Automated COCOA Disease Detection Using Convolutional Neural Networks: A Case Study of VSD and Other Pathogens

Sachin Kumar[1], Avadhesh Kumar Sharma[2], Saurabh Srivastava[3], Vivek Tiwari[4], Praveen Kumar Patidar[5], Sangeeta Rai[6]

[1, 4, 5] Department of CSE, Parul University, Vadodara, Gujarat, India
[2]Department of ECE, GL Bajaj Institute of Technology & Management, Greater Noida, U.P, India,
[3]Department of CSE, United University, Rawatpur, Prayagraj-U.P, India,
[6]Sangeeta Rai, School of CSE, Lovely Professional University, Phagwara, Panjab,
jaiswalsachin009@gmail.com

Abstract: Theobroma cacao L., commonly known as cocoa, is a plantation crop of significant economic value, renowned for its dried fruits. The high market demand for cocoa is not negatively correlated with its low production output. The high prevalence and quick spread of illness is the main problem in cocoa farms. A majority Vascular Streak Dieback (VSD) is a prevalent illness. To maintain productivity, appropriate treatment must be administered promptly. The diagnosis of cocoa leaf disease diseases can be sped up and made simpler by utilizing a "Convolutional Neural Network (CNN)" to identify diseases based on leaf images. The main objective of this research article is to distinguish VSD-infected cocoa plants, we have used total 1200 image for classification of VSD disease. DenseNet-19 shows the best result with accuracy of 99.1% in 7.48 minutes only.

Keywords: Cocoa leaf disease, Vascular streak dieback, Convolutional neural network.

1.   Introduction

Cocoa is one of the core agricultural products of Indonesia. It plays a significant role in the economy Indonesia. The quality of chocolate produced by Indonesia is good in comparable to that of global cocoa since it has the benefit of resist of melting. Due to these benefits, both the domestic and export markets for cocoa in Indonesia are fairly open. However, due to complicated issues including illnesses, Indonesia's cocoa production is still low [1]. Vascular Streak Disease (VSD) is one of the conditions that affect coffee crops the most. The fungus Oncobasidiumtheobromae (Ceratobasidiales: Ceratobasidiaceae) is responsible for VSD disease [2]. Yellowing of the leaves is often the first symptom to appear. These leaves develop a scattering of tiny, distinct green specks against a yellow background. This disease can be fatal and spreads rapidly. Increasing cocoa productivity is a difficult for cocoa plants. To reach the desired production rate, precision agriculture is necessary answer.Deep learning has been

utilized in precision agriculture [3-4]. One of the most crucial components of artificial intelligence is neural network. It is an adaptable system that draws inspiration from the human brain. A neural network may learn from previous data to perform a variety of tasks, including pattern recognition, data classification, and event forecasting. A component of ANN, which is currently widely utilized for classification of image [5], is convolutional neural networks. Any type of data can be input into a CNN, including quantitative, graphical, audiovisual, and natural language data [6-7]. CNN is frequently utilized for image classification. CNNs have been widely used across various industries, including agriculture, healthcare, mail services, web services, and more.

Numerous studies have been conducted to automatically classify and identify plant diseases, including a study by Brahimi et al. Using AlexNet and GoogleNet, they accurately diagnosed nine infections in tomato leaves with a 99.23% rate of success. In a comparison of four CNN models for classifying three diseases affecting tomato leaves, Baihaki et al. [8] found that VGG-16 achieved the highest accuracy rate at 99.67%. Darwis et al. [9] developed a model for recognizing 3 infections of maize and discovered an accuracy of 98.2% using VGG-16 and VGG-19. With an accuracy of 97.62%, Liu et al. developed a novel CNN architecture to recognize apple leaf infections. The authors used CNN architecture in similar study to find and identify diseases in maize [10, 11]. The authors also employ the CNN architecture in other plants. According to a study by Tugrul et al., no one has yet explored image-based classification of leaf diseases affecting cocoa plants.

CNN are used as a feature extractor & classifier in this study. We employ the DenseNet, SqueezeNet, AlexNet& Modified-CNN architectures. We divide leaves of cocoa into two groups: the healthy group and the VSD group.

