IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Enhanced Potato Pest Identification: A Deep learning approach for identifying potato pests

AMIR SOHEL [1], MD. SHAHRIAR SHAKIL [1], SHAH MD TANVIR SIDDIQUEE [1], AHMED AL MAROUF [2], (Member, IEEE), JON G. ROKNE [2], (Life Fellow, IEEE), AND REDA ALHAJJ.[2, 3, 4], (Senior Member, IEEE)

[1]Multidisciplinary Action Research Lab, Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh (e-mail: amir.cse@diu.edu.bd; shakil.cse@diu.edu.bd, Tanvir.cse@diu.edu.bd)
[2]Department of Computer Science, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: ahmedal.marouf@ucalgary.ca; rokne@ucalgary.ca; alhajj@ucalgary.ca
[3]Department of Computer Engineering, Istanbul Medipol University, Istanbul 34810, Turkey
[4]Department of Heath Informatics, University of Southern Denmark, 5230 Odense, Denmark

Corresponding author: Ahmed A. Marouf (e-mail: ahmedal.marouf@ucalgary.ca).

ABSTRACT Potato crops and their salability are influenced by potato pests in that both crop yield and quality are reduced. This in turn reduces the income for potato farmers due to lower prices for the crop, lower crop yield, trade restriction and reduced market access. Agricultural viability over the long run therefore depends on sustainable pest management. In order to efficiently detect potato pests, a dataset was constructed which contains eight prevalent potato species that were taken from several sources. Image pre-processing techniques were employed enhance image quality for compatibility with deep learning models. Among InceptionV3, VGG-16, and MobileNetV2 models, VGG-16 attained the highest accuracy of 94.44%, outperforming others. Inception-V3 achieved 58% accuracy, while MobileNetV2 reached 75%. Pre-processing has a major influence on improving result accuracy, which emphasizes its significance in enhancing model performance, according to an evaluation of its effects. These findings might lead to the development of pest management strategies for potato farming that are more effective. The efficient use of VGG-16 in potato pest identification systems is demonstrated by its excellent performance. Using deep learning models can therefore reduce financial losses and promote sustainable potato production. This study provides an approach for further investigation into the best ways to control pests in potato production, allowing farmers to overcome the obstacles and take advantage of valuable market prospects even in the face of pest threats.

INDEX TERMS Potato Pest, Classification, Deep Learning, VGG-16

I. INTRODUCTION

POTATO which is one of the world's most vital staple crops, plays a pivotal role in global food security [1]. However, a major challenge to potato growing that affects yields and quality is the constant risk of pests and illnesses. Specifically, insect pests pose a serious threat to potato growers and can result in significant harm if neglected [2]. In potato farming, conventional approaches to pest detection and monitoring have shown to be time-consuming and labor-intensive, frequently resulting in response delays and financial losses [3].

The use of deep learning models presents a revolutionary and important response for the distinct agricultural environment of Bangladesh' and for potato farming worldwide. By merging deep learning with computer vision technology for detecting potato pests has made it possible to produced effective cutting-edge tools for farmers for pest detection and surveillance thus bringing forth a new age in agriculture [4]. This work therefore aims to achieve previously unheard-of levels of efficacy and efficiency in potato pest management using deep learning approaches.

This study relies in particularly on state-of-the-art Convolutional Neural Network (CNN) architectures, such as VGG16, InceptionV3, and ResNet50, to identify insect pests in potato farming. The goal in this project is to provide an extensive and proactive strategy to reduce the adverse impacts of pests on potato yields and quality by emphasizing key issues in potato pest control, such as early detection, precise species identification, and real-time monitoring. It aims to solve the fundamental issues associated with potato agriculture and pave the way for a more resilient and sustainable future for potato producers throughout the world by utilizing cutting-edge technology and creative techniques. We are able to develop automated systems that can identify insect pests with previously unheard-of precision and efficiency by utilizing the capabilities of these neural networks. With the use of this technical development, farmers and other agricultural experts may now take prompt and focused action against pest infestations, minimizing crop losses and lowering the necessity for chemical treatments. Additionally, by com-paring several CNN architectures insightful information is obtained that may be used to choose the best model for certain pest identification tasks. In addition to improving potato pest monitoring in Bangladesh, this research also provides a global viewpoint that will aid in the development of general, effective and sustainable pest control techniques elsewhere.

There are many studies concerned with potato pest identification. The purpose of this study is to lay out the most important discoveries made so far and then outline possible new avenues of research in this area.

Talukder et al. (2023) presented an advanced potato pest identification system, CTInceptionV3-RS with a customized Inception V3 model, fine-tuned through random search (RS) optimization. The authors curated a reliable dataset containing eight types of potato pests, employing five pre-trained transfer learning models: MobileNetV2, NASLargeNet, Xception, DenseNet201, and InceptionV3. This dataset was divided into training, testing, and validation sets with a ratio of 70:15:15. The CTInceptionV3-RS model showcased outstanding performance metrics, achieving an accuracy of 91%, along with high precision, recall, and F1-score. This highlights its effectiveness in au-tomating the identification of common potato pests. The study strategically addressed data limitations through augmentation techniques and transfer learning, providing a robust solution to enhance potato production and pest treatment [5].

The paper of Roldán-Serrato et al, (2018) describes an autonomous pest detection system using artificial neural networks. The method detects two defoliating pests on potato and bean crops: Mexican Bean Beetle (MBB) and Colorado Potato Beetle (CPB) in their adult stages. The neural classifiers used to detect beetles are RSC (Random Sub-space Classifier) and LIRA (Limited Receptive Area). Using two image databases, they examined the effectiveness of pest recognition systems. The first picture database comprises 75 CPB photos, while the second has 200 MBB beetle images. There are two classes: beetles and backdrop. Results were obtained, with 89% for CPB detection and 88% for MBB detection in this research. The RSC classifier achieved a higher recognition rate of 89%, compared to LIRA's 88% [6].

Anim-Ayeko et al. (2023) developed a ResNet-9 model to detect potato and tomato leaf blight diseases. Trained on the "Plant Village Dataset" with 3,990 samples, the model achieved a remarkable 99.25% test accuracy. The dataset includes 54,309 ex-pert-categorized images covering various plant diseases. The ResNet-9 model, trained on Kaggle NVIDIA TESLA P100 GPUs, outperforms a VGG-16 baseline, demonstrating superior accuracy and precision. Data augmentation played a crucial role in enhancing learning. The study concludes that the ResNet-9 model excels in blight disease detec-tion, surpassing the VGG-16 baseline, potential future exploration of additional agri-cultural and environmental data for further advancements [7].

Tiwari et al. (2020) used the pre-trained model VGG19 for fine-tuning to extract the relevant features from a dataset to detect potato blight problems. They utilized a dataset containing 2152 potato leaf images from the Kaggle Plant Village Dataset, with three classes: Early Blight, Late Blight, and healthy potato leaf images. Multiple classifiers, including KNN, SVM, Neural Network, and Logistic Regression, were employed for classification. Logistic Regression emerged as the top-performing classifier, achieving an impressive 97.8% classification accuracy over the test dataset. Computational tasks were executed using the orange data mining tool, complemented by an NVIDIA Ge-Force N16V GMR1 graphic card. The study underscores the efficacy of utilizing the pre-trained VGG19 model in conjunction with logistic regression to address potato leaf diseases [8]. The research of Tarik et al., (2021) centers on the application of Convolutional Neural Networks (CNN) for automating the detection of potato diseases. Utilizing advanced machine learning and image processing, the study analyzed 2034 images of unhealthy potatoes and leaves sourced from an open database, covering seven potato diseases, including potato leaf roll virus, hollow heart of potato, scab of potato, soft rot of potato, tuber worm, and more. The use of pre-trained models contributed to disease recognition, achieving an impressive accuracy of 99.23%, resulting in an overall accuracy of 99%. Data collection involved directly capturing images from potato fields, emphasizing seven disease types. Python's OpenCV library and a seven-level CNN model was instrumental in the study, presenting a robust model for precise potato disease diagnosis, with the ultimate goal of digitizing and enhancing potato production [9].

