

Advanced Disease Detection via Transfer Learning and Convolutional Neural

Sanket Gupta , Suresh Jain , Kailash Chandra Bandhu* 

Department of Computer Science and Engineering, Medi-Caps University, Indore 453331, India

Corresponding Author Email: kailashchandra.bandhu@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290618>

ABSTRACT

Received: 7 March 2024
Revised: 22 October 2024
Accepted: 7 December 2024
Available online: 25 December 2024

Keywords:

ResNet50, classification, deep learning, soybean disease, CNN, agriculture

Soybean is a vital agricultural crop, but its yield is often threatened by diseases affecting soybean leaves, including septoria, frogeye, bacterial blight, brown spots, and others. Previous detection of the diseases is essential to boost outcome and minimize agricultural losses. A major challenge in leaf image classification is misclassification due to the similar disease patterns. Effective feature extraction is critical for building high-performance image classifiers. Convolutional neural networks (CNNs) are best at extracting image features, significantly enhancing classifier accuracy. This study proposes an efficient method for soybean leaf image classification, utilizing transfer learning and CNNs to identify diseases in soybean leaves. The ResNet-50 convolution model is employed for feature extraction, and the extracted features are fed into a fully connected neural network classifier. The proposed model is trained on approximately 6804 images of diseased soybean leaves from Kaggle and Plant Village. The effectiveness of the recommended techniques in comparison to previous approaches in similar experimental setups. F1 score, precision, recall for each class, and overall model accuracy are calculated to evaluate the proposed model's performance.

1. INTRODUCTION

Major ecological and financial losses in the agriculture sector are being caused by diseases of plant and pests. Diseases of plant, weeds, and insects cause 14% of the world's crop yields to be lost, which reduces agricultural industry profits and increases crop treatment costs [1-4]. Moreover, effective regulations must be created to improve crop yield in order to satisfy the growing demands of the world's rapidly expanding populace [5]. By 2050, it is predicted that the world's crop output would need to double, making loss reduction even more crucial. In addition to its financial cost, plant infestation treatment has an environmental consequence. Chemical, biological, and cultural methods are used to control plant infestation [6, 7]. Cultural activities include soil solarisation and crop rotation, for example. But frequently, this is inadequate, therefore chemical and biological techniques are utilised along with. Jeon Gwanggil was the associate editor in charge of organising the manuscript's assessment and granting publication approval. Consequently, the employment of chemicals remains crucial. Overuse of pesticides may lead to contamination of the soil, water, and air. Furthermore, the poisons in pesticides affect healthy plants, animals, and microorganisms [8]. The health of people is also affected by this, since pesticides can enter the body through contaminated air [9].

In central India, soybean is a major crop whose cultivation has increased recently. One of the main causes of crop output losses is the susceptibility of soybean leaves to various diseases. In order to minimize crop losses, early disease

detection is crucial. Based on their respective climates, India's five main zones for soybean growth are separated. The distinct growing environment and climate of each zone can have an impact on the occurrence of particular diseases. Andhra Pradesh, Maharashtra, Karnataka, Rajasthan, Chhattisgarh, and Madhya Pradesh are among the major states that produce soybeans. According to the agroclimatic conditions of each zone, particular cultivars are produced for it.

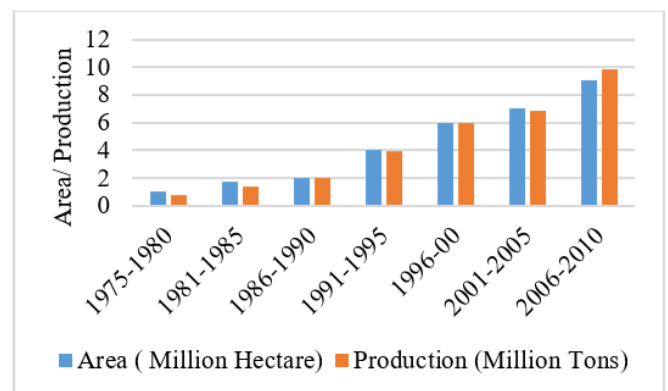


Figure 1. Production of soybean and decade wise changes in India

Soybean production has expanded dramatically over the years, from 0.03 million hectares of land assigned to this crop in 1970 to an incredible 9.30 million hectares in 2010. Furthermore, as shown in Figure 1, the average yield of

soybeans across the country increased dramatically from 0.43 tons per hectare in 1969 to 1.37 tons per hectare in 2010 [10, 11].

Modern technologies play a critical role in early detection of soybean disease, which reduces crop losses and ensures food security. Farmers are able to swiftly and precisely identify crop diseases and take preventive action to minimise damage by utilising deep learning, picture identification, and remote sensing technology. It can also help decrease the usage of dangerous pesticides, which will benefit the country's farmers and economy in addition to enhancing the quality of soybean harvests and overall production quality.

Among these methods, CNNs have been extensively used for image recognition applications. CNN automatically extracts features from photos, and then uses these elements to make accurate classifications of new images. Recently, a large number of researchers used deep learning and machine learning to classify plant leaves. All earlier techniques were trained on a smaller set of photos and had misclassification problems. A plant leaf categorisation model with reasonable accuracy was proposed by numerous researchers; however, it is not a memory-efficient solution.

The aim of this paper was to offer an accurate and efficient approach for classifying soybean leaf diseases. This technique uses two fully connected layers in a convolutional neural network that has been already trained using the ResNet-50 model. Using the pre-trained ImageNet weights, this work applied transfer learning to the model in question and extracted the features. In order to assess the effectiveness of the suggested model, these extracted features are used as input to a fully connected layer, from which the F1 score, Precision, and Recall for each class are derived.

The rest of the essay is structured as follows: the literature review is covered in Section 2, the suggested methodology in Section 3, and the experimental setup in Section 4. Model training is covered in Section 5, and outcome analysis and discussions are presented in Section 6. Conclusion is covered in Section 7, along with the most pertinent references.

2. LITERATURE REVIEW

One significant field of agricultural research is the use of Artificial Intelligence (AI) to detect plant diseases. Numerous methods have been developed for identifying and detecting leaf diseases, including neural networks, clustering algorithms, and leaf colour and disease pattern analysis.

