Tomato Research Project

Tharmithan Nanthakumar¹ Department of InformationTechnology Sri Lanka Institute of InformationTechnology New Kandy Road, Malabe, Sri Lanka

Nirushan Pathmarajan³ Department of InformationTechnology Sri Lanka Institute of InformationTechnology New Kandy Road, Malabe, SriLanka

Aruna Ishara Gamage⁵ Department of InformationTechnology Sri Lanka Institute of InformationTechnology New Kandy Road, Malabe, Sri Lanka

Abstract:- This research aimed to harness the power of machine learning techniques for comprehensive analysis of tomato plants in the context of disease identification, seed type prediction, and NPK estimation. Deep learning models, including ResNet and VG19, were employed for disease identification and NPK prediction, while an Artificial Neural Network (ANN) model was utilized for seed type prediction. The results revealed that ResNet achieved a superior overall accuracy of approximately 0.71 compared to MobileNet for disease identification. The VG19 model showcased impressive accuracy with 0.95 overall accuracy for NPK prediction, while the ANN model achieved an accuracy of 0.37. These findings highlight the potential of deep learning models and transfer learning for accurate disease identification and NPK estimation in tomato plants. The research contributes valuable insights and guidance for the development of intelligent systems to enhance tomato cultivation practices and empower farmers with effective tools for decision-making.

Keywords:- Image Processing, Image Classification, Machine Learning, Deep Learning, Computer Vision, Regression.

I. INTRODUCTION

The tomato, belonging to the Solanaceae family, holds immense significance as one of the most widely cultivated vegetable plants in Sri Lanka. Its export potential has attracted attention, with a notable increase in interest and exports in recent years. Tomatoes are not only a cash crop but also a rich source of essential vitamins A and C, along with minerals. Despite its versatility and adaptability to various agroclimatic zones in Sri Lanka, tomato production faces challenges due to climatic variations and soil conditions. The impact of climate change on agricultural yields has been observed globally, highlighting the crucial link between temperature and crop productivity. To ensure Ishalini Senthamilpalan² Department of InformationTechnology Sri Lanka Institute of InformationTechnology New Kandy Road, Malabe, Sri Lanka

Kaviseshan Santhirapiragasam⁴ Department of Information Technology Sri Lanka Institute of Information Technology New Kandy Road, Malabe, Sri Lanka

Jeewaka Perera⁶ Department of Information Technology Sri Lanka Institute of Information Technology New Kandy Road, Malabe, SriLanka

higher yields, proper fertilization practices are vital in tomato cultivation. Environmental factors and soil nutrition play a pivotal role in the fertilization process, with organic fertilizer demonstrating significant yield enhancements [1]. However, the cultivation of tomatoes in Sri Lanka encounters threats such as diseases and nutrient deficiencies, which affect the overall yield and quality of the crop. Disease identification poses a significant challenge in tomato farming. In Sri Lanka, typical diseases include anthracnose, leaf mold, early and late blight, and powdery mildew. These diseases, caused by bacteria, fungi, and viruses, can inflict substantial damage on the crop, leading to decreased output and quality losses. Disease-related losses to tomato production have been estimated at 22% in Sri Lanka [2]. Timely and accurate disease identification is crucial for implementing effective control measures and mitigating crop losses. Nutrient deficiencies are another significant challenge in tomato cultivation. Insufficient levels of nitrogen, phosphorus, potassium, calcium, magnesium, and iron can result in stunted growth, yellowing leaves, reduced yields, and even plant mortality. Excessive moisture levels in the soil can lead to oxygen deficits, hindering the efficient uptake of nutrients by the plant roots [6]. Furthermore, imbalances in nutrient levels can disrupt nutrient absorption mechanisms. Thus, monitoring and addressing nutritional deficiencies are vital for ensuring a productive tomato harvest. In recent years, machine learning and deep learning techniques have gained prominence in various fields, including agriculture. These techniques provide valuable automated disease identification. tools for seed recommendation based on soil conditions, and NPK (nitrogen, phosphorus, potassium) level detection in plants[3]. This research paper aims to investigate the application of machine learning and deep learning techniques to tackle these challenges in tomato cultivation.

