IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, March 2025
Impact Factor: 7.67

Deep Learning Approach for Seed Disease Prediction Using VGG16 with Transfer Learning and Optuna Optimization

Deepa Bhadana[1] and Aravindhan Kurunthachalam[2]
Chaudhary Charan Singh University, Meerut, India[1]
School of Computing and Information Technology, REVA University, Bangalore[2]
drdeepabhadana@gmail.com and aravindhan03@gmail.com

Abstract*: Early detection of seed diseases helps produce healthy crops and create less agricultural loss. Seed disease detection methodologies currently in existence are time-consuming, labour-intensive, and heavily dependent on expert knowledge. Traditional techniques under threat of being unavailable for large-scale agricultural practices. They do not avoid challenges and quick and accurate results hindering their timely interventions. These methods also lack scalability and are poorly adaptable to diverse datasets or indeed of the complexity of disease patterns. Hence, there is a necessity of automating more efficient solutions on scale and thus bring the much-needed shift in seed disease detection from early on. This research proposes a model that makes efficient seed disease prediction using VGG16 architecture under transfer learning and Optuna optimization. The VGG16 model was fine-tuned with trained data from big datasets toward the classification of such seed diseases as fungal and bacterial infections. For optimal performance, Optuna hyperparameter optimization was applied to fine-tune parameters such as learning rate, batch size, and training epochs. Among key performance metrics are accuracy, recall, precision, and F1-score, which would establish the effectiveness of the model in classifying healthy to diseased seeds. The ROC curve yielded high true positive and low false positive, indicating model robustness and adaptability for real-world purposes in agriculture. Overall, this results into accurate, quick, and scalable approaches toward early seed disease detection, which will contribute more to good management of crops and food security.*

Keywords: VGG16, transfer learning, Optuna optimization, and hyperparameter tuning. Additional key terms are deep learning, precision agriculture, image classification, fungal infections, and bacterial infections

I. INTRODUCTION

The prevention of seed diseases helps promote healthy growth of crops and minimize any losses due to the plant disease. Once seed disease is detected early, the spreading of pathogens will be contained with timely intervention, leading to better health for plants [1]. Given the increased demand for food globally, there is a dire need to implement efficient and accurate methods for early detection of seed diseases. Traditional techniques employed for seed disease identification usually take time, are labour-intensive, and based on the knowledge of experts; thus, making it unsustainable for large-scale agricultural practices [2]. Hence, the advanced image recognition and machine learning techniques are gaining acceptance in the automation of the detection process. Recently, deep learning, especially Convolutional Neural Networks (CNN), has emerged as a promising prospect for image-based classification tasks including seed disease prediction [3]. The VGG16 model has emerged as one of the best architectures in the field of computer vision that has been significantly successful in image recognition contests. This is due to its depth of layers being able to capture complex characteristics in an image. Transfer learning can be applied to the pre-trained VGG16 model, which serves to enhance seed disease prediction through fine-tuning for the task in hand and enables a great level of generalizability over various datasets without incurring an extreme cost of computational resources for

DOI: 10.48175/IJARSCT-23688

generation [4]. Optuna hyperparameter optimization framework helps in model performance improvement. Fine-tuning the hyperparameters of the VGG16 model like learning rate, batch size, and the number of training epochs with the help of Optuna for the best performance. The optimization process searches for that hyperparameter combination which maximizes the accuracy of the model while minimizing prediction error, essentially making the process of seed disease prediction more accurate and efficient [5].

The prevention of seed diseases helps promote healthy growth of crops and minimize any losses due to the plant disease. Once seed disease is detected early, the spreading of pathogens will be contained with timely intervention, leading to better health for plants [6]. Given the increased demand for food globally, there is a dire need to implement efficient and accurate methods for early detection of seed diseases. Traditional techniques employed for seed disease identification usually take time, are labour-intensive, and based on the knowledge of experts; thus, making it unsustainable for large-scale agricultural practices [7]. Hence, the advanced image recognition and machine learning techniques are gaining acceptance in the automation of the detection process. Recently, deep learning, especially Convolutional Neural Networks (CNN), has emerged as a promising prospect for image-based classification tasks including seed disease prediction. The VGG16 model has emerged as one of the best architectures in the field of computer vision that has been significantly successful in image recognition contests. This is due to its depth of layers being able to capture complex characteristics in an image [8]. Transfer learning can be applied to the pre-trained VGG16 model, which serves to enhance seed disease prediction through fine-tuning for the task in hand and enables a great level of generalizability over various datasets without incurring an extreme cost of computational resources for generation. Optuna hyperparameter optimization framework helps in model performance improvement [9]. Fine-tuning the hyperparameters of the VGG16 model like learning rate, batch size, and the number of training epochs with the help of Optuna for the best performance [10]. The optimization process searches for that hyperparameter combination which maximizes the accuracy of the model while minimizing prediction error, essentially making the process of seed disease prediction more accurate and efficient.