Researches employed various deep learning models to identify soybean leaf diseases, obtaining impressive accuracy rates of 99.04% (Inceptionv3), 99.02% (Resnet-50), 99.02% (VGG19), and 98.56% (Xception) [12]. This study examined five deep learning models - Inception-v3, Resnet-50, VGG16, VGG19, and Xception - for classification of soybean pest images, yielding accuracies of 91.873%, 93.82%, 91.80%, 91.33%, and 90.52%, respectively [13].

The methodology outlined in the paper demonstrates a systematic approach to using deep learning for pest detection in soybean crops. The VGG19 model achieved a notable accuracy of 93.71%, setting a new benchmark for the detection of soybean leaf infestations [14].

A total of 38 deep transfer learning models were evaluated, including popular architectures such as EfficientNet, Inception, VGG, ResNet, and MobileNet. The models were pre-trained on the ImageNet dataset, which contains a wide variety of images, allowing them to leverage learned features for the new task of plant disease detection. The accuracies of some of them were 78.87%, 72.62%, 84.43%, 92.43% for MobileNetV2 ResNet50, DenseNet121, EfficientNetB2V3 respectively [15].

The performance of various models was evaluated for recognizing leaf diseases of different crops. Notably, state-of-the-art models achieved impressive testing accuracies exceeding 90% for maize leaf disease identification. Similarly, all models except ResNet 50 and ResNet 101 demonstrated superior performance with testing accuracies above 90% for rice leaf disease recognition. For wheat leaf disease classification, MobileNet, MobileNetV2, Inception V3, and InceptionResNetV2 consistently delivered testing accuracies above 90%. Furthermore, our proposed model, trained from scratch on the developed datasets, yielded remarkable testing accuracies surpassing 95% [16].

The work demonstrates the growing interest in applying deep learning algorithms to identify leaf diseases in agriculture. The author provides an in-depth examination of the most current advancements and issues with models focused on deep learning for the diagnosis of plant diseases. It's also suggested that deep learning systems like CNN can accurately identify plant illnesses if they have access to enough training data. All things considered, the study provides a useful review of recent research on plant disease identification using computer vision techniques and provides a solid foundation for deep learning concepts [17].

This research has made it possible to precisely and successfully identify and classify plant diseases. The application of deep learning techniques has greatly increased the accuracy of disease detection [18]. Several CNNs are employed by researchers to improve and get high accuracy in the classification model [19]. Nevertheless, a variety of factors influence the effectiveness of various CNNs, such as inadequate annotated data, the quality of the training data, kernel size, activation function, optimizer, and loss function [20]. This work used MFL-DCNN-RSF model which combined multi-dimensional feature learning with CNN for pesticide recommendation for different leaf disease with accuracy of 98.93% [21]. In this work different pre-trained deep learning models (GoogleNet, AlexNet, and ResNet-50) used with SVM classifier on tomato leaf disease dataset. ResNet-50 with SVM achieved the highest accuracy of 95.96% among all other models [22]. Summary of the studied literature are given in Table 1.

Table 1. Literature summary

Sr. No.	Title of Work Done	Image Acquisition	Methodology Applied	Accuracy
1	Automatic identification of illnesses in soybean leaves [12]	Acquisition images with natural conditions.	Inception-v3	99.04%
			Resnet-50	99.02%
			VGG-19	99.02%
			Xception	98.56%
2	Identification and categorization of different pests of soybeans [13]	Data collection using a UAV under natural conditions.	Inception-v3	91.87%
			Resnet-50	93.82%

Sr. No.	Title of Work Done	Image Acquisition	Methodology Applied	Accuracy
3	Automatic identification of infested soybean leaves [14]	Collection using two smartphones featuring 48MP AI triple cameras and a UAV in natural weather and field conditions.	VGG-16	91.80%
			VGG-19	91.33%
			Xception	90.52%
4	Identifying plant diseases based on the type of pathogen [15]	Images captured in different settings, under varying lighting conditions, and using various cameras.	VGG-19	Between 93.71% and 94.16%
			MobileNetV2	78.87%
			ResNet50	72.62%
			DenseNet121	84.43%
5	Creation and development of a real-time dataset and detection system for automatic identification of plant diseases [16]	The dataset comprises manually curated images from online sources (Google, Ecosia, Bing, Flickr) and supplements from Plant Village and Kaggle for healthy maize, rice, and wheat classes.	EfficientNetB2V3	92.43%
			Xception	95.80%
			MobileNet	94.64%
6	Detection and assessment of soybean leaf disease using multiclass SVM and KNN classifiers [23]	Images captured in Conditional environment.	MobileNetV2	96.32%
			InceptionV3	96.20%
			CNN	87%
7	Identification based on inception V3 [24]	Images of tobacco dataset are used.	InceptionV3	90.80%
8	A classification method for soybean leaf diseases based on an improved ConvNeXt model [25]	Images captured from Grapevine leaf dataset.	ResNet50	72.22%
			MobileNetV3	67.27%
			ConvNeXt	66.41%
			CBAM-ConvNeXt	85.42%

3. PROPOSED METHOD

In this age of AI-based technologies, everything has been influenced by all the attention paid to better identification and control of the diseases in the farm sector. Computer vision and machine learning are recently applied majorly for the automation of crop disease detection. The popular option regarding the object detection is CNNs, showing superior performance in the detection of illness. In order to help farmers find and treat the diseases in an early period, the proposed work represents a CNN-oriented method for locating soybean leaf diseases.

The flow chart shows that the process kicks off with a pre-trained ResNet-50 model, which has already been trained on a large dataset like ImageNet, serving as the foundation because it understands a wide variety of features from images. The soybean dataset consists of images categorized into nine different disease types, aiding the model in learning to differentiate between various diseases effectively. The dataset undergoes several pre-processing steps, including rescaling to match the input size expected by ResNet-50, augmentation through transformations like rotation, flipping, generalize better, converting images to grayscale or HSI format to emphasize specific features relevant enhancement techniques.

