ISSN: 2993-8759

Iris Journal of Educational Research

DOI: 10.33552/IJER.2025.04.000595

Iris Publishers

Research Article

Revolutionizing Crop Disease Detection: Harnessing Ensemble Learning and Computer Vision for Enhanced Accuracy

Tagel Aboneh*, Abebe Rorissa and Ramasamy Srinivasagan

Software Engineering, HPC Centre of Excellence, Founder of divergent technology solution, AAST, Addis Ababa, Ethiopia

***Corresponding author**: Tagel Aboneh, Software Engineering, HPC Centre of Excellence, Founder of divergent technology solution, AAST, Addis Ababa, Ethiopia

Received Date: January 24, 2025

Published Date: February 05, 2025

Abstract

Crop diseases also remain a major threat to food security around the world. Pathogens are the main cause of for different crop diseases. However, rapid identification of the pathogens remains difficult in many developing world and countries such as Ethiopia due to a lack of technologies and infrastructure. In this research work, we proposed ensemble-based early crop disease detection model using computer vision. We have employed different pre-trained models like Inceptionv3, Resnet50, and VGG19 as a base model to conduct the experiments. More than 23,000 crop image datasets collected from different sources to train our model. The models are trained to classify the dataset into their respective categories. The selected base model is further combined to build ensemble learning model. Our main objective is to build more generable and robust machine learning model that can efficiently classify crop diseases. We have considered model explainability, scalability and drifting issues to improve the generalizability capability of the proposed model. From the experimental results, the ensemble-based learning model classified the training and validation with 96.14 %, and 96.85% accuracy respectively.

Keywords: Crops disease; Agriculture; Computer vision; Image process; Ensemble learning; Deep learning mode

Introduction

World Population growth estimated to be 9 billion by 2050 and food security issues remains a critical challenge for many nations around the globe. On the other hand, urbanization, fragmented arable land sizes, and dynamic weather and crop diseases are the main constraints to realize food security agenda. As a result, it is very difficult for smallholder farmers to feed themselves and their families let alone to supply the market according to [1]. Crop diseases are the primarily cause for yield loss and a factor for food security issue around the globe. Pathogens are the main cause of such diseases. In Ethiopia, agricultural domain experts and farmers uses a field survey and inspection method to identify plant diseases using their naked eyes [4]. The challenges encountered in the fields of data collection [6] and interpretation persist due to several factors that contribute to their time-consuming and intricate nature. Additionally, ongoing efforts to streamline processes, adopt advanced technologies, and improve collaboration among stakeholders can help to mitigate the challenges and improve the efficiency of crop diseases detection process. Research finding reveal that 20% to 40% crop diseases and weather variability dynamic changes. Plant diseases can impact various aspects of agricultural production, leading to reduced yields, economic losses, and disruptions in the food supply chain.

Thus, the sector demands an AI enabled system at least to minimize its consequence on food security. Furthermore, lack of timely and sufficient market information; the low price of the product at harvest time; weak market links between value chain actors and traders, price cheating and less negotiation power of farmers in the market; and unfair competition from illegal traders are the main marketing constraints facing wheat farmers and traders. The state of the art in agriculture employed AI enable crop diseases early detection technologies to support the agriculture domain area. Currently machine learning based solutions plays significant role to detect and classify crop diseases as early as possible. In this study, we proposed ensemble-based deep learning approaches for crop diseases classification purpose.

Ensemble learning is an approach that involves the use of multiple machines learning models, combining their results as a cohesive committee of decision makers [1-3]. The underlying principle is that the collective decision of the committee tends to exhibit superior overall accuracy compared to any individual member. Ensemble Deep Learning is a cutting-edge technique in machine learning that combines the strengths of multiple deep learning models to achieve superior performance compared to any individual model. It's like forming a team of experts, each with unique strengths, to tackle a complex problem. Lior Rokach [4-6] defines ensemble methodology as the construction of a predictive model through the integration of multiple based models. The current study main focus is to gain competitive advantages of Ensemble Deep Learning to handle complex data sources:

- Improved accuracy and generalization capability of machine learning models. The combined predictions are often more accurate than any individual model.
- Reduced model variance due to data quality and feature engineering challenges. This would help to mitigate the over-fitting issue common in deep learning, making the model more reliable on unseen data.
- The combined features of different based enhance robustness of the proposed model. Therefore, ensemble model is less sensitive to noise and outliers in the data.

