

A Model for Smart Farming through Cloud-Based Plant Leaf Disease Detection using Supervised Machine Learning

Levis Makurumure¹, F. D Mukoko²

¹Computer Science Department, Harare Institute of Technology
levismak[at]gmail.com

²Computer Science Department, Harare Institute of Technology
fnfungied82[at]gmail.com

Abstract: *Plant diseases reduce both quality and quantity, preventing a high yield. Leaf diseases must be identified at all plant stages so as to preserve the expected yield from the plants. However, agriculture specialists are not often available to identify plant diseases thereby delaying the plant disease management process slow and costly. With smart agriculture, digital image processing and deep learning have made significant breakthroughs in plant disease management. A real-time disease prediction model will cover the time gap caused by human involvement, offering early diagnosis and treatments to plant leaf diseases and greater yields. This study implemented CNN to detect and classify diseases in tomato plants. The results of the study showed that the model loss reduced from 3.0 at the beginning to 0.5 at epoch 1 and remained falling until the final epoch, when training and validation had the least loss. On the other hand, the accuracy of the model also increased as the training progressed registering a higher accuracy of 90% for both the training and validation scores. The classification result shows that the model accurately detected Tomato Mosaic Virus on infected leaves and healthy plants. Predicting late blight was 92% accurate. Infected eaves scored 100%, while late blight leaves got 91%. The model detected health leaves with a 99% accuracy. Late blight was predicted with 91% and tomato mosaic virus with 100%. The model predicted all 1000 sample leaves classes with 96% accuracy.*

Keywords: Cloud computing, Tomato plant leaf disease, Machine learning.

1. Introduction

The advent of cloud computing has fundamentally altered how individuals use technology in their daily lives, especially organizations that seek out more powerful computers as well as mobile applications that were previously only available through traditional computing methods [1]. This study capitalises the services of cloud computing in enabling smart farming through provision of platforms and services for enhancing smart farming solutions among which includes plant disease detection. Plant diseases are the greatest obstacle towards achieving a high yield among other factors creating a reduction in both quality and quantity [2]. Identification of plant leaf diseases is critical at all stages of the plant. With the movement towards smart agriculture, technologies such as digital image processing, deep learning has made considerable breakthrough and more effectively than the traditional methods [3]. Using deep learning in plant disease identification has made it interesting and productive area of study.

2. Background to the Study

The majority world over has relied on agriculture as their main source of income, as seen by the rising levels of agriculture commercialization [4]. However, the proliferation of diseases, particularly in horticulture-based agriculture, has greatly adversely impacted the success of agriculture, necessitating the detection of plant diseases [5]. Plant diseases are a factor in production losses; however, they can be reduced with constant observation. Plant

diseases on horticultural crops have been highlighted as the most difficult problems affecting agricultural output in Europe. The most prevalent infectious diseases in the area have been rots, fruit spots, wilts, and leaf spots/blights, which have reduced output and decreased the quality of produce.

In Asia, Fungi make up the majority of pathogens, followed by bacterial and viral diseases, with different symptoms depending on the host [6]. In any case, diseases that affect tomato plants have decreased the predicted output of tomatoes by 37% in 2019. Similar to this, plant diseases that survive despite the use of control techniques, primarily in the form of pesticides, have drastically reduced the seasonally available tomato yield in Asia by a margin of 20% to 30% [3].

Africa has had a high incidence of tomato diseases recently. Concern has been raised about their effects on the productivity of small farmers, especially in Cameroon and Sudan, two of Africa's major producers of tomatoes [7]. An increase in trans-border plant pests and diseases in South Africa, according to a study by the Food and Agriculture Organization (FAO), is mostly attributable to diminished resistance brought on by decades of agricultural intensification, as well as globalization, trade, and climate variability. Global epidemics are caused by exotic diseases and pests that have been introduced into other countries. The FAO indicates that epidemics result in significant output losses, putting farmers' livelihoods as well as the nation's food and nutritional security at jeopardy [8].

Volume 11 Issue 12, December 2022

www.ijsr.net

Licensed Under Creative Commons Attribution CC BY

While Zimbabwe does not produce a lot of tomatoes on a global scale, small-scale production has been limited in some places, especially those with high temperatures by diseases and pests [9]. Numerous diseases and pests affect tomatoes, which results in leaf defoliation and the development of black patches on the fruit. Older leaves are first affected by the diseases, which then spread as the number of lesions rises, leaves drop, and fruit is exposed to sunscald risks [10]. The fruit contracts an infection at the stem or calyx end, expanding to show concentric rings on lesions, turning velvet black, and experiencing substantial internal flesh dryness. This will ultimately result in a reduction of production of tomatoes of 20% to 30%, with some extreme cases exceeding 50% [11].

