International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN (Online): 2347-5552, Volume-12, Issue-5 September 2024
https://doi.org/10.55524/ijircst.2024.12.5.15
Article ID IRP-1565, Pages 110-116
www.ijircst.org

Classification of Healthy and Diseased Broccoli Leaves Using a Custom Deep Learning CNN Model

Saikat Banerjee[1], Soumitra Das[2], and Abhoy Chand Mondal[3]

[1] State Aided College Teacher, Department of Computer Applications, Vivekananda Mahavidyalaya, Haripal, Hooghly, West Bengal, India
[2] Research Scholar, Department of Computer Science, The University of Burdwan, Golapbag, West Bengal, India
[3] Professor, Department of Computer Science, The University of Burdwan, Golapbag, West Bengal, India

Correspondence should be addressed to Saikat Banerjee; saikat.banerjee56@gmail.com

Received 5 September 2024; Revised 19 September 2024; Accepted 30 September 2024

ABSTRACT- Agriculture is essential for sustaining the global population and is a crucial element in economic returns and food supply. However, plant leaf diseases are a major threat to agriculture and economy since they retard yield and increase cost of production. Because of high demand for broccoli, a wonderful and profitable crop, in the market, it has tremendous business opportunities for farmers. Nevertheless, similar to many other food crops, it is vulnerable to diseases which may affect its production and quality. Prevent losses from these diseases requires early detection of the disease affecting the leaves and with further enhancement of the technology, especially deep learning. In this research, an application of a new, specifically designed CNN model for the differentiation of healthy and diseased broccoli leaves is proposed. Data were collected directly from the field using mobile cameras and the images were sorted under healthy and unhealthy classes respectively. A new CNN model with an architecture specific to this dataset was designed and trained in this project using Keras. As evaluated from the result, the model proved efficient providing an accurate prediction on the health status of the leaf. The use of deep learning in disease diagnosis in crops enables farmers to make timely interventions thus protecting their crops, their potential economic value and nutritional value. This research acknowledges the possibilities of applying advanced technological improvement on the practice of agriculture.

KEYWORDS- Broccoli, Deep Learning, Convolutional Nural Networks, Leaf Disease Classification, TensorFlow, Keras.

I. INTRODUCTION

The contribution of agriculture goes beyond the provision of food; it is crucial in supporting countries' economies and feeding the world. To meet the growing population densities of the world, it is necessary to produce quality crops. However, plant diseases are a real problem for agricultural production, and diseases affecting leaves are hazardous since they result in lower yields, reduced quality and production efficiency, and substantial economic repercussions. Early identification and control of these diseases is therefore critical in maintaining agriculture productivity and farmers' business profitability. Broccoli is among the most grown and marketed crops today due to the rapidly increasing nutrient demand in the global market. For instance, farming broccoli has nutritional and business value to farmers since broccolis are nutritious and have a ready market.

Nevertheless, things are similar for this plant and many other crops to which diseases, particularly those of the leaves, are always a potential threat. Hemilea and Uettwilian mirids are diseases that, if not diagnosed early, can otherwise cut yields dramatically. Detecting diseases mainly involves inspection, which is usually tiresome and ambiguous. This supports the fact that different farming methods require mechanical means to assist farmers in noting diseases early on. Recently, deep learning especially Convolutional Neural Networks has shown great potential to be used in image classification tasks. CNNs are able to localize miniscule features specific to images, which makes it manageable to detect diseases in plant leaves. In this study, we develop a new convolutional neural network model uniquely trained to diagnose healthy or unhealthy broccoli leaves using images captured through mobile cameras in the field. It is in this regard that our deep learning approach can offer a swift and precise solution to disease detection; therefore, enabling the farmers to act - swiftly and optimally on the yields, with respect to both quality and quantity. Following this, the data collection and preparation, proposed CNN model, and outcome of implementing the model for the classification are discussed in this paper. The demonstrated efficacy of deep learning in timely detecting diseases in broccoli farming is one way to boost farming productivity and income.