In this study, our main focus was to support the agricultural sector by automating the existing wheat disease identification process. The data processing time and interpretation dependency on domain experts were the main limitations in the domain area. Thus, we implemented various deep learning approaches to assess the best fitting framework(s) for the proposed study. The majority of the utilized models were promising in terms of classifying crop diseases into their respective target classes.

The proposed model is obviously advantageous, and the contributions made in this paper are summarized as follows:

- An ensemble-based deep learning approaches are used to design generalizable model to improve the limitations of manual early plant disease identification challenges.
- The proposed approaches have made significant contributions to various fields by leveraging the meta-data features of multiple base models. By combining diverse deep learning architectures or training strategies, these approaches enhance model robustness, generalization, and predictive performance. Their ability to capture complementary patterns and mitigate individual model biases has been pivotal in achieving state-of-the-art results across tasks such as image classification, natural language processing, and medical diagnosis.
- A more general crop disease identification deep learning model is created, which can be applied to other crop disease image disease datasets and, at the same time, provides a reference for other disease control and management research work.
- Compared with deep learning models, this model achieves high accuracy in wheat disease image classification.

Materials and Methods

An experimental research approach has been used to implement ensemble-based deep learning framework. In this study, we have made an attempt to understand important variables and identified their effects on the performance of the proposed model. In this regard, detailed experiment analysis has been made to improve plant diseases classification accuracy. Therefore, in the subsection we have discussed the experiment setting and its results respectively [40 –43].

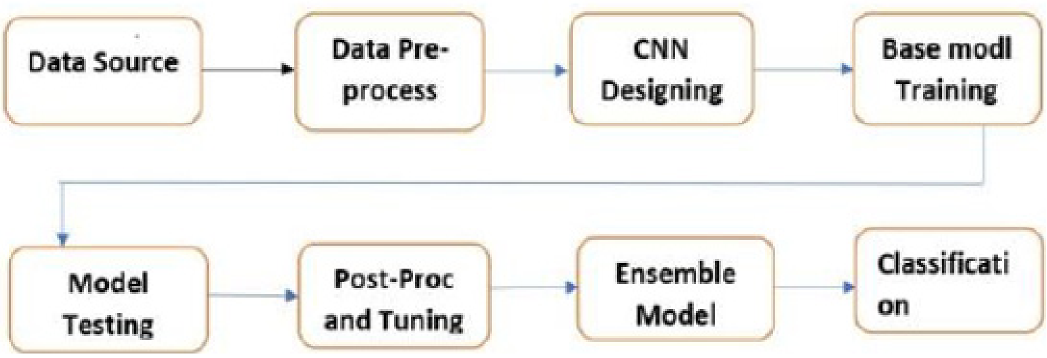

Figure 1: Proposed Ensemble Learning Approach System Architecture.

Datasets

To implement the experiments, we have collected more than 25 thousand image datasets from Kaggle repository and around 1500 wheat image (RGB) datasets were collected from Bishoftu agriculture research Institute. We have utilized more than 25 thousand image data are classified into four major categories namely corn, wheat, potatoes and tomato. These categories are further classified into 20 classes namely tomato diseases such as yellow-leaf-curv-virus, Mosaic-virus, target-spot, spider-mites, septoria-leaf-spot, leaf-mold, late-blight, and healthy-tomato. Corn disease such as north-leaf-blight, health-corn, common-rust, cercosporaleaf-spot. Potato diseases such as late-blight, healthy, early-blight, and Wheat disease such as stem-rust, yellow-rust, leaf-rust. The data-sets are well labelled and divided into three categories namely training, testing and validation. We have selected the crop to ensemble learning model on the basis its economic significance at national level. Figure 1 presents a general architecture of the proposed model, which includes its end-to-end implementation and explicitly defines the expected data processing requirements at each stage.