- Specifically, the Objectives of this Research are as follows:
- Develop a disease identification model using a CNN based deep learning techniques to accurately detect and classifycommon diseases in tomato plants.
- Design a seed recommendation system that suggests suitable seed types based on soil characteristics such as soil type, pH, moisture, temperature, humidity, rainfall, light exposure, fertilizer type, pest control, and nutrient levels.
- Implement a CNN and an artificial neural network (ANN) to detect NPK deficiencies in tomato plants using images and predict the required NPK levels for optimal growth.

By addressing these objectives, this research aims to contribute to the advancement of tomato cultivation practices in Sri Lanka. The utilization of machine learning and deep learning techniques in disease identification, seed recommendation, and NPK level detection will provide farmers with valuable insights and enable them to make informed decisions for improved crop management and higher yields.

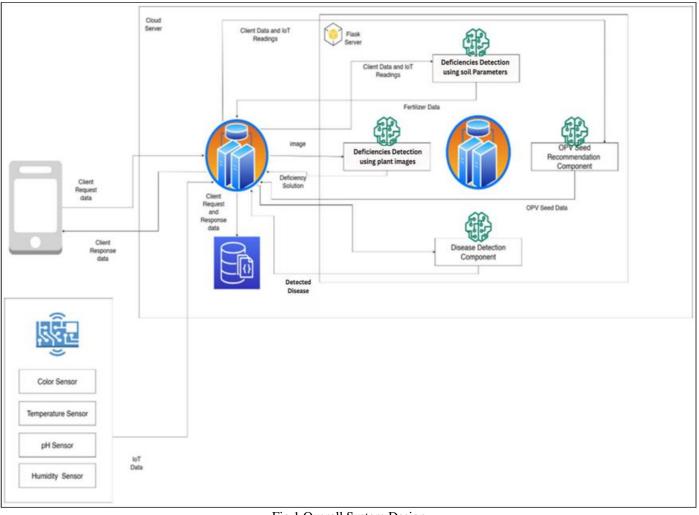


Fig 1 Overall System Design

II. LITERATURE REVIEW

The field of agriculture has greatly benefited from the application of deep learning and machine learning techniques in various aspects of crop management. Several research papers have focused on utilizing these technologies to address critical issues such as disease identification, fertilizer application, yield prediction, and pest management in tomato plants. This literature review aims to explore and discuss the key findings and methodologies presented in the selected research papers [4]. The first paper in this review focuses on the early blight identification in tomato leaves using deep learning. By leveraging deep learning algorithms, the authors aimed to develop an automated system capable of accurately detecting and classifying early blight disease in tomato plants. The paper highlights the importance of using deep learning techniques for disease identification and emphasizes the potential of these methods in early disease detection to prevent significant crop losses [3]. Another research paper investigated the effects of organic fertilizer application on tomato yield and quality through a metaanalysis. The authors aimed to synthesize and analyze existing studies to provide insights into the impact of organic fertilizer on tomato crops. The meta-analysis approach allowed for a comprehensive examination of various studies, enabling the identification of trends, patterns, and potential factors influencing tomato yield and quality. This paper contributes to the understanding of the benefits and limitations of organic fertilizer application in tomato cultivation [5]. The third paper in this review addresses the experimental determination of convolutional neural network (CNN) hyperparameters for tomato disease detection using leaf images. The authors aimed to optimize the hyperparameters of a CNN model to achieve accurate disease classification. By conducting experimental trials, they identified the optimal combination of hyperparameters for the specific task of tomato disease detection. This study sheds light on the importance of fine-tuning CNN hyperparameters to enhance the performance of disease detection systems[5].

