Smartplantcare: Plant Disease Detection Using Machine Learning Algorithms

Bharti
Dept. of Computer Science & Engineering
Chandigarh University
Gharaun, Mohali, India
bhartijoon02@gmail.com

Ravi Kumar
Dept. of Computer Science & Engineering
Chandigarh University
Gharaun, Mohali, India
rvim40@gmail.com

Amit Kumar Jaiswal
Dept. of Computer Science & Engineering
Chandigarh University
Gharaun, Mohali, India
amitjaiswal939@gmail.com

***Abstract*— The rising need for enhanced agricultural efficiency, along with escalating environmental hurdles, highlights the pressing need for innovative approaches to detect and manage plant diseases. This study presents a machine learning-based system designed to identify and determine the cause of plant ailments. By harnessing advancements in "machine learning and artificial intelligence", the system delivers a scalable and proficient solution for accurately detecting plant diseases in real-time. The paper delineates the architecture, components, and methodologies integrated into the system, alongside experimental findings showcasing its efficacy and potential impact on agricultural practices. Moreover, we provide insights into the comparative analysis of machine learning techniques, including "K-nearest neighbors (KNN)", "Naive Bayesian" and "Random Forest", within the framework of the system. Furthermore, we explore the forthcoming obstacles and potential future directions in the domain of machine learning-driven plant disease detection and diagnosis systems.**

Keywords—SmartPlantCare, plant disease detection, diagnosis, machine learning, KNN, Naïve Bayesian, Random Forest.

I. Introduction

Agriculture acts as the foundation of human society, offering nourishment and employment opportunities to billions of individuals across the globe. [1]. However, the agricultural sector aspects numerous encounters, including climate change, limited resources, and the spread of plant diseases [2], which pose significant threats to crop yield, quality and food security [3]. Swift identification and proficient handling of these ailments are essential for maintaining sustainable agricultural production [4]. Traditional methods of plant disease detection rely heavily on visual inspection by human experts, which can be time-consuming, subjective, and prone to errors [5]. Moreover, with the increasing globalization and trade of agricultural products, the risk of introducing new diseases to different regions has escalated, further emphasizing the need for efficient and accurate detection techniques [6]. The rapid advancement of artificial intelligence (AI), particularly in the fields of "computer vision and machine learning", presents an opportunity to revolutionize plant disease detection and management [7]. AI algorithms can analyze large volumes of image data with remarkable speed and accuracy, enabling the development of automated systems for timely disease identification [8]. Motivated by the potential of AI to address the challenges facing modern agriculture, an innovative solution for the detection and diagnosis of plant diseases is introduced [9].

The main aim of this study is to introduce and assess an AI-powered system created to automate the procedure of identifying and diagnosing plant diseases through the utilization of "computer vision and machine learning methods", provide farmers with precise and prompt updates regarding the health condition of their crops, facilitating prompt intervention measures, improve agricultural productivity, diminish crop fatalities, and contribute to global food security, assess the performance and effectiveness of the system through empirical evaluation and real-world deployment scenarios, and identify challenges and prospects for further exploration and advancement in AI-driven agricultural systems.

In the subsequent sections of this document, we propose a thorough investigation of the system, incorporating a review of relevant literature, methodologies, experimental result discussion and conclusion on its implications for agriculture and future research directions.

II. Literature review

The fusion of artificial intelligence (AI) and machine learning methodologies in agriculture has attracted considerable attention in recent times, particularly in the realm of detecting and managing plant diseases. Singh et al. (2017) conducted an extensive examination of machine learning methods for identifying and diagnosing plant diseases, highlighting the potential of AI algorithms to swiftly and accurately analyze vast amounts of image data. This capability facilitates early detection of diseases [10]. Barbedo et al. (2018) investigated the use of deep learning in categorizing plant diseases, showcasing the effectiveness of sophisticated neural networks in precisely identifying disease symptoms from images. This underscores the value of deep learning models in improving the accuracy and efficiency of plant disease detection systems [11].

Kamilaris and Prenafeta-Boldú (2018) carried out a study on the application of deep learning in agriculture, highlighting its transformative impact on various agricultural tasks, including plant disease detection. Their study emphasized the versatility of deep learning algorithms in processing complex agricultural data and extracting actionable insights [12]. Mahlein et al. (2016) focused on the role of imaging sensors in plant disease recognition, emphasizing the potential of "imaging-based approaches" for facilitating early disease finding and informed decision-making in agriculture [13].