Feature selection and data understanding is very important step in the modelling process, especially for high-dimensional data sources. The removal of the least relevant features often improves generalization performance and minimizes over-fitting issue. During the data preparation stage, we perform image normalization, formatting, removal of poor-quality images, re-scaling or image resizing, and cropping of irrelevant parts of the image. Re-scaling pixel intensities values ranging from 0 to 255. Re-scaling these values to a standard range (between 0-1) improves numerical stability and training efficiency. Furthermore, we transformed the data by rescaling and setting the dimensions of the images at 224 by 224 and channel=3 to standardize the data set. We have used well annotated crop image data to train our model and the data sources are organized into training, testing and validation data-set. Figure 2 illustrates the sample input dataset and its respective deep learning model performance in classifying the input dataset.

To successfully implement the proposed model system, customizing a pretrained model has been done to leverage the knowledge gained by a model trained on a large dataset to perform crop diseases classification purpose for the above-mentioned dataset. The first step was selecting pretrained models such as ResNet50, Inception V3, denseNet121, and VGG19 to conduct the experiment. Figure 3 below summarized the experiment output of pretrained models on the given data.

From the experimental results, Inceptionv3, Resnet50, VGG16, and VGG19 produced 95.65%, 81.57%, 96.48%, and 99.38% classification accuracies, respectively. Based on the model's classification performance, we selected VGG19 to combine with base models. Computational efficient, diversity, scalability, explainability, and drifting issues has been considered to further ensemble the base models. The following architecture has been used to designed ensemble based deep learning architecture.

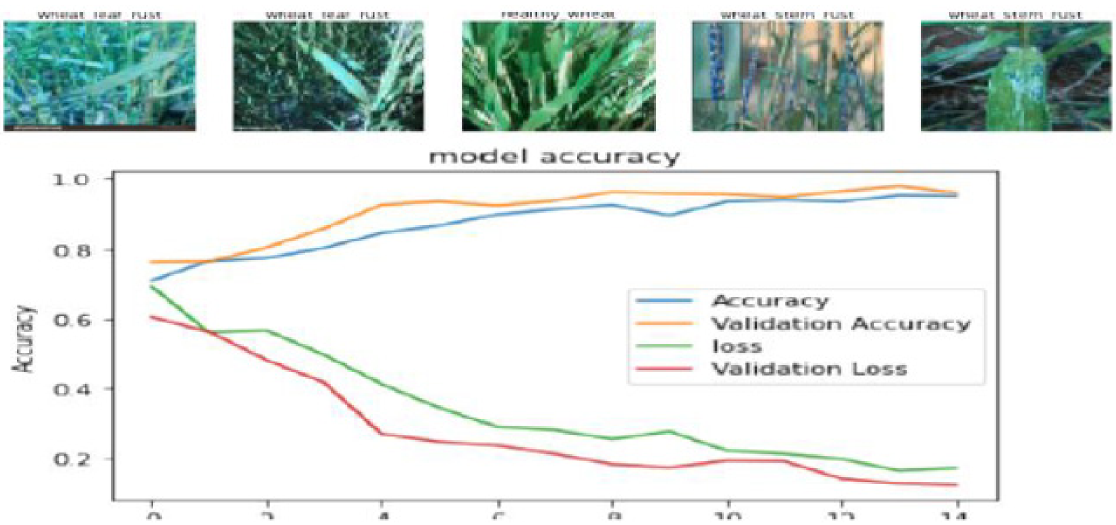

Figure 2: Sample input crop dataset and proposed model output.