1.1   Steps for VSD Disease Recognition in COCOA leaf

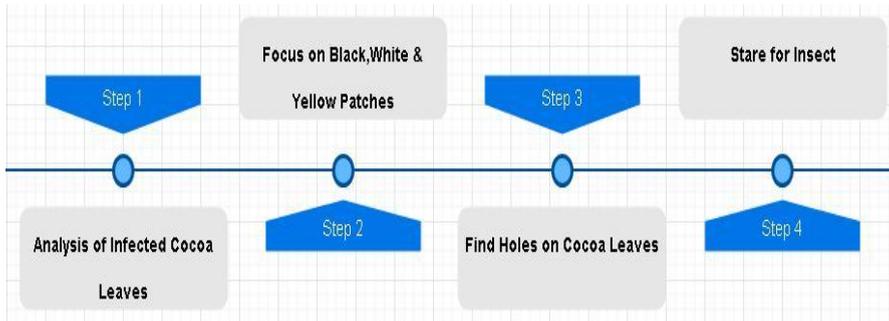There are few steps to identify VSD disease in cocoa leaves. These steps are mentioned in Fig 1.

Fig. 1. Steps for VSD Disease Recognition

1.2   VSD Infection in Cocoa Leaf

Fig. 2 shows images of healthy and VSD infected cocoa leaf.

Healthy Cocoa Leaf   Vascular Streak Dieback (VSD)

Fig. 2. Images of Healthy and VSD Infected Cocoa Disease

The main objective of this research article is to classify VSD infection using deep convolutional neural networks. This research paper proposed two key contributions to the classification of cocoa leaf:

-Development of the model for classification of cocoa leaves utilizing CNN.
-Recognition of VSD infection in cocoa leaves

This suggested article's remaining content is arranged as follows:

-The literature review on the recommended methodology is covered in this section.
-The research methodology is covered in this section.
-The collection of datasets on cocoa illnesses and the creation of datasets with images of cocoa diseases are covered in this section.
-This section provides an explanation of the DenseNet 19 analysis and result discussion.

-The conclusion of the proposed experiment and upcoming research are covered in this section.

2   Literature Review

A computational model designed to diagnose diseases in cocoa pods using deep learning techniques is presented in the research article "Deep Learning-Based Computational Model for Disease Identification in Cocoa Pods ". In order to precisely identify different diseases that afflict cocoa pods, the study uses convolutional neural networks (CNNs), a sort of deep learning architecture, to analyse photos of the pods. Through model training on a dataset of photos of both healthy and diseased cocoa pods, the researchers were able to create a robust algorithm that could differentiate between various disorders, including black pod, frosty pod, and swelling shoot. The outcomes show how well the suggested deep learning technique works to automate the identification of diseases in cocoa plants, which may eventually help farmers put early disease control measures into practice [12].

The research paper "Cocoa Companion: Deep Learning-Based Smartphone Application for Cocoa Disease Detection" describes a deep learning-based smartphone application that may be used to identify diseases in cocoa plants. Convolutional Neural Networks (CNNs) are used by the Cocoa Companion app to analyse smartphone photos of cocoa plants and identify a variety of diseases, such as frosty pod and black pod. Because of the system's architecture, which is designed for real-time processing on mobile devices, farmers may identify plant health problems in the field with greater speed. Using a dataset of photos of cocoa plants, the study tests Cocoa Companion's effectiveness and reports encouraging findings regarding its ability to correctly identify and categorise various diseases. [13]

An enhanced VGG19 Convolutional Neural Network (CNN) model is used in the research paper "An Image-Based Cocoa Diseases Classification Based on an Improved VGG19 Model" to present a novel method of categorising cocoa diseases. Using image-based classification approaches, the study aims to address the problem of reliably recognising various illnesses affecting cocoa plants. By adding transfer learning and fine-tuning techniques, the researchers improve the VGG19 model's performance and make it more capable of extracting discriminative characteristics from photos of cocoa disease. Utilising a dataset containing photos of several cocoa illnesses, such as swollen shoots and black pods, the suggested method is assessed. The experimental findings show that when compared to conventional CNN architectures, the enhanced VGG19 model obtains better classification accuracy. [14]

"Classification of Cacao Pod if Healthy or Attacked by Pest or Black Pod Disease" is the title of the study paper. It describes a study that employed deep learning algorithms to classify cocoa pods into three categories: healthy, afflicted by pests, or impacted by black pod disease. In order to effectively manage disease and maintain quality control in the production of cacao, it is imperative that a dependable and efficient method for determining the health condition of cacao pods be developed. The study makes use of a deep learning system, most likely a convolutional neural network (CNN) that was trained on a dataset that included pictures of cocoa pods that were pest-affected, disease-affected, and in good condition. The algorithm gains the ability to accurately distinguish between the various categories by examining the visual features of these photographs. The findings of the experiment show how well the deep learning method works to accurately identify cacao pods according to their health state, providing farmers and other agricultural professionals with an invaluable tool for better monitoring and management of cacao plants. [15, 20]

3   Research Methodology

This section outlines the research's methodology, which includes loading of RGB images, segmentation, feature extraction, and classification of cocoa leaf disease using 4 architectures of CNN i.e SqueezeNet, AlexNet, & DarkNet modified CNN for model evaluation. We have proposed a model whose output shows accuracy with time.