II. LITERATURE REVIEW

For disease detection in leaves, this paper compares new techniques, which are ML and DL, with the previous approaches, such as SVM and Decision Tree, along with the latest methods, CNN and RNN. [1] Concerns have been identified regarding some usual issues, namely the data quality and demand for computational power; future use of various data types and utilization of more than one model will be discussed. Based on the improved images of the plant leaves, this study employs ML mechanisms such as

SVM, k-NN, and Decision Tree to diagnose leaf diseases[2]. SVM achieved 97% accuracy. Issues include data quality, and the next steps will focus on developing new methods of creating hybrid models and field testing models [3]. SVM and feature extraction are used in this study to improve the identification of diseases on leaves, but the work faces challenges with data quality and online processing. The plan laid down in the proposed framework under future work is to use DL models and more extensive data sets [4]. In their respective works, Yao et al. discussed the performance of traditional ML techniques and CNN approaches with an essential note for obtaining robust datasets. Constraints include image quality, size, and complexity of the model. [5] The present study employs relatively straightforward image analysis procedures for identifying the various kinds of diseased leaves, which yields good results through the SVM and k-means clustering methods. Limitations on datasets are discussed explicitly. [6] A highly accurate convolutional neural network-based model for the detection of diseases in leaves is needed for significant, diverse inputs. ; Further studies may look into transfer learning or other types of data.[7] SVM, Random Forest, and Decision Trees are employed to classify the diseases affecting the leaves. Future work concerns DL, feature extraction techniques, and data set enlargement. [8] The author discusses differences between ML and DL when applied to plant disease diagnosis, where issues such as overfitting, data size, and computational requirements are identified. They mentioned that future research should enhance scalability and real-world application of observing behavior.

III. COMMON LEAF DISEASE OF BROCCOLI

Alternaria leaf spot: pots are small and dark in color. They grow, then turn into tiny circles with a diameter of 1mm. White Rust: White, shiny, raised, blister-like spots or pustules often found at the underside of the leaves, stem, and flower. Black rot: First, they appear as chlorotic or yellow (angular) spots, usually around the edges of the leaves. The yellow area gets extended to veins and midrib and forms a ‘V’ shaped chlorotic spot which blackened later stage. Downy mildew: Small purplish brown spots on the surface of leaves. On the upper side of the leaves angular, small, pale yellow spots, and on the underside, it has down-like growth. The spots join together, and the leaves roll up and drop off too young.

IV. RESEARCH METHODOLOGY

Broccoli leaf classification presupposes data accumulation, data cleansing, feature extraction, model development, training the model, assessing the given CNN model, and classification with this CNN model.

A. Data Collection

In the interest of constructing a specialized Convolutional Neural Network (CNN) model that would accurately detect diseases in broccoli leaves, many images of the broccoli leaves were obtained manually from agricultural fields. It was important to ensure that the dataset was diverse and of high quality, which is critical to the training of a robust model with good generalization capabilities from the training data set. Altogether, ten thousand images were gathered. Data collection was therefore conducted in two phases to capture dynamism due to seasonal variations. The first of these phases was conducted in December, during which we acquired 5000 images. Figure 1 depicts the Broccoli Field. The second phase was carried out in February, and here we collect the remaining 5000 images. Consequently, this temporal division allows us to look at the differences in environmental conditions and the influence of those conditions on the health of the broccoli leaves.

Figure 1: Broccoli Field

In December 2023, we collected images of fields at the geographical coordinates 23.9802629 N, 86.9551737 E. Most of the leaves collected in this period were healthy, the standard form of healthy leaves. Figure 2 illustrates the gathering of data in the field. On the other hand, in February, we gathered pictures from fields at the geographical latitude of 23.83276 and the geographic longitude of 86.90298. This phase mainly provided unhealthy leaves since environmental conditions at this stage are suitable for disease formation.

Figure 2: On field data collection

Apart from that, the methodical approach to selection was useful for obtaining a balanced collection of data and considering various stages of diseases and healthy conditions necessary for obtaining efficient CNN for disease detection.

B. Data Preprocessing:

The following central process was subdivided into two stages of image preprocessing: initial assessment of the collected data and material preparation to train the CNN model. The initial task was to categorize the images into two primary folders: healthy and unhealthy. This manual separation process was very intensive since the feeder had to go through the images and sort them according to the identified category for each image. Figure.3. shows the methods for splitting and preprocessing the train test data.