It has been difficult to maintain tomato output in the face of these diseases, despite the fact that several manual processes were put in place to contain the diseases. Manual plant disease monitoring takes time and is prone to human mistake [12]. Most people consider these techniques to be cultural practices. The removal of old plant detritus from the field prior to the planting of new crops is one of the cultural practices utilized. Fungal spores can help infected plant debris and weeds associated to tomatoes including horse nettle, ground cherry, and night shade endure the winter [13]. The disease cycle is started when overwintering fungal spores from sick tomatoes or weed debris in the soil attach to newly planted tomatoes the following planting season.

The limitations of constant human monitoring can be overcome while simultaneously enabling early detection of plant diseases utilizing machine learning techniques. These days, Artificial and Deep Neural Networks (ANNs and DLs), as well as Support Vector Machines, are the models most frequently utilized in agriculture (SVMs) [14]. ANNs models are extensively used for regression and classification tasks, indicating their potential for use in managing crops, predicting yields, controlling weeds, diagnosing diseases, and identifying distinctive traits. The goal of the study is to create a machine learning model that will classify tomato plant diseases using picture recognition [15]. The process of creating software that can learn from data and/or experience without having to be explicitly coded is known as machine learning. These machine learning models learn the characteristics of individuals (covariates), which are then utilized to predict the future using the mapping function that they have learned [8].

The study creates a deep learning model for identifying tomato plant diseases by looking at images of tomato leaves. Producers might use this technology to classify diseases affecting tomato growing by simply shooting sick leaves rather than shelling out for pricey expert research. Photographs of tomato plants collected by farmers will be used to create synthetic image datasets that will be added to the dataset for training with the suggested technique. In order to detect tomato disease from leaf photos, the model will then be trained using tomato plant leaf images.

3. Problem Statement

Current plant disease identification is largely relying on Agritex officers who often are not available in time to assess

plants. Farmers have not been able to realize their full potential and contribute to the national GDP. Such continuous failure will render the agriculture sector uncompetitive at the national level. A real-time disease prediction model will cover the gap in time caused by human intervention providing early diagnosis and providing solutions to the leaf diseases on plants and ultimately providing better yields.

4. Literature Review

This section presents a high-level overview of methods for machine learning-based deep learning plant disease diagnosis. Deep learning detection techniques may be seen as an application of pertinent classification networks in the realm of agriculture since the ultimate goal of this research is to use machine learning to diagnose plant ailments [16]. The network can be divided into classification, detection, and segmentation networks based on the various network topologies, as shown in Figure 1.

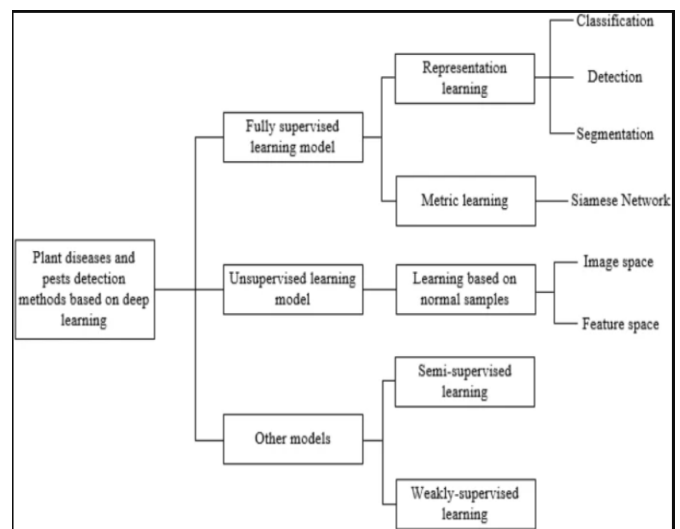


Figure 1: An illustration of Deep learning-based framework for detecting plant disease [17]

a) Classification network

After receiving a test image, the classification network evaluates it and then returns a label that categorizes it. Based on the function it serves, the classification technique may be categorized into three groups: using the network to extract features, direct classification, and sickness detection.