II. LITERATURE REVIEW

A blockchain framework has been introduced for the secure management of employee data in HRM, utilizing distributed control and tensor decomposition techniques. The application of AI and ML algorithms allows for the transformation of chronic disease management and fall prevention processes in geriatric care, advancing predictive healthcare applications for elder care [11]. Enhancements to software solutions have been reported through memory-augmented neural networks, hierarchical multi-agent learning, and concept bottleneck models to improve performance and explainability [12]. Cloud computing is being utilized for healthcare data collection and security optimization to enable real-time access while ensuring data privacy. In smart manufacturing, data security is addressed through an AI-blockchain hybrid framework for IIoT that employs self-sovereign identity and sidechains to enhance privacy [13]. Blockchain and Hierarchical Verifiable Tokens are also used to ensure data integrity in multi-cloud storage environments. Further applications show the expansion of AI and blockchain technologies across areas like security, healthcare, and software development [14]. In cloud networks, security enhancement is achieved through intrusion detection and alert correlation using CNN and autoencoders, while another approach employs a Tab-Transformer-supported intrusion detection system. The performance and adaptability of AI-based software models have also been improved by integrating Particle Swarm Optimization (PSO) with Quadratic Discriminant Analysis (QDA) [15].

Advanced deep learning techniques are applied for traffic classification and attack detection, optimizing traffic management and cloud security [16]. Disease diagnosis in healthcare has seen improvements using the Firefly Algorithm (FA) with CNN and Differential Evolution (DE) with Extreme Learning Machine (ELM). A novel hybrid GWO and DBN model has been presented for accurate disease prediction [17]. An AI-powered anomaly detection system has been proposed for the secure sharing of data in multi-cloud healthcare networks. For improved cyber threat detection in federated learning, a framework integrates K-Nearest Neighbors (KNN), GANs, and IOTA [18]. Further optimization of AI systems is achieved using A3C, TRPO, and POMDP for better decision-making and performance. Robotic software quality assurance benefits from AI-based cloud load testing combined with automated UI/UX testing for smart robotics systems [19]. Recruitment processes have been revolutionized by integrating AI and blockchain

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, March 2025
Impact Factor: 7.67

technologies to enhance talent acquisition, streamline candidate screening, and ensure secure, transparent hiring [20]. Stronger AI software development is supported through Non-Orthogonal Multiple Access (NOMA), UVFA, and dynamic graph neural networks for scalable decision-making. AI-enabled customer relationship management (CRM) solutions are also being developed to improve customer experience through Fuzzy Cognitive Systems (AI-FCS), enabling personalized services [21]. These advancements collectively demonstrate how merging AI with deep learning, blockchain, and optimization techniques significantly enrich healthcare, recruitment, traffic management, security, and software quality across diverse domains [22].

III. PROBLEM STATEMENT

This research problem seeks to improve scalability, decision-making, and personalization across various sectors, such as healthcare, recruitment, traffic management, and security. Traditional systems find it problematic to handle large-scale data, real-time adaptability, or personalization [23]. Decision-making is enhanced by combining Non-Orthogonal Multiple Access (NOMA), Universal Value Function Approximators (UVFA), and dynamic graph neural networks, while AI-assisted customer relationship management (CRM) is achieved using Artificial Intelligence and Fuzzy Cognitive Systems (AI-FCS) [24]. These innovations are intended to deliver improved performance, service quality, and security and efficiency in operation throughout various domains [25].