Another study focuses on fertilizer estimation using a deep learning approach. The authors proposed a method that employs deep learning techniques to estimate fertilizer requirements for tomato plants. By analyzing leaf images and utilizing deep learning models, they aimed to provide accurate and efficient fertilizer recommendations based on plant nutrient needs. This research contributes to the development of intelligent agricultural systems that optimize fertilizer usage and promote sustainable farming practices[6]. One research paper investigates the identification of an efficient deep learning architecture for tomato disease classification using leaf images. The authors aimed to identify a deep learning model that achieves high accuracy in classifying different tomato diseases. Through comparative analysis, they evaluated and compared the performance of different deep learning architectures, ultimately identifying an efficient model for accurate disease classification. This study provides valuable insights into the selection and optimization of deep learning architectures for crop disease classification tasks [7]. A research paper explores the intelligent agricultural modeling of soil nutrients and pH classification using ensemble deep learning techniques. The authors proposed an ensemble approach that combines multiple deep learning models to predict soil nutrient levels and pH. By utilizing ensemble techniques, they aimed to enhance the accuracy and reliability of soil nutrient and pH classification models. This study highlights the potential of ensemble deep learning methods in agricultural modeling and their contribution to precision farming practices [8].

Another research paper focuses on the development of an intelligent insecticide and fertilizer recommendation system based on TPFCNN for smart farming. The authors aimed to create an automated system that recommends insecticides and fertilizers based on pest detection and crop nutrient requirements. By integrating deep learning techniques with insect detection and nutrient analysis, they aimed to provide accurate and timely recommendations for pest control and fertilization. This research contributes to the advancement of smart farming systems that optimize crop management practices[9]. The literature review also includes a study on the intensity of leaf spot and rough bark diseases in cinnamon accessions collected from major cinnamon growing areas of Sri Lanka. The authors aimed to assess the prevalence and severity of leaf spot and rough bark diseases in different cinnamon accessions. By analyzing disease incidence data, they identified variations in disease intensity among accessions and growing areas. This research provides insights into disease management strategies for cinnamon cultivation, contributing to improved yield and quality[10]. The strategies used in this paper translate well into the tomato cultivation disease detection section.

Another paper investigates the prediction of crop yield and fertilizer recommendation using machine learning algorithms. The authors aimed to develop models that predict crop yield and recommend optimal fertilizer application based on crop characteristics and environmental factors. By utilizing machine learning algorithms, they aimed to provide accurate yield predictions and optimize fertilizer usage. This study contributes to the development of precision agriculture techniques that enhance productivity and resource efficiency[11]. The relationship between pest and disease incidences and agronomic operations implemented by farmers in cinnamon fields in southern Sri Lanka is explored in another research paper. The authors aimed to understand the impact of agronomic practices on pest and disease incidences in cinnamon cultivation. By analyzing farming practices and disease data, they identified relationships between agronomic operations and pest/disease occurrences[12]. This research highlights the importance of appropriate agricultural practices in managing pest and disease risks, and aided in understanding of what factors of a cultivation must be considered when investigating the effects of pests and diseases in similar crops. A study examines the screening of tomato varieties against damping off disease under the climatic conditions of Batticaloa District, Sri Lanka. The authors aimed to evaluate different tomato varieties for their resistance to damping off disease. By conducting field trials and assessing disease incidence, they identified resistant varieties that can withstand the prevalent climatic conditions. This research contributes to the selection and cultivation of diseaseresistant tomato varieties, promoting sustainable tomato production [13]. One paper focuses on the classification of tomato fertilizer deficiency and the development of a fertilization decision model based on leaf images and deep learning. The authors aimed to develop an automated system capable of identifying fertilizer deficiencies in tomato plants through leaf image analysis. By utilizing deep learning models, they aimed to classify nutrient deficiencies and provide appropriate fertilization recommendations. This research contributes to precision fertilization practices and nutrient management in tomato cultivation[14]. Transfer learning for fine-grained crop disease classification based on leaf images is explored in another study. The authors aimed to leverage transfer learning techniques to classify finegrained crop diseases using leaf images. By utilizing pretrained deep learning models, they aimed to enhance disease classification accuracy and enable efficient disease management. This research highlights the potential of transfer learning in crop disease identification and its implications for precision agriculture [15]. Lastly, a research

paper presents the use of transfer learning for multi-crop leaf disease image classification using the VGG convolutional neural network. The authors aimed to develop a transfer learning-based system that can classify leaf diseases in multiple crops, including tomato plants. By leveraging the VGG network and transfer learning, they aimed to achieve accurate and efficient disease classification across different crops. This study contributes to the development of generalized disease detection systems applicable to various agricultural contexts[16].