b) Using a network to extract features

In some early experiments with deep learning-based plant disease and pest classification algorithms, CNN's improved feature extraction capabilities were utilized, and the results were integrated with conventional classifiers. In order to develop image recognition capabilities, a CNN network that has already been pre-trained receives images. A typical machine learning classifier is then used to categorize the collected features. Yalcin [18] created a CNN architecture for extracting image features, and SVM classifiers with different kernels and feature descriptors like LBP were used to show the method's efficacy. The meta-architecture was then discussed by Fuentes [19], who proposed a CNN-based meta-architecture with numerous feature extractors that

classified healthy and sick plants into the proper groups based on input of images of healthy and sick plants.

c) Related studies

As early as the 1990s, several researchers used information technology to recognize and diagnose plant diseases. Zhang [20] used genetic algorithms to construct identification criteria based on spectrum reflection characteristics and shape characteristics in order to diagnose the disorder in 1999. However, due to a lack of usage of color and texture information from the diseases, the identification impact was not as high as anticipated. Bodria et al. conducted multispectral identification for wheat afflicted by several fungi in four bands in 2002 using a 200 W hernia light source (radiation wavelength: 360 nm–430 nm). (The wavelengths of the light are 450 nm, 550 nm, 690 nm, and 740 nm).

The study of Wang. [21] used neural networks to identify cucumber powdery mildew, downy mildew, and leaf miner-damaged leaves. In order to estimate the prevalence of diseases and insect pests in apple plants, Liu [22] used back propagation neural networks to analyse data from the previous 11 years. The study of [17] optimized the structure and parameters of neural networks using genetic algorithms. Based on the perspective of geometry idea of the scene, Chang. [23] employed Hough transform and Gabor filtering to identify weeds between rows of crops in the field and solved the weed identification problem under various perspectives and conditions.

In the past, deep convolutional neural networks have proven to be successful at identifying plant diseases. Deep convolutional neural networks are unable to identify the locations of ill spots when there are multiple diseases present in a tomato image; they can only identify the sorts of diseases present. In this study, two object detection architectures are combined with deep convolutional neural networks to address these issues. These techniques can be used to identify not only the types of tomato diseases but also the locations and shapes of affected patches.

5. Methodology

The study employed Convolutional Neural Networks (CNN) to predict survivability among cancer patients. The

a) Dataset

The visual appearance of each disease in the dataset used is as shown in Figure 2.

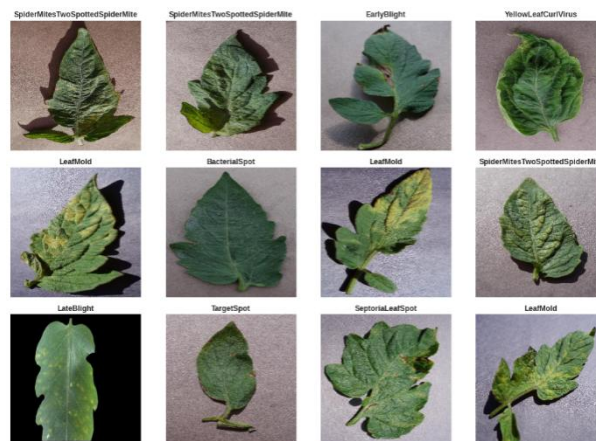


Figure 2: Disease categories

The visualisation show that the infected leaves have varying signs from the image visuals. Ten diseases categories are presented with the corresponding impact they have on the infected leaves

b) Machine Learning Algorithm Used

The study implemented CNN for the image classification. The model is as shown in Figure 3.

```

model = Sequential()
model.add(Conv2D(filters=32, kernel_size=(3,3), input_shape = image_shape, activation = 'relu'))
model.add(MaxPool2D(pool_size=(2,2)))
model.add(Conv2D(filters=64, kernel_size=(3,3), input_shape = image_shape, activation = 'relu'))
model.add(MaxPool2D(pool_size=(2,2)))
model.add(Dropout(0.3))
model.add(Conv2D(filters=128, kernel_size=(3,3), input_shape = image_shape, activation = 'relu'))
model.add(MaxPool2D(pool_size=(2,2)))
model.add(Dropout(0.4))
model.add(Flatten())
model.add(Dense(128, activation = 'relu'))
model.add(Dropout(0.5))
model.add(Dense(10, activation = 'softmax'))
model.compile(loss='sparse_categorical_crossentropy',
              optimizer = 'adam',
              metrics = ['accuracy'])

```

Figure 3: Models used for comparison

Evaluating the model was done using the classification report, model loss and confusion matrix.