3.1 Objective

The research will aim to develop solutions that can be scalable, adaptive, personalized, and augment enhanced decision-making across different environments like healthcare, recruitment, traffic management, and security. The techniques include NOMA and UVFA, dynamic graph neural networks, and AI-driven CRM with Fuzzy Cognitive Systems using advanced AI, which will solve problems of large-scale data handling, real-time adaptability, and personalized service delivery. With these measures, the research aims to improve performance, user experience, as well as secure and efficient systems that can cater to the different needs of all sectors.

IV. PROPOSED SEED DISEASE PREDICTION USING VGG16 WITH TRANSFER LEARNING AND OPTUNA OPTIMIZATION

The method introduced for predicting seed disease is one following the structure for seed disease prediction using VGG16 with transfer learning and Optuna optimization. This is to direct seed disease detection and classification efficiently. It includes the collection and pre-processing image of seeds, which is sometimes resized, normalized, and enhanced using rotation, flipping, and scaling augmentation techniques so that the training dataset becomes diversified. Initially, the pre-trained model, VGG16, is fine-tuned for the classification of seed diseases. Transfer learning is then opted for adapting the features learned by the model to those seed disease images, even as the complex patterns specific to seed health can be recognized without requiring very many, or maybe even without much training, by the model itself. For optimal performance, the model is also fine-tuned using Optuna for hyperparameter tuning specific to the VGG16 model. Optuna searches in the hyperparameter space, tuning parameters like learning rate, batch size, and number of epochs to maximize accuracy and minimize error. Finally, performance metrics such as accuracy, precision, recall, or F1 score are utilized to evaluate the optimized model regarding its capability to predict seed diseases reliably. This integrated approach of deep learning along with optimization techniques would efficiently, e, and most importantly scalable feature in the early detection of seed diseases in agriculture.

4.1 Data Collection

The collection stage obtains a complete seed sickness database that holds images of seeds marked with diverse health conditions. This database will consist of seeds dividing into three main healthy classes: fungal-infected, bacterial-infected, and healthy. In addition, it is important to make the seeds diverse in types and diseases for robustness and generalization of the model. This database will be very crucial for training the disease classification with the VGG16 model.

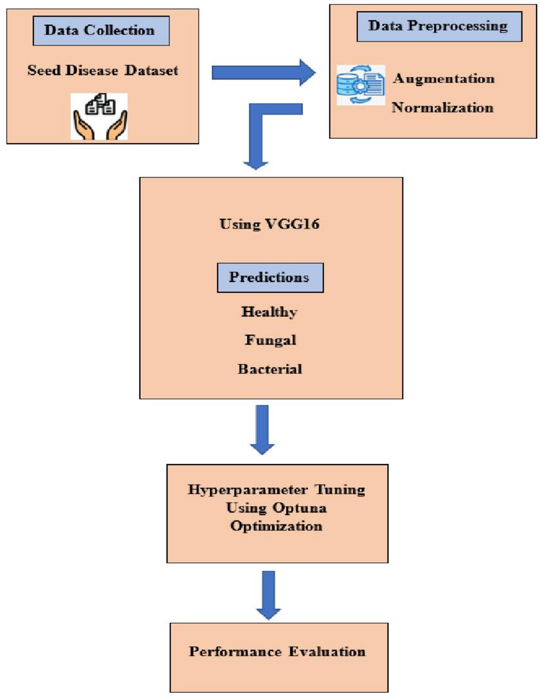

Figure 1: Seed Disease Prediction Using VGG16 with Transfer Learning and Optuna Optimization

4.2 Data Preprocessing

Data pre-processing procedure is to incorporate the seed images through rotation, flipping, and scaling to augment the training data artificially to make it more diversified for better generalization of the model. Normalization is there to scale image pixel values to the same range, usually between 0 to 1, to enable the model to converge more stably and much faster during its training. This entire process of preprocessing greatly boosts the performance and success of the developed VGG16 model in seed disease identification.

4.2.1 Augmentation

Augmentation is a process by which the training sets transform original images by a variety of manipulations so as to artificially increase the size and diversity of the training set. These manipulations are varied in nature: rotation, flipping, cropping, scaling, shifting, monochrome or brightness adjustments, etc. The end goal of augmentation is to prevent overfitting by supplying the model with many different variations of the same input, promoting its generalization ability to unseen data. Augmentation increases the training data variability so that the model can withstand and not be sensitive to minor changes in input data.