The subject of the study is cocoa, which is extensively cultivated in South and Southeast Asia. The research institute in Indonesia grows cocoa plants from which the leaf image data is derived. Two classes of leaves were taken: VSD & healthy. Identification of infection in cocoa leaves was based on the state of the leaves and knowledge from local farmers & literature. Each photograph is saved as a jpg file with a 240 x 240 pixel image size.

3.1   Convolutional Neural Network

A type of ANN typically employed in image processing and recognition includes convolutional neural networks. The operation in the hidden layer modifies the data in order to discover unique features in the data. Convolution, activation, and pooling layers are the three most utilised layers in CNN. Fig. 3 represents a convolutional neural network.
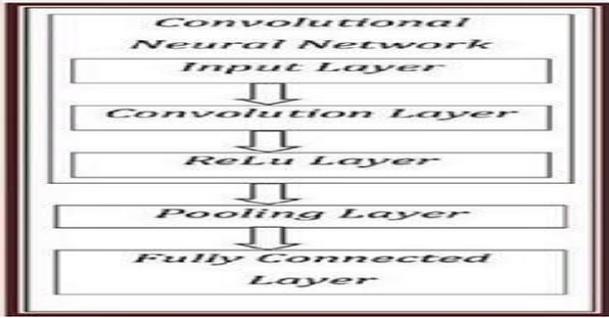
Fig. 3. Convolutional Neural Network

3.2   Rectified Linear Unit (ReLU)

It is a neural network that uses an activation function. It promotes sparse activation where negative values are set to zero. Fig 4 represents a graph of the ReLU layer. Table 1 represents the output of a function after applying the ReLU layer.

Table 1 : Rectified Linear Unit

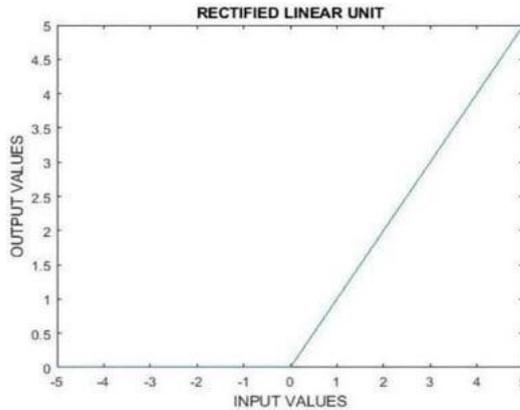| Y	-99	-98	-86	-74	-21	0	11	29	38
f(y)	f(-99)	f(-98)	f(-86)	f(-74)	f(-21)	f(0)	f(11)	f(29)	f(38)
F(y)	0	0	0	0	0	0	11	29	38

Fig. 4. Rectified Linear Unit

3.3   Pooling Layer

It uses nonlinear down-sampling on the output to lower the amount of parameters the network needs to learn. Fig. 5 represents a max pooling layer with 2×2 filters and a stride of 2.
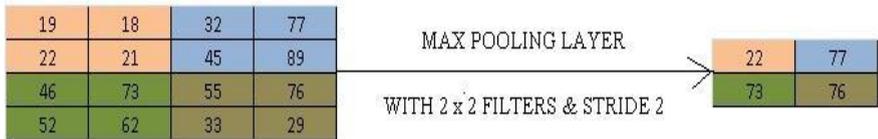

Fig. 5. Max Pooling Layer

4   Dataset Collection and Preparation

Data from cocoa plants cultivated at the "Indonesian Coffee and Cocoa Research Institute in Jember, East Java," were used in this study. The seized leaves were healthy and VSD-infected. This data collection consists of two data classes: the healthy class and the VSD class. The data set contains 1200 snapshots, of which 840 were used for training & 360 for testing. The goal of the training phase is to teach the neural network how to recognise the input images. Training accuracy and a training model are the results of network training. This approach is employed during testing to identify new data.

