks, Adv. Anim. Biosci. 8 (2017) 244–249.

[57] G. Wang, Y. Sun, J. Wang, Automatic image-based plant disease severity estimation using deep learning, Comput. Intell. Neurosci. 2017 (2017).

[58] T. Subetha, R. Khilar, and M. S. Christo, "WITHDRAWN: A comparative analysis on plant pathology classification using deep learning architecture–Resnet and VGG19," ed: Elsevier, 2021.