One commonly used augmentation technique is random rotation. If an image I is rotated by an angle θ, the new image I' can be represented by the following equation:

$$I' = R_\theta(I) \tag{1}$$

Where:

$R_\theta(I)$ denotes the rotation of the image I by an angle θ.

4.2.2 Normalization

Normalization is a preprocessing step for ensuring the pixel values of an image are scaled to produce a small, fixed range with either 0 to 1 or -1 to 1, whichever one needs it for the model. This ensures that the scale of all input features is the same, thus stabilizing and accelerating the training process. Normalizing input data allows the model to learn quickly and converge during training, preventing some features from entirely dominating others based on little differences in magnitude. Normalization also eliminates any unfairness towards specific characteristics in the dataset instantly on a large number of data points.

A common normalization formula used for scaling pixel values to the range [0,1] is:

$$I_{\text{normalized}} = \frac{I_{\text{original}} - \min(I)}{\max(I) - \min(I)} \tag{2}$$

Where:

I_{original} is the original pixel value of the image.

$\min(I)$ and $\max(I)$ are the minimum and maximum pixel values in the image, respectively.

$I_{\text{normalized}}$ is the normalized pixel value scaled to the range $[0,1]$.

4.3 Seed Disease Prediction Using Visual Geometry Group 16 (VGG16)

VGG16 (Visual Geometry Group 16) is the architecture for a deep convolutional neural network (CNN) developed by the Visual Geometry Group in the University of Oxford. Due to its simple yet efficient architecture, the model can suit several purposes in image classification. It consists of 16 weight layers, 13 convolutions, and 3 fully connected. VGG16 uses a small filter at 3×3 across the whole network making it very efficient and able to handle scoring in both lower-level and higher-level features. This is where length along with the aid of small filters contributes to such great accuracy in tasks such as image classification and object detection. VGG16 is frequently applied to transfer learning where it is previously trained by a large dataset; e.g. ImageNet, and later refined into specific cases like seed disease detection or probably other tasks within some areas of interest.

A common normalization formula used for scaling pixel values to the range [0,1] is:

$$I_{\text{normalized}} = \frac{I_{\text{original}} - \min(I)}{\max(I) - \min(I)} \tag{3}$$

Where:

I_{original} is the original pixel value of the image.

$\min(I)$ and $\max(I)$ are the minimum and maximum pixel values in the image, respectively.

$I_{\text{normalized}}$ is the normalized pixel value scaled to the range $[0,1]$.

4.4 Optuna Optimization

Optuna is an open-source hyperparameter optimization framework that automates the best hyperparameter selection process for machine learning models. From the standpoint of techniques, it uses bayesian optimization, tree-structured Parzen estimator (TPE), and Pruner to proficiently search through the hyperparameter space and find the best optimal. Optuna's biggest strength lies in automatic pruning; that is, Optuna can terminate unpromising trials early, thus saving computational resources and time. Flexibility is one of the major features of Optuna. This means that it can easily plug-and-play with different machine-learning libraries like TensorFlow, PyTorch, and Scikit-learn.

In other words, the optimization loop is similar to various commercially available optimization-based commercial implementations. Here, the reward function (model accuracy or loss, for example) is sampled on separate hyperparameter sets, after which the optimization process improves the search for the hyperparameters.

The optimization process can be expressed mathematically as the minimization of an objective function $f(\mathbf{x})$:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \tag{4}$$

Where:

$\mathbf{x}^*$ represents the optimal set of hyperparameters.

$f(\mathbf{x})$ is the objective function (such as model loss or accuracy) that is evaluated for different hyperparameter sets.

$\mathbf{x}$ is a vector representing the hyperparameters.

Optuna iteratively explores and evaluates different values for $\mathbf{x}$, adjusting based on past evaluations to find the set that minimizes or maximizes the objective function.

V. RESULTS AND DISCUSSION

The findings of the seed disease prediction model show a remarkable increase in accuracy and classification performance after the application of VGG16 in transfer learning and Optuna optimization. Upon fine-tuning the hyperparameters, the model attained high precision, recall and F1-scores, proving its utility for classifying seeds into healthy and diseased classes. The learning curves demonstrated the stable convergence of the model with minimal overfitting incident, thereby highlighting the robustness of the model. A comparison with competing methods further evidences the supremacy of this approach in seed disease classification.