The research of Iqbal et al., (2020) presents an image processing and machine learning-based automatic system (Random Forest Classifier) for identifying and classifying potato leaf diseases. The information was compiled from the publicly available plant village database, which comprises over 450 photos of potato leaves. Among the seven classifiers, the Random Forest classifier had the highest accuracy (97%) in de-tecting and classifying potato leaf disease [10].

Fenu et al, (2020) unveiled an AI-driven approach using Support Vector Machine (SVM) to forecast potato late blight in Sardinia. Collaborating with ARPAS, they harnessed a 4-

year dataset from 50 locations and 36,034 instances, emphasizing temperature, humidity, and wind speed's significance. The model aligns with the LANDS Decision Support System and extends applicability beyond Sardinia. With 82% accuracy, feature sets 5 and 6 (including humidity, temperature, wind, and rainfall) prove pivotal for predicting late blight outbreaks, and promising practical applications in crop disease management [11].

The research of Nishad et al, (2022) seeks to detect and categorize potato leaf dis-eases using a deep learning algorithm and K-means clustering segmentation to improve model efficiency. Here collected dataset from two individual sources: PlantVillage and Mendeley. VGG16, VGG19, and ResNet50 network models were chosen, with VGG16 achieving 97% accuracy, the best performance among the three networks [12].

Zhao, Y. et al, (2021) proposed using DoubleGAN (a two-stage GAN) to generate high-resolution images of unhealthy leaves using fewer samples. The paper combines WGAN and ARGAN to get high-resolution images with clear and detailed information. Additionally, DCGAN was employed to generate 256x256 images and then comparing them with the originals and those generated by DoubleGAN. In this paper, 31361 leaf images were collected from the PlantVillage dataset. Unhealthy leaf images served as training data for balancing the dataset. VGG16, ResNet50, and DenseNet121 were then applied to classify the dataset before and after expansion. Classification accuracy of each leaf after the data set expansion was 98.58%, 99.14%, and 99.70% (VGG16, ResNet50, DenseNet121) affirming its effectiveness [13].

Arshaghi et al. (2022) used image processing and Convolutional Neural Network (CNN) methods to detect and classify surface defects in potatoes, focusing on five spe-cific classes of diseases: Healthy, Black Scurf, Common Scab, Black Leg, and Pink Rot. The software method proposed involves machine vision, feature extraction, and SVM for categorizing potato diseases using a dataset of 5000 potato images. Comparative analysis against established methods like Alexnet and Googlenet reveals promising outcomes, with the proposed approach achieving up to 100% accuracy in categorizing surface potato defects across the specified disease classes [14].

Table 1 provides a comparative study of model performance over several datasets. A study or research effort is represented by each row, which also includes the dataset used, the particular model used, and the accuracy that was attained. The table lists several examples of successful model implementations across various domains and datasets, ranging from Talukder et al.'s use of CTInceptionV3 with an accuracy of 91% on an open-source dataset to Arshaghi et al.'s CNN model attaining 99% accuracy on potato fields' footage. The potential application cross-crop transfer learning and the smooth integration of several pest identification models is lacking in the current research. The missing development of long-term monitoring models, resource optimization, and addressing real-world difficulties are critical gaps.

TABLE 1: Performance comparison of various models on different datasets

Author Name	Dataset	Model	Accuracy
Talukder et al., (2023)	Open Source(web)	CTInceptionV3 -RS	91%
Roldán-Serrato et al., (2018)	CPP-MBB Image database	RSC & LIRA	RSC: 89%, LIRA:88%
Anim-Ayeko et al., (2023)	PlantVillage Dataset	ResNet-9	99.25%
Tiwari, D et al., (2020)	Kaggle Plant Dataset	VGG19	97.8%
Tarik et al., (2021)	Open dataset - 2034 images	CNN	99%
Iqbal & Talukder (2020)	Open dataset-450 images	RF (Random Forest)	97%
Fenu & Malloci (2020).	ARPAS	SVM	82%
Nishad et al., (2022)	PlantVillage and Mendeley	VGG16	97%
Zhao et al., (2021)	PlantVillage dataset	DoubleGAN	99%
Arshaghi et al., (2022)	CFIA , USDA And potato farms in Ardabil city, Iran.	CNN	99%

The scope suggested here includes developing ethical frameworks, guaranteeing worldwide adaptability, and designing user-friendly interfaces. Further prospects for comprehensive breakthroughs in potato pest control include investigating ensemble models, optimizing for different hardware, using predictive analytics, and researching socio-economic consequences. The following contri-butions of our proposed methodology aims to closed some of these gaps:

- Comprehensive Dataset Construction: Through the combination of eight distinct datasets to provide a large, varied dataset for model training.
- Optimized Pre-processing Techniques: Applying and improving pre-processing methods to the combined dataset will enhance the quality of images. Verify the quality of the pre-processed image using mathematical equations.
- Comparative Analysis of Pre-trained Models: Using both raw and preprocessed datasets, thoroughly analyze several pre-trained models that have been fine-tuned (e.g., InceptionV3, VGG-16, MobileNetV2). Compare their performance measures in-depth to help academics and practitioners choose the best model for comparable tasks.
- Open-Source Implementation: Perhaps consider making our dataset, preprocessing methods, and refined models available to the scientific community as an open-source project. Collaboration, openness, and more developments in the area are encouraged by this.

II. MATERIALS AND METHODS

The main objective of potato insect identification is to identify the insect, classify it, and determine the negative effects it has on potato crops. In Figure 1 the workflow for achieving this objective is divided into seven segments.