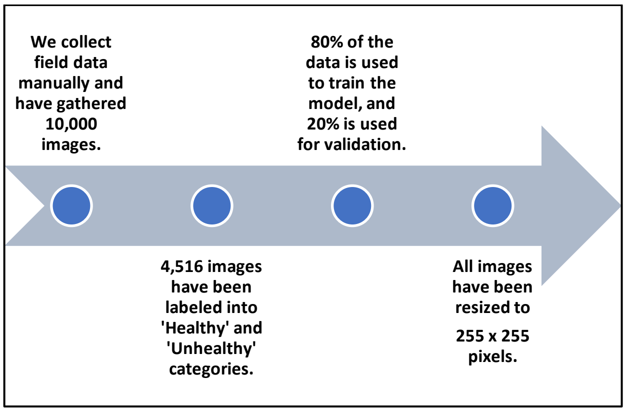

Figure 3: Train Test Data split and preprocessing methodology

Overall, we could tag 5516 photos with classifications of healthy and unhealthy leaves. To organize better the data for the training process, we established two primary folders: Test and train. Overall, we could tag 5516 images with the labels of healthy and unhealthy leaves. To further structure the dataset for the training process, we created two main folders: train and test. The train folder was filled with 4516 images, and the test folder with 1000 images. Each folder was further divided into two subfolders: healthy and unhealthy. Namely, the train folder contains 2016 healthy images and 2500 unhealthy images. Likewise, it was just as split with 500 healthy test images and 500 unhealthy test images. This was done to maintain close to an 80/20 split of the data, where nearly 80% of data is used to train the model while a minimum of 20% is used to test the trained model.Given that the images were captured using mobile cameras, their sizes varied significantly. To ensure uniformity and facilitate effective model training, we resized all the images to a standard dimension of 255x255 pixels. This resizing was accomplished using the Python Pillow library, a powerful tool for image processing. This step was vital for maintaining consistency in the input data, thereby enhancing the performance and accuracy of the CNN model in detecting broccoli leaf diseases.

C. CNN Model Creation and Evaluation

This study used TensorFlow and Keras to develop the CNN model for identifying broccoli leaf diseases. Here is a pictorial representation of the CNN model's workflow, broken down into individual steps displayed in Figure 5. The process