4.1 Data Enhancement

Each image used in this investigation had a "white balance colour augmenter" added to it. This type of enhancement can replicate realistic colour constancy decline. Existing color enhancement techniques frequently produce strange colours, which happens infrequently. White balance colour augmenters improve the accuracy of "image categorisation& image semantic segmentation" procedures by imitating various white balance effects. The augmentation procedure generates 1200 images in total, 600 of which are for the healthy class and 600 of which are for the VSD class. Figure 4 represents implementation on an image dataset of cocoa leaves. Figure 6 represents implementation on cocoa image dataset.
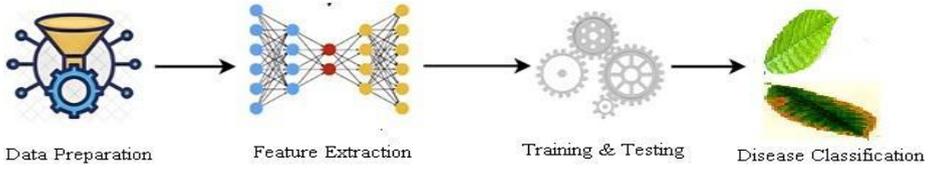
Fig.6. Implementation on Cocoa Image Dataset

5   Result Discussion

The Table 2 shown below represents the different deep learning models previously applied for image classification.

Table 2. Different Deep Learning Model Previously Applied for Image Classifications

Sr. No	Author	No. of images (Approx)	No. of classes	Measures	Deep Neural Network Architecture	Accuracy %
1	Kawasaki,Y. et al. (2015)	3	799	Specificity, Accuracy, Sensitivity	Custom - built	93.98
2	S Sladojevic . et al. (2016)	15	4480	Correctness	1 GPU-version of AlexNet	95.99
3	Mohanty et al. (2016)	54K	38	F1Score, Mean Position	AlexNet&GoogleNet	99.36
4	Nachtingall et al. (2016)	14K	5	Sensitivity, Accuracy, Exactness	Alexnet	96.97
5	Fujita,.et al. (2016)	14K	7	True Negative Rate, True Positive Rate	Custom - built	82.30
6	Brahminiet al. (2017)	14K	9	Correctness, Macro Recall	GoogleNet	98.99
7	DeChantet al.(2017)	1.8K	2	Exactness, Accuracy	Cutomized	95.96
8	Lu,Yet al. (2017)	.5K	10	Accuracy	Alexnet	96.97
9	Wang et al.(2017)	2.1 K	4	Accuracy	Inception V3	89.998
10	Brahimi et al (2018)	54 K	38	Accuracy	InceptionV3	98.99
11	Wang et al. (2018)	2.5 K	5	Accuracy	Custom-built	89.97
12	Rangarajanet al. (2018)	13K	7	Accuracy	Alexnet	96.98
13	I, Khandelwalet al. (2019)	86K	5	Correctness	InceptionV3, Residual Network-50	98.99
14	This Work	1.2K	2	Accuracy	DenseNet-19	99.1

In this study, using cocoa leaf analysis, we propose a novel method for detecting infestation in cocoa leaves. By working together, farmers' issues with recognizing cocoa plant infections will resolve without involving cocoa botanists. As a result, it will help them treat the cocoa plant infections quickly, increasing the quantity and quality of cocoa plant crops produced and assisting in raising farmer's incomes. We gathered pictures of cocoa leaves from a plant community in order to conduct the experiment. Once the leaf dataset was gathered, we identified cocoa leaf image using deep CNN model. We have discussed the results on behalf of training and testing.

5.1   Testing and Data Training

The data generated were split in a ratio of 80:20. 960 pictures are utilized for training & 240 pictures are utilized for assessment. To create a suitable model to assess data validation, training data is used. To check the models accuracy based on the training outcomes, data testing is used. The creation of the most precise identification was the focus of the training model. Table 3 represents accuracy and time taken by DenseNet-19. Figure 7 represents accuracy and time elapsed by DenseNet19.