6. Results of Algorithm Performance

The model loss is a measure of how horrendously erroneous the model's prediction was for a specific case [17]. The goal of model training is to identify a collection of weights and biases with a low average loss across all cases. As a result, when a model is being trained, the optimal loss occurs when the loss approaches zero more frequently than the floor limit deviates from it. Loss analysis for the model is shown in Figure 4.

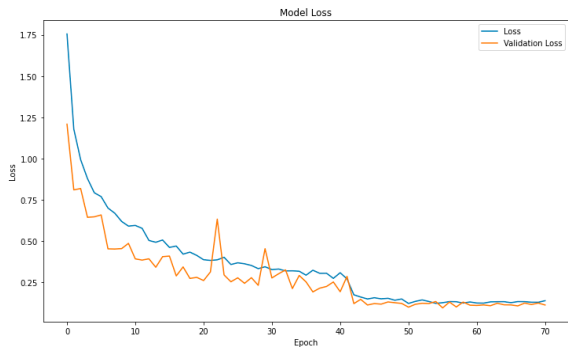


Figure 4: History of model loss on both train and validation

During model training, model and ideal loss trends were similar. During training, loss approached zero. During training, the loss fell from 3.0 at the beginning to 0.5 at epoch 1 and kept reducing until the final epoch, when both the training and validation tasks had the least loss. The model's failure shows that it was adequately trained and could understand data patterns, proving its value and efficiency.

Figure 5 shows model training accuracy on training and validation sets.

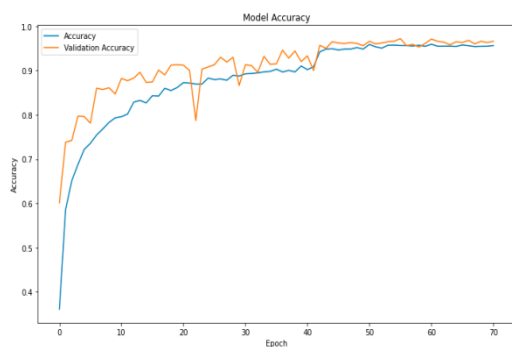


Figure 6: Model accuracy

Model accuracy shows that the model improved learning efficiency from epoch 2, when prediction accuracy increased sharply. The model's prediction accuracy rose, as it should as the prediction process advances. The model's validation accuracy also scored above 0.9 after the prediction process. The model's accuracy shows that its predictions are accurate.

The best state of the saved model was loaded for detecting and predicting the diseases on the loaded images which produced the following outcomes. Figure 6 depicts the prediction results for testing the model.

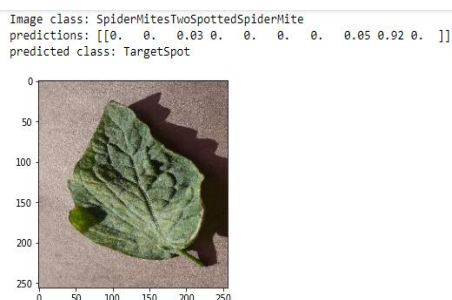


Figure 6: First predicted image

The first image loaded had the Spider Mites Two Spotted Spider Mite disease, which corresponded to the initial list of disease classifications. But the model misidentified the disease's classification as TargetSpot. Inaccurate predictions may be caused by the model's inability to successfully learn the characteristics of the pixels belonging to the erroneous disease category, which it failed to recognise and accurately categorise, according to a study by Dube [7].

A second image loaded for prediction belonged to the category of LateBlight diseases. The outcome of the prediction is shown in Figure 7.



Figure 7: Second predicted image disease category

The real class of the disease category is the same as the projected disease category, as shown in Figure 4.9, demonstrating that the model was successful in properly predicting the disease on the loaded leaf image. This demonstrates that the categorization model correctly predicted a positive outcome.

A third image was loaded, showing a leaf affected by the Yellow Leaf Curl Virus. Figure 8 depicts the prediction's result.

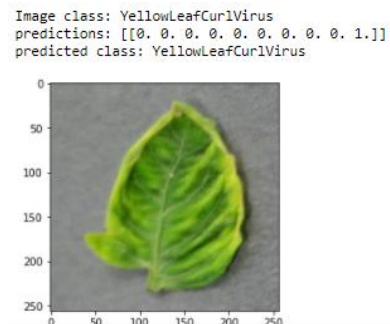


Figure 8: Third predicted image disease category

The result of the third prediction also suggests that the classification given to the leaf illness corresponds to the category of the disease that was actually identified on the image during the prediction procedure. This proves that the model was successful in accurately predicting the illness on the picture leaf.