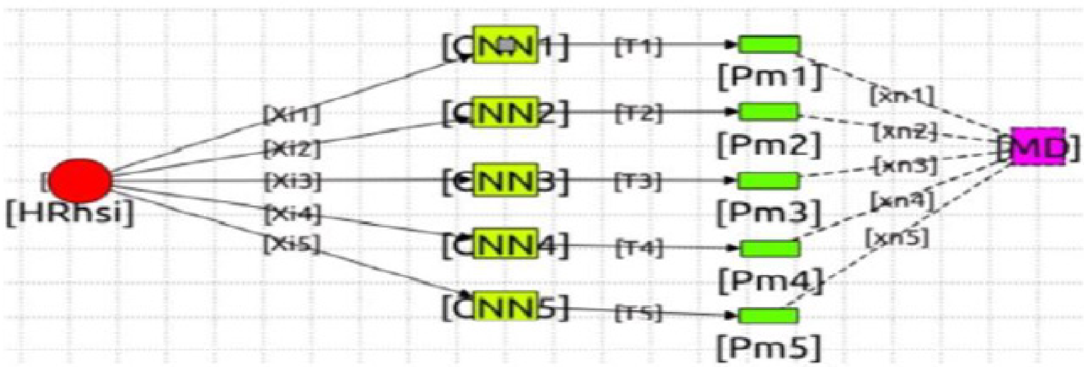

Figure 3: Ensemble Based Deep Learning Model.

Where training datasets were feed to individual base model as input data to perform model training task. In this regard, about five different CNN architectures have been used to train Xi input data. The performance of each base classifier in the training data set has been evaluated using the data set T = [t1...t5] to evaluate the overall performance of each base classifier. What are specific characteristics of individual base model has before built the ensemble learning model.

Result and Discussion

In this study, we implemented an ensemble-based deep learning approach by combining three different pre-trained models. The experimental results demonstrated that the ensemble method effectively leverages the strengths of each base model to handle complex image datasets. However, computational challenges remain a significant concern in deep learning model training, requiring adequate time and memory resources for parameter execution. During training, our primary objective was to minimize categorical cross-entropy loss. By fine-tuning model parameters (α), the model aimed to maximize the probability of the true class while minimizing the probabilities of incorrect classes, ultimately reducing the overall loss. To address the computational demands, we utilized GPU infrastructure, specifically NVIDIA Tesla T4 and NVIDIA Tesla K80 accelerators, along with a Core i7 processor and 1 TB hard disk for model training.

Additionally, we conducted a comparative analysis of VGG19 against other state-of-the-art models in the field. The results suggest that further optimization techniques could enhance the model's classification accuracy. Nevertheless, certain limitations were identified that impacted the performance of the deep learning models used:

- The quality of image datasets significantly influenced model performance. Applying various preprocessing and post-processing techniques can enhance the features extracted from the data.
- Factors such as data format inconsistencies, image rotations, size variations, dark objects in the background, and the use of different camera types adversely affected the model's accuracy. Standardizing the data and employing high-quality cameras could mitigate these bottlenecks.

Table 1: Comparison of the proposed model accuracy with other models.

Author	Crop Diseases	Model	Training	Valid
Arun Pandian J [6]	Different crops	VGG16	87.03%	-
Helal Sheikh [59]	Maize' and 'Corn'	CNN	98.29%	99.29%
Divyansh Tiwari [60]	Potato (plant village)	VGG19	97.8%	97.8%
Xihai Zhang [61]	maize leaves	GoogLeNet	89.6%	98.9%
Anshuman Singh [12]	Wheat disease	VGG19	96.6%	91.3%
Ashok [40]	Tomato leaf	CNN		98.12%
Huiqun H [15]	Tomato disease	DXception		97.10%
Mikhail G [30]	Wheat rust	Densenet		98%
Sholihati R [7]	Potato disease	VGG19		91.%
Ours model	Wheat disease	VGG19	99.38%	98.23%

Ensemble learning entails combining multiple models, such as through averaging or voting, surpassing the performance of individual models. Despite the advantages of deep learning models, such as deep architectures, they face challenges like vanishing/exploding gradients and degradation problems, hindering optimal performance. Theoretical and empirical justifications demonstrate that ensemble approaches exhibit superior generalization compared to individual models. Ensemble learning has emerged as a potential strategy for enhancing the performance of deep learning model.