Table 3. ACCURACY % & TIME ELAPSED BY DENSENT19

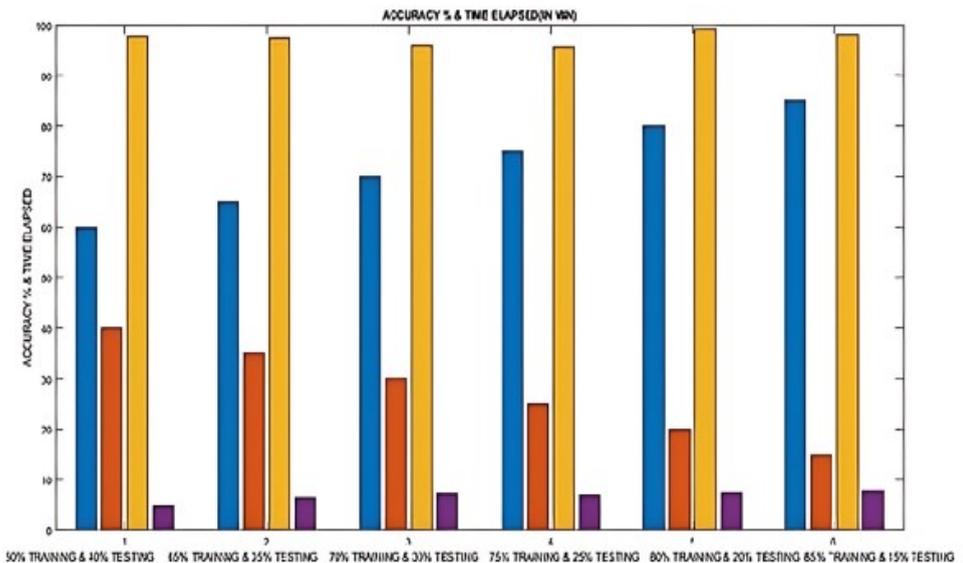| SR.NO	TESTING %	TRAINING %	TIME ELAPSED (IN MIN)	ACCURACY %
1	40	60	4.81	97.72
2	35	65	6.50	97.41
3	30	70	7.37	95.97
4	25	75	7.04	95.77
5	20	80	7.48	99.1
6	15	85	7.83	97.99

Fig.7. Accuracy& Time Elapsed by Densenet 19

5.2   Confusion Matrix:

This confusion matrix can be utilized to evaluate the model's performance. Accuracy displays the model's degree of accuracy. Table 4 represents the confusion matrix of the model.

Table 4. CONFUSION MATRIX

60% TRAINING & 40% TESTING		65% T RAINING & 35% TESTING		70% TRAINING & 30% TESTING	
VCD	HEALTHY	VCD	HEALTHY	VCD	HEALTHY
84	2	74	2	63	2
2	86	2	74	5	61
75% TRAINING & 25% TESTING		**80% TRAINING & 20% TESTING**		**85% TRAINING &15% TESTING**	
VCD	HEALTHY	VCD	HEALTHY	VCD	HEALTHY
49	3	44	0	32	0
1	54	1	43	1	32

The quantity of data from the true class that is classified as true is compared to the amount of data from the true class that is classified as true and false. Recall compares the true and false data from the true class to the true and false data that are not from the true class.

The agro industrial sector plays a significant role in boosting the local economy. Agro industrial export commodity cocoa is a product. However, due to the rapid development of the deadly disease VSD, cocoa production is poor. Because of this, it is essential to develop a prototype that can swiftly & accurately diagnose disorders so that cure can start right away. Artificial neural networks are now often employed in precision agriculture. Convolutional neural networks, in particular, were used in this study to categorize VSD-infected & healthy cocoa leaves using deep learning. We employed 4 CNN prototypes for comparison in order to determine which one was more successful in resolving our issue. Modified CNN, Alex Net, Squeeze Net, and DarkNet-19 are the four models we employ. The DarkNet-19 model produced the best results, with a test accuracy of 99.1%.

6.   Conclusion

With cocoa being a major export commodity, the agro-industrial sector is essential to bolstering the local economy. However, cocoa production has decreased as a result of the catastrophic VSD disease's quick spread. This circumstance emphasizes the pressing need for a prototype that can identify illnesses fast and precisely so that prompt treatment can begin. Artificial neural networks are being used more and more in precision agriculture. CNN was used in this study to use deep learning techniques to categorize cocoa leaves as either healthy or diseased with VSD. We examined four CNN models—Modified CNN, AlexNet, SqueezeNet, and DarkNet-19—in order to determine which one worked best. With a test accuracy of 99.1%, the DarkNet-19 model performed the best out of all of them.