Evaluating the performance of the model was done using the classification report which shows the model's precision, recall, F1 score and accuracy. The confusion matrix was also used to assess the effectiveness of the model performance. The matrix is a computation of the error scores of the model which reflect its false predictions counted as false positive and false negatives.

Figure 9 depicts the classification report of the model.

```

=====Classification Report=====

```

	precision	recall	f1-score	support
Tomato__Bacterial_spot	0.94	0.98	0.96	100
Tomato__Early_blight	0.93	0.95	0.94	100
Tomato__Late_blight	0.92	0.91	0.91	100
Tomato__Leaf_Mold	0.95	0.96	0.96	100
Tomato__Septoria_leaf_spot	0.96	0.95	0.95	100
Tomato__Spider_mites_Two-spotted_spider_mite	0.96	0.96	0.96	100
Tomato__Target_Spot	0.98	0.97	0.97	100
Tomato__Tomato_Yellow_Leaf_Curl_Virus	0.98	0.94	0.96	100
Tomato__Tomato_mosaic_virus	0.99	1.00	1.00	100
Tomato__healthy	1.00	0.99	0.99	100
accuracy			0.96	1000
macro avg	0.96	0.96	0.96	1000
weighted avg	0.96	0.96	0.96	1000

Figure 9: Classification report

The classification report demonstrates that the model detected Tomato Mosaic Virus with 99% accuracy on infected leaves and 100% accuracy on healthy plants. Late blight prediction scored 92% for precision. Detecting Tomato mosaic virus-infected eaves scored 100%, whereas late blight-infected leaves scored 91%. 99% of healthy leaves were identified. Tomato mosaic virus was predicted with a 100% F1 score, while late blight was predicted with 91%. The model correctly predicted the classes of all 1000 sample leaves with an average accuracy of 96%.

Figure 10 depicts the confusion matrix of the model.

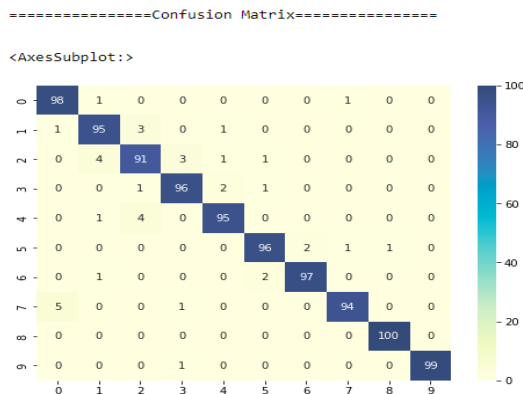


Figure 10: Confusion matrix

The confusion matrix shows the true and false predictions done by the model during the testing process. The model made few errors in predicting disease on category 8 while the largest number of error false predictions were made on disease category 2 which recorded 9% false predictions. However, the overall performance of the model show that the model successfully predicted the tomato diseases with high scores as measured using the metrics in its classification report.

The overall performance of the model is summarised in Figure 11.

Overall Accuracy: 0.961
 Overall Precision: 0.9612026825985766
 Overall Recall: 0.961

Figure 11: Overall performance of the model

The overall performance of the model shows that the model predicted tomato diseases with an overall accuracy of 96%,

precision of 96% and Recall score of 96%. The model performed better as compared to the results of the study of [17] and [11] whose studies attained highest scores of 93% and 91% respectively.

7. Conclusion

Ten diseases in the dataset are equally frequent in the area. Spot, early blight, healthy, late blight, leaf mould, mosaic virus, septorial leaf spot, twospotted spidermite, targetspot, and Yellow Leaf Curl Virus are prevalent. The CNN model's performance was improved by using the pre-trained MobileNetV2 model for feature extraction and data augmentation. Testing showed the model's ability to learn from supervised data. Using the model, three leaves from the testing set were tested.

The analysis properly predicted and categorised three of the four plant leaves that were sampled and appropriately diagnosed as having illnesses. However, the model gave a wrong prediction on one of the sampled testing leaves.