III. METHODOLOGY

> Disease Detection

The disease detection component in this research project aimed to address the challenge of identifying diseases in tomato plants using machine learning techniques. To begin the process, a dataset containing images of tomato plants affected by different diseases was gathered and organized into distinct disease classes. This dataset was then divided into two subsets: a training set used to train the models and a validation set to evaluate their performance. As a crucial step in the data preparation phase, the tomato plant images were resized to a standardized dimension of 224x224 pixels. This resizing ensures uniformity and compatibility across the dataset, facilitating effective model training and evaluation.



Fig 2 Tomato Data set Sample

To leverage the power of pre-trained models, two wellestablished models, MobileNetV2 and ResNet50, were chosen as the base models for disease detection. These models had previously been trained on large-scale image datasets and demonstrated strong performance in various computer vision tasks. The pre-trained MobileNetV2 model was loaded without its top classification layer, which is responsible for assigning labels to images. By excluding this layer, the model's ability to recognize specific disease classes could be customized. To retain the knowledge and features learned by the pre-trained models, the layers of MobileNetV2 and ResNet50 were set as non-trainable. This

decision ensured that the models retained their expertise in general image recognition while enabling the addition of a custom classification head specifically designed for tomato plant disease identification. The custom head consisted of a flatten layer, followed by a dense layer with rectified linear unit (ReLU) activation and a final dense layer with softmax activation. The softmax activation allowed the model to assign probabilities to each disease class, indicating the likelihood of a particular disease being present in a given tomato plant image. With the model architectures in place, the next step involved compiling the models using the Adam optimizer. The optimizer adjusts the model's parameters during training to minimize the loss function, which was chosen as the sparse categorical cross-entropy. This loss function is well-suited for multi-class classification tasks like disease identification in tomato plants. The models were trained over multiple epochs, allowing them to iteratively learn and refine their internal parameters. Throughout the training process, accuracy and loss metrics were continuously recorded for both the training and validation sets. These metrics provided insights into how well the models were performing and whether they were overfitting or generalizing to unseen data.

Seed Type Recommendation

The seed type prediction component focused on classifying tomato plant seeds into different types using an Artificial Neural Network (ANN) model. The process began with the preparation of the dataset, which contained information about tomato plant seeds and their corresponding types. The dataset was split into training and testing sets, with 80% of the data allocated for training and 20% for testing. To ensure consistent scaling, the features in the training and testing sets were normalized using the StandardScaler. The ANN model was constructed using the Sequential class from the Keras library. It consisted of three dense layers: the first layer with 64 neurons and the ReLU activation function, the second layer with 32 neurons and ReLU activation, and the final layer with three neurons representing the seed types and using the softmax activation function. The model was compiled with the sparse categorical crossentropy loss function, suitable for multi-class classification tasks. The Adam optimizer was employed to optimize the model's parameters. Training was performed on the training set for 30 epochs, with a batch size of 32, allowing the model to adjust its parameters and improve accuracy. Following training, the model was evaluated using the testing set, and the loss and accuracy metrics were computed. The accuracy of the model in predicting the correct seed types was printed as a measure of its performance.

> Nutrition Deficiency Detection

The NPK component consisted of two models: an image-based model and an Artificial Neural Network (ANN) model. These models aimed to predict the nutrient levels (NPK) required for optimal tomato plant growth based on image analysis and soil characteristics. The first step involved preparing the dataset, which contained images of tomato plants categorized into different NPK levels. The dataset was split into training and validation sets, and image preprocessing techniques, such as resizing and normalization, were applied. For the image-based model, a VGG19 architecture was used. The pre-trained VGG19 model was loaded, and its layers were set as non-trainable to preserve the learned features. The model was then augmented with additional layers, including a flatten layer, a dense layer with ReLU activation, and a final dense layer with softmax activation representing the three NPK levels. The model was compiled using the Adam optimizer with a learning rate of 0.001 and the sparse categorical crossentropy loss function. Training was performed on the augmented dataset, and the model's performance was evaluated using accuracy and loss metrics. The accuracy and loss values were plotted over multiple epochs to assess the model's learning progress and convergence. The trained model was saved for future use and deployment. The second part of the NPK component involved an ANN model. The dataset used for this model contained soil characteristics such as soil type, pH, soil moisture, temperature, humidity, rainfall, and light exposure, along with the corresponding NPK levels. The dataset was split into features (X) and the target variable (y). The features represented the soil characteristics, while the target variable represented the NPK levels. The data was then split into training and testing sets. To ensure consistent scaling, the features were normalized using the StandardScaler. An ANN model was built, comprising dense layers with ReLU activation and a final dense layer with softmax activation representing the three NPK levels. The model was compiled using the Adam optimizer and the sparse categorical cross-entropy loss function. The model was trained using the training set, and the training process involved multiple epochs. The model's accuracy and loss metrics were recorded during training. After training, the model was evaluated on the testing set, and the accuracy was computed to assess its performance.