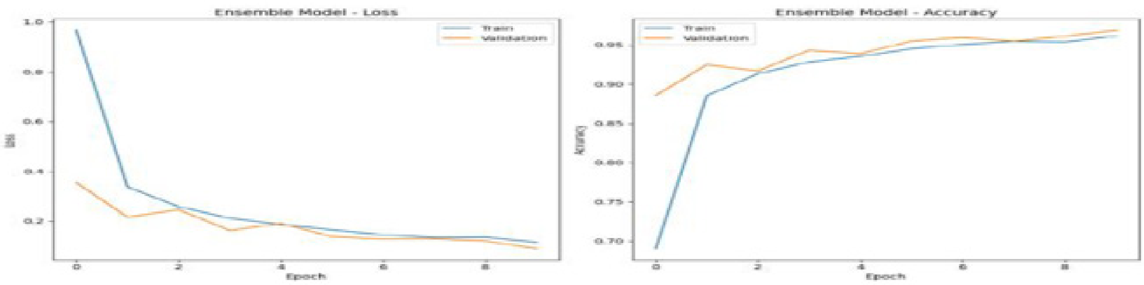

Figure 4: Ensemble Based Deep Learning Model Accuracy on training dataset.

The weighted average of model predictions is a statistical technique used to combine multiple predictions from different models or sources, taking into account the reliability or importance of each prediction. This is particularly useful in ensemble modelling, where multiple models are combined to improve overall predictive performance. The weighted average of model predictions can be calculated using the formula:

$$\text{Weighted Average} = \frac{\sum_{i=1}^{n} w_i \cdot y_i}{\sum_{i=1}^{n} w_i}$$

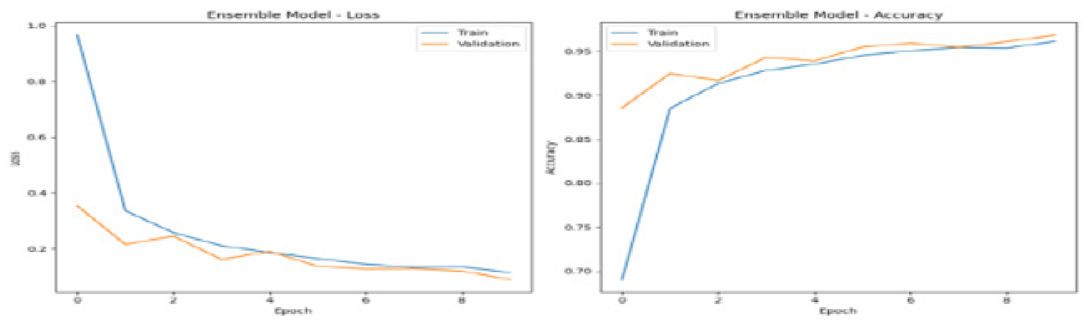

Figure 5: Ensemble Based Deep Learning Model Loss function.

The experiment results on Figure 3 and 4 showed that, ensemble based deep learning model scored optimal classification performance on crop diseases. As we mentioned in the methodology section, four different base model are combined together to create the proposed ensemble approach. By using different optimization techniques, it is possible to further improve the classification of the proposed ensemble learning model.

Conclusion and Recommendation

Currently, rapid population growth, exponential decline of arable land, and dynamic environmental change are the main challenge for many developing countries. Similarly, plant disease early detection and identification challenging and time-consuming process. Pathogens are the main cause of for different crop diseases. Crop diseases also remain a major threat to food security around the world. However, rapid identification of the pathogens remains difficult in many developing world and countries such as Ethiopia due to a lack of technologies and infrastructure. In this study, we have developed deep learning based system to detect crop diseases as early as possible.

Image-based data processing provides detailed features to discriminate one object from the other better than other data types. Implementing deep learning frameworks is promising for extracting relevant features from image data. Agriculture is one of the hot research areas that demands the application of state-of-the art technology for automating the sector. To conduct the experiment, we have collected crop data from Bishoftu Agriculture Research Institute in case of wheat diseases and open source repository. The type of crops is selected on the basis of their significance for national economy.

Then, we utilized different pre-trained deep learning models such as InceptionnetV3, MobileNet, ViT, VGG19, ResNet to build the model. The experiment results, reveal that most of the deep learning model scored a best crop diseases classification result with slight difference. We employed GPU infrastructure to handle computation complexity, as a result the deep learning model are compiled in short time.

Then, we further extended our experiment to build ensemble learning model using the above pre-trained deep learning model as a base learner. The main aim is to build more generalizable deep learning model that can handle a complex image data with efficient performance. Ensemble learning approach is one of the best solutions to handle the domain specific related limitation of different deep learning models.