References

1. Kouassi, K. S., Diarra, M., Edi, K. H., &Koua, B. J. C. Detection of Cocoa Leaf Diseases Using the CNN-Based Feature Extractor and XGBOOST Classifier. *Open Journal of Applied Sciences*, *14*(10), 2955-2972 (2024).
2. Soh, K. S., Moung, E. G., Danker, K. J. J. Dargham, J. A., &Farzamnia, A. Cocoa Diseases Classification using Deep Learning Algorithm. In ITM Web of Conferences (Vol. 63, p. 01014). EDP Sciences (2024).
3. Kumar, Sachin, Pal, Saurabh, Singh, VijendraPratap and Jaiswal, Priya. "Performance evaluation of ResNet model for classification of tomato plant disease" *Epidemiologic Methods*, vol. 12, no. 1, pp. 20210044. https://doi.org/10.1515/em-2021-0044(2023)
4. Darwish, A., Ezzat, D., &Hassanien, A. E. An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis. Swarm and evolutionary computation, 52, 100616 (2020).
5. Kumar, Sachin, Singh, VijendraPratap, Pal, Saurabh and Jaiswal, Priya. "Energy-efficient model "DenseNet201 based on deep convolutional neural network" using cloud platform for detection of COVID-19 infected patients" *Epidemiologic Methods*, vol. 12, no. 1, pp. 20210047. https://doi.org/10.1515/em-2021-0047 (2023)
6. Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., &Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360 (2016).
7. Kamilaris, A., &Prenafeta-Boldú, F. X. Deep Learning in Agriculture: A Survey. Computers and Electronics in Agriculture, 147, 70-90. https://doi.org/10.1016/j.compag.2018.02.016 (2018).
8. Andreas Kamilaris, Francesc X. Prenafeta-Boldú.Disaster monitoring using unmanned aerial vehicles and deep learning. arXiv preprint arXiv:1807.11805. https://doi.org/10.48550/arXiv.1807.11805(2018)
9. Keane, P. J., & Prior, C. Vascular-streak dieback of cocoa, Phytopathological Papers, No. 33, v + 39 pp. ref. 87. Commonwealth Mycological Institute (1991).
10. Kumar, S., Singh, V. P., Pal, S., & Jaiswal, P. Energy-efficient model "DenseNet201 based on deep convolutional neural network" using cloud platform for detection of COVID-19 infected patients. Epidemiologic Methods, 12(1), 20210047 (2023).
11. Afifi, M., & Brown, M. S. What else can fool deep learning? Addressing color constancy errors on deep neural network performance. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 243-252) (2019).
12. Nachtigall, L. G., Araujo, R. M., &Nachtigall, G. R. Classification of apple tree disorders using convolutional neural networks. In 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 472-476). IEEE (2016, November).
13. Ploetz, R. The impact of diseases on cacao production: a global overview. Cacao diseases: a history of old enemies and new encounters, 33-59. DOI: 10.1007/978-3-319-24789-2_2(2016).
14. Harvyanti, A. F. M., Baihaki, R. I., Ridlo, Z. R., & Agustin, I. H. Application of Convolutional Neural Network for Identifying Cocoa Leaf Disease. In 1st International Conference on Neural Networks and Machine Learning 2022 (ICONNSMAL 2022) (pp. 283-304). Atlantis Press(2023, May).
15. Baihaki, R. I., Agustin, I. H., Ridlo, Z. R., &Kurniawati, E. Y. The Comparison of Convolutional Neural Networks Architectures on Classification Potato Leaf Diseases. In Proceedings of the 1st International Conference on Neural Networks and Machine Learning 2022 (ICONNSMAL 2022) (Vol. 177, p. 125). Springer Nature (2023, May).
16. Kumar, S., Shahi, A. K., Patel, K., & Jaiswal, P. Energy and Cost Efficient Pre-trained Convolutional Neural Network (ResNet-50) Model for Tomato Disease Recognition Using Cloud Platform. NeuroQuantology, 20(9), 2009 (2022).
17. Kumar, D. G., Subbarao, M. V., Pratibha, M. S., Swarna, M., Varshini, K., &Prasanthi, N. D. Comparative Analysis of Deep Learning Architectures and Optimizers for Paddy Leaf Disease Classification. In 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-5). IEEE (2024, February).
18. Demilie, W.B. Plant disease detection and classification techniques: a comparative study of the performances. J Big Data 11, 5. https://doi.org/10.1186/s40537-023-00863-9 (2024).
19. Ngugi, H. N., Ezugwu, A. E., Akinyelu, A. A., &Abualigah, L. Revolutionizing crop disease detection with computational deep learning: a comprehensive review. Environmental Monitoring and Assessment, 196(3), 302. (2024).

20. Kumar, S., Pal, S., Singh, V. P., &SinghPriya, S. K. S. S. K. COVID-19 Disease Detection using Pre-Trained Deep Convolutional Neural Network (GoogleNet) on Cloud Platform. NeuroQuantology, 20(9), 1989.(2022).