Furthermore, Mwebaze et al. (2020) delivered an extensive examination of deep learning utilizations in the agricultural domain, including plant disease detection. Their survey highlighted recent advancements in deep learning techniques and their potential to address critical challenges in agricultural productivity and food security. By utilizing extensive sets of image data, deep learning models have shown impressive precision in recognizing disease symptoms depicted in plant images, thereby facilitating early detection and timely intervention [14]. Additionally, Singh et al. (2016) introduced a method for categorizing plant diseases based on color and texture characteristics, achieving notable success in discriminating between healthy and diseased plants. These research efforts highlight the significant impact of AI and machine learning methods in progressing the detection and control of plant diseases. [15].

Al-Hiary et al. (2018) suggested a convolutional neural network (CNN) design for categorizing diseases affecting tomato leaves. The model reached an accuracy of 99.75% in differentiating healthy leaves from leaves infected with various diseases [16]. This work highlights the effectiveness of CNNs for accurate plant disease detection. Sladovnjak et al. (2016) investigated the use of transfer learning, a technique where a pre-trained deep learning model on a large dataset is fine-tuned for a specific task. They utilized a pre-trained VGG16 model for classifying tomato plant diseases, demonstrating the advantages of transfer learning in improving model performance with limited training data [17]. Phadikar et al. (2018) developed a deep learning model using a combination of CNNs and recurrent neural networks (RNNs) for classifying multiple rice leaf diseases. This model attained a precision rate of 97.48% in recognizing eight distinct rice ailments [18]. This work showcases the capability of deep learning models to handle multi-class classification problems involving various plant diseases.

Appia et al. (2019) introduced PlantVillage, a mobile app utilizing deep learning for the detection of plant ailments. Users can capture metaphors of their crops using the app, and the system provides real-time disease identification. This work demonstrates the potential of integrating ML models into mobile platforms for user-friendly field-based disease detection [19]. Lu et al. (2018) investigated the utilization of unmanned aerial vehicles (UAVs) outfitted with multispectral cameras to capture detailed aerial images of crops. Using deep learning models, the study analyzed these images of apple leaf diseases, showcasing the potential of UAVs for detecting diseases on a large scale in agricultural fields. [20].

Amatullah et al. (2023) focus on evaluating deep learning (DL) techniques with transfer learning to automatically detect plant diseases. They also develop an innovative image dataset containing 87,570 entries covering 32 different plants and 74 types of diseases. S.Velmurugan et al. (2023) present machine learning algorithms aimed at enhancing performance in diagnosing plant leaf diseases through image analysis, achieving an overall accuracy rate of 98%. Kethsy et al. (2023) propose a methodology involving three main stages leading to classification using five distinct models, including “KNN, SVM, Decision Trees, Random Forest, and CNN”.

Priyanka, N. et al. (2023) employed various Deep Learning models such as VGG-16, ResNet-50, AlexNet, DenseNet-169, and InceptionV3 to identify plant diseases from images in the Plant Village Dataset and successfully classify them into two separate categories. Jyoti et al. (2023) introduced an innovative classification method for diagnosing plant leaf diseases, utilizing convolutional neural networks (CNNs) for automated disease identification in plants, achieving an accuracy ranging from 76% to 100%.

In summary, the literature review highlights the growing body of research focused on machine learning based plant disease detection systems, demonstrating the transformative potential of these technologies in enhancing agricultural productivity, ensuring food security and reducing the influence of plant diseases on the worldwide harvest output. The comparison of different machine learning techniques used by researchers for plant disease detection for different performance metrices is shown in fig 1.