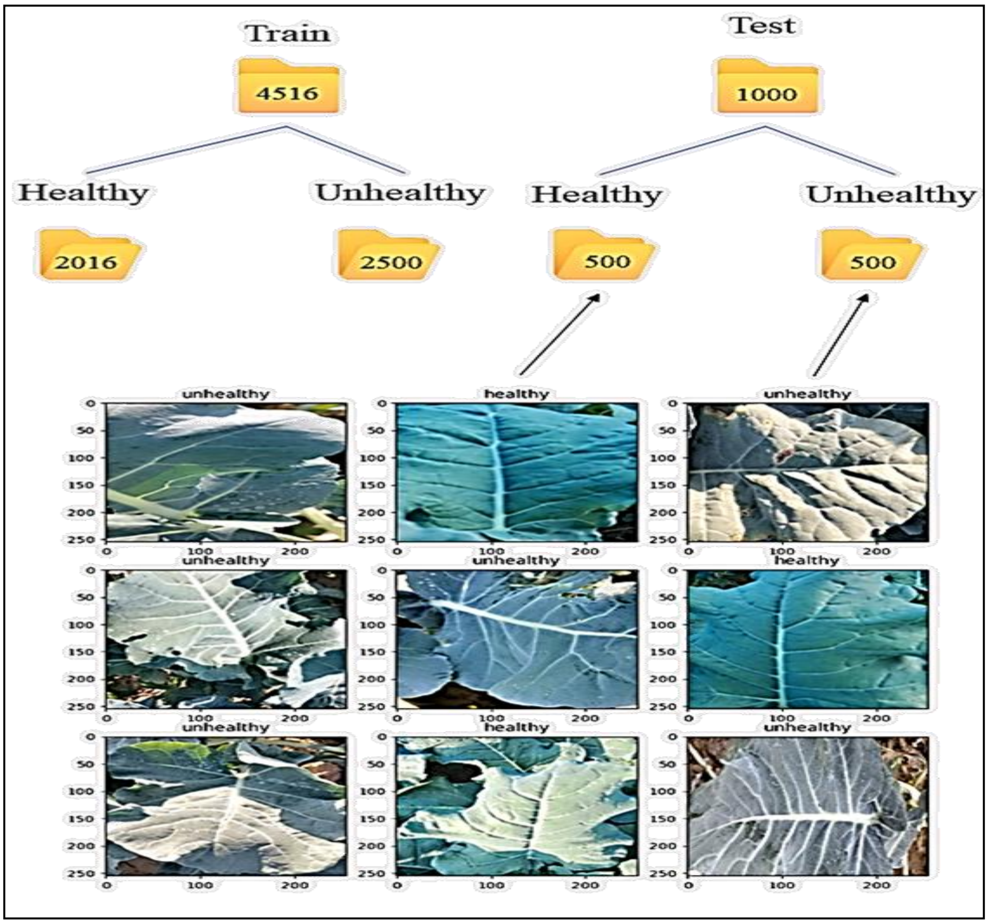

Figure 4: Sample Leaves and Folder Structure

started with loading the images in the 'train' and 'test' directories using the `image_dataset_from_directory` tool from Keras. To make the best of the data, it was split into training and validation with a batch size of 32 and an image size of 255*255 pixels. The display of Sample Leaves and Folder Structure is shown in Figure 4. This method inferred the labels from the director structure.

After that, we normalized the pixel values and divided every pixel by 255 to bring the values into a range of 0 to 1. A particular function was created to change the pixel values to float for this normalization. They normalized the images, which made this input data uniform for the model.

CNN was developed using the Sequential API from Keras. The architecture components included the following: We also had the Conv2D layer with 32 filters followed by ReLU activation with a kernel size 3*3. An activation layer succeeded. It is called the MaxPooling2D layer, with a pool size of 2 nodes in a row. This pattern was followed by the use of more filter sizes of 64 and then 128 in other layers of Conv2D, each followed by MaxPooling2D layers. These CNN and pooling layers were helpful for feature extraction from the images. After convolutional layers, the model was composed of a flattened layer that flattened the feature maps into 1D feature vectors. It was overcome by two Dense layers with a unit parameter of 128 and 64 with ReLU activation. The last layer for this model was a Dense layer with one neuron and sigmoid activation to classify the given leaves as either healthy or unhealthy. The model was trained using the Adam optimizer, binary cross-entropy lossfunction, and accuracy was used as the metric. The training process was carried out for ten epochs, and the model's accuracy was tested based on the validation dataset. Training epoch analysis demonstrated a consistent increase in accuracy from the initial 61.43% in the initial epoch to 100% in the last epoch of training. The graphical representations of train accuracy, train loss, train loss and validation accuracy, train loss and validation loss, and validation accuracy and validation loss, which displayed within Figures 6, 7, 8, and 9, are directly relevant to our analysis. As can be observed in the graph, a similar trend was observed for validation accuracy, which rose rapidly from 72.80% to 100% by the tenth epoch. The final model that was created can accurately distinguish healthy from unhealthy broccoli leaves, and the accuracy, as expected, was 100% on both the training and validating sets.

