Leveraging Deep Learning for Accurate Tea Leaf Disease Identification

Dr. David Raj Micheal
Division of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
davidraj.micheal@vit.ac.in

R.Bhuvanesh
Division of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
bhuvanesh.2023@vitstudent.ac.in

***Abstract*—-Identifying diseases in tea leaves is crucial for maintaining the quality and yield of tea crops. This study leverages deep learning techniques to enhance the detection of diseases in tea leaves through image analysis. We utilize three different models—VGG, ResNet , and Sequential—to classify images of tea leaves and identify potential diseases. By evaluating these models on a dataset of tea leaf images, we aim to determine their accuracy and effectiveness in disease detection. This research provides insights into the application of advanced deep learning algorithms for agricultural diagnostics, with the goal of improving disease management and crop health in tea production.**

***Index Terms*—Keywords- Tea Leaf Disease Detection, Deep Learning, Image Classification, VGG ResNet Sequential Models, Agricultural Diagnostics**

I. Introduction

Tea leaf diseases pose a significant threat to the quality and yield of tea crops, affecting both small and large-scale tea plantations. Common diseases, such as Gray Blight, Algal Spot, Red Spot, and Brown Blight, can lead to severe crop damage if not detected and treated early. Timely and accurate identification of these diseases is crucial for managing tea plant health and maintaining productivity. The paper on leveraging deep learning for accurate tea leaf disease identification explores the application of advanced machine learning techniques to enhance disease detection. By employing convolutional neural networks (CNNs) and training them on tea leaf images, the study aims to surpass traditional methods with more precise, rapid, and scalable solutions. This research showcases the potential of deep learning in revolutionizing disease diagnosis, offering critical insights for improving crop health and yield in tea agriculture.

II. Objectives

The objective of this study is to conduct a comprehensive comparison of the VGG, ResNet, and Sequential deep learning models in identifying diseases in tea leaves through image analysis. By leveraging these models, the study aims to evaluate their classification accuracy, efficiency, and robustness in detecting common tea leaf diseases such as Gray Blight, Algal Spot, and Brown Blight. Through extensive experimentation on a dataset of tea leaf images, we seek to determine the in most effective model for early disease identification, which is critical for timely intervention and reducing crop losses. Ultimately, the goal is to provide a reliable and scalable solution for disease management in tea cultivation, contributing to improved crop health and productivity.

III. Literature Review

Datta, S., Gupta, N. (2023), this study focuses on the detection of tea leaf diseases using a deep Convolutional Neural Network (CNN) for classifying diseased leaves into categories such as Gray Blight, Algal Spot, Brown Blight, Helopeltis, Red Spot, and Healthy Leaves. A dataset of 5,867 tea leaf images was created and uploaded on Kaggle. The proposed model achieves an overall accuracy of 96.56 percentage, outperforming existing methods in detecting specific diseases with high precision. The model's architecture is adaptable for deployment in IoT devices and can be extended to classify diseases in other crops. **Gayathri, S., Wise, D. J. W., Shamini, P. B., Muthukumaran, N. (2020)**, this paper explores the use of deep learning, specifically the LeNet CNN model, to detect diseases in tea plant leaves from a set of leaf images. The goal is to accurately identify various tea leaf diseases that can negatively impact crop growth.

Hu, G., Yang, X., Zhang, Y., Wan, M. (2019), this study presents an improved deep CNN model for the identification of tea leaf diseases, incorporating a multiscale feature extraction module and depth wise separable convolution to enhance feature detection and reduce computational complexity. The proposed model, based on CIFAR10-quick, achieves an identification accuracy of 92.5%, surpassing traditional machine learning and classical deep learning methods. It also demonstrates fewer parameters and faster convergence compared to VGG16 and AlexNet, making it more efficient for real-time disease detection.**Hu, G., Wu, H., Zhang, Y., Wan, M. (2019)**, this paper introduces a low-shot learning approach for identifying tea leaf diseases, aiming for timely prevention and control. It utilizes Support Vector Machines (SVM) to segment disease spots based on colour and texture features. To address data scarcity, the study employs an improved Conditional Deep Convolutional Generative Adversarial Network (C-DCGAN) for data augmentation, generating new training samples. These samples train a VGG16 deep learning model for disease identification. The method achieves an average identification accuracy of 90%, significantly outperforming traditional low shot learning techniques.