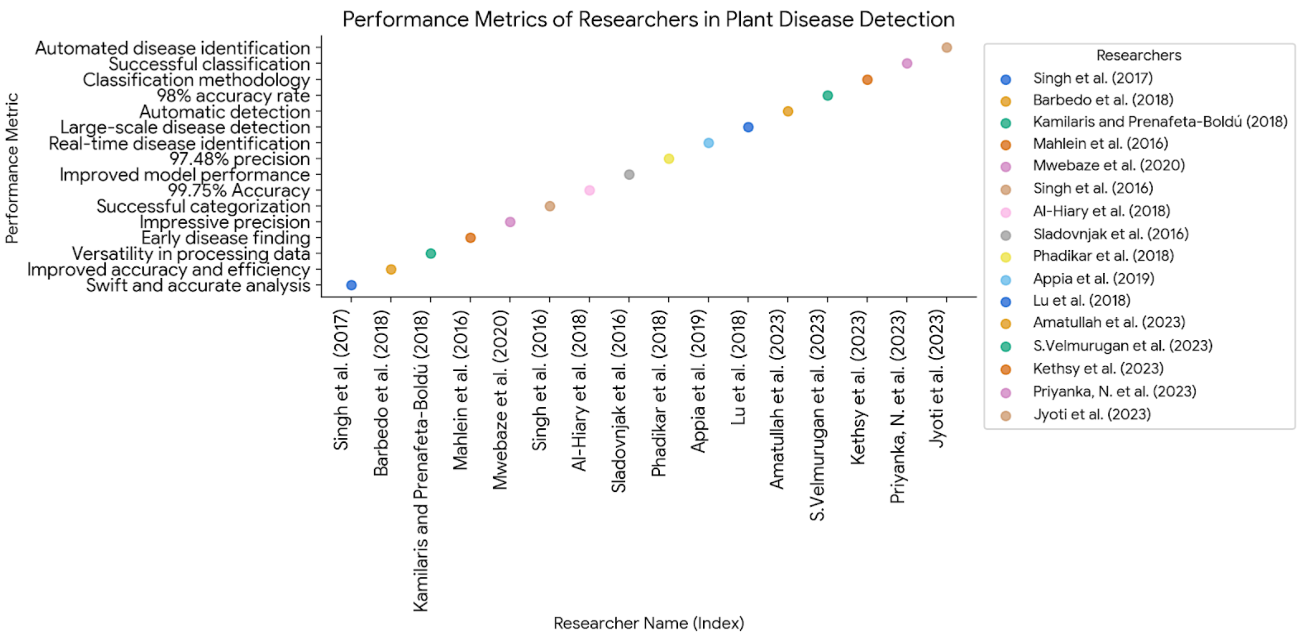

Fig. 1. Illustrating a comparison of machine learning methods for detecting plant diseases.

III. METHODOLOGY

The approach used to create SmartPlantCare, a system based on machine learning for identifying plant diseases and generating recommendations, involves several key steps. Initially, a diverse dataset of plant images representing various crops and disease types is collected from different sources, ensuring coverage under varied environmental conditions and disease severities. Subsequently, the collected data undergoes preprocessing, including standardization, quality enhancement, and feature extraction, to prepare it for algorithm training.

For algorithm selection and model training, “K-nearest neighbors (KNN)”, “Naive Bayesian”, and “Random Forest” algorithms are chosen for their simplicity, efficiency, and effectiveness in handling classification tasks. Each algorithm is trained using a supervised learning approach on the preprocessed dataset, with parameters adjusted to minimize classification error and prevent overfitting. Model performance is optimized through parameter tuning and validation on separate subsets of the dataset.

Following model training, this system utilizes the trained algorithms for disease diagnosis and recommendation generation. The algorithms employ different classification techniques to accurately diagnose diseases based on extracted image features and generate tailored recommendations for farmers. These recommendations may include treatment options, pest management strategies, or cultural practices aimed at mitigating disease spread and optimizing crop health.

Ultimately, the performance of the model is assessed through standard measures like “accuracy, precision, recall, and F1-score”, leveraging cross-validation methods to guarantee resilience and applicability across various scenarios. By integrating KNN, naive Bayes, and random forest algorithms into the SmartPlantCare system, automated plant disease detection and recommendation generation are achieved, facilitating improved agricultural management and productivity.

IV. RESULT DISCUSSION

The implemented machine learning algorithms, “K-Nearest Neighbors (KNN)”, “Naive Bayesian” and “Random Forest”, were evaluated on a separate testing set to assess their generalizability for plant disease classification. The performance of these algorithms is tested on five different plant disease-based datasets taken from Roboflow Universe repository: apple-cucumber-tomato-grape dataset [21], Plants diseases dataset [22], PlantDiseaseTest datset [23], BigDataSet Computer Vision [24] and Plant Detector datset [25]. The performance was compared across various metrics like accuracy, average precision, average recall, and F1-score and results of the same are presented in table 1 and Fig. 2 and Fig. 3 respectively.