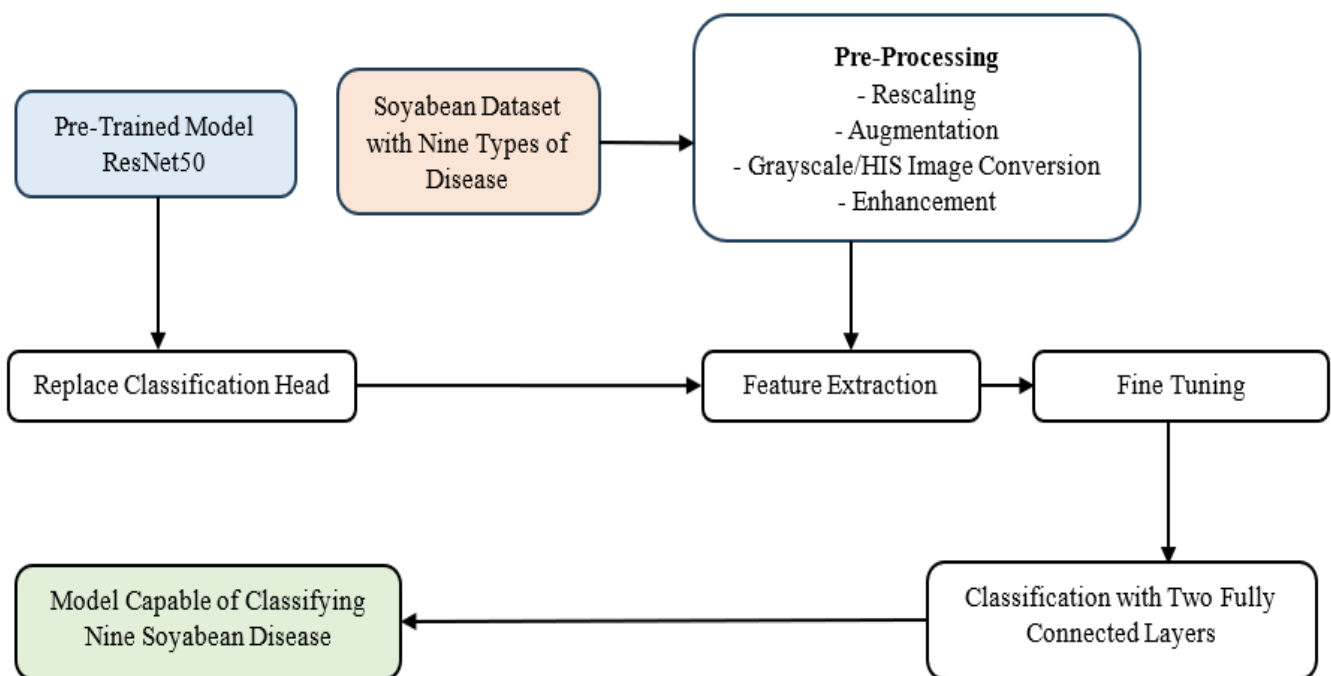


Figure 2. Transfer learning flow chart

like noise reduction or contrast adjustment to make key features more distinguishable. The original output layer of the ResNet-50 model is replaced to tailor it for classifying soybean leaf diseases. The modified model then extracts important features from the pre-processed images, identifying patterns crucial for distinguishing the different diseases. Fine-tuning involves adjusting the network's weights on the specific soybean leaf dataset to ensure the pre-trained features adapt to the new dataset's nuances. The fine-tuned model incorporates two fully connected (FC) layers to process the extracted features and make the final classification. The first FC layer reduces the dimensionality of features, while the second one outputs the probabilities for each disease class using a SoftMax activation function. The end result is a robust model capable of accurately classifying soybean leaves into one of the nine disease categories, providing valuable insights to farmers for timely and effective disease management. The work depicted in Figure 2 employs a disease detection system of soybean leave based on CNN model. In order to help farmers manage soybean leaf diseases in a timely manner, the algorithm seeks to properly categorise using CNN. and diagnosis of these diseases [26, 27]. Large datasets are necessary to improve model performance, which presents a big problem for agricultural applications. Transfer learning is used to solve this, leveraging pre-trained models and fine-tuning them on soybean datasets for improved accuracy and reliability. This method achieves higher accuracy with less training data and time. For feature extraction, the pre-trained ResNet-50 model is employed, known for its effective feature extraction and classification capabilities through its 50 deep layers and multiple convolutional and identity blocks, supported by skip connections to maintain performance.

Since, ResNet-50 is a deep neural network, it may face the Problem of vanishing gradient decent. To combat this, ResNet50 incorporates skip connections (also known as residual connections). These skip connections allow gradients to flow more directly through the network, bypassing certain layers. By doing so, they maintain stronger gradients throughout the network, thereby reducing the impact of the vanishing gradient problem. The only foundation for these relationships is identity mapping. Thus, Eq. (1) provides the mapping function.

$$H(x) = F(x) + x \tag{1}$$

where, $H(x)$ represents the output of the residual block. $F(x)$ is the residual mapping learned by the network's layers and x is the identity mapping passed directly to the next layer. This structure helps preserve the gradient's strength, ensuring that even deeper layers can learn effectively.

The proposed model is build using the following steps:

Data Collection and preprocessing:

Step 1: Dataset contains infected soybean leaves along with healthy leaves.

Step 2: Normalize the pixel values to be in the range [0, 1] by preprocessing the images by resizing them to the required input size of the ResNet50 model, 224×224 pixels.

Transfer Learning:

Step 3: Initialized the ResNet50 model with pre-trained weights on ImageNet.

Step 4: Freeze the weights of the convolutional layers to prevent them from being updated during training.

Step 5: Two dense layers are added after ResNet50 networks with ReLU and SoftMax activation functions.

Training:

Step 6: Split the dataset into training, validation, and test sets.

Step 7: Train the modified ResNet50 model on the training set.

Step 8: To fine-tune the model for its particular goal of classifying soybean leaf diseases, release a few of the top layers of ResNet50 and carry on training at a lower learning rate.

Evaluation:

Step 9: Accuracy, precision, recall, and F1-score are used to measure the performance of proposed model.

4. EXPERIMENTAL SETUP

An analysis of soybean leaf disease utilising the suggested model was reported in this paper. This work outlines the several stages of our experiment, which begin with gathering the high-quality dataset and conclude with evaluating the performance of the model as depicted below.