Soeb, M. J. A., Jubayer, M. F., Tarin, T. A., Al Mamun, M. R., Ruhad, F. M., Parven, A., ... Meftaul, I. M. (2023), the dataset includes five types of tea leaf diseases, with images manually annotated and data-augmented to address sample size limitations. YOLOv7 outperformed several other models like CNN, YOLOv5, and Deep CNN, achieving high detection accuracy (97.3%), precision (96.7%), recall (96.4%), mAP (98.2%), and F1-score (0.965). The method is expected to streamline disease identification, reduce workload for entomologists, and prevent economic losses in the tea industry.**Hu, G., Fang, M. (2022)**, this study introduces a multi convolutional neural network (CNN) model, Merge Model, for the automatic identification of tea leaf diseases, particularly in small sample sizes. The model integrates diseased leaf segmentation to reduce background interference and combines multiple CNN modules to enhance feature extraction. A weight initialization method helps the model focus on critical disease features early in training. Data augmentation is achieved using the SinGAN model, generating additional training samples. Experimental results show that Merge Model effectively identifies tea diseases like white scab, leaf blight, red scab, and sooty mould, outperforming existing methods in accuracy for small datasets.

Rahman, H., Ahmad, I., Jon, P. H., Salam, A., Rabbi, M. F. (2024), this study addresses the early detection of tea leaf diseases using a convolutional neural network (CNN) model, with a focus on tea leaves affected by red rust, brown blight, grey blight, and healthy leaves. A dataset of 3,330 images was collected from the Sylhet division in Bangladesh, known for its tea production. The proposed CNN model achieved an accuracy of 96.65%, which was further validated through laboratory tests, including microbial culture and microscopic analysis. This model offers a practical solution for detecting tea foliage diseases and is tailored for the Bangladesh tea industry, contributing to improved disease management and crop yield. **Zou, X., Ren, Q., Cao, H., Qian, Y., Zhang, S. (2020)**, this research proposes a tea disease identification methodology using spectral reflectance data and machine learning techniques. A decision tree-based feature selector and a random forest recognizer are employed to handle high-dimensional hyperspectral data, which can lead to overfitting. Experimental results show significant improvements in recall rate and F1 score, with increases of 15%, 7%, and 11%, respectively. The methodology enhances feature selection and efficiently learns from high-dimensional data, enabling non-destructive and accurate tea disease identification. This approach offers a new strategy for handling large datasets in crop disease detection.

Ihsan, C. N., Agustina, N., Naseer, M., Gusdevi, H., Rusdi, J. F., Hadhiwibowo, A., Abdullah, F. (2024), this study focuses on improving tea leaf production in Indonesia by early detection of diseases. It evaluates several machine learning algorithms—Random Forest (RF), Support Vector Classifier (SVC), Extra Trees Classifier (ETC), Decision Tree (DT), XGBoost Classifier (XGB), and Convolutional Neural Networks (CNN)—for classifying tea leaf images. The Extra Trees Classifier (ETC) achieved the highest average accuracy at 77.47%, surpassing other methods like SVC (76.57%), RF (76.12%), XGB (71.62%), and DT (65.31%), with CNN performing the lowest at 59.08%. The study concludes that ETC is the most effective algorithm for detecting tea leaf diseases, offering a promising approach for enhancing tea production quality and efficiency. **Nath, M., Mitra, P., Kumar, D. (2023)**, this study addresses the challenge of identifying tea leaf diseases by developing an automated model that leverages deep learning techniques to improve accuracy and reduce computational costs. The proposed model integrates a Convolutional Neural Network (CNN) architecture with depth wise separable convolutions, residual networks, and a Support Vector Machine, enhanced by an attention module for precise feature extraction. The model is trained on a dataset of healthy and diseased tea leaves affected by blister blight, grey blight, and red rust. It achieves an impressive overall accuracy of 99.28%, outperforming eight other state-of-the-art deep learning models, highlighting its effectiveness for early and accurate disease detection.

IV. METHODOLOGY

The methodology for the project ”Leveraging Deep Learning for Accurate Tea Leaf Disease Identification” involves several key steps, from data preparation and model development to training, evaluation, and visualization. Below is a detailed explanation of each step:

A. *Data Preparation*

The dataset consists of images of tea leaves, which are categorized into different classes based on the type of disease. The images are preprocessed to ensure they are suitable for input into the deep learning models. This includes resizing the images to a consistent shape (256x256 pixels) and normalizing the pixel values.

B. *Model Development*

Three different deep learning models are developed for this project: ResNet50, VGG16, and a custom Sequential model.