IV. RESULTS AND DISCUSSION

Disease Detection

The results of the tomato disease identification section revealed varying levels of accuracy among the different models used. The MobileNet model achieved an overall accuracy of approximately 0.34, indicating moderate performance in classifying tomato diseases. On the other hand, the ResNet model demonstrated higher accuracy, with an overall accuracy of around 0.71. This suggests that the ResNet model outperformed the MobileNet model in accuratelyidentifying tomato diseases.

➢ Seed Type Recommendation

In terms of the seed ANN model, the accuracy obtained was 39%. While this accuracy is relatively low, it is important to note that the seed ANN model might have faced challenges in accurately classifying tomato diseases due to the complexity and variability of disease symptoms. Further improvements and optimizations may be required to enhance the accuracy of the seed ANN model for tomato disease identification.

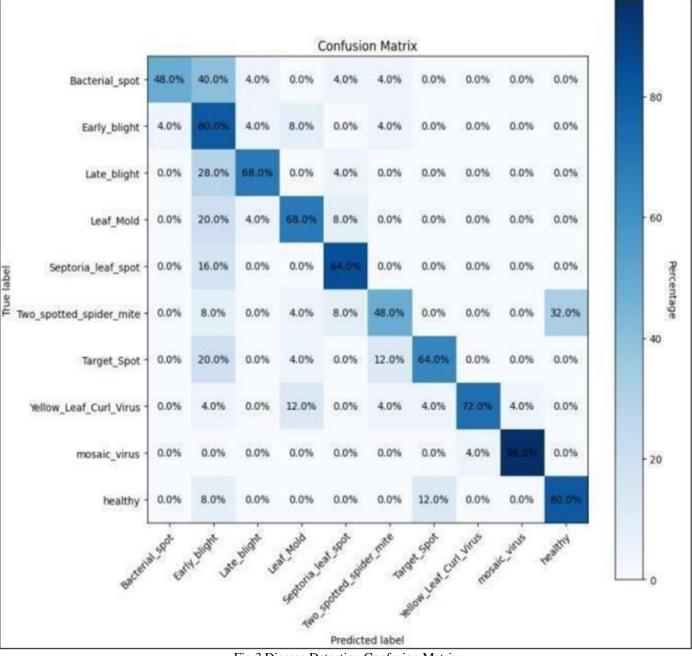


Fig 3 Disease Detection Confusion Matrix

> Nutrition Deficiency Detection

Moving on to the NPK prediction models, the VG19 model achieved an impressive accuracy of 0.95. This indicates that the VG19 model effectively classified and predicted the NPK (nitrogen, phosphorus, and potassium) levels in tomato plants based on the provided dataset. In contrast, the ANN model designed for NPK prediction had a lower accuracy of 0.37. This suggests that the VG19 model performed significantly better in NPK prediction compared

to the ANN model. The high accuracy achieved by the VG19 model in NPK prediction can be attributed to the utilization of a pre-trained VGG19 model and transfer learning. The VGG19 model, which was pre-trained on a large dataset, learned intricate features that are beneficial for accurate NPK prediction. On the other hand, the ANN model might have faced challenges due to the limited dataset and the complexity of capturing the complex relationships between soil characteristics and NPK levels.