4.1 Dataset acquisition

The soybean leaf dataset utilized in this study is employed to train the model. Images of soybean leaves were gathered from the Plant Village collection [28] and some actual images were recorded under various environmental conditions in Figure 3.

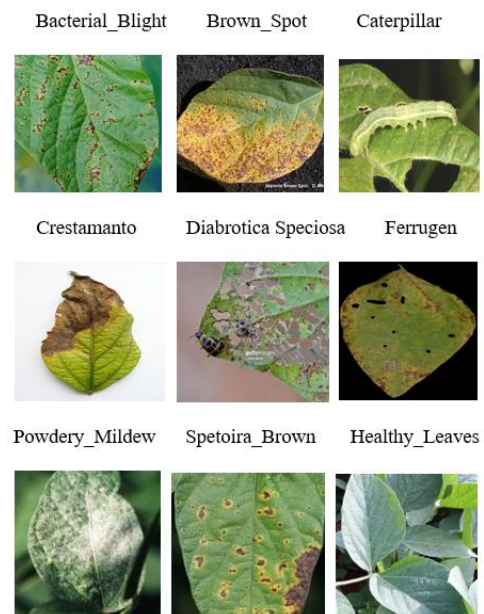


Figure 3. Soybean diseased leaf images dataset

Table 2. Images distribution

Disease Type	Number of Images
Crestamento	750
Bacterial Blight	850
Brown Spot	900
Diabrotica Speciosa	700
Frogeye	800
Septoria	750
Powdery Mildew	900
Caterpillar	654
Healthy Leaves	500

In addition to one healthy set of soybean leaves, it includes 6804 pre-processed and labelled images of soybean leaves that show eight different types of diseases, including Brown Spot, Diabrotica Speciosa, Frogeye, Septoria, Powdery Mildew, and Caterpillar. Table 2 shows the distribution of images across the different types of diseases.

4.2 Pre-processing and augmentation

Building an efficient CNN model for image classification requires both picture pre-processing and augmentation [29, 30]. So, to improve the excellence of the dataset, we performed a number of pre-processing procedures in this research study. Next, we processed the images by augmentation. Augmenting data is crucial to expanding the dataset.

To ensure uniformity in the input size, we scaled each image to a constant size ratio of 224 by 224 pixels. This procedure also lowers the computational complexity of the model in order to eliminate variations in brightness and contrast, we first normalized the photos by scaling down each pixel's value to the same value. Normalization not only lowers overfitting but also model error. Following that, the collected data was split into three parts: testing, validation, and training sets, each containing 10%, 10%, and 80% of the photos [23]. Consequently, the model can be assessed on fresh, untested data in addition to being trained on a suitable volume of data. Next, we added to our training set to make it larger. Preprocessing used different techniques, including scaling, flipping horizontally and vertically, zooming, and height shifting to a specific position.

5. MODEL TRAINING

This work utilised a transfer learning strategy for our research article since it is well known for working well on a variety of tasks, such as computer vision. We integrated our transfer learning methodology with a convolutional neural network. We were able to identify our soybean dataset effectively because to this combination.

5.1 Model architecture of Restnet50

The ResNet V2-50 feature vector model from TensorFlow Hub, which was trained on the ImageNet dataset, is the pre-trained model utilised in this paper [31]. The suggested

approach is frequently utilised as a foundation model for transfer learning in computer vision problems and is intended to extract high-level [32] features from images. More than a million photos [33] of diverse things from 1000 distinct types can be found in the ImageNet dataset. The network can gain the learned feature of the previously trained model by employing transfer learning and reusing the ResNet-v2-50 model's pre-trained weights. This can increase training accuracy and improve the model's performance for the particular task of detecting soybean leaf disease, as shown in Table 1.

The ResNet-v2 design, which was first presented in 2015 by [34], is an enhanced version of the original ResNet architecture, as Table 2 shows. Using residual blocks with skip connections, the ResNet-v2 is an effective image recognition model that tackles the vanishing gradient issue in neural networks. The design is comprised of phases featuring layers for down-sampling and an altered identity mapping function that adds a residual connection to boost speed. In addition, a bottleneck structure is employed to save computational costs while increasing accuracy, and batch normalisation is used to prevent internal covariate shifts. The ResNet-v2 is a suitable option for semantic segmentation and object recognition applications since it can train longer networks (up to 152 layers) without running into the vanishing gradient issue. A fully linked layer in the last stage outputs the class probabilities. ResNet-50-v2 is a very successful image recognition model that has demonstrated strong performance over a wide range of image datasets.

The model architecture utilized in this paper, which includes several trainable parameters, an output structure, and different types of layers, is summarized in Table 3. The 224×224×3 input images are accepted by the input layer, which is the first layer. With pre-trained ResNet50 model weights, all layers are Keras Layers and produce a 2048 feature vector. A Dropout layer to stop overfitting and a Dense layer with 128 neurons make up the subsequent layers. The output gives the probability for each of the nine types of soybean leaf diseases and is a dense layer with nine neurons. Including the pre-trained ResNet50 model parameters, the model comprises a total of 23,828,233 parameters.

Overall, for the particular objective of soybean leaf disease detection, the combination of transfer learning with the pre-trained ResNet50 model enables enhanced precision and accelerated training timeframes.

Table 3. RestNet-v2 architecture [34]

Layer Name	Input Size	18 Layers	34 Layers	50 Layers	101 Layers	152 Layers
Conv	112×112	7×7, 64, stride 2				
		3×3 max pool, 64, stride 2				
Conv2_x	56×56	$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$
Conv3_x	28×28	$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 8$
Conv4_x	14×14	$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 36$
Conv5_x	7×7	$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2028 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2028 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2028 \end{bmatrix} \times 3$

Table 4. Soybean diseased leaf detection model summary

Layer	Output Shape	Parameter	Activation Function	Optimizer
Input I	(None, 224, 224,3)	0	--	--
Resnet50v2	(None, 2048)	23564800	--	--
Dense Layer1	(None, 128)	262272	Adam	ReLU
Dropout Layer	(None, 128)	0	--	--
Dense Layer1	(None, 9)	1152	Adam	Softmax
Total Params	23,849,224	--	--	--

5.2 Fine-tuning process

Additional layers of classification, consisting of already processed soybean sick leaves with 9 categories, were introduced to the ResNet50v2 model after it was frozen. Two thick layers make up the model. ReLU activation was utilised in the first dense layer, and SoftMax activation was utilised in the second dense to avoid overfitting, regularisation in the form of L2 weight decay was performed to the first Dense layer.