Performance Metrics

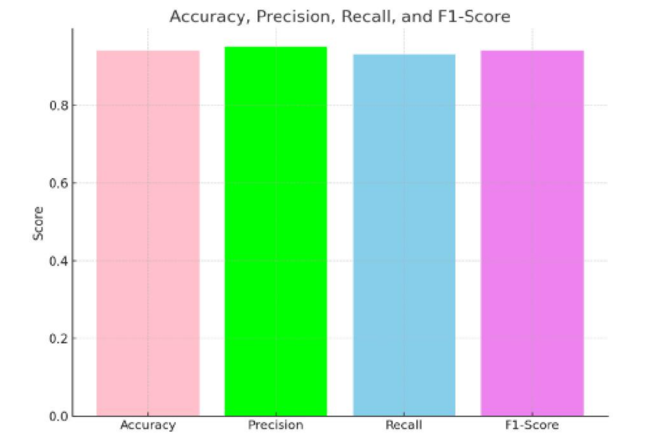

Figure 2: Performance Metrics

Figure 2 shows this bar graph metaphorically shows the multiple metrics related to your model performance: Accuracy, Precision, Recall, and F1-Score. The performance scores are very high in terms of metric, which shows that this model has a very good potential in classifying seeds against disease. In order to compare the results easily, all metrics are colour coded.

ROC Curve

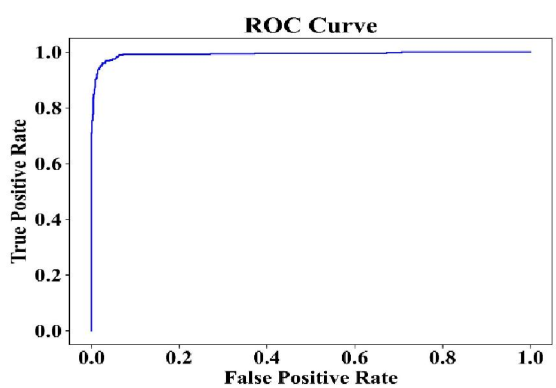

Figure 3: ROC Curve

Figure 3 Shows that the ROC curve just shows the performance of a classifier model with very high true positive rate and very low false positive rate and steeply rises up to a very high point in top-left corner which indicates the accuracy of its performance. With sharp rising the model seems to have good capability to separate positive and negative classes, which means it has a very high True Positive Rate.

DOI: 10.48175/IJARSCT-23688

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, March 2025 Impact Factor: 7.67

VI. CONCLUSION

This work shows an effort toward an efficient seed disease classification system using VGG16 with transfer learning and Optuna optimization for diagnosis purposes. The models presented impressive results in terms of metrics such as the accuracy and AUC to classify the healthy, fungal, and the bacterial seeds. Thereby enabling early detection, the application of a lesion detection method for better crop management now arises.

REFERENCES

[1]. Chug, A., Bhatia, A., Singh, A. P., & Singh, D. (2023). A novel framework for image-based plant disease detection using hybrid deep learning approach. Soft Computing, 27(18), 13613-13638.

[2]. Delnevo, G., Girau, R., Ceccarini, C., & Prandi, C. (2021). A deep learning and social iot approach for plants disease prediction toward a sustainable agriculture. IEEE Internet of Things Journal, 9(10), 7243-7250.

[3]. Haque, M. A., Marwaha, S., Deb, C. K., Nigam, S., Arora, A., Hooda, K. S., ... & Agrawal, R. C. (2022). Deep learning-based approach for identification of diseases of maize crop. Scientific reports, 12(1), 6334.

[4]. Ushadevi, G. (2020). A survey on plant disease prediction using machine learning and deep learning techniques. Inteligencia Artificial, 23(65), 136-154.

[5]. Ahmed, I., & Yadav, P. K. (2023). Plant disease detection using machine learning approaches. Expert Systems, 40(5), e13136.

[6]. Kanna, G. P., Kumar, S. J., Kumar, Y., Changela, A., Woźniak, M., Shafi, J., & Ijaz, M. F. (2023). Advanced deep learning techniques for early disease prediction in cauliflower plants. Scientific Reports, 13(1), 18475.