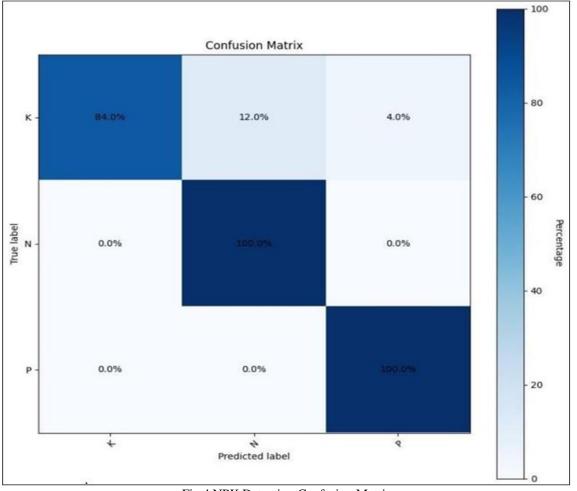


Fig 4 NPK Detection Confusion Matrix

Overall, the results indicate that the choice of model architecture and training approach significantly impacts the accuracy of disease identification and NPK prediction in tomato plants. The ResNet model outperformed the MobileNet model in disease identification, suggesting that deeper and more complex architectures are advantageous for accurate disease classification. Additionally, the VG19 model showcased superior performance in NPK prediction compared to the ANN model, emphasizing the effectiveness of pretrained models and transfer learning for such tasks. It is worth mentioning that the reported accuracies are based on the specific datasets and training configurations used in the experiments. Further research and model optimizations may be necessary to improve the accuracy and generalizability of these models. Additionally, considering the dynamic nature of plant diseases and soil nutrient levels, continuous updates and refinements of the models would be essential to ensure their effectiveness in real-world agricultural applications.

V. CONCLUSION

In conclusion, the conducted study focused on the development and evaluation of various machine learning models for different aspects of tomato plant analysis, including disease identification, seed type prediction, and NPK prediction. The results obtained shed light on the performance and capabilities of these models, providing valuable insights for the agricultural community and researchers involved in tomato cultivation. For tomato disease identification, the ResNet model outperformed the MobileNet model, achieving an overall accuracy of approximately 0.71 compared to 0.34. This highlights the importance of utilizing deep and complex architectures for accurate disease classification. However, it is essential to acknowledge that further improvements and optimizations are necessary to enhance the accuracy of the seed ANN model for disease identification due to the complexities and variabilities associated with disease symptoms. Regarding the prediction of NPK levels in tomato plants, the VG19 model exhibited remarkable accuracy with an overall accuracy of 0.95, while the ANN model achieved an accuracy of 0.37. These results emphasize the effectiveness of pre-trained models and transfer learning, as the VG19 model leveraged learned features from a large dataset to make accurate predictions. Conversely, the ANN model faced challenges due to limited data and the complexity of capturing intricate relationships between soil characteristics and NPK levels. Therefore, further research and model refinements are warranted to improve the accuracy and generalizability of NPK prediction models.

The findings of this study highlight the significance of selecting appropriate model architectures and training methodologies for different agricultural tasks. Deep learning models, such as ResNet and VG19, demonstrate their

superiority in disease identification and NPK prediction, respectively. Moreover, transfer learning from pre- trained models proves to be a valuable approach for leveraging existing knowledge and improving model performance. However, it is crucial to consider the dynamic nature of plant diseases and soil nutrient dynamics, necessitating continuous model updates and refinements to keep pace with evolving agricultural conditions. The outcomes of this research contribute to the field of agricultural technology, offering insights and guidance for the development of intelligent systems to support tomato cultivation. By accurately identifying diseases, predicting seed types, and estimating NPK levels, these models can assist farmers and researchers in making informed decisions regarding disease management, seed selection, and nutrient optimization. Nevertheless, it is essential to recognize that the reported accuracies are specific to the datasets and experimental setups employed, and further investigations are necessary to validate and enhance the models' performance in realworld scenarios.