- **ResNet50 Model:**The ResNet50 model is a pre-trained convolutional neural network (CNN) that is fine-tuned for the task of tea leaf disease identification. The top classification layer is excluded, and custom classification layers are added on top of the base model. The base model's layers are frozen to retain the pre-trained weights, and only the custom layers are trained.
- **VGG16 Model:** Similar to the ResNet50 model, the VGG16 model is a pre-trained CNN that is fine-tuned for the task. The top classification layer is excluded, and custom classification layers are added. The base model's layers are frozen, and only the custom layers are trained.
- • **Sequential Model:** A custom sequential model is developed from scratch. It consists of multiple convolutional

layers followed by max-pooling layers, a flatten layer, and dense layers. Dropout layers are added for regularization to prevent overfitting.

C. Model Compilation

Each model was configured with the Adam optimizer, a popular choice in machine learning for its efficiency and adaptive learning rate capabilities, which helps achieve faster convergence. The loss function used was sparse categorical cross-entropy, suitable for multi-class classification tasks where labels are integers rather than one-hot encoded vectors, making it memory-efficient and straightforward for classification problems. During training, model performance was monitored using the accuracy metric, allowing for an intuitive measure of the model's success in correctly predicting class labels. This setup, balancing effective optimization with appropriate loss and evaluation metrics, aims to provide robust training dynamics and facilitate accurate model evaluation across epochs.

D. Model Evaluation

The trained models are assessed on a separate validation dataset to gauge their performance beyond the training data. This step involves calculating the validation loss and accuracy metrics, which indicate how well the models generalize to new, unseen data. A lower validation loss alongside high accuracy suggests effective generalization, implying that the models have learned meaningful features rather than memorizing the training data. These evaluation metrics provide a benchmark for the models' predictive capabilities and can highlight potential overfitting or underfitting issues, guiding further refinement if necessary

E. Visualization

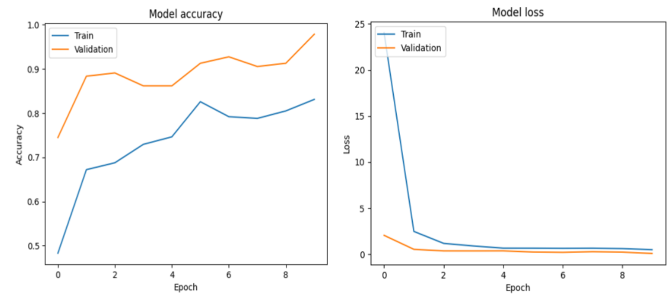
To gain insight into the models' performance over time, the training and validation accuracy and loss metrics are visualized. This involves plotting accuracy and loss curves for both the training and validation datasets across epochs, which provides a clear picture of how the models' performance evolves during training. These visualizations help in identifying trends and potential issues, such as overfitting—where the model performs well on training data but poorly on validation data—or underfitting, where the model fails to capture underlying patterns. By analyzing these curves, adjustments can be made to improve the model's ability to generalize, ensuring better performance on unseen data.

V. RESULTS AND FNDINGS

The project "Leveraging Deep Learning for Accurate Tea Leaf Disease Identification" involved training and evaluating three different deep learning models: ResNet50, VGG16, and a custom Sequential model. Below are the detailed results and findings for each model:

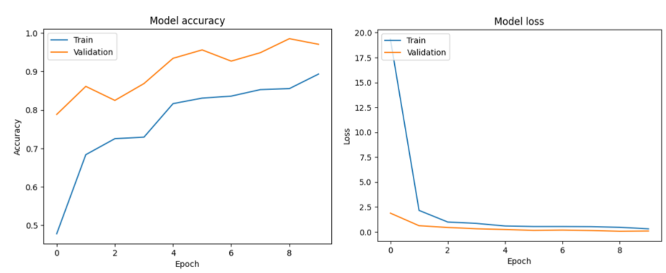
ResNet50 Model:

- **Epoch 1:**The model started with a training accuracy of 37.31% and a validation accuracy of 74.45 The training loss was 35.8840, and the validation loss was 2.0419.
- **Epoch 10:**The model achieved a training accuracy of 84.11% and a validation accuracy of 97.81%. The training loss decreased to 0.4474, and the validation loss to 0.0802.
- **Final Evaluation:**The model started with a training accuracy of 37.31% and a validation accuracy of 74.45 The training loss was 35.8840, and the validation loss was 2.0419.



