

Crop Disease Detection Using Machine Learning: A Comprehensive Review

Sachin Sharma

Department of Engineering and Technology, Jagannath University, Jaipur, India

ABSTRACT: Agriculture plays a vital role in ensuring global food security and economic development. However, crop diseases remain one of the major challenges affecting agricultural productivity, causing significant yield losses and economic damage worldwide. Early and accurate identification of plant diseases is essential for implementing timely treatment measures and minimizing crop losses. Recent advancements in artificial intelligence, computer vision, and machine learning have transformed traditional disease diagnosis methods by enabling automated and intelligent crop health monitoring systems. This review paper presents a comprehensive analysis of machine learning approaches used for crop disease detection. The study discusses the complete disease detection workflow, including image acquisition, image preprocessing, feature extraction, feature selection, classification, and disease prediction. Various traditional machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Artificial Neural Networks (ANN) are examined along with advanced deep learning techniques including Convolutional Neural Networks (CNNs), Deep Neural Networks (DNNs), Transfer Learning models, and Vision Transformers (ViTs). The advantages, limitations, and performance characteristics of these approaches are analyzed based on existing research studies. Furthermore, the paper highlights the role of machine learning in improving disease diagnosis accuracy, reducing human intervention, enabling real-time monitoring, and supporting precision agriculture practices. Key challenges such as limited datasets, environmental variability, computational requirements, model generalization, and field deployment issues are also discussed. Finally, future research directions focusing on explainable artificial intelligence, Internet of Things (IoT) integration, drone-based monitoring, edge computing, and multimodal learning are presented. The review demonstrates that machine learning has significant potential to revolutionize crop disease management and contribute to sustainable agricultural development through accurate, efficient, and scalable disease detection systems.

Keywords — Crop Disease Detection, Machine Learning, Deep Learning, Precision Agriculture, Computer Vision, Convolutional Neural Networks, Artificial Intelligence.

1. Introduction

Agriculture is the backbone of many economies and plays a crucial role in ensuring global food security. With the world's population projected to exceed 9 billion by 2050, the demand for agricultural production is expected to increase significantly. However, agricultural productivity is continuously threatened by various biotic and abiotic factors, among which crop diseases are one of the most critical challenges. Plant diseases caused by fungi, bacteria, viruses, pests, and environmental stress conditions result in substantial reductions in crop yield and

quality, leading to economic losses for farmers and affecting food supply chains worldwide.

According to the Food and Agriculture Organization (FAO), plant diseases are responsible for the loss of approximately 20–40% of global crop production annually. Such losses not only impact agricultural profitability but also threaten food security, particularly in developing countries where agriculture remains the primary source of income and employment. Early and accurate detection of crop diseases is therefore essential for implementing timely treatment strategies, minimizing crop damage, and improving

overall agricultural productivity.

Traditionally, crop disease diagnosis relies on visual inspection performed by farmers or agricultural experts. Although experienced agronomists can often identify diseases effectively, manual diagnosis is time-consuming, labor-intensive, and highly dependent on expert knowledge. Furthermore, in rural and remote areas, access to agricultural specialists may be limited, resulting in delayed disease identification and treatment. Variations in disease symptoms, environmental conditions, and crop varieties further complicate the diagnosis process and increase the possibility of human error.

Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), Computer Vision, and Deep Learning have created new opportunities for automating crop disease detection [1]. These technologies enable intelligent systems to analyze plant images, identify disease symptoms, and classify diseases with high accuracy. Machine learning algorithms can learn complex patterns from large datasets and make predictions based on visual characteristics such as leaf color, texture, shape, lesions, and discoloration. By automating disease diagnosis, these systems can assist farmers in making informed decisions and reducing dependency on manual inspection [2].

Machine learning techniques have gained considerable attention in agricultural research due to their ability to process large volumes of data and generate accurate predictions. Traditional machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Naïve Bayes (NB), and K-Nearest Neighbors (KNN) have been widely used for crop disease classification. These methods typically involve image preprocessing, feature extraction, feature selection, and classification stages. Although effective, their performance often depends on the quality of handcrafted features extracted from plant images [3]-[6].