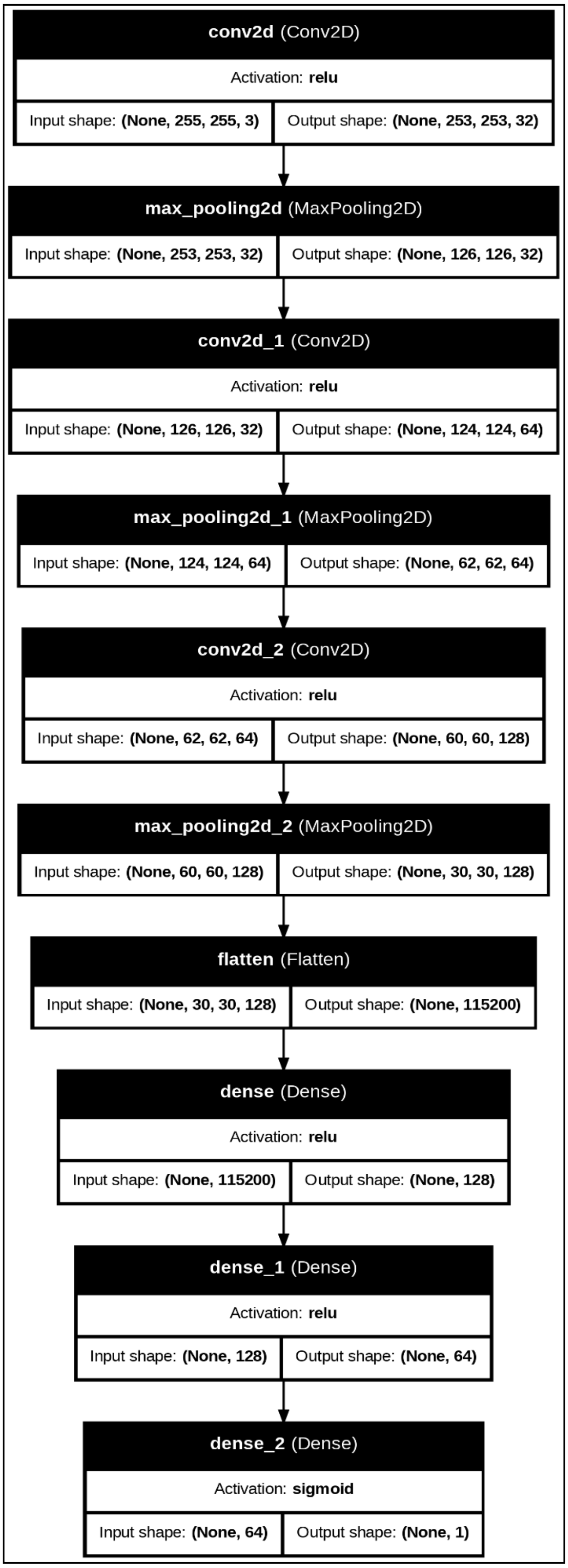

Figure 5: Pictorial Representation of CNN model

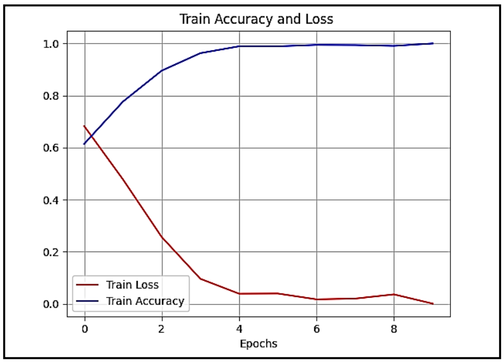

Figure 6: Train Accuracy and Train Loss

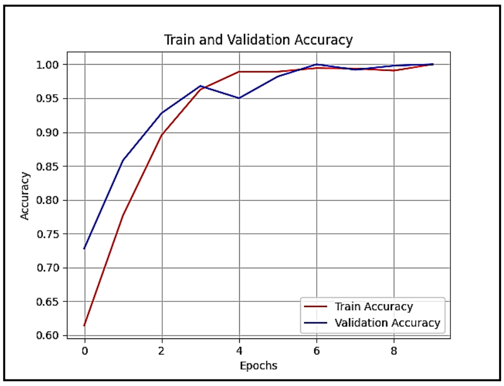

Figure 7: Train Accuracy and Validation Accuracy

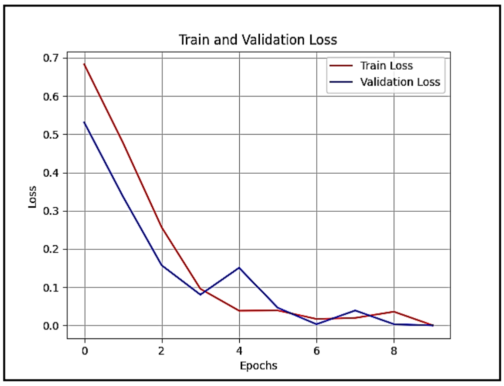

Figure 8: Train Loss and Validation Loss

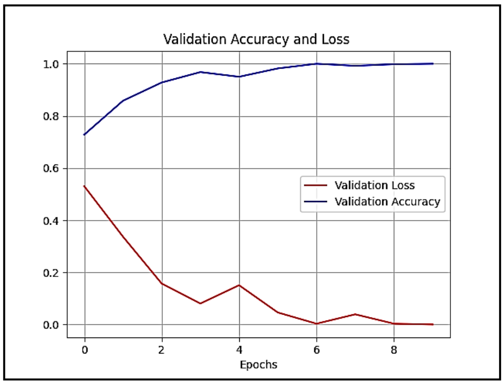

Figure 9: Validation Accuracy and Validation Loss

This visualization method not only helps to increase the model's interpretability but also allows one to spot potential