5.3 Hyperparameter tuning

To maximise the performance of our classification model, we adjusted some hyperparameters which is mentioned below. To get the most out of our model, it's important to select the right hyperparameters from Table 4. Hyperparameters are the set of parameters that control how a model learns and have a big impact on how accurate the model is. Even with the best model architecture, poor results are possible. Initially, the learning rate was set to 0.001, and the Adam optimiser was chosen for the suggested task. In order to prevent overfitting, this work added implemented L2 weight decay to the first dense layer with a coefficient of 0.0001.

In selecting the hyperparameters, we aimed to achieve a harmonious balance between performance, computational efficiency, and model stability. The batch size of 36 ensures that the model can process a reasonable number of samples in each training step, optimizing memory usage and maintaining stable gradient estimates.

Training the model for 80 epochs provides sufficient iterations for the network to learn the underlying patterns without falling into overfitting, where the model performs well on training data but poorly on new data. Setting the initial learning rate at 0.001 is a common practice to ensure that weight updates during training are neither too drastic, risking instability, nor too small, hindering convergence. The dropout rate of 0.5 helps keep the model from depending too much on a few particular neurones, thereby improving its capacity for generalisation to unseen data. Incorporating a weight decay of 0.0002 serves as a regularization technique, discouraging excessively large weights and promoting a simpler model structure. The patience value of 5 allows the training process to continue for several epochs without improvement, avoiding premature termination and ensuring thorough learning. Finally, reducing the learning rate by a factor of 0.6 when the model's performance plateaus help fine-tune the model's parameters, facilitating convergence to a more optimal solution. These decisions are grounded in empirical evidence and best practices, aimed at developing a robust and effective model for detecting diseases in soybean leaves.

5.4 Training

The model trains on 80 epochs with a batch size of 36 during

the training phase. Two callbacks were used to keep an eye on the training process and avoid overfitting. The initial one involved Early Stopping, wherein the 'patience' parameter was set at 5. This meant that after five consecutive epochs, if the error did not reduce, the training process would end. In addition, the weights from the epoch with the best validation loss were retrieved by activating the 'restore best weights' parameter.

Additionally, we used the 'ReduceLROnPlateau' callback with a 'factor' argument set to 0.6 to cut the learning rate in half in the event that the validation loss does not improve after the 'patience' number of epochs. Ultimately, a sparse categorical cross-entropy loss function was initiated during the modelling process, and accuracy served as the assessment metric.

6. RESULT ANALYSIS AND DISCUSSION

Promising results were obtained using the proposed model for the identification of sick leaves using CNN and transfer learning ResNet50. The model, which was trained using the ImageNet dataset, demonstrated efficacy in feature extraction and pattern recognition—two critical processes in the identification of damaged leaves in soybean plants. For the suggested work experiment, a Jupyter Notebook is utilised since it offers a better interface for executing Python code and evaluating the outcomes. The AMD Ryzen 5 CPU-3350h @ 2.5GHz and Windows 10 operating system combined to provide a strong and capable hardware arrangement for the research. We used the Nvidia Cuda 10.1 version, which has 1024 cores for GPU acceleration, to greatly improve the training procedure.

It performed quite well during training on both the training and validation sets. Figure 4 illustrates that the test set reaches 89 percent and the training set achieves 91 percent, indicating the approach was successful in generalising using validation data that had unseen. It indicates model made progress towards developing a useful method for agricultural disease diagnosis and treatment by correctly classifying damaged soybean leaves into the appropriate group.

The application of ResNet-50 and transfer learning for diagnosing soybean leaf diseases is very much prevalent in recent studies; however, its prominence emerges from some significant perspectives. Due to its applications to the description of latent features underpinning totally different spots, the nearly flawless fine-tuning of the ResNet-50 model with thorough preprocessing and augmentation, involving rescaling, normalisation, and transformations, improves modelling robustness and accuracy.

In contrast to previous research, the suggested approach contains a comparative assessment based on multiple additional measures, such as precision, recall, and F1-score, hence removing performance bias about the model's continuous development across various illness classes.

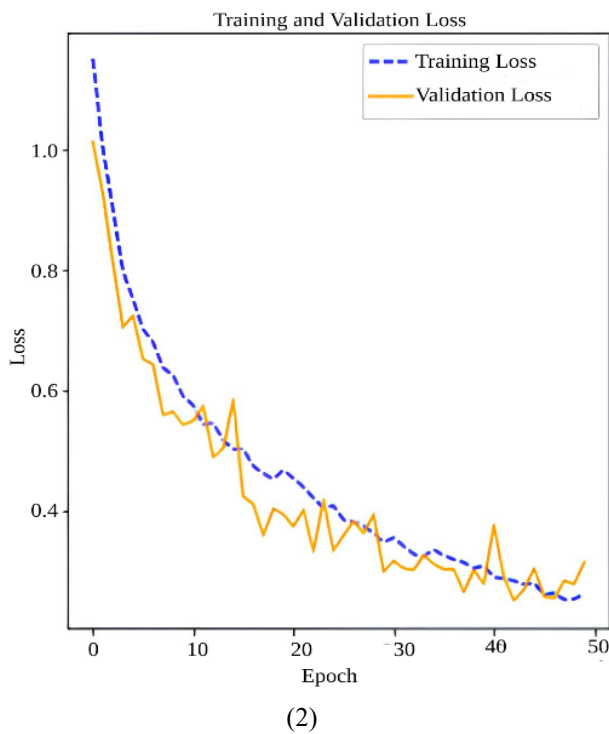
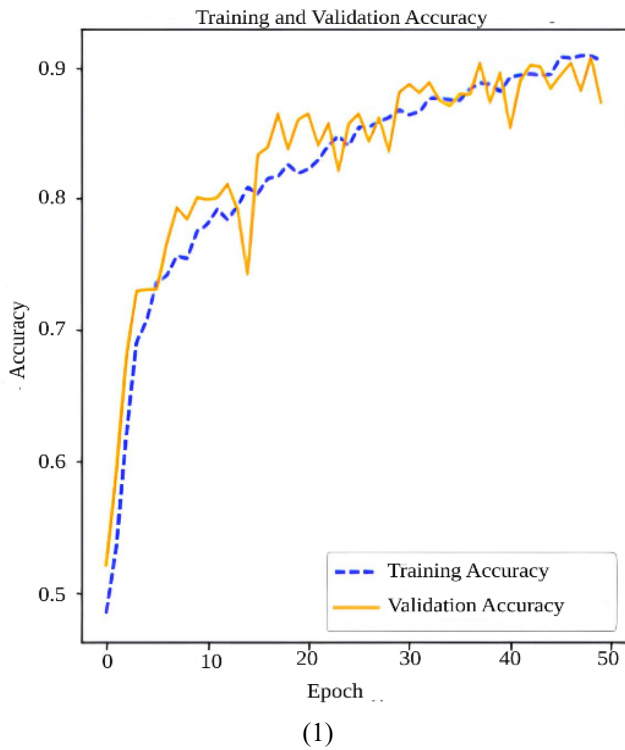