The emergence of deep learning has significantly transformed the field of crop

disease detection. Deep learning models, particularly Convolutional Neural Networks (CNNs), can automatically learn hierarchical features directly from raw images, eliminating the need for manual feature engineering. Advanced architectures such as AlexNet, VGGNet, ResNet, DenseNet, InceptionNet, EfficientNet, and Vision Transformers (ViTs) have demonstrated remarkable performance in disease classification tasks. The availability of publicly accessible datasets, such as PlantVillage, has further accelerated research and development in this area.

In addition to image-based disease detection, modern agricultural systems increasingly integrate machine learning with emerging technologies such as the Internet of Things (IoT), Unmanned Aerial Vehicles (UAVs), drones, cloud computing, edge computing, and multispectral imaging. These technologies enable real-time crop monitoring, large-scale field surveillance, and precision agriculture applications. Smart farming systems can continuously collect environmental and crop health data, allowing farmers to detect diseases at an early stage and optimize resource utilization.

Despite the significant progress achieved in recent years, several challenges remain in developing robust crop disease detection systems. Variations in illumination, background complexity, occlusion, image quality, disease similarity, and limited availability of real-world datasets can affect model performance. Furthermore, many deep learning models require large amounts of labeled data and substantial computational resources, limiting their deployment in resource-constrained agricultural environments. Issues related to model interpretability, scalability, and generalization across different geographical regions also require further investigation.

Given the rapid advancement of machine learning technologies and their growing applications in agriculture, a comprehensive review of existing approaches is essential. This review paper aims to examine the current state of machine learning-based crop disease

detection systems by analyzing traditional machine learning techniques, deep learning architectures, publicly available datasets, performance evaluation methods, practical applications, and existing challenges. Additionally, recent developments and future research directions are discussed to provide insights into the next generation of intelligent agricultural disease management systems.

The findings presented in this review highlight the transformative potential of machine learning for improving crop health monitoring, enhancing agricultural productivity, reducing economic losses, and supporting sustainable farming practices. As artificial intelligence continues to evolve, machine learning-based crop disease detection is expected to become an integral component of precision agriculture and smart farming ecosystems worldwide.

2. Crop Diseases and Their Impact

Crop diseases affect plant growth, yield quality, and productivity. Common diseases include:

- Late Blight in Potato
- Early Blight in Tomato
- Leaf Rust in Wheat
- Bacterial Spot in Pepper
- Powdery Mildew
- Rice Blast Disease
- Citrus Canker
- Maize Leaf Blight

The impact of crop diseases includes:

- Reduced agricultural productivity
- Economic losses for farmers

- Increased pesticide usage
- Environmental degradation
- Threats to food security

Early disease detection is therefore essential for sustainable agricultural management.

3. Machine Learning for Crop Disease Detection

Machine learning has emerged as a powerful tool for automated crop disease detection, enabling computer systems to learn patterns from historical data and make accurate predictions without explicit programming. In agriculture, machine learning techniques are increasingly being utilized to identify plant diseases at an early stage, thereby reducing crop losses and improving productivity. By analyzing images of crops and extracting meaningful information from them, machine learning models can distinguish between healthy and diseased plants with high accuracy. The general workflow of a machine learning-based crop disease detection system consists of image acquisition, image preprocessing, feature extraction, feature selection, classification, and disease prediction [4]-[6].

The general workflow consists of:

- Image Acquisition
- Image Preprocessing
- Feature Extraction
- Feature Selection
- Classification
- Disease Prediction