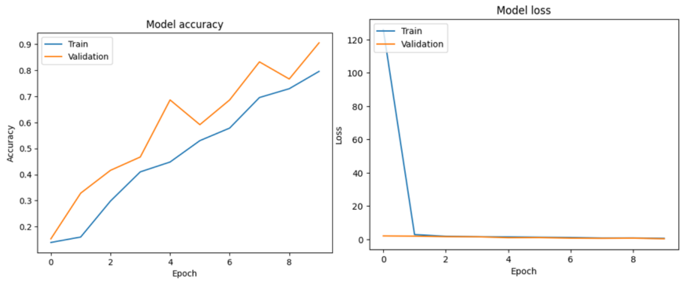
VGG16 Model:

- **Epoch 1:** The model started with a training accuracy of 34.59% and a validation accuracy of 78.83%. The training loss was 28.1159, and the validation loss was 1.8764.accuracy of 84.11% and a validation accuracy of 97.81%. The training loss decreased to 0.4474, and the validation loss to 0.0802.
- **Epoch 10:** The model achieved a training accuracy of 88.71% and a validation accuracy of 97.08%. The training loss decreased to 0.3152, and the validation loss to 0.0797.
- **Final Evaluation:** The model achieved a validation accuracy of 97.06% and a validation loss of 0.0630.



Sequential Model:

- **Epoch 1:** The model started with a training accuracy of 13.35% and a validation accuracy of 15.33%. The training loss was 298.3332, and the validation loss was 2.0815.
- **Epoch 10:**The model achieved a training accuracy of 79.43% and a validation accuracy of 90.51%. The training loss decreased to 0.6130, and the validation loss to 0.3645.
- **Final Evaluation:** The model achieved a validation accuracy of 92.55% and a validation loss of 0.3111.



VI. CONCLUSION

The project "Leveraging Deep Learning for Accurate Tea Leaf Disease Identification" successfully demonstrated the effectiveness of deep learning models in accurately identifying diseases in tea leaves. By employing pre-trained models like ResNet50 and VGG16, as well as a custom Sequential model, the study achieved high accuracy and robust performance. The VGG16 model, in particular, outperformed the others, achieving the highest validation accuracy and the lowest validation loss. These findings highlight the potential of deep learning techniques in enhancing disease management and crop health in tea production. The results underscore the importance of using advanced models to handle complex image data, paving the way for practical applications in agricultural disease detection and management

VII. REFERENCES

Datta, S., Gupta, N. (2023). A novel approach for the detection of tea leaf disease using deep neural network. Procedia Computer Science, 218, 2273-2286.

Gayathri, S., Wise, D. J. W., Shamini, P. B., Muthukumaran, N. (2020, July). Image analysis and detection of tea leaf disease using deep learning. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 398-403). IEEE.

Hu, G., Yang, X., Zhang, Y., Wan, M. (2019). Identification of tea leaf diseases by using an improved deep convolutional neural network. Sustainable Computing: Informatics and Systems, 24, 100353.
Hu, G., Wu, H., Zhang, Y., Wan, M. (2019). A low shot learning method for tea leaf's disease identification. Computers and Electronics in Agriculture, 163, 104852.

Soeb, M. J. A., Jubayer, M. F., Tarin, T. A., Al Mamun, M. R., Ruhad, F. M., Parven, A., ... Meftaul, I. M. (2023). Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). Scientific reports, 13(1), 6078.

Hu, G., Fang, M. (2022). Using a multi-convolutional neural network to automatically identify small-sample tea leaf diseases. Sustainable Computing: Informatics and Systems, 35, 100696.

Rahman, H., Ahmad, I., Jon, P. H., Salam, A., Rabbi, M. F. (2024). Automated detection of selected tea leaf diseases in Bangladesh with convolutional neural network. Scientific Reports, 14(1), 14097.

Zou, X., Ren, Q., Cao, H., Qian, Y., Zhang, S. (2020). Identification of tea diseases based on spectral reflectance and machine learning. Journal of Information Processing Systems, 16(2), 435-446.

Ihsan, C. N., Agustina, N., Naseer, M., Gusdevi, H., Rusdi, J. F., Hadhiwibowo, A., Abdullah, F. (2024). Comparison of Machine Learning Algorithms in Detecting Tea Leaf Diseases. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 8(1), 135-141.

Nath, M., Mitra, P., Kumar, D. (2023). A novel residual learning-based deep learning model integrated with attention mechanism and SVM for identifying tea plant diseases. International Journal of Computers and Applications, 45(6), 471-484.