[7]. Nischitha, K., Vishwakarma, D., Ashwini, M. N., & Manjuraju, M. R. (2020). Crop prediction using machine learning approaches. International Journal of Engineering Research & Technology (IJERT), 9(08), 23-26.

[8]. Haridasan, A., Thomas, J., & Raj, E. D. (2023). Deep learning system for paddy plant disease detection and classification. Environmental monitoring and assessment, 195(1), 120.

[9]. Ali, A., Qadri, S., Mashwani, W. K., Brahim Belhaouari, S., Naeem, S., Rafique, S., ... & Anam, S. (2020). Machine learning approach for the classification of corn seed using hybrid features. International Journal of Food Properties, 23(1), 1110-1124.

[10]. Bharadiya, J. P., Tzenios, N. T., & Reddy, M. (2023). Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches. Journal of Engineering Research and Reports, 24(12), 29-44.

[11]. Oh, M., & Zhang, L. (2020). DeepMicro: deep representation learning for disease prediction based on microbiome data. Scientific reports, 10(1), 6026.

[12]. Zhao, T., Hu, Y., Peng, J., & Cheng, L. (2020). DeepLGP: a novel deep learning method for prioritizing lncRNA target genes. Bioinformatics, 36(16), 4466-4472.

[13]. Sangeetha, R., Logeshwaran, J., Rocher, J., & Lloret, J. (2023). An improved agro deep learning model for detection of Panama wilts disease in banana leaves. AgriEngineering, 5(2), 660-679.

[14]. Shaikh, T. A., Mir, W. A., Rasool, T., & Sofi, S. (2022). Machine learning for smart agriculture and precision farming: towards making the fields talk. Archives of Computational Methods in Engineering, 29(7), 4557-4597.

[15]. Nusinovici, S., Rim, T. H., Li, H., Yu, M., Deshmukh, M., Quek, T. C., ... & Cheng, C. Y. (2024). Application of a deep-learning marker for morbidity and mortality prediction derived from retinal photographs: a cohort development and validation study. The lancet Healthy longevity, 5(10).

[16]. Kundu, N., Rani, G., Dhaka, V. S., Gupta, K., Nayak, S. C., Verma, S., ... & Woźniak, M. (2021). IoT and interpretable machine learning based framework for disease prediction in pearl millet. Sensors, 21(16), 5386.

[17]. Kumar, D., & Kukreja, V. (2022). Deep learning in wheat diseases classification: A systematic review. Multimedia Tools and Applications, 81(7), 10143-10187.

[18]. Zhang, Y., Li, G., Muskat, B., & Law, R. (2021). Tourism demand forecasting: A decomposed deep learning approach. Journal of Travel Research, 60(5), 981-997.

ISSN: 2581-9429

Volume 5, Issue 1, March 2025

Impact Factor: 7.67

[19]. Medeiros, A. D. D., Silva, L. J. D., Ribeiro, J. P. O., Ferreira, K. C., Rosas, J. T. F., Santos, A. A., & Silva, C. B. D. (2020). Machine learning for seed quality classification: An advanced approach using merger data from FT-NIR spectroscopy and X-ray imaging. Sensors, 20(15), 4319.

[20]. Genaev, M. A., Skolotneva, E. S., Gultyaeva, E. I., Orlova, E. A., Bechtold, N. P., & Afonnikov, D. A. (2021). Image-based wheat fungi diseases identification by deep learning. Plants, 10(8), 1500.

[21]. RL, M., & Mishra, A. K. (2021). Forecasting spot prices of agricultural commodities in India: Application of deep-learning models. Intelligent Systems in Accounting, Finance and Management, 28(1), 72-83.

[22]. Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2019). Machine learning approach for prescriptive plant breeding. Scientific reports, 9(1), 17132.

[23]. Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2021). Crop yield prediction integrating genotype and weather variables using deep learning. Plos one, 16(6), e0252402.

[24]. Kumar, M., Kumar, A., & Palaparthy, V. S. (2020). Soil sensors-based prediction system for plant diseases using exploratory data analysis and machine learning. IEEE Sensors Journal, 21(16), 17455-17468.

[25]. Niazian, M., & Niedbała, G. (2020). Machine learning for plant breeding and biotechnology. *Agriculture*, *10*(10), 436

DOI: 10.48175/IJARSCT-23688

ISSN 2581-9429 IJARSCT