ML for Crop Disease Detection Workflow

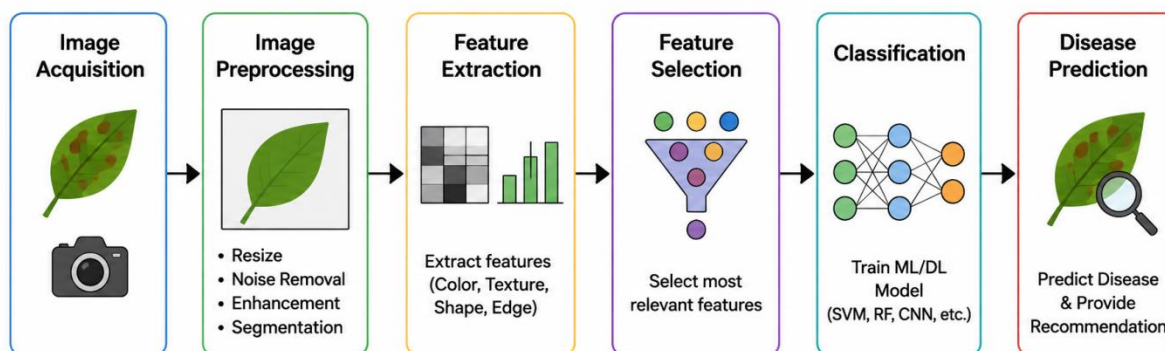


Figure 1: Workflow of Machine Learning for Crop Disease Detection

The process begins with image acquisition, where images of crops are collected using various imaging devices such as digital cameras, smartphones, drone, satellites, and multispectral sensors. These images may contain visible symptoms of diseases, including discoloration, lesions, spots, wilting, and abnormal growth patterns. The quality and diversity of the collected images play a crucial role in determining the effectiveness of the disease detection system. Large datasets containing images of both healthy and infected plants are typically used to train machine learning models to recognize different disease categories.

Once the images are collected, image preprocessing is performed to improve image quality and remove unwanted distortions. Agricultural images captured under field conditions often contain noise, shadows, varying illumination levels, and complex backgrounds that may affect the accuracy of disease detection. Preprocessing techniques such as image resizing, noise filtering, contrast enhancement, normalization, and color space transformation are applied to standardize the images and enhance important disease-related features. Segmentation techniques may also be used to isolate diseased regions from the background and healthy plant tissues, allowing the model to focus on the most relevant areas of the image.

After preprocessing, feature extraction is carried out to identify important characteristics

that can be used to differentiate healthy plants from diseased ones. Features are numerical representations of image properties and typically include color, texture, shape, and edge information. Color features help identify changes in leaf pigmentation caused by disease infection, while texture features capture variations in surface patterns and lesion structures. Shape and edge features provide additional information about the size, geometry, and boundaries of diseased regions. Effective feature extraction reduces data complexity while preserving essential information required for accurate disease classification.

The next stage involves feature selection, which aims to identify the most relevant features and eliminate redundant or irrelevant information. Since large numbers of features may increase computational complexity and lead to overfitting, selecting an optimal subset of features improves model efficiency and classification performance. Various statistical and optimization techniques are used to select features that contribute most significantly to disease discrimination. By reducing dimensionality, feature selection helps machine learning algorithms learn more effectively and improves their ability to generalize to unseen data.

Once the most informative features have been selected, classification algorithms are employed to categorize the crop images into different disease classes or identify them as

healthy. Several machine learning algorithms have been widely used for this purpose, including Support Vector Machines (SVM), Random Forests (RF), Decision Trees (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB), and Artificial Neural Networks (ANN). These algorithms learn relationships between extracted features and disease categories during the training phase and subsequently use this knowledge to classify new images. More recently, deep learning approaches such as Convolutional Neural Networks (CNNs) have gained significant popularity because they can automatically learn discriminative features directly from raw images, eliminating the need for manual feature engineering.

The final stage of the workflow is disease prediction, where the trained model analyzes a new crop image and predicts the presence and type of disease. The output may include the disease name, confidence score, severity level, and recommendations for treatment or preventive measures. The accuracy of the prediction system is typically evaluated using performance metrics such as accuracy, precision, recall, F1-score, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve. High-performing machine learning models can support farmers and agricultural experts in making timely decisions, reducing crop damage, and improving overall farm productivity.

Overall, machine learning provides an efficient and intelligent framework for crop disease detection by automating the analysis of crop images and enabling rapid, accurate, and scalable disease diagnosis. As advancements in artificial intelligence, computer vision, and agricultural sensing technologies continue, machine learning-based disease detection systems are expected to play an increasingly important role in precision agriculture and sustainable crop management.