REFERENCES

- [1]. B. P. Luitel, P. B. Adhikari, C. S. Yoon, and W. H. Kang, "Yield and fruit quality of tomato (Lycopersicon esculentum Mill.) cultivars established at different planting bed size and growing substrates," *Hortic Environ Biotechnol*, vol. 53, pp. 102–107, 2012.
- [2]. D. H. S. Komahan and K. Prasannath, "Screening of tomato (Solanum lycopersicum L.) varieties against damping off disease under the climatic conditions of Batticaloa District, Sri Lanka," *Sch J Agric Vet Sci*, vol. 4,no. 7, pp. 264–266, 2017.
- [3]. F. Gao *et al.*, "Effects of Organic Fertilizer Application on Tomato Yield and Quality: A Meta-Analysis," *Applied Sciences*, vol. 13, no. 4, p. 2184, 2023.
- [4]. A. S. Chakravarthy and S. Raman, "Early blight identification in tomato leaves using deep learning," in 2020 International conference on contemporary computing and applications (IC3A), IEEE, 2020, pp. 154–158.
- [5]. M. M. Gunarathna and R. Rathnayaka, "Experimental determination of CNN hyperparameters for tomato disease detection using leaf images," in 2020 2nd International Conference on Advancements in Computing (ICAC), IEEE, 2020, pp. 464–469.
- [6]. S. R. JuhiReshma and D. J. Aravindhar, "Fertilizer estimation using deep learning approach," *NVEONATURAL VOLATILES & ESSENTIAL OILS Journal*/ *NVEO*, pp. 5745–5752, 2021.
- [7]. M. M. Gunarathna, R. Rathnayaka, and W. M. W. Kandegama, "Identification of an Efficient Deep Learning Architecture for Tomato Disease Classification Using Leaf Images," *Journal of Food and Agriculture*, vol. 13, no. 1, pp. 33–53, 2020.

- [8]. J. Escorcia-Gutierrez, M. Gamarra, R. Soto-Diaz, M. Pérez, N. Madera, and R. F. Mansour, "Intelligent agricultural modelling of soil nutrients and ph classification using ensemble deep learning techniques," *Agriculture*, vol. 12, no. 7, p. 977, 2022.
- [9]. T. Thorat, B. K. Patle, and S. K. Kashyap, "Intelligent insecticide and fertilizer recommendation system based on TPF-CNN for smart farming," *Smart Agricultural Technology*, vol. 3, p. 100114, 2023.
- [10]. R. Azad, K. L. W. Kumara, G. Senanayake, R. Ranawaka, Pushpakumara, and S. Geekiyanage, "Intensity of leaf spot and rough bark diseases in cinnamon accessions collected from major cinnamon growing areas of Sri Lanka," *J Natl Sci Found*, vol. 47, no. 3, pp. 297–305,2019.
- [11]. D. A. Bondre and S. Mahagaonkar, "Prediction of crop yield and fertilizer recommendation using machine learning algorithms," *International Journal* of Engineering Applied Sciences and Technology, vol. 4, no.5, pp. 371–376, 2019.
- [12]. G. G. Jayasinghe, K. Ketakumbura, K. G. G. Wijesinghe, K. S. Hemachandra, and S. Weligamage, "Relationship between pest and disease incidences and agronomic operations implemented by farmers in cinnamon (Cinnamomum zeylanicum Blume) fields in southern Sri Lanka," *Sri Lanka Journal of Food and Agriculture*, vol.2, no. 1, pp. 35–40, 2016.
- [13]. D. H. S. Komahan and K. Prasannath, "Screening of tomato (Solanum lycopersicum L.) varieties against damping off disease under the climatic conditions of Batticaloa District, Sri Lanka," *Sch J Agric Vet Sci*, vol. 4,no. 7, pp. 264–266, 2017.
- [14]. W. B. Jia, H. R. Wei, and Z. Y. Wei, "Tomato fertilizer deficiency classification and fertilization decision model based on leaf images and deep learning," in *International Conference on Computer*, *Artificial Intelligence, and Control Engineering* (CAICE 2022), SPIE, 2022, pp. 170–174.
- [15]. K. C. Kamal, Z. Yin, B. Li, B. Ma, and M. Wu, "Transfer learning for fine-grained crop disease classification based on leaf images," in 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), IEEE, 2019, pp. 1–5.
- [16]. A. S. Paymode and V. B. Malode, "Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG," *Artificial Intelligence in Agriculture*, vol. 6, pp. 23–33, 2022.