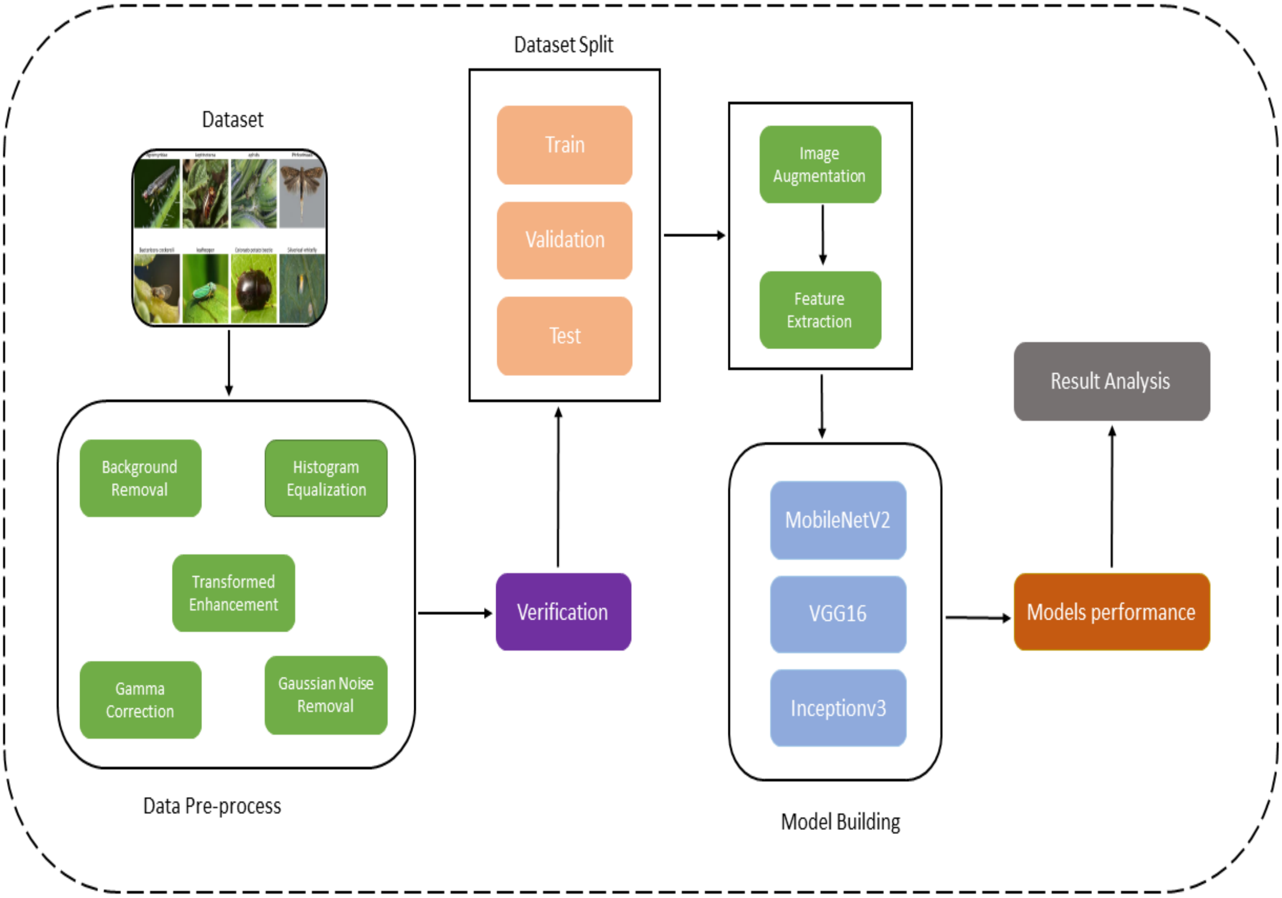


FIGURE 1: Proposed methodology for identifying, classifying and determining negative effects of insect of potato crops.

A. DATASET COLLECTION PROCEDURE

During the first stage, a wide range of publicly accessible platforms, such as websites, scholarly repositories, and open-access picture databases, were thoroughly searched. In total, 753 images were compiled from these sources, ensuring that each was relevant for the respective insect species under study. The insects are photographed in a diversity of environmental contexts and angles to maximize those that would represent most of the variations in lighting conditions, backgrounds, and angles, all of which are critical in trying to make the model more generalizable to different real-world situations. These images were afterward labeled manually into eight various species of insects. This process was overseen by a sub-assistant agricultural officer, Department of agricultural extension, Peoples Republic of Bangladesh through expert comments to make correct labeling. The species found in the dataset are as follows: Leptinotarsa – 102 images; Colorado Potato Beetle – 97 images; Aphids – 88 images; Leafhoppers – 97 images; Agromyzidae – 97 images; Bactericera Cockerelli – 92 images; Phthorimaea – 90 images; Silverleaf Whitefly – 90 images, in total 753 images. The distribution of the images for different species is given in detail in Table 2.

Thus, no name, age, or gender of any human being has been used to identify him or her for the dataset, hence not violating any privacy act.

These were then reviewed for correctness in the representation of the target species and checked against their relevance to the study. All photographs were thus checked under this quality assurance mechanism, ensuring that the species in the photograph were correctly represented and relevant. An effort was also made to see that the dataset contained those images with characteristics that were as diverse as possible. These would include conditions of surroundings and angles of anatomy. This diversity makes the dataset quite powerful.

Such images were collected from various online sources, which imposed very different terms at the moment of collection. On the whole, such diversity, coupled with the preprocessing techniques that will follow, greatly helps in balancing out any potential biases and contributes to making the model more general, thus escaping overfitting for certain environmental conditions. Figure 2 illustrates visual examples of the eight insect species in the dataset, highlighting their unique characteristics. This visual representation, along with the detailed inspection process, ensures the reliability and diversity of the dataset.

There is no human related data (i.e., name, age, gender, etc.) included in the dataset. With a priority on acquiring representative and diverse samples from various sources, the collecting method required painstaking attention to detail. Every picture was put through a thorough inspection process to make sure it accurately represented the desired bug species and was relevant.

In the following the description of the attributes of the various insect species now is provided.

Nuisance in agriculture, [15] leptinotarsa has yellow-orange bodies with black stripes and, when it reaches adulthood, zebra-like patterning on its exoskeleton. The bright yellow-orange oval body, black stripes, and unique antennae, which measure around 10–12 mm make the Colorado potato beetle stand out [15]. Aphids are tiny, pear-shaped insects that are usually green or yellow in color and have soft bodies and long antennae. They are frequently seen in bunches on plant leaves [16]. Leafhoppers are little, wedge-shaped insects with brightly colored bodies and long, rear legs that are designed for leaping [17].

The Agromyzidae family of insects is tiny (2–5 mm), with a thin body, trans-parent wings with characteristic vein patterns, and a yellowish to black hue [18]. Approxi-mately 2-3 millimeters in size, Bactericera cockerelli, commonly called the potato psyllid, has a unique wedge-shaped body with transparent wings and a mottled brown or yellowish hue [19]. With a wingspan of 9–12 millimeters, Phthorimaea operculella, often referred to as the potato tuberworm, is a tiny, brownish-gray moth that may be identified by its black markings on the forewings, which include a Y-shaped pattern near the center [20]. Belonging to the Aleyrodidae family, the silverleaf whitefly is modest in size, with females having yellowish bodies and males having smaller, slenderer bodies [21]. Its white, waxy wings are held tent-like over the body while at rest. This comprehensive overview captures the key characteristics of each insect species, providing a basis for the description of the corresponding figure.

B. IMAGE PRE-PROCESSING

In pre-processing our image dataset for analysis, we performed numerous essential data pre-treatment activities to assure the quality and consistency of our data as we prepared our image collection for analysis.

Figure 3 demonstrates an example of how the key pre-processing steps have been applied on the images.

TABLE 2: Distribution of species in the dataset

Class	Number of Original Images
Leptinotarsa	102
Colorado potato beetle	97
Aphids	88
Leafhoppers	97
Agromyzidae	97
Bactericera cockerelli	92
Phthorimaea	90
Silverleaf whitefly	90

1) Background Removal

Background removal is a technique of deletion of useless visual noises and distractions in images by segmenting the key subjects from their backgrounds. Focusing only on insects during this step will aid in scaling the model performance. This is because it reduces the influence of unimportant parts of images, and so the model focuses its attention only on those features appertaining to the target species. Removing the background leaves a pure grayscale image in which the information provided visually is further reduced [22]. In figure 3 an example of background removal applied to an image is provided, showing enhancing object isolation for improved visual clarity.

2) Histogram Equalization

It increases the contrast and makes features in an image more clearly visible by redistributing more pixel intensity values in wider tones. The use of enhanced contrast via histogram equalization allows the identification of finer details needed to assure the species. This way, potential enhancement in image quality could be achieved, hence making it easier for the features—like in the anatomy of insects—to be easily detected and ensure high model accuracy [23]. Figure 3 provides an example where the equalized image demonstrating improved contrast and brightness through histogram equalization.

3) Image enhancement

It refines some of the key features in the dataset, hence improving its quality. Enhanced images viewed at this step present the structures of insects more clearly, hence better model predictions. Techniques that can be used at this step include sharpening filters and adjusting the contrast since it will be easy to recognize different species by the model [24]. Figure 3 shown a transform-enhanced image showcasing optimized features after image processing trans-formation.

C. IMAGE AUGMENTATION

Image augmentation that was used included a random flip, rotation, width shift, height shift, shear, zoom, and brightness-altering to increase further variety in the dataset. This hugely increases the number of training data, hence reducing overfitting by increasing the generalization ability of the model on any unseen data. The augmentations are simulating various real-world scenarios; hence, this makes the model robust toward operating under different conditions [25]. For our image-based research to produce solid and trustworthy results, these augmentations were essential.