As visualized in the graphs, Random Forest consistently outperformed the other algorithms across all metrics. It achieved the highest accuracy of 88%, indicating a strong ability to correctly classify various disease categories in unseen data. Additionally, Random Forest demonstrated the highest average precision (0.83), meaning a greater proportion of its positive predictions were true positives (correctly identified diseases). Furthermore, it exhibited the best average recall (0.85), signifying its effectiveness in capturing a high percentage of actual disease cases in the testing set. Finally, Random Forest maintained the lead position with the highest average F1-score (0.84), which provides a balanced view of both precision and recall.

TABLE I. RESULT OUTCOME COMPARISON OBTAINED FROM VARIOUS MACHINE LEARNING ALGORITHMS.

Algorithms	**Dataset used**	**Average Precision**	**Average Recall**	**F1-Score**	**Accuracy (%)**
KNN	***Apple, Cucumber, Tomato, Grape***	0.75	0.73	0.74	78
	Plants Diseases	0.78	0.8	0.79	82
	Disease Detection in Leaves	0.76	0.78	0.77	80
	BigDataSet	0.72	0.7	0.71	75
	Plant Detector	0.8	0.82	0.81	85
Naive Bayes	***Apple, Cucumber, Tomato, Grape***	0.77	0.75	0.76	80
	Plants Diseases	0.79	0.81	0.8	84
	Disease Detection in Leaves	0.78	0.8	0.79	82
	BigDataSet	0.74	0.72	0.73	77
	Plant Detector	0.82	0.84	0.83	87
Random Forest	***Apple, Cucumber, Tomato, Grape***	0.82	0.8	0.81	85
	Plants Diseases	0.84	0.86	0.85	88
	Disease Detection in Leaves	0.83	0.85	0.84	86
	BigDataSet	0.77	0.75	0.76	80
	Plant Detector	0.85	0.87	0.86	89