V. VISULIZATION AND MODEL ACTIVATION

Further information was obtained from the convolutional layers' heat maps to analyze the training process of the CNN model in detail and determine how it recognizes healthy and unhealthy broccoli leaves. This visualization helps decipher which areas of the images these models focus on when making their predictions. To do so, we created an activation model using TensorFlow and Keras that, given the same input as the original model, outputs the activations from the first convolutional layer. In this experiment, we applied this activation model by passing a sample image from the test dataset to the model to acquire the activations. In the feature maps below, the observed blobs signify the features the model has learned to identify in the first convolutional layer activations.

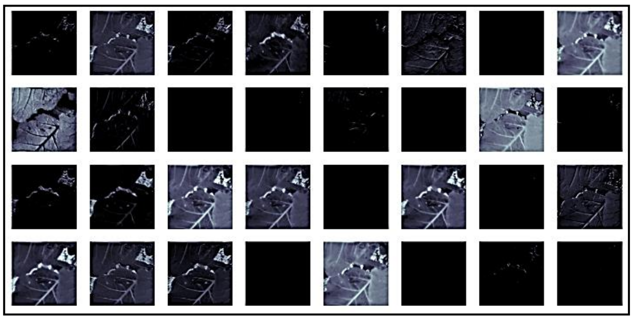

Figure 10: Heatmap of Unhealthy leaf using bone colormap

These feature maps were illustrated in a figure composed of several sub-figures; each figure represented a filter of the convolutional layer. Utilizing a bone colormap, Figure 10 displays a heatmap of an unhealthy leaf. Heatmap of a healthy leaf is displayed in Figure 11 using a grayscale colormap. The learned activations were shown using the 'grey' and 'bone' colormap for improved contrast/clarity. This process gave a clear representation of how the filters of the convolutional layer interpreted the various areas of the input image to arrive at the classification of the pictures of leaves.

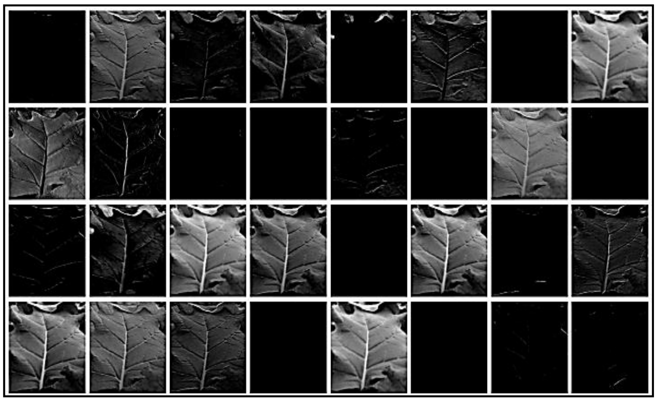

Figure 11: Heatmap of Healthy leaf using gray colormap

architectural or training process improvement opportunities.

VI. RESULT AND DISCUSSION

A. *Confussion Matrix*

Confusion matrix is a robust measure that can used to determine the effectiveness of classifiers. It evaluates the model performance by stating the number of correctly and wrongly classified groups. It aids in determining not just the errors of the model and the degree of success but the discrepancies, too.BY generating a confusion matrix concerning the predictions of the CNN model on the validation set, we were able to assess the model's performance in detecting diseases affecting the broccoli leaf. The matrix shows that the actual labels are compared with the predicted labels to display the model on each class. It is the confusion matrix for the model mentioned above:

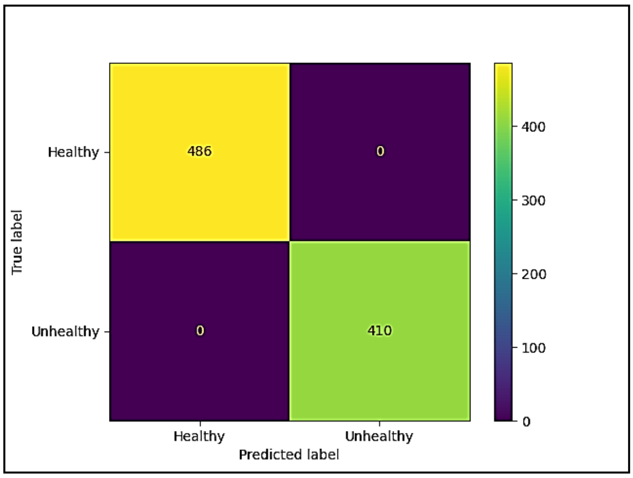

Figure 12: Confusion Matrix

The actual class is represented along the rows of the matrix, and the predicted class is defined along the columns of the matrix. Diagonal elements show the number of correct classifications, and off-diagonal elements show the number of incorrect classifications.

- True Positives (TP): 486 The model correctly forecasted the scenarios over the healthy leaves.
- True Negatives (TN): The value of being right indicates 410 cases where the model gave the right outcomes on unhealthy leaves.
- False Positives (FP): No false positives about the classification of unhealthy leaves as healthy by the model.
- False Negatives (FN): No case was identified where the model classified a healthy leaf as an unhealthy one.

B. *Accuracy:*

$$\frac{TP+TN}{TP+FP+TN+FN} = \frac{486+410}{486+0+410+0} = 1.0$$

C. *Performance Metrics*

Since accuracy and other objectives following GOAL were set equal to 100% and the number of recalled pictures was also 100%, the model can differentiate between the healthy and unhealthy leaves:

- **Precision:** Indicates the accuracy of positive predictions.
  Precision = $\frac{TP}{TP+FP} = \frac{486}{486+0} = 1.0$
- **Recall (Sensitivity):** Measures the model's ability to correctly identify positive instances.
  Recall = $\frac{TP}{TP+FN} = \frac{486}{486+0} = 1.0$
- **F1 Score:** Harmonic mean of precision and recall, providing a balanced metric for evaluation.
  F1 Score = $2 \cdot \frac{precision \cdot recall}{precision+recall} = 2 \cdot \frac{1 \cdot 1}{1+1} = 1.0$

The F1 scores in the above calculation are high at 1.00, representing the summative performance of F1 that reflects the class balance. Figure 12 depicts the confusion matrix. In the confusion matrix and derived metrics, we observe that:

- The model performs exceptionally well in identifying healthy and unhealthy leaves with perfect precision and recall.
- There are no misclassifications, indicating that the model is highly accurate and reliable.
- The overall F1 score 1.00 suggests a balanced and perfect performance across different classes.

VII. CONCLUSION

In this study, we created a feature extraction model based on the Convolutional Neural Network (CNN) for disease diagnosis on broccoli leaves. In the present method, our strategy entailed the collection of a vast dataset, where we collected 10000 images from different regions and various seasons to make the selection broad-based and inclusive. The raw data was thoroughly preprocessed, accompanied by a healthy and an unhealthy leaves dataset, which was also separated into training and testing data. As for the artificial neural network, the CNN model constructed with TensorFlow and Keras achieved perfect scores for predictive accuracy equal to 100% in the training and validation data after ten iterations. The presented architecture with several convolutional, pooling and dense layers helped the model capture all required features to provide relevant classification. Also, the transfer of images into heatmaps used for visualization helped to comprehend which features were learned by the model and interpret this result more efficiently. They found that broccoli farming is economically profitable to farmers and makes more profits than other crops, such as cabbage and cauliflower. Using image processing methods to detect diseases early can minimize damage to crops and financial loss. In this respect, our model looks highly promising and may be used as an efficient means of disease early detection.