We carefully performed these primary steps to produce a clean and uniform dataset that is prepared for further analysis. These steps guarantee that our dataset looks good and perfectly ready for the goals of our methodology. The generated dataset will help produce accurate and reliable

FIGURE 2: Sample dataset having six categories of potato pests.

TABLE 3: Number of augmented images

Class	Number of Original Images	Number of Augmented Images
Leptinotarsa	102	3060
Colorado potato beetle	97	2970
Aphids	88	3000
Leafhoppers	97	3000
Agromyzidae	97	3000
Bactericera cockerelli	92	3000
Phthorimaea	90	3000
Silverleaf whitefly	90	3000

results in our research or analysis. It is represented in Table 3.

D. IMAGE VERIFICATION

In image processing, verification is the process of figuring out whether an image is real or has been altered. There are 10 images randomly taken from the dataset for image verification. Figure 4 represents the images.

In Table 4, the evaluation of image similarity in our dataset involves three distinct metrics. The Structural Similarity Index (SSIM) have scores ranging from approximately 0.198 to 0.587, where a higher score signifies greater similarity. Notably, the pair with a score of 0.198 exhibits significant differences in structure and perception, while the pair scoring 0.587 indicates a close match in terms of both structure and perception [26]. The Root Mean Square Error (RMSE) scores, ranging from about 46.78 to 103.93, measure average pixel-wise differences, with lower scores indicating higher similarity. Specifically, a pair with an RMSE score of 46.78 demonstrates minimal pixel-wise distinctions, suggesting substantial similarity, whereas a pair with a score of 103.93 exhibits larger differences, indicating lower similarity [27]. Additionally, the Mean Squared Error (MSCE) scores, spanning from approximately 2055.66 to 7622.15, serve as another measure of squared pixel-wise differences. A lower MSCE score corresponds to higher similarity, with the pair scoring 2055.66 displaying minimal squared pixel-wise differences, contrasting with the pair scoring 7622.15, which reflects larger squared pixel-wise distinctions and lower similarity [28].

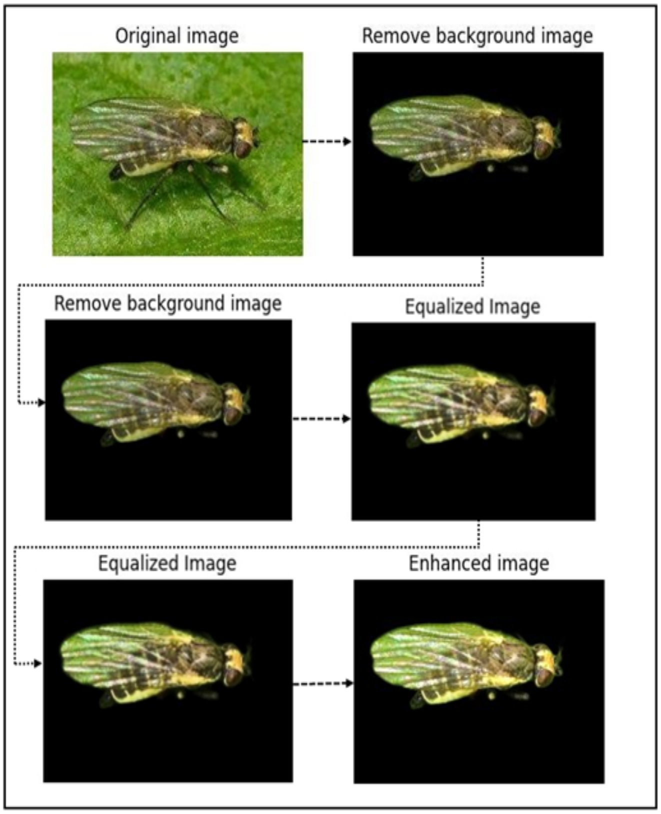


FIGURE 3: Pre-processing steps for the proposed methodology.

TABLE 4: Verfication parameters SSIM, RMSE, and MSCE

Number of Image	SSIM	RMSE	MSCE
Image-1	0.29	86.69	6003.95
Image-2	0.54	90.45	6426.36
Image-3	0.58	46.77	2055.65
Image-4	0.44	87.65	6177.40
Image-5	0.5	84.69	5784.63
Image-6	0.44	103.7	76221.02
Image-7	0.49	59.43	3222.69
Image-8	0.52	59.42	3222.912
Image-9	0.52	59.41	3220.39
Image-10	0.52	59.40	3219.28

E. DATA SPLITTING

A split in any machine learning or data analysis task is necessary, where a model should have appropriate generalization to unseen data. Thus, after the preprocessed data step was finished, the dataset was divided into three parts: 80% for training, 10% for testing, and 10% for validation. This will ensure most of the data go into training the model while the test and validation sets remain blinded during model performance evaluation and generalization.

The training set is used for model learning, where the model adjusts its parameters. The validation set is used in the training phase of a model to get an idea about performance and fix hyperparameters accordingly. The test set is utilized after training and that too for the purpose of having a very fair and objective look at various models' performances. This separation ensures that testing of any model is done on completely unseen data for proper estimation of real-world performance.

Besides that, the augmentation was performed only on the training data; such data became more diverse due to random transformations: flipping, rotation, and zooming. This kind of augmentation helps a model avoid overfitting so that it can see more variations in input, which may generalize better on new unseen data.

F. MODEL DESCRIPTION

1) Mobile Net V2

MobileNetV2 is a deep neural network architecture designed for mobile and em-bedded vision applications [29]. It is an improvement over the original MobileNet architecture and it is specifically optimized for running on resource-constrained devices such as smartphones, Internet of things (IoT) devices, and edge computing platforms. MobileNetV2 is designed for real-time or near-real-time inference, making it suitable for applications where quick identification is crucial, such as identifying potato insects in the field. Its efficient architecture allows it to process images rapidly. Despite its small size, MobileNetV2 still maintains a good level of accuracy in image classification tasks. It leverages techniques like depth-wise separable convolutions and inverted residual blocks to capture meaningful features in images efficiently [29].

2) VGG-16

VGG16 is a deep convolutional neural network (CNN) architecture that consists of 16 weight layers, including 13 convolutional layers and 3 fully connected layers [30]. The convolutional layers use small receptive fields (3x3) with a stride of 1, and maxpooling is applied with a 2x2 window. The deep and hierarchical structure of VGG16 enables it to automatically learn hierarchical features from images. In insect identification, this is beneficial for capturing intricate patterns and details. VGG-16 is often used as a pre-trained model for transfer learning. Pre-trained versions of VGG-16 on large image datasets like ImageNet are available, allowing researchers and developers to fine-tune the model for specific image classification tasks with relatively small datasets [30].

3) Inception-V3

Inception-V3 is a deep convolutional neural network (CNN) architecture that is designed for image classification and object recognition tasks [31]. Inception V3 introduces auxiliary classifiers at intermediate layers during training. These classifiers provide additional gradients for the backpropagation process and help mitigate the vanishing gradient problem. Insect images may contain objects of various sizes and scales. Inception V3's use of Inception modules facilitates multi-scale feature extraction, allowing the model to capture information at different resolutions [31]. That's why we choose this model for insect identification.