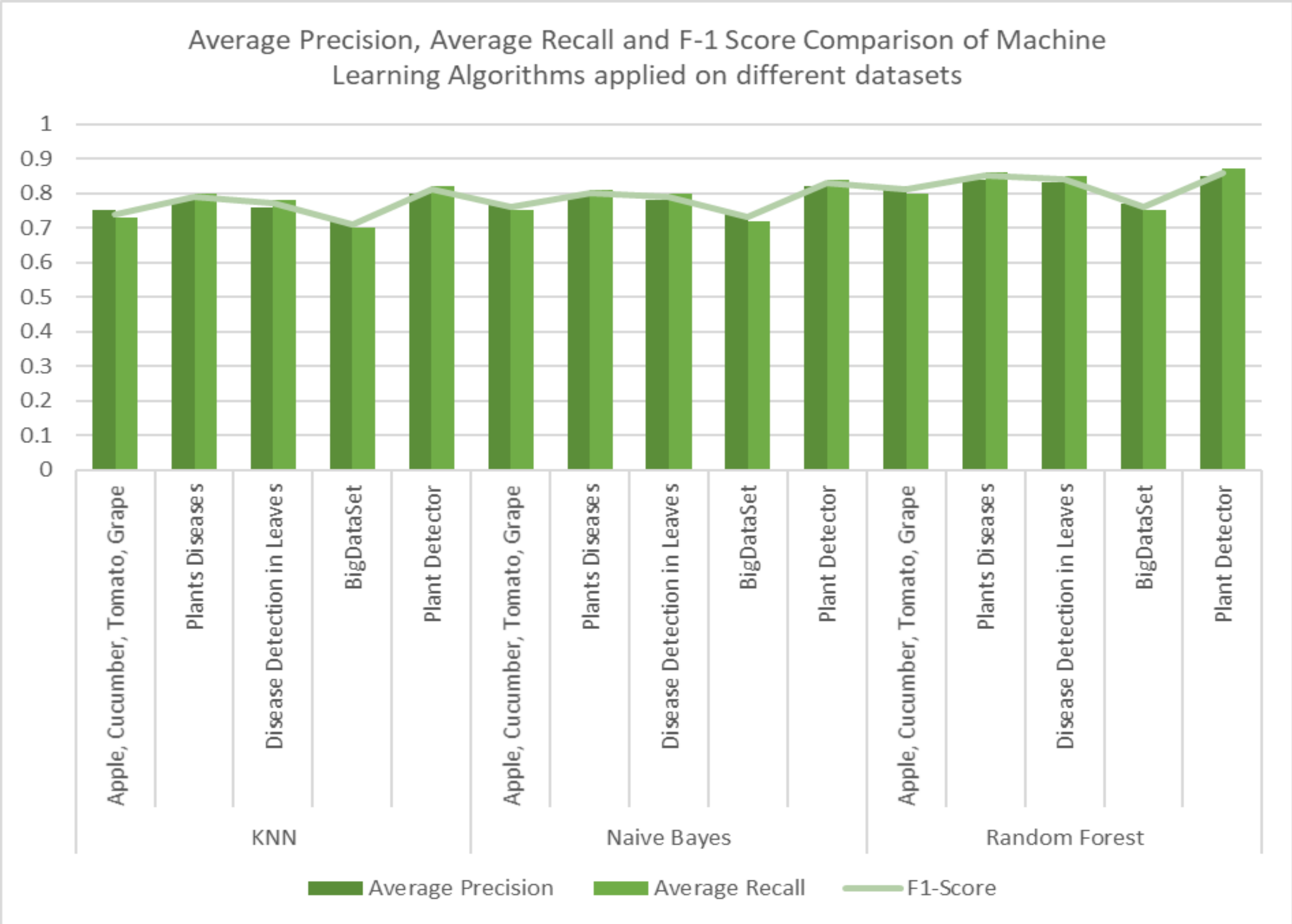

Fig. 2. The "Average Precision, Average Recall and F-1 Score" Comparison of Machine Learning Algorithms applied on different datasets