Nevertheless, there are specific directions for further research and improvement of the model, due to which such high accuracy has been obtained. Firstly, expanding the dataset's size and variability by photos from different zones taken in various weather conditions also improves the model's performance. Secondly, more elaborate procedures, including transfer learning and data augmentation, could be studied which would enhance the result and minimize overfitting risks. However, adopting this model into real-world applications like mobile apps or automated field monitoring could be beneficial to farmers in diagnosing diseases early enough. In practical agriculture, integrating this system with IoT devices and remote sensing could provide real-time monitoring and decision-making capability, improving food systems' sustainability and output. Thus, the trained CNN model is a possible approach to recognizing Broccoli leaf diseases with several purposes in discriminating against other crops and plant diseases. More work in this field would help to support precision agriculture, increase agricultural output, and lower crop failure via diseases.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] C. Sarkar, D. Gupta, U. Gupta, and B. B. Hazarika, "Leaf disease detection using machine learning and deep learning: Review and challenges," Applied Soft Computing, vol. 145, p. 110534, 2023, doi: 10.1016/j.asoc.2023.110534. Available from: https://doi.org/10.1016/j.asoc.2023.110534

[2] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," Global Transitions Proceedings, vol. 3, no. 1, pp. 305-310, 2022, Available from: https://doi.org/10.1016/j.gltp.2022.03.016

[3] K. Prabavathy, M. Bharath, K. Sanjayratnam, N. S. S. R. Reddy, and M. S. Reddy, "Plant Leaf Disease Detection using Machine Learning," in 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2023, pp. 378-382, https://doi.org/10.1109/ICCCNT45670.2019.8944556

[4] J. Yao, S. N. Tran, S. Sawyer, et al., "Machine learning for leaf disease classification: data, techniques and applications," Artificial Intelligence Review, vol. 56, Suppl. 3, pp. 3571-3616, 2023, Available from: https://doi.org/10.1007/s10462-023-10610-4

[5] K. Joshi, R. Awale, S. Ahmad, S. Patil, and V. Pisal, "Plant Leaf Disease Detection Using Computer Vision Techniques and Machine Learning," ITM Web Conf., vol. 44, p. 03002, 2022. Available from: https://doi.org/10.1051/itmconf/20224403002

[6] J. Andrew, J. Eunice, D. E. Popescu, M. K. Chowdary, and J. Hemanth, "Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications," Agronomy, vol. 12, no. 10, p. 2395, 2022. Available from: https://doi.org/10.3390/agronomy12102395

[7] U. Fulari, R. Shastri, and A. Fulari, "Leaf Disease Detection Using Machine Learning," in 15th IEEE International Conference, 2020, pp. 1828-1832. Available from: https://doi.org/10.1109/ICCCNT45670.2019.8944556

[8] W. B. Demilie, "Plant disease detection and classification techniques: a comparative study of the performances," Journal of Big Data, vol. 11, p. 5, 2024. Available from: https://doi.org/10.1186/s40537-023-00863-9

ABOUT THE AUTHORS

Saikat Banerjee is currently employed as a teacher at Vivekananda Mahavidyalaya, a state-aided college in the department of BCA. He is also pursuing his Ph.D. in Computer Science at the University of Burdwan, located in Burdwan, West Bengal, India. He obtained his Bachelor of Science degree with a specialization in Computer Science and his Master of Computer Application (MCA) award in 2013 from the University of Burdwan in West Bengal, India. He possesses more than 11 years of teaching experience. He has published several articles in various reputable journals and conferences. His research interests encompass a variety of topics, such as deep learning, soft computing, artificial intelligence, and machine learning.

Soumitra Das completed MSc in Computer Science at the University of Burdwan in Burdwan, West Bengal, India. He earned his Bachelor of Computer Application in 2022 from the Kazi Nazrul University in West Bengal, India. He has more than two years of development and data analysis experience. He has authored numerous publications in several esteemed journals and conferences. He is presently a software developer and machine learning enthusiast. His research interests include deep learning, soft computing, artificial intelligence, and machine learning.

Dr. Abhoy Chand Mondal is presently a Professor and Head of the Department of Computer Science at the University of Burdwan in Burdwan, West Bengal, India, where he also serves as the Head of the Department of Computer Science. In 1987, he graduated with a Bachelor of Science in Mathematics with honors from The University of Burdwan. In 1989 and 1992, he earned a Master of Science in Mathematics and MCA from Jadavpur University. In 2004, he obtained a doctoral degree from Burdwan University. He also has 28 years of experience teaching and researching and one year of work experience in the sector. More than 120 articles and more than 60 journals were published. His areas of study interest include fuzzy logic, soft computing, document processing, natural language processing, natural language processing, big data analytics, machine learning, deep learning, and other areas.