4. Machine Learning Algorithms for Crop Disease Detection

Machine learning algorithms play a crucial

role in the automated detection and classification of crop diseases. These algorithms learn patterns from historical agricultural data and image features to distinguish between healthy and diseased plants. Over the past decade, various machine learning techniques have been applied to crop disease diagnosis, demonstrating significant improvements in accuracy, efficiency, and scalability compared to traditional manual inspection methods. The selection of an appropriate algorithm depends on factors such as dataset size, feature complexity, computational requirements, and desired prediction accuracy.

A. Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the most widely used machine learning algorithms for crop disease detection. SVM is a supervised learning technique that constructs an optimal hyperplane to separate different classes in a feature space. It is particularly effective when dealing with high-dimensional data and limited training samples. In crop disease detection, SVM utilizes features such as color, texture, shape, and lesion characteristics extracted from plant images. Several studies have reported high classification accuracy using SVM for diseases affecting crops such as rice, wheat, tomato, maize, and grape. The algorithm is known for its robustness and ability to generalize well on unseen data.

B. Random Forest

Random Forest (RF) is another popular supervised learning algorithm extensively used in agricultural applications. It is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce overfitting. Each tree in the forest is trained using a random subset of training samples and features, and the final prediction is determined through majority voting. Random Forest performs exceptionally well when handling large datasets and complex feature relationships. In crop disease detection, it has been successfully applied to

classify various plant diseases based on image-derived features and environmental parameters. Its ability to provide feature importance rankings also helps researchers identify the most influential disease indicators.

C. Decision Tree

Decision Tree (DT) is a simple yet effective classification algorithm that represents decisions in a tree-like structure. The algorithm recursively partitions data into smaller subsets based on feature values until classification rules are established. Decision Trees are easy to interpret and visualize, making them suitable for agricultural decision-support systems. Although they may suffer from overfitting when used individually, they provide valuable insights into disease classification processes and are often employed as components of ensemble methods such as Random Forest.

D. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a non-parametric machine learning algorithm that classifies a new sample based on the majority class of its nearest neighbors in the feature space. The algorithm relies on distance measures such as Euclidean distance to identify similar instances within the training dataset. KNN is relatively simple to implement and performs well for small to medium-sized datasets. In crop disease detection, KNN has been applied to classify diseases based on color and texture features extracted from leaf images. However, its computational complexity increases significantly with larger datasets.

E. Naïve Bayes (NB)

Naïve Bayes (NB) is a probabilistic classification algorithm based on Bayes' theorem. It assumes that all features are independent of each other, which simplifies computation and enables fast classification. Despite this assumption being unrealistic in many practical situations, Naïve Bayes often produces satisfactory results in agricultural disease diagnosis. It is particularly useful

when dealing with large datasets and can efficiently classify diseases based on statistical relationships among image features.

F. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) are inspired by the structure and functioning of the human brain. ANNs consist of interconnected processing units called neurons that learn complex nonlinear relationships between input features and output classes. These networks can model intricate disease patterns that may not be captured by traditional machine learning algorithms. In crop disease detection, ANNs have been used to identify diseases in various crops by analyzing image features, environmental conditions, and sensor data. Their adaptive learning capability makes them suitable for handling diverse agricultural datasets.

G. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) represent one of the most significant advancements in crop disease detection. CNNs are deep learning architectures specifically designed for image analysis tasks. Unlike traditional machine learning algorithms that require manual feature extraction, CNNs automatically learn hierarchical features directly from raw images through convolutional and pooling operations. This capability enables CNNs to capture complex disease characteristics such as lesion shapes, color variations, and texture patterns with high accuracy. Popular CNN architectures used in crop disease detection include AlexNet, VGGNet, ResNet, DenseNet, InceptionNet, MobileNet, and EfficientNet. Numerous studies have reported classification accuracies exceeding 95% when using CNN-based models on publicly available datasets such as PlantVillage.