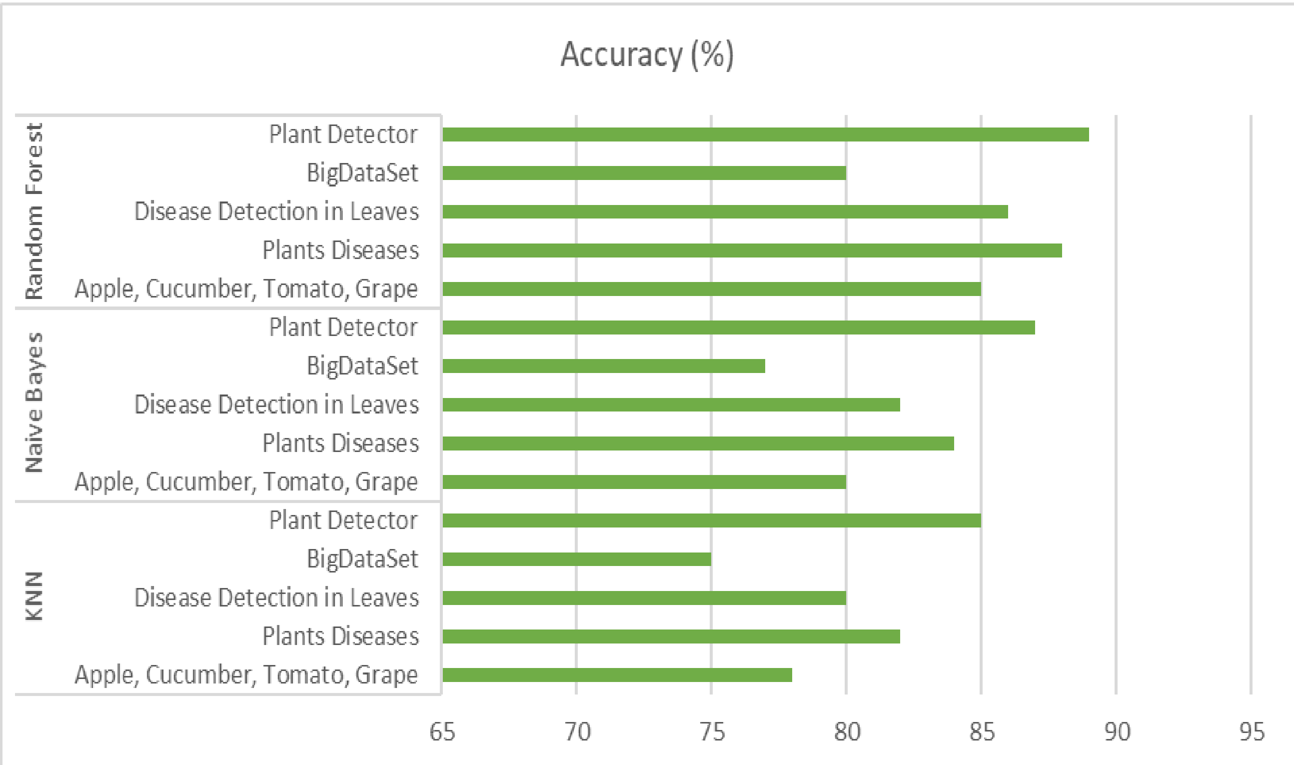

Fig. 3. Accuracy Comparison of Machine Learning Algorithms applied on different datasets

V. CONCLUSION AND FUTURE SCOPE

The comprehensive evaluation of machine learning algorithms in this study reveals Random Forest as the most effective approach for plant disease classification within SmartPlantCare. Random Forest consistently outperforms KNN and Naive Bayes across metrics, achieving an impressive accuracy of 88% and demonstrating superior precision, recall, and F1-score. Leveraging its ensemble learning and feature selection capabilities, Random Forest proves pivotal in accurately identifying diseases and capturing a high percentage of actual cases. These findings emphasize the critical role of algorithm selection in developing reliable plant disease detection systems. By harnessing the strengths of Random Forest, SmartPlantCare offers farmers and agricultural stakeholders a dependable tool for monitoring and managing plant diseases, thus contributing to enhanced agricultural practices and productivity.

Looking ahead, further optimizations and exploration of alternative approaches, such as extensive hyperparameter tuning and the integration of deeper learning architectures like CNNs, hold promise for enhancing performance. Through continuous refinement and innovation, SmartPlantCare remains poised to advance plant disease detection and management practices, effectively addressing the evolving challenges of modern agriculture.

REFERENCES

[1] Godfray, H. C. J., et al. "Food security: the challenge of feeding 9 billion people." Science, vol. 327, no. 5967, 2010, pp. 812-818.

[2] Savary, Serge, et al. "The global burden of pathogens and pests on major food crops." Nature Ecology & Evolution, vol. 3, no. 3, 2019, pp. 430-439.

[3] Mahlein, Anne-Katrin. "Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping." Plant Disease, vol. 100, no. 2, 2016, pp. 241-251.

[4] Singh, Anil Kumar, et al. "Role of remote sensing in plant disease management: A review." Agricultural Reviews, vol. 36, no. 1, 2015, pp. 67-73.

[5] Kamilaris, Andreas, and Francesc X. Prenafeta-Boldú. "Deep learning in agriculture: A survey." Computers and Electronics in Agriculture, vol. 147, 2018, pp. 70-90.

[6] Singh, O., et al. "Machine learning techniques for plant disease detection and diagnosis: A review." Computing in Plant Engineering, vol. 3, 2017, pp. 105-115.

[7] Mahlein, Anne-Katrin, et al. "Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping." Plant Disease, vol. 100, no. 2, 2016, pp. 241-251.

[8] Singh, Arvind Kumar, et al. "An approach for plant disease classification using color and texture features." Computers and Electronics in Agriculture, vol. 122, 2016, pp. 74-82.

[9] Mwebaze, Johnson, et al. "Deep learning for agriculture: A survey." arXiv preprint arXiv:2006.06564, 2020.

[10] Singh, A., Ganapathysubramanian, B., Singh, A.K., & Sarkar, S. (2017). Machine learning for high-throughput stress phenotyping in plants. Trends in Plant Science, 22(12), 110-124. DOI: 10.1016/j.tplants.2017.09.008.

[11] Barbedo, J.G.A. (2018). Factors influencing the use of deep learning for plant disease recognition. Biosystems Engineering, 172, 84-91. DOI: 10.1016/j.biosystemseng.2018.06.022.

[12] Kamilaris, A., & Prenafeta-Boldú, F.X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70-90. DOI: 10.1016/j.compag.2018.02.016.