Figure 4. Performance analysis of training and validation datasets

Moreover, the current focus addresses a huge agricultural problem, in which the solution has dire practical implications for maximally benefiting crop health and yield to farmers. Also, detailed selection and justification of hyperparameters like batch size, learning rate, dropout rate, weight decay, and patience do show profound understanding of optimization models, hence a reliable detection system of a highly performing detection system. All these unique aspects as a collective make this research a significant contribution to the agricultural disease-detection field.

6.1 Significance of the study

The worth of this study relates to its outcome for farmers and agricultural workers. The research helps to improve the management of crops through its reliable gadget for quick and precise detection of diseases thereby helping avoid incurrences of losses and improving production. Careful metrics and analysis of misclassifications and others are guaranteed to ensure that the model performs accurately and consistently over many diseases. This has added value to the model in its application in practice, where the model may have to be used in different and quite changing conditions.

6.2 Future research directions

Though this research presents a useful method of identifying soybean leaf diseases, here are some avenues to explore which will improve the impact:

6.2.1 Novel structures

Investigate newer and increasingly advanced deep learning architectures to enhance both feature extraction and classification. One could efficiently try EfficientNet or Vision Transformers.

6.2.2 Increased amount and variety of datasets

Also, incorporate more images taken in several environmental conditions into the dataset in order to achieve better model generalization and possible robustness.

6.2.3 Real-time monitoring and IoT

Other IoT devices can be incorporated alongside this model so as to provide a real time monitoring and detecting of the diseases hence providing farmers with instant help and strategies.

6.2.4 Identifying cross-crop diseases

Further extend the study to many more crops and diseases so that a multi-crop disease detection system can be designed to help as many farmers as possible.

6.2.5 Multidisciplinary collaboration

Work with plant pathologists and agriculture specialists to obtain more diverse datasets and inform the model for higher accuracy and practicality.

7. EVALUATION AND CLASSIFICATION

In addition to its excellent accuracy, our suggested model demonstrated encouraging outcomes on particular classes of soybean leaf diseases. Figure 5 illustrates the model's ability to correctly identify and categorize photos of soybean leaves infested with caterpillar leaf, a widespread and damaging disease that affects soybean harvests. In order to prevent the model from incorrectly classifying healthy leaves as diseased, it is also crucial that the model show accurate predictions for the healthy class.

Overall, our suggested model's high accuracy across a wide spectrum of soybean leaf illnesses emphasises its potential as a useful tool for crop disease early detection and monitoring. This may result in more productive and environmentally friendly farming methods that aid in lowering crop losses and raising yields.

Actual: Healthy
 Predicted: healthy
 Confidence: 100%

Actual: Bacterial_Blight
 Predicted: Bacterial_Blight
 Confidence: 57.13%

Actual: Caterpillar
 Predicted: Caterpillar
 Confidence: 52.75%



Figure 5. Estimates of the submitted model for identifying soybean leaf disease on the test dataset

Table 5. Hyperparameter matrix

Hyperparameter	Value
Batch Size	36
Epochs	80
Initial Learning Rate	0.001
Dropout Rate	0.6
Weight Decay	0.0002
Patience	5
Factor	0.6

Table 6. Performance matrix

Class	Precision%	Recall%	F1-Score%
Caterpillar	71	70	78
Diabrotica Speciosa	82.5	62	67.5
Healthy	63.3	81.1	72
Bacterial Blight	78	90	83.3
Brown Spot	95	96	96
Crestamento	70	94	78
Frogeye	97.7	84	94
Powdery Mildew	74	83	76
Septoria	98	95	97

Confusion matrix based on model predictions on testing dataset was developed to evaluate the performance of the proposed model. In order to assess how well the ResNet50 model can identify soybean leaf diseases, the recall and precision metrics are computed for each disease class. Recall measures the percentage of accurate outcomes in the test set relative to the actual accurate results, which are shown in Table 5, whereas precision measures the percentage of relevant outcomes in the test set among all the retrieved outputs. The F1 score was also utilised in this work to assess the model's overall performance. Eqs. (2)-(4) [35] of the established formulas were used to calculate these measures which is given in Table 6.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

However, there may be misclassification of disease type. Including an analysis of which disease types are most prone to misclassification and the possible reasons for these errors is crucial for enhancing the scientific value of your paper. Crestamento might be mistaken for other fungal diseases like Brown Spot due to subtle visual similarities. Bacterial Blight could be confused with Frogeye because both exhibit water-soaked spots that turn brown. Brown Spot and Septoria both cause dark spots, leading to frequent misclassifications. Differentiating between Diabrotica Speciosa and Caterpillar damage is challenging as both result in physical leaf damage. Frogeye might be misidentified as Bacterial Blight due to their shared symptom of dark spots with lighter halos. Septoria and Brown Spot have similar dark lesions, complicating their distinction. Powdery Mildew, characterized by white, powdery spots, could be confused with the early stages of other fungal diseases. Finally, Caterpillar damage and Diabrotica Speciosa both leave visible damage, making them hard to distinguish. This detailed analysis of misclassifications enhances transparency regarding the model's performance and highlights areas for further improvement, thereby significantly contributing to the reliability of the disease detection model. The proposed work and the previous work are evaluated is shown in Table 7.