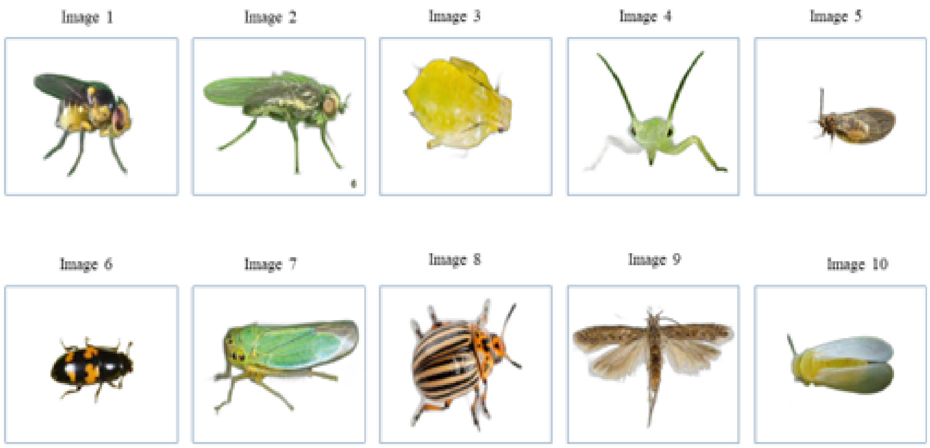


FIGURE 4: Sample images for verification.

4) Hyper-parameters of the Models

The training was designed based on some key parameters (Table 5) concerning efficiency and performance. Whereas at first, the number of epochs was set to 10, early stopping was used to avoid overfitting which would have cut down the time taken for training. It then uses this in monitoring validation loss, stopping the training when it no longer sees any considerable improvements for a certain number of epochs, which allows the models to be optimally converged even before reaching the maximum number of epochs. In the selected batch size for the experiments, it is set as 32, considering that this offers a good trade-off between memory efficiency and generalization quite well when training deep learning models on image data. Here, the learning rate was 0.001 because this is the ideal starting point for deep learning models; weights converge smoothly and do not face substantial fluctuations in their updates. An Adam optimizer because it is an adaptive learning rate optimizer that works perfectly well for the training of complex models such as InceptionV3, VGG-16, and MobileNetV2. Here, Adam's dynamic adjustment of the learning rate assured faster and more reliable convergence in contrast to traditional optimizers such as SGD. Thirdly, since the nature of the classification problem was multiclass, categorical cross-entropy was used as a loss function. That would drive the models quite effectively toward making proper predictions. All these parameters remain the same across all the models to keep the tuning of models consistent and their performances comparable.

TABLE 5: Comparison of performance metrics of models

Parameter	Value
Number of Epochs	10
Batch Size	32
Learning Rate	0.001
Optimizer	Adam
Loss Function	Categorical Cross-Entropy

III. RESULTS AND DISCUSSION

A. PERFORMANCE METRICS

Performance measurement metrics in the context of datasets refer to quantitative measures used to evaluate the quality, characteristics, and effectiveness of a dataset. These metrics help assess various aspects of a dataset's suitability for specific data analysis or machine learning tasks.

We use the abbreviations TP for True Positive, TN for True Negative, FP for False Positive. and FN for False negative and we have the following metrics:

Eq. 1 the accuracy [32] is a measure of how close the predicted/generated value is to the true value of the item.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Eq. 2 the recall [32] is a measure of how many true positives were found in all.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Eq. 3 the precision [32] is an indication of the reproducibility of the predicted value, if the same value is fed to the model twice, how far apart would the results be for each case.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Lastly, in Eq. 4 the F1-score [32] is calculated, which is the harmonic mean of precision and recall.

$$F1 - score = \frac{2(Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3488730

TABLE 6: Comparison of performance metrics of models

Model	ACC	Recall	Precision	F1 Score
InceptionV3	66.67%	0.67	0.64	0.63
VGG-16	94.44%	0.96	0.96	0.96
Mobile Net V2	86.11%	0.86	0.93	0.84

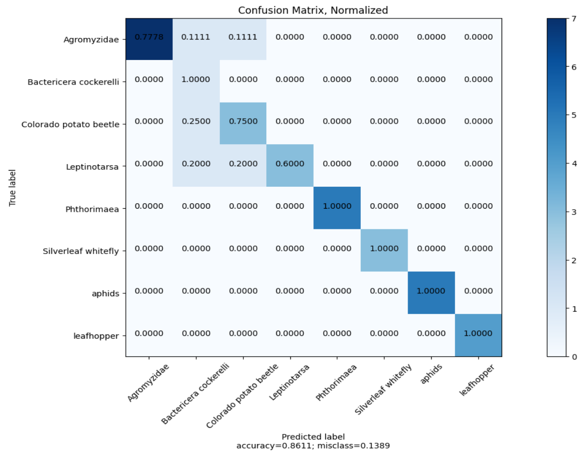
The performance study of three transfer learning models is given in Table 5. From this table one can see that, VGG-16 has the best performance overall, with an accuracy of 94.44%, recall of 0.96, precision of 0.96, and F1 score of 0.96. Mobile Net V2 has the second-best performance, with an accuracy of 86.11%, recall of 0.86, precision of 0.93, and F1 score of 0.89. InceptionV3 has the worst performance, with an accuracy of 66.67%, recall of 0.67, precision of 0.64, and F1 score of 0.63.

B. RESULT ANALYSIS OF PRE-PROCESSED DATASET

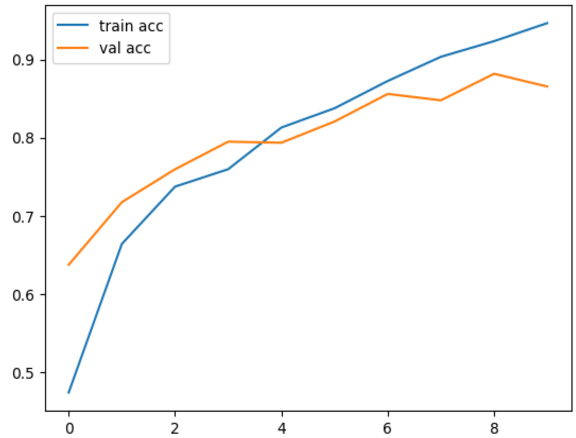
The MobileNetV2 algorithm has achieved the accuracy of 86.11%. for this dataset. Here, 23 images were perfectly predicted. However, when applying this model, seven images were misclassified such that the model predicted the image of a Colorado potato beetle as a leafhopper, 3 images of Leptinotarsa were predicted as Agromyzidae, 2 im-ages of Phthorimaea were predicted as Leptinotarsa and 1 image of Silverleaf whitefly was misclassified as Agromyzidae. The ACC score was 0.78 and ROC score was 0.61 in the dataset. Figure 5, and 6 shows confusion metrics, ACC curve and ROC curve of Mobile Net V2.

The VGG-16 algorithm obtained an accuracy of 94.44%. The mis-classification rate was 0.0556, which means that 5.56% of the predictions were incorrect. Applying this model, basically two classes were occasionally misidentified such as the image of Agromyzidae were predicted as Bactericera cockerelli and the image of aphids were predicted as Agromyzidae. The ACC score was 0.97 and ROC score was 0.85 in the dataset. Most classes were predicted with perfect accuracy, two classes had some instances of being misclassified, and the overall model accuracy was quite high. Figure 7, and 8 shows confusion metrics, ACC curve and ROC curve of VGG-16.

The InceptaV3 algorithm achieved the accuracy of 66.67%. for this dataset. The misclassification rate is 0.3333, indicating 33.33% of predictions were incorrect. How-ever, when applying this model 9 classes of image were misclassified such that Agromyzidae was predicted as Bactericera cockerelli Colorado potato beetle and leaf-hopper, Bactericera cockerelli was predicted as Colorado potato beetle, Colorado potato beetle was predicted as Bactericera cockerelli, Leptinotarsa was predicted as Bactericera cockerelli , Phthorimaea was predicted as Agromyzidae, aphids were misclassified as Silverleaf whitefly and leafhopper and leafhopper was misclassified as aphids. The ACC score was 0.68 and ROC score was 0.59 for the dataset. Figure 9, and 10 shows confusion metrics, ACC curve and ROC curve for InceptaV3.