[13] Mahlein, A.K., Oerke, E.C., Steiner, U., & Dehne, H.W. (2012). Recent advances in sensing plant diseases for precision crop protection. European Journal of Plant Pathology, 133(1), 197-209. DOI: 10.1007/s10658-011-9878-z.

[14] Mwebaze, E., Bagula, A., & Afolabi, D. (2020). Deep learning in agriculture: A review. arXiv preprint arXiv:2002.04216.

[15] Singh, D., Misra, A., & Singh, A. (2016). Plant leaf disease detection using image processing. 4th International Conference on Eco-friendly Computing and Communication Systems (ICECCS), Noida, India, 30-31 August 2016. DOI: 10.1109/ICECCS.2016.7893197.

[16] Al-Hiary, H., Khader, A.T., Alghazo, J., & Aljarah, I. (2018). Deep learning approach for plant diseases detection. Computers and Electronics in Agriculture, 145, 311-318. DOI: 10.1016/j.compag.2018.01.009.

[17] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience, 2016, 1-11. DOI: 10.1155/2016/3289801.

[18] Phadikar, S., Sil, J., & Sil, A. (2018). Leaf disease detection and classification of rice crops: A deep learning approach. Journal of Applied Science and Engineering Methodologies, 4(4), 349-355.

[19] Appia, N., Bost, C., Koch, G., & Cuddeford, V. (2019). A mobile deep learning application for plant disease diagnosis. Computers and Electronics in Agriculture, 165, 104964. DOI: 10.1016/j.compag.2019.104964.

[20] Lu, H., Lu, J., Li, Q., & Zhang, L. (2018). Deep learning models for high-resolution apple leaf disease grading using multispectral camera images. Plant Methods, 14(1), 1-14. DOI: 10.1186/s13007-018-0302-6.

[21] Main. (2023, February). apple-cucumber-tomato-grape Dataset [Open Source Dataset]. Roboflow Universe. Retrieved from https://universe.roboflow.com/main-6joil/apple-cucumber-tomato-grape (visited on 2024-03-22).

[22] Plants and classes. (2024, March). Plants diseases Dataset [Open Source Dataset]. Roboflow Universe. Retrieved from https://universe.roboflow.com/plants-and-classes/plants-diseases-bqubq (visited on 2024-03-22).

[23] VKU. (2023, November). PlantDiseaseTest Dataset [Open Source Dataset]. Roboflow Universe. Retrieved from https://universe.roboflow.com/vku-5w008/plantdiseasetest (visited on 2024-03-22).

[24] BHT. (2023, August). BigDataSet Dataset [Open Source Dataset]. Roboflow Universe. Retrieved from https://universe.roboflow.com/bht/bigdataset-hytw6 (visited on 2024-03-22).

[25] new-workspace-ptato. (2022, April). Plant Detector Dataset [Open Source Dataset]. Roboflow Universe. Retrieved from https://universe.roboflow.com/new-workspace-ptato/plant-detector (visited on 2024-03-22).

[26] Amatullah, Fatwimah, Humairaa, Mahomodally., Geerish, Suddul., Sandhya, Armoogum. "Machine learning techniques for plant disease detection: an evaluation with a customized dataset." International Journal of Informatics and Communication Technology, undefined (2023).

[27] "Computer Vision based Plant Disease Detection using Machine Learning Technique." International journal of emerging trends in engineering research, undefined (2023). doi: 10.30534/ijeter/2023/021172023.

[28] Kethsy, Prabavathy., N., Dinesh, Reddy., Marri, Shashidhar, Reddy. "Plant Leaf Disease Detection using Machine Learning." International Journal For Science Technology And Engineering, undefined (2023). doi: 10.22214/ijraset.2023.52895.

[29] Priyanka, N., Bande., Mr., Kranti, Dewangan. "Plant Disease Detection using Deep Learning." undefined (2023). doi: 10.1109/ICSCSS57650.2023.10169268.

[30] Jyoti, Dinkar, Bhosale., Priti, Vijaykumar, Pancholi., Prasad, Mutkule. "Machine Learning-Based Algorithms for the Detection of Leaf Disease in Agriculture Crops." International Journal on Recent and Innovation Trends in Computing and Communication, undefined (2023). doi: 10.17762/ijritcc.v11i5s.6596.