Therefore, to build robust crop diseases classification and detection deep learning model, data is acquisition is issue are very important. We recommend inspired research to work on building data repository system that enable further study in the domain area. Similarly, data quality challenges are common factor that affect the performance your model. We encourage different stakeholder to contribute their role by providing more advanced image acquisition technology for research institutions. Finally, we would like to motivate interested researcher to further explore the area to bring a best possible solutions for our farmer. Overall contribution is to improve crop yield production which enable us to assure food security issue at national and international level.

Acknowledgment

None.

Conflict of Interest

No conflict of interest.

References

1. Tadesse W, Bishaw Z, Assefa S (2018) Wheat production and breeding in sub-Saharan Africa: Challenges and opportunities in the face of climate change. International Journal of Climate Change Strategies and Management.
2. Bachewe FN, Berhane G, Minten B, Taffesse AS (2018) Agricultural transformation in Africa? assessing the evidence in Ethiopia. World Development 105: 286-298.
3. Sara L, Jakob S, Saumya S (2014) The state of food and agriculture 2014. FAO: Rome, Italy.
4. Anteneh A, Asrat D (2020) Wheat production and marketing in Ethiopia: Review study. Cogent Food & Agriculture 6(1): 1778893.
5. Kedir U (2017) The effect of climate change on yield and quality of wheat in Ethiopia: A review. Journal of Environment and Earth Science 7(12).
6. Rashid S, Getnet K, Lemma S (2019) Maize value chain potential in Ethiopia: Constraints and opportunities for enhancing the system. Gates Open Res 3(354): 354.
7. Negassa A, Shiferaw B, Koo J, Sonder K, Smale M, et al. (2013) The potential for wheat production in Africa: Analysis of biophysical suitability and economic profitability.
8. Jiang P, Chen Y, Liu B, He D, Liang C (2019) Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. IEEE Access 7: 59069-59080.
9. Kawatra M, Agarwal S, Kapur R (2020) Leaf disease detection using neural network hybrid models. In: 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), IEEE pp. 225-230.
10. Mukti IZ, Biswas D (2019) Transfer learning-based plant diseases detection using resnet50. In: 2019 4th International Conference on Electrical Information and Communication Technology (EICT), IEEE pp. 1-6.
11. Pham TN, Van Tran L, Dao SVT (2020) Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection. IEEE Access 8: 189960-189973.
12. Pirttioja N, Carter TR, Fronzek S, Bindi M, Hoffmann H, et al. (2015) Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. Climate Research 65: 87-105.
13. Alrayyes WH (2018) Nutritional and health benefits enhancement of wheat-based food products using chickpea and distiller's dried grains.
14. Arya S, Singh R (2019) A comparative study of cnn and alexnet for detection of disease in potato and mango leaf. In: 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), IEEE 1: 1-6.
15. Tidiane Sall A, Chiari T, Legesse W, Seid-Ahmed K, Ortiz R, et al. (2019) Durum wheat (Triticum durum desf.): Origin, cultivation and potential expansion in sub-Saharan Africa. Agronomy 9(5): 263.
16. Durmu H, Gune EO, Kırcı M (2017) Disease detection on the leaves of the tomato plants by using deep learning. In: 2017 6th International Conference on Agro-geoinformatics, IEEE pp. 1-5.
17. Hasan MZ, Ahamed MS, Rakshit A, Hasan KZ (2019) Recognition of jute diseases by leaf image classification using convolutional neural network. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE pp. 1-5.
18. Chakravarthy AS, Raman S (2020) Early blight identification in tomato leaves using deep learning. In: 2020 International Conference on Contemporary Computing and Applications (IC3A), IEEE pp. 154-158.
19. Mengistu AD, Alemayehu DM, Mengistu SG (2016) Ethiopian coffee plant diseases recognition based on imaging and machine learning techniques. International Journal of Database Theory and Application 9(4): 79-88.
20. Sheikh MH, Mim TT, Reza MS, Rabby ASA, Hossain SA (2019) Detection of maize and peach leaf diseases using image processing. (2019) 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE pp. 1-7.