(a) Confusion Matrix



(b) ACC Curve

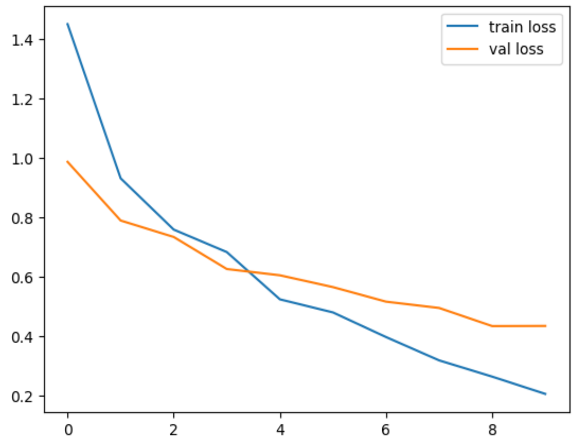
FIGURE 5: (a) Confusion metrics for Mobile Net V2; (b) ACC curve for Mobile Net V2.

C. RESULT ANALYSIS OF RAW DATASET

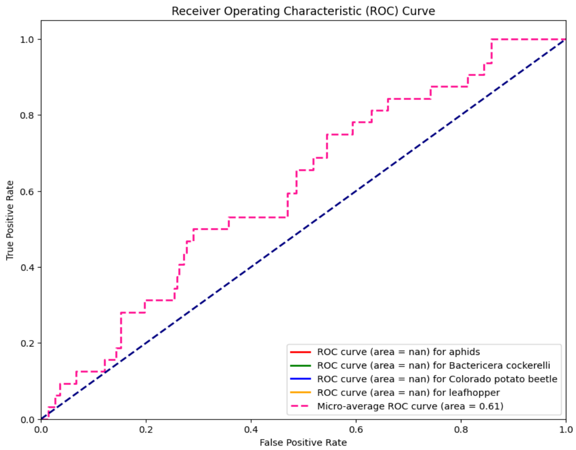
The performance measures findings of for a range of models that use raw photos as input are shown in Table 6. Accuracy (ACC), Recall, Precision, and F1 Score are among the parameters evaluated. As noted, the percentage of properly categorized examples is known as accuracy, while the percentage of real positive occurrences that the model correctly recognized is known as recall. The F1 Score strikes a compromise between Precision and Recall. Precision denotes the percentage of positive cases that are actually positive. With the highest precision (94%), and the highest accuracy (92%), VGG-16 is the best-performing model in this comparison. InceptionV3 shows similar percentages for Recall and Precision, despite its decreased accuracy. MobileNet V2 provides a decent trade-off between Recall and Precision, however it is not as exact as VGG-16.

Figure 11, and 12 shows confusion matrix of Inception-V3, VGG-16, and Mobile Net V2.

For the detection of pests, this research concentrated mainly on deep learning models (InceptionV3, VGG-16, and

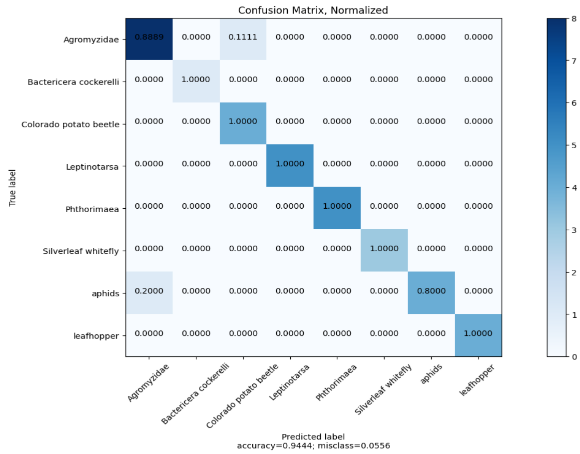


(a) Loss curve

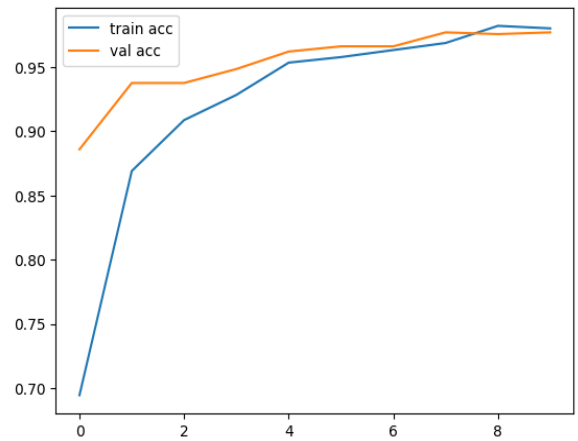


(b) ROC Curve

FIGURE 6: (a) Loss curve of Mobile Net V2; (b) ROC curve of Mobile Net V2.



(a) Confusion Matrix



(b) ACC Curve

FIGURE 7: (a) Confusion metrics of VGG-16; (b) ACC curve of VGG-16.

TABLE 7: Comparison of performance metrics of models on raw image dataset

Model	ACC	Recall	Precision	F1 Score
InceptionV3	58%	0.58	0.61	0.58
VGG-16	92%	0.92	0.94	0.88
Mobile Net V2	75%	0.75	0.71	0.70

MobileNetV2). This is because they are more effective when it comes to dealing with demanding aspects of image identification than others. In contrast to old-fashioned ways where people came up with features on their own, deep learning machines can do this procedure on their own which makes identification easier since these machines can easily adapt to different environments including different insects' growth stages and surrounding habitats. Large-scale, practical applications are less suited for traditional models like SVM or k-NN since they frequently suffer from such unpredictability. The selection of these models is further supported by the increasing use of deep learning in agricultural research and recent improvements in computing capacity. The goal of this study is to provide pertinent insights into contemporary pest detection techniques while being in line with state-of-the-art methodologies.

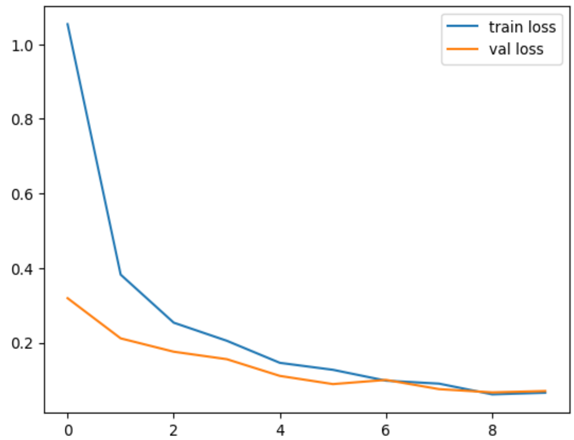
IV. CONCLUSION

Our objective was to identify the potato beetle in potato crop images. In that case, we worked with 8 species of insects. We have used 3 models for insect identification, where the accuracy of the VGG-16 model is the highest. Overall, our study shows that VGG-16 is a useful tool for pest management in potato farming since it can identify potato insects with accuracy. In our work, we had a vision to implement the real-life detection model to detect those insects in potatoes that can be easily detected with AI or mobile apps, although that is our future plan for this work. In this work, we wanted to work on detection which we succeeded in accomplishing.