Table 7. Comparison of proposed work with existing work

Pre-Trained Network Model	Accuracy	No. of Images
InceptionV3 [24]	90.80	800
ResNet50 [25]	72.22	
MobileNetV3[25]	67.27	1296
ConvNeXt [25]	66.41	
CBAM-ConvNeXt [25]	85.42	
Proposed ResNet 50 Model	91.00	6804

InceptionV3 achieved an accuracy of 90.80% using 800 images [24], while ResNet50 had a lower accuracy of 72.22% with the same number of images presents MobileNetV3 with an accuracy of 67.27% and ConvNeXt with 66.41%, both trained on 1296 images [25]. However, CBAM-ConvNeXt, also from [25], performed significantly better with an accuracy of 85.42% using 1296 images. In contrast, the proposed model utilizing ResNet 50 stands out, achieving the highest accuracy of 91.00% with a substantially larger dataset of 6804 images. This comparison underscores the effectiveness of the proposed

model in utilizing a larger dataset to achieve enhanced performance in soybean leaf disease detection, highlighting its potential for real-world agricultural applications.

8. CONCLUSION

In conclusion, applying transfer learning and CNN for diagnosing soybean leaf disease proves to be an effective and efficient technique that demonstrates the application of contemporary technology. Beforehand diagnosis of the illness is essential since soybeans are important crop and a major source of reliance for a large number of countries. Making sure food security and keeping agricultural losses at bay are vital. Modern technological developments like computer vision and deep learning methods have improved the outcomes throughout the previous ten years. Convolutional neural networks (CNNs) and transfer learning have been shown to function well together to diagnose diseases with high accuracy. This lessens the need for physical labour while also enabling faster and more accurate diagnosis. The proposed model demonstrated the potential of technology to revolutionize agriculture and boost productivity, achieving an overall accuracy of 91% on the training dataset and 89% on the testing dataset.

REFERENCES

[1] Khalili, E., Javed, M.A., Huyop, F., Wahab, R.A. (2019). Efficacy and cost study of green fungicide formulated from crude beta-glucosidase. *International Journal of Environmental Science and Technology*, 16: 4503-4518. <https://doi.org/10.1007/s13762-018-2084-1>

[2] Khalili, E., Kouchaki, S., Ramazi, S., Ghanati, F. (2020). Machine learning techniques for soybean charcoal rot disease prediction. *Frontiers in Plant Science*, 11: 590529. <https://doi.org/10.3389/fpls.2020.590529>

[3] Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L.R., Davis, C.E., Dandekar, A.M. (2015). Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development*, 35: 1-25. <https://doi.org/10.1007/s13593-014-0246-1>

[4] Wilson, C., Tisdell, C. (2001). Why farmers continue to use pesticides despite environmental, health and sustainability costs. *Ecological Economics*, 39(3): 449-462. [https://doi.org/10.1016/S0921-8009\(01\)00238-5](https://doi.org/10.1016/S0921-8009(01)00238-5)

[5] Kofsky, J., Zhang, H., Song, B.H. (2018). The untapped genetic reservoir: The past, current, and future applications of the wild soybean (*Glycine soja*). *Frontiers in Plant Science*, 9: 949. <https://doi.org/10.3389/fpls.2018.00949>

[6] Panizzi, A.R. (2013). History and contemporary perspectives of the integrated pest management of soybean in Brazil. *Neotropical Entomology*, 42(2): 119-127. <https://doi.org/10.1007/s13744-013-0111-y>

[7] Barzman, M., Bärberi, P., Birch, A.N.E., Boonekamp, P., Dachbrodt-Saaydeh, S., Graf, B., Hommel, B., Jensen, J.E., Kiss, J., Kudsk, P., Lamichhane, J.R., Messéan, A., Anna-Camilla, M., Ratnadass, A., Ricci, P., Jean-Louis, S., Sattin, M. (2015). Eight principles of integrated pest management. *Agronomy for Sustainable Development*, 35: 1199-1215. <https://doi.org/10.1007/s13593-015-0327-9>

[8] Sharma, A., Kumar, V., Shahzad, B., Tanveer, M., Sidhu, G.P.S., Handa, N., Kohli, S.K., Yadav, P., Bali, A.S., Parihar, R.D., Dar, O.I., Singh, K., Jasrotia, S., Bakshi, P., Ramakrishnan, M., Kumar, S., Bhardwaj, R., Thukral, A.K. (2019). Worldwide pesticide usage and its impacts on ecosystem. *SN Applied Sciences*, 1: 1446. <https://doi.org/10.1007/s42452-019-1485-1>

[9] Bhandari, S., Paneru, S., Pandit, S., Rijal, S., Manandhar, H. K., Ghimire, B.P. (2020). Assessment of pesticide use in major vegetables from farmers' perception and knowledge in Dhading district, Nepal. *Journal of Agriculture and Natural Resources*, 3(1): 265-281. <https://doi.org/10.3126/janr.v3i1.27180>

[10] Agarwal, D.K., Billore, S.D., Sharma, A.N., Dupare, B.U., Srivastava, S.K. (2013). Soybean: Introduction, improvement, and utilization in India-problems and prospects. *Agricultural Research*, 2(4): 293-300. <https://doi.org/10.1007/s40003-013-0088-0>

[11] Singh, A., Dutta, M. K., Jennane, R., Lespessailles, E. (2017). Classification of the trabecular bone structure of osteoporotic patients using machine vision. *Computers in Biology and Medicine*, 91: 148-158. <https://doi.org/10.1016/j.combiomed.2017.10.011>

[12] Tetila, E.C., Machado, B.B., Menezes, G.K., Oliveira, A.D.S., Alvarez, M., Amorim, W.P. (2019). Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 17(5): 903-907. <https://doi.org/10.1109/LGRS.2019.2932385>

[13] Tetila, E.C., Machado, B.B., Astolfi, G., de Souza Belete, N.A., Amorim, W.P., Roel, A.R., Pistori, H. (2020). Detection and classification of soybean pests using deep learning with UAV images. *Computers and Electronics in Agriculture*, 179: 105836. <https://doi.org/10.1016/j.compag.2020.105836>