H. Deep Neural Networks (DNNs)

Deep Neural Networks (DNNs) extend traditional neural networks by incorporating multiple hidden layers, allowing the model to

learn more complex representations of data. DNNs have been employed in crop disease diagnosis to analyze both image and non-image data, including climatic conditions, soil properties, and sensor measurements. Their ability to learn intricate patterns contributes to improved disease prediction performance.

I. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are deep learning architectures designed for sequential data analysis. While less common in image-based disease detection, these algorithms are useful for analyzing temporal agricultural data such as weather conditions, disease progression, and crop growth patterns. By capturing temporal dependencies, RNNs and LSTMs can support predictive disease monitoring and forecasting applications.

J. Transfer Learning

More recently, Transfer Learning has gained considerable attention in crop disease detection research. Transfer learning utilizes pre-trained deep learning models that have already learned useful image features from large-scale datasets. These models are then fine-tuned using agricultural datasets, significantly reducing training time and improving performance, especially when labeled crop disease images are limited. Popular transfer learning models include ResNet, VGG16, MobileNet, DenseNet, and EfficientNet.

K. Vision Transformers (ViTs)

Vision Transformers (ViTs) have emerged as a promising alternative to convolutional neural networks for image classification tasks. Unlike CNNs, Vision Transformers utilize self-attention mechanisms to capture global relationships within images. Recent studies have demonstrated that ViTs can achieve state-of-the-art performance in crop disease detection by effectively learning complex disease patterns and contextual information

from plant images.

A comparative analysis of machine learning algorithms reveals that traditional methods such as SVM, Random Forest, Decision Tree, KNN, and Naïve Bayes are computationally efficient and perform well on small and medium-sized datasets. However, their effectiveness often depends on carefully designed feature extraction and selection processes. In contrast, deep learning approaches, particularly CNNs and Vision Transformers, automatically learn features from raw images and generally achieve higher classification accuracy. Nevertheless, these models require larger datasets, greater computational resources, and longer training times.

Overall, machine learning algorithms have transformed crop disease detection by enabling accurate, rapid, and automated diagnosis of plant diseases. Continuous advancements in artificial intelligence, deep learning, and computer vision are expected to further enhance the performance of disease detection systems and support the development of intelligent precision agriculture solutions.

5. Conclusion

Crop diseases continue to pose a significant threat to agricultural productivity, food security, and economic sustainability across the world. Traditional disease identification methods largely depend on manual inspection by agricultural experts, which can be time-consuming, labor-intensive, subjective, and often inaccessible to farmers in remote regions. The emergence of machine learning and artificial intelligence technologies has provided effective solutions for automating disease diagnosis and enabling timely intervention.

This review has examined the fundamental concepts, workflow, and machine learning algorithms used in crop disease detection systems. The complete process, including image acquisition, preprocessing, feature

extraction, feature selection, classification, and disease prediction, forms the foundation of intelligent disease diagnosis frameworks. Various traditional machine learning algorithms such as Support Vector Machine, Random Forest, Decision Tree, K-Nearest Neighbor, Naïve Bayes, and Artificial Neural Networks have demonstrated promising performance in disease classification tasks. However, their effectiveness often depends on carefully engineered features and domain expertise.

Recent developments in deep learning, particularly Convolutional Neural Networks, Transfer Learning, and Vision Transformers, have significantly improved disease detection accuracy by automatically learning complex visual features directly from plant images. These approaches have achieved state-of-the-art results across multiple crop species and disease categories while reducing the need for manual feature engineering. Furthermore, advancements in computer vision, cloud computing, IoT devices, drones, and smart sensors have expanded the practical applications of machine learning-based disease monitoring systems in precision agriculture. Despite substantial progress, several challenges remain, including limited availability of high-quality annotated datasets, environmental variations in field conditions, computational complexity, model interpretability, and deployment constraints in resource-limited agricultural environments. Addressing these challenges will require the development of robust, scalable, and explainable machine learning models capable of operating effectively under real-world conditions.

Future research should focus on integrating machine learning with emerging technologies such as IoT-enabled sensing platforms, unmanned aerial vehicles, edge computing, explainable AI, and multimodal data fusion. These innovations have the potential to further enhance disease detection accuracy, support real-time decision-making, and improve crop management practices. Overall, machine learning-based crop disease detection

represents a transformative approach toward sustainable agriculture, enabling early disease diagnosis, reducing crop losses, improving productivity, and contributing to global food security.