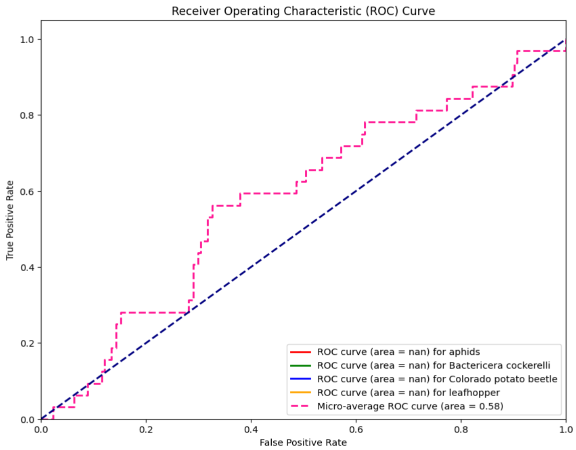
REFERENCES

[1] Food and Agriculture Organization of the United Nations., International Fund for Agricultural Development., UNICEF., World Food Programme., and World Health Organization., The state of food security and nutrition in the world: building climate resilience for food security and nutrition.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3488730

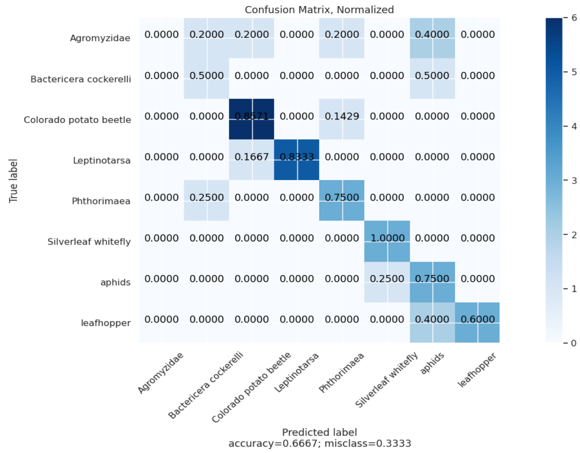


(a) Loss curve

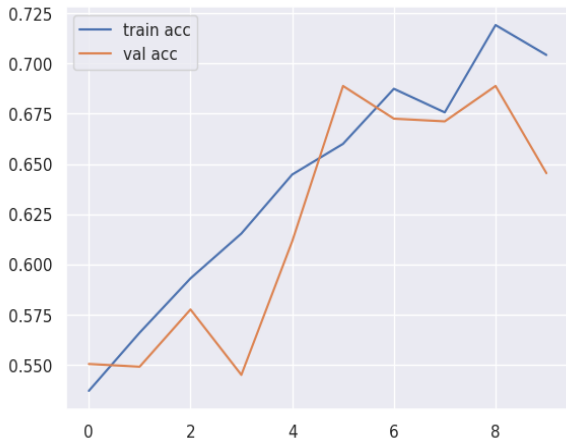


(b) ROC Curve

FIGURE 8: (a) Loss curve of VGG-16; (b) ROC curve of VGG-16.



(a) Confusion Matrix



(b) ACC Curve

FIGURE 9: (a) Confusion metrics of Inception V3; (b) ACC curve of Inception V3.

[2] Z. Bashir, M. Bowling, and L. H. S. Lelis, "Assessing the Interpretability of Programmatic Policies with Large Language Models," Nov. 2023, [Online]. Available: http://arxiv.org/abs/2311.06979

[3] S. and S. M. A. Chakrabarti S. K. and Sharma, "Potato Pests and Diseases: A Global Perspective," in Sustainable Management of Potato Pests and Diseases, S. and S. M. A. Chakrabarti Swarup Kumar and Sharma, Ed., Singapore: Springer Singapore, 2022, pp. 1–23. doi: 10.1007/978-981-16-7695-6_1.

[4] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," Comput Electron Agric, vol. 151, pp. 61–69, Aug. 2018, doi: 10.1016/J.COMPAG.2018.05.012.

[5] M. Simul Hasan Talukder, R. Bin Sulaiman, M. Raziuddin Chowdhury, M. Saberin Nipun, T. Islam, and C. Author, "PotatoPestNet: A CTInceptionV3-RS-Based Neural Network for Accurate Identification of Potato Pests."

[6] K. L. Roldán-Serrato, J. A. S. Escalante-Estrada, and M. T. Rodríguez-González, "Automatic pest detection on bean and potato crops by applying neural classifiers," Engineering in Agriculture, Environment and Food, vol. 11, no. 4, pp. 245–255, Oct. 2018, doi: 10.1016/j.eaef.2018.08.003.

[7] 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, vol. 4, p. 100178, Aug. 2023, doi: 10.1016/j.atech.2023.100178.

[8] 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), IEEE, May 2020, pp. 461–466. doi: 10.1109/ICICCS48265.2020.9121067.

[9] M. I. Tarik, S. Akter, A. Al Mamun, and A. Sattar, "Potato Disease Detection Using Machine Learning," in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, Feb. 2021, pp. 800–803. doi: 10.1109/ICICV50876.2021.9388606.

[10] M. Islam, Anh 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), IEEE, Apr. 2017, pp. 1–4. doi: 10.1109/CCECE.2017.7946594.

[11] G. Fenu and F. M. Malloci, "Artificial Intelligence Technique in Crop Disease Forecasting: A Case Study on Potato Late Blight Prediction," 2020, pp. 79–89. doi: 10.1007/978-981-15-5925-9_7.

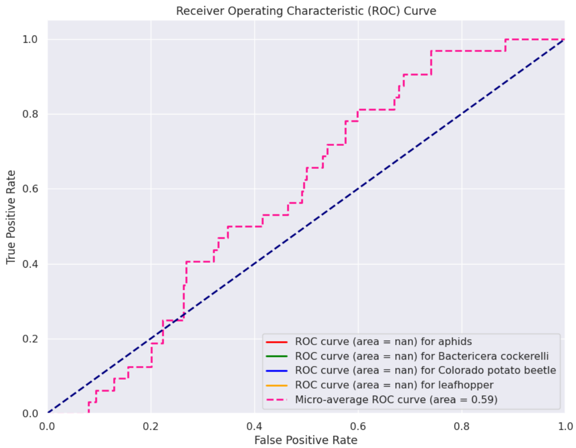
[12] Md. A. R. Nishad, M. A. Mitu, and N. Jahan, "Predicting and Classifying Potato Leaf Disease using K-means Segmentation Techniques and Deep Learning Networks," Procedia Comput Sci, vol. 212, pp. 220–229, 2022, doi: 10.1016/j.procs.2022.11.006.

[13] Y. Zhao et al., "Plant Disease Detection using Generated Leaves Based on DoubleGAN," IEEE/ACM Trans Comput Biol Bioinform, pp. 1–1, 2021, doi: 10.1109/TCBB.2021.3056683.

[14] A. Arshaghi, M. Ashourian, and L. Ghabeli, "Potato diseases detection and classification using deep learning methods," Multimed Tools Appl, vol. 82, no. 4, pp. 5725–5742, Feb. 2023, doi: 10.1007/s11042-022-13390-1.

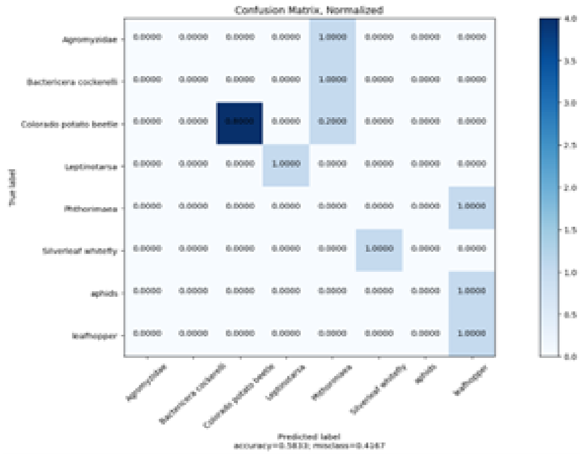
[15] Available online: "https://picryl.com/media/colorado-potato-beetle-leptinotarsa-decemlineata-insect-175bdc " (accessed on 4 June 2024).

(a) Loss curve



(b) ROC Curve

FIGURE 10: (a) Loss curve of InceptionV3; (b) ROC curve of InceptionV3.