[14] Farah, N., Drack, N., Dawel, H., Buettner, R. (2023). A deep learning-based approach for the detection of infested soybean leaves. *IEEE Access*, 11: 99670-99679. <https://doi.org/10.1109/ACCESS.2023.3313978>

[15] Asha Rani, K.P., Gowrishankar, S. (2023). Pathogen-based classification of plant diseases: A deep transfer learning approach for intelligent support systems. *IEEE Access*, 11: 64476-64493. <https://doi.org/10.1109/ACCESS.2023.3284680>

[16] Joseph, D.S., Pawar, P.M., Chakradeo, K. (2024). Real-time plant disease dataset development and detection of plant disease using deep learning. *IEEE Access*, 12: 16310-16333. <https://doi.org/10.1109/ACCESS.2024.3358333>

[17] Li, L., Zhang, S., Wang, B. (2021). Plant disease detection and classification by deep learning-a review. *IEEE Access*, 9: 56683-56698. <https://doi.org/10.1109/ACCESS.2021.3069646>

[18] Singh, A., Kaur, H. (2022). Comparative study on identification and classification of plant diseases with the support of transfer learning. In *International Conference on Innovative Computing and Communications*, pp. 375-386. https://doi.org/10.1007/978-981-16-2594-7_31

[19] Rahman, M.A. (2020). Deep learning approaches for tomato plant disease detection. *International Journal of Hybrid Information Technology*, 13(2): 71-78. <https://doi.org/10.21742/IJHIT.2020.13.2.06>

[20] Agarwal, M., Singh, A., Arjaria, S., Sinha, A., Gupta, S.

- (2020). ToLeD: Tomato leaf disease detection using convolution neural network. *Procedia Computer Science*, 167: 293-301. <https://doi.org/10.1016/j.procs.2020.03.225>
- [21] Saleem, J.B.M., Shanmugam, K. (2023). Pesticide recommendation for different leaf diseases and related pests using multi-dimensional feature learning deep classifier. *Ingénierie des Systèmes d'Information*, 28(1): 133-140. <https://doi.org/10.18280/isi.280113>
- [22] Babu, P.R., Krishna, A.S. (2023). Deep learning-assisted SVMs for efficacious diagnosis of tomato leaf diseases: A comparative study of GoogleNet, AlexNet, and ResNet-50. *Ingénierie des Systèmes d'Information*, 28(3): 639-645. <https://doi.org/10.18280/isi.280312>
- [23] Jadhav, S.B., Udup, V.R., Patil, S.B. (2019). Soybean leaf disease detection and severity measurement using multiclass SVM and KNN classifier. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(5): 4077. <https://doi.org/10.11591/ijece.v9i5.pp4077-4091>
- [24] Zhang, W.J., Sun, X.M., Qiao, Y.L., Bai, P., Jiang, H.H., Wang, Y.J., Du, C.Y., Zong, H. (2021). Identification based on inception V3. *Chinese Journal of Tobacco*, 27(5): 61-70. <https://doi.org/10.16472/j.chinatobacco.2021.T0061>
- [25] Wu, Q., Ma, X., Liu, H., Bi, C., Yu, H.L., Liang, M.J., Zhang, J.C., Li, Q., Tang, Y., Ye, G.S. (2023). A classification method for soybean leaf diseases based on an improved ConvNeXt model. *Scientific Reports*, 13(1): 19141. <https://doi.org/10.1038/s41598-023-46492-3>
- [26] Sutaji, D., Rosyid, H. (2022). Convolutional Neural Network (CNN) models for crop diseases classification. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 7(2): 187-196. <https://doi.org/10.22219/kinetik.v7i2.1443>
- [27] Saleem, M.H., Potgieter, J., Arif, K.M. (2022). A performance-optimized deep learning-based plant disease detection approach for horticultural crops of New Zealand. *IEEE Access*, 10: 89798-89822. <https://doi.org/10.1109/ACCESS.2022.3201104>
- [28] Pujari, J.D., Yakkundimath, R., Byadgi, A.S. (2013). Classification of fungal disease symptoms affected on cereals using color texture features. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(6): 321-330. <https://doi.org/10.14257/ijsp.2013.6.6.29>
- [29] Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., Menaka, R. (2020). Attention embedded residual CNN for disease detection in tomato leaves. *Applied Soft Computing*, 86: 105933. <https://doi.org/10.1016/J.ASOC.2019.105933>
- [30] Sujatha, R., Chatterjee, J.M., Jhanjhi, N.Z., Brohi, S.N. (2021). Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocessors and Microsystems*, 80: 103615. <https://doi.org/10.1016/J.MICPRO.2020.103615>
- [31] Mao, Y., Yang, Y., Ma, Z., Li, M., Su, H., Zhang, J. (2020). Efficient low-cost ship detection for SAR imagery based on simplified U-net. *IEEE Access*, 8: 69742-69753. <https://doi.org/10.1109/ACCESS.2020.2985637>
- [32] Caldeira, R.F., Santiago, W.E., Teruel, B. (2021). Identification of cotton leaf lesions using deep learning techniques. *Sensors*, 21(9): 3169. <https://doi.org/10.3390/s21093169>
- [33] Bevers, N., Sikora, E.J., Hardy, N.B. (2022). Soybean disease identification using original field images and transfer learning with convolutional neural networks. *Computers and Electronics in Agriculture*, 203: 107449. <https://doi.org/10.1016/j.compag.2022.107449>
- [34] Milone, D., Longo, F., Merlino, G., De Marchis, C., Risitano, G., D'Agati, L. (2024). MocapMe: DeepLabCut-enhanced neural network for enhanced markerless stability in sit-to-stand motion capture. *Sensors*, 24(10): 3022. <https://doi.org/10.3390/s24103022>
- [35] Hossain, M.I., Haq, I., Talukder, A., Suraiya, S., Rahman, M., Saleheen, A.A.S., Methun, M.I.H., Habib, M.J., Hossain, M.S., Nayan, M.I.H., Hussain, S. (2023). Performance evaluation of machine learning-based algorithms to predict the early childhood development among under five children in Bangladesh. *Journal of Computer Science*, 19(5): 641-653. <https://doi.org/10.3844/jcssp.2023.641.653>