References

- [1] P. Jha, D. Dembla, and W. Dubey, "Deep learning models for enhancing potato leaf disease prediction: Implementation of transfer learning based stacking ensemble model," *Multimedia Tools and Applications*, vol. 83, pp. 37839–37858, 2024.
- [2] P. Jha, D. Dembla, and W. Dubey, "Comparative Analysis of Crop Diseases Detection Using Machine Learning Algorithm," in *Proceedings of the Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, pp. 569–574, 2023.
- [3] P. Jha, D. Dembla, and W. Dubey, "Crop Disease Detection and Classification Using Deep Learning-Based Classifier Algorithm," in *Emerging Trends in Expert Applications and Security, Proceedings of the International Conference on Emerging Trends in Expert Applications and Security (ICETEAS 2023)*, Lecture Notes in Networks and Systems, vol. 682. Singapore: Springer, 2023.
- [4] P. Jha, D. Dembla, and W. Dubey, "Implementation of Machine Learning Classification Algorithm Based on Ensemble Learning for Detection of Vegetable Crops Disease," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, pp. 1–10, 2024.
- [5] M. Francis and C. Deisy, "Disease Detection and Classification in Agricultural Plants Using Convolutional Neural Networks—A Visual Understanding," in *Proceedings*

- of the 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, pp. 1063–1068, 2019.
- [6] P. Jha, G. K. Soni, H. Dogra, D. Goswami, K. Choudhary, and H. Vaishnav, "Plant Disease Detection and Classification using Convolutional Neural Network," in Proceedings of the 4th International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, pp. 1442–1446, 2025.
- [7] I. Yadav, V. Shekhawat, K. Gautam, G. K. Soni, and R. Yadav, "Artificial Intelligence for Cybersecurity: Emerging Techniques, Challenges, and Future Trends," in Proceedings of the 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, pp. 1176–1180, 2025.
- [8] P. Jha, D. Dembla, and W. Dubey, "Implementation of Transfer Learning Based Ensemble Model using Image Processing for Detection of Potato and Bell Pepper Leaf Diseases," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, pp. 69–80, 2024.
- [9] M. K. Jha, G. K. Soni, G. Jain, S. Tiwari, K. Gupta, and B. Singhal, "Comparative Analysis of Classical Machine Learning Models for Twitter Sentiment Classification," in Proceedings of the 5th International Conference on Communication, Computing and Electronics Systems (ICCCES), pp. 1949–1954, 2026.
- [10] M. Kumar, R. Ajmera, and D. Kumar, "Statistical Analysis and Accuracy Assessment of Improved Machine Learning-Based Opinion Mining Framework," *Advances in Nonlinear Variational Inequalities*, vol. 27, no. 1, 2024.
- [11] A. Kumar and N. Hemrajani, "Comparative Analysis of Different Transport Layer Protocol Techniques in Cognitive Network," *Recent Advances in Computer Science and Communications*, Bentham Science Publishers, vol. 17, 2024.
- [12] A. Johari, R. Sharma, A. Meena, and V. Tiwari, "Advancements in Pre-Trained Language Models and Their Impact on Various Natural Language Processing Tasks," *International Journal of Engineering Trends and Applications (IJETA)*, vol. 11, no. 3, pp. 201–209, 2024.
- [13] R. Ajmera, A. Johari, A. Goyal, A. Purohit, A. Kumar, and J. A. Ashok, "Multilingual Sentiment Analysis Based on Fine-Tuned Transformer Architectures," in Proceedings of the 5th International Conference on Communication, Computing and Electronics Systems (ICCCES), pp. 1589–1592, 2026.
- [14] A. Gautam, R. Ajmera, D. K. Dharamdasani, S. Srivastava, and A. Johari, "Improving climate change predictions using time series analysis and deep learning," *Global and Stochastic Analysis*, vol. 12, no. 4, Jul. 2025.