(a) Confusion Matrix

(b) Confusion Matrix

FIGURE 11: (a) Confusion metrics of Inception V3 (Raw Images); (b) Confusion metrics of Mobile Net V2 (Raw Images)

[16] Available online: "https://influentialpoints.com/Blog/Identifying_aphids_to_genus_from_photos.htm " (accessed on 4 June 2024).
[17] Available online: "https://www.pbase.com/tmurray74/leafhoppers_cicadellidae " (accessed on 4 June 2024).
[18] Available online: "https://tasmaniansinsectfieldguide.com/hexapoda/insectsoftasmaniadiptera/suborderbrachycera/infraorder-cyclorrhapha/agromyzidae-leaf-miner-flies/ " (accessed on 4 June 2024).
[19] Available online: "https://www.inaturalist.org/taxa/372093-Bactericera-cockerelli/browse_photos" (accessed on 4 June 2024).
[20] Available online: "https://www.alamy.com/stock-photo/phthorimaea.html?sortBy=relevant" (accessed on 4 June 2024).
[21] Available online: "https://alchetron.com/Silverleaf-whitefly " (accessed on 4 June 2024).
[22] J. Liang, Y. Liu, and V. Vlassov, "The Impact of Background Removal on Performance of Neural Networks for Fashion Image Classification and Segmentation," Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.09764
[23] Z. Bashir, M. Bowling, and L. H. S. Lelis, "Assessing the Interpretability of Programmatic Policies with Large Language Models," Nov. 2023, [Online]. Available: http://arxiv.org/abs/2311.06979
[24] N. Bhardwaj, G. Kaur, and P. K. Singh, "A Systematic Review on Image Enhancement Techniques," 2018, pp. 227–235. doi: 10.1007/978-981-10-6614-6_23.
[25] S. Shakil, A. A. K. Akash, N. Nabi, M. Hassan, and A. Haque, "PithaNet: A transfer learning-based approach for traditional pitha classification," International Journal of Electrical and Computer Engineering (IJECE), vol. 13, no. 5, p. 5431, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5431-5443.
[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
[27] M. Z. Naser and A. H. Alavi, "Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences," Architecture, Structures and Construction, vol. 3, no. 4, pp. 499–517, Dec. 2023, doi: 10.1007/s44150-021-00015-8.
[28] M. BÜYÜKKEÇECİ and M. C. OKUR, "A Comprehensive Review of Feature Selection and Feature Selection Stability in Machine Learning," Gazi University Journal of Science, vol. 36, no. 4, pp. 1506–1520, Dec. 2023, doi: 10.35378/gujs.993763.
[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.04381
[30] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.1556
[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.
[32] M. Hasan, M. J. Zobair, S. Akter, M. Ashef, N. Akter, and N. B. Sadia, "Ensemble Based Machine Learning Model for Early Detection of Mother's Delivery Mode," in 2023 International Conference on Electrical,

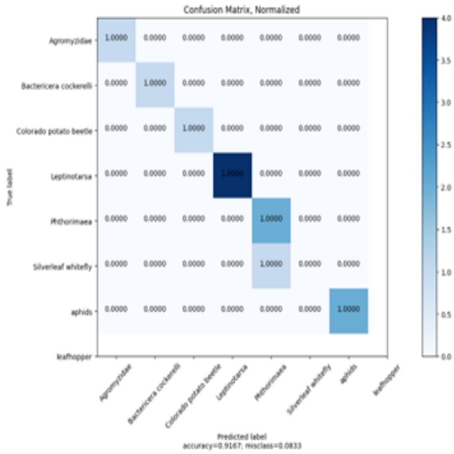


FIGURE 12: Confusion metrics of VGG-16 (Raw Images)

Computer and Communication Engineering (ECCE), 2023, pp. 1–6. doi: 10.1109/ECCE57851.2023.10101558.

AMIR SOHEL is working as a Senior Lecturer at the Department of Computer Science and Engineering at the Daffodil International University, Dhaka, Bangladesh. He works as a principal investigator in the Multidisciplinary Action Research (MARS) Lab at the Daffodil International University, Dhaka, Bangladesh. He completed his Masters in Computer Science and Engineering from Jahangirnagar University and Bachelor in Computer Science and Engineering from Daffodil International University. His research interests are computer vision, machine learning, deep learning.

MD. SHAHRIAR SHAKIL is a passionate machine learning developer. He has both academic and industrial professional experiences. Currently, He is working as a Lecturer in CSE department at Daffodil International University. Previously, he worked as a Teaching Apprentice Fellow at Daffodil International University and a Machine Learning Trainee Programmer at Expert Consortium Ltd. He works as a principal investigator in the Multidisciplinary Action Research (MARS) Lab at the Daffodil International University, Dhaka, Bangladesh. With his theoretical knowledge and practical experience, he want to build my career in this competitive world. He has published some results of research in international peer-reviewed publications. He published several courses at the International Online University (IOU) related to Python. My research interests are Machine learning, Deep learning, Computer vision, Natural language processing, and Data mining.

SHAH MD. TANVIR SIDDIQUEE is currently working as an assistant professor at the Daffodil International University, Dhaka, Bangladesh. He works as a principal investigator in the Multidisciplinary Action Research (MARS) Lab at the Daffodil International University, Dhaka, Bangladesh. He completed his Masters in computer science from South Asian University, New Delhi, India and B. Sc. in computer science and engineering from Hajee Mohammad Danesh Science and Technology University, Dinajpur. His research interests are machine learning, artificial neural network, computer vision, deep learning, and image processing.

AHMED AL MAROUF (Member, IEEE) is currently a Ph.D. Candidate and graduate assistant – teaching at the Department of Computer Science of University of Calgary, Alberta, Canada. He works with the Advanced Database Systems and Applications (ADSA) lab. He received his bachelor's and master's degree from the Department of Computer Science and Engineering (CSE), Islamic University of Technology (IUT), Gazipur, Bangladesh. He is currently on study leave from the Department of Computer Science and Engineering (CSE), Daffodil International University (DIU), Dhaka, Bangladesh, where he acts as the Technical Lead of the Human-Computer Interaction (HCI) Research Lab and Multidisciplinary Action Research (MARS) Lab. His research interest lies within data science, health informatics, bioinformatics, machine learning and computational social science.

JON G.ROKNE (Life Fellow, IEEE) is a professor and former Head in the Computer Science department at the University of Calgary. Educated in Norway and Canada he obtained a Ph. D. in Mathematics from the University of Calgary. He chaired the Computer Science Department at the University of Calgary from 1989 to 1996. He has published extensively in mathematics, and computer science including co-authorship of the encyclopedia of Social Network and Mining. He is also co-organizer of the IEEE/ACM ASONAM Conference Series. He has been extensively involved with IEEE activities including having been the IEEE Publications Services and Products Board Vice-President 2009-2010, membership on a number of boards and committees and developing several new IEEE transactions and magazines.

REDA ALHAJJ (Senior Member, IEEE) is a tenured professor in the Department of Computer Science at the University of Calgary, Alberta, Canada. He is also affiliated with Medipol University, Istanbul, Turkey, and University of Southern Denmark, Odense, Denmark. He published over 500 papers in refereed international journals, conferences, and edited books. He served on the program committee of several international conferences. He is founding editor in chief of the Springer premier journal “Social Networks Analysis and Mining”, founding editor-in-chief of Springer Series “Lecture Notes on Social Networks”, founding editor-in-chief of Springer journal “Network Modeling Analysis in Health Informatics and Bioinformatics”, founding co-editor-in-chief of Springer “Encyclopedia on Social Networks Analysis and Mining” (ranked top 3rd in most downloaded sources in computer science in 2018), founding steering chair of the flagship conference “IEEE/ACM International Conference on Advances in Social Network Analysis and Mining”, and three accompanying symposiums FAB (for big data analysis), FOSINT-SI (for homeland security and intelligence services) and HI-BI-BI (for health informatics and bioinformatics). He is member of the editorial board of the Journal of Information Assurance and Security, Journal of Data Mining and Bioinformatics, Journal of Data Mining, Modeling and Management; he has been guest editor of a number of special issues and edited a number of conference proceedings. Dr. Alhajj’s primary work and research interests focus on various aspects of data science and big data.