References

- [1] M. Butt, "A review of machine learning algorithms for cloud computing security," *Electronics*, vol. 9, no. 9, p. 1379, 2020.
- [2] T. Alam, "Cloud Computing and its role in the Information Technology," *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, vol. 1, no. 2, pp. 108-115, 2021.
- [3] S. & G. B. Namani, "Smart agriculture based on IoT and cloud computing. In 2020 3rd International Conference on Information and Computer Technologies (ICICT)," *IEEE*, vol. 1, no. 2, pp. pp. 553-556), 2020.
- [4] J. M. Balogh, "The role of agriculture in climate change: a global perspective," *IEEE*, vol. 2, no. 3, p. 670216917, 2020.
- [5] T. Alam, "Cloud Computing and its role in the Information Technology," *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, vol. 3, no. 2, pp. 108-115., 2021.
- [6] S. C. P. & S. R. Yule, "Detection of Tomato yellow leaf curl Thailand virus transmitted by Bemisia tabaci Asia I in tomato and pepper," *Phytoparasitica*, vol. 47, no. 1, pp. 143-153, 2019.
- [7] M. Dube, "Tomato breeding in Sub-Saharan Africa-Challenges and opportunities: A review," *African Crop Science Journal*, vol. 28, no. 1, pp. 131-140, 2020.
- [8] T. Aliyu, "Journal of Bangladesh Agricultural University," *Bangladesh Agril Univ*, vol. 18, no. 2, pp. 266-271, 2020.
- [9] A. Karavina, "Genome Sequence of a Tomato-Infecting Tomato Mosaic Virus Isolate from Zimbabwe," *Genome Announcements*, vol. 6, no. 11, pp. e01457-17, 2018.
- [10] M. Dube, "Tomato breeding in Sub-Saharan Africa-Challenges and opportunities: A review," *African Crop Science Journal*, vol. 28, no. 1, pp. 131-140, 2020.
- [11] R. Lett, "Tomato leaf curl Kunene virus: a novel tomato-infecting monopartite begomovirus from

Namibia,” *Archives of Virology*, vol. 165, no. 8, pp. 1887-1889, 2020.

- [12] M. Muroyiwa, “Evaluation of FAO AquaCrop Model for ability to simulate attainable yields and water use for field tomatoes grown under deficit irrigation in Harare, Zimbabwe,” *African Crop Science Journal*, vol. 30, no. 2, pp. 245-269, 2022.
- [13] F. Suryawati, “Deep structured convolutional neural network for tomato diseases detection. In 2018 international conference on advanced computer science and information systems (ICACSIS),” *IEEE*, vol. 3, no. 2, pp. (pp. 385-390), 2020.
- [14] S. Ashqar, “Image-based tomato leaves diseases detection using deep learning,” *International Journal of Academic Engineering Research (IJAER)*, vol. 2, no. (12), pp. 113-127, 2019.
- [15] U. Anubhove, “Machine learning algorithm based disease detection in tomato with automated image telemetry for vertical farming,” *In 2020 International Conference on Computational Performance Evaluation (ComPE)*, vol. 1, no. 1, pp. (pp. 250-254), 2020.
- [16] S. Sujatha, “Performance of deep learning vs machine learning in plant leaf disease detection,” *Microprocessors and Microsystems*, vol. 8, no. 2, p. 103615, 2021.
- [17] L. Annabel, “Machine Learning for Plant Leaf Disease Detection and Classification—A Review. In 2019 International Conference on Communication and Signal Processing (ICCSP),” *IEEE*, vol. 1, no. 2, pp. (pp. 0538-0542), 2019.
- [18] A. Yalcin, “The use of multi-criteria decision-making methods in business analytics: A comprehensive literature review,” *Technological Forecasting and Social Change*, vol. 121193, p. 174, 2022.
- [19] A. Fuentes, “Impact of selenium and copper nanoparticles on yield, antioxidant system, and fruit quality of tomato plants,” *Plants*, vol. 8, no. 10, p. 355, 2019.
- [20] Y. Zhang, “Deep learning-based object detection improvement for tomato disease,” *IEEE Access*, vol. 8, pp. 56607-56614, 2020.
- [21] Q. Wang, “Tomato diseases recognition based on faster RCNN. In 2019 10th International Conference on Information Technology in Medicine and Education (ITME),” *IEEE*, vol. 1, no. 1, pp. (pp. 772-776), 2019.
- [22] Y. Liu, “A stochastic computational multi-layer perceptron with backward propagation,” *IEEE Transactions on Computers*, vol. 67, no. 9, pp. 1273-1286, 2018.
- [23] C. Chang, “A feature extraction approach based on Hough transform to estimate growth responses of leaf lettuce. In 2019 ASABE Annual International Meeting,” *American Society of Agricultural and Biological Engineers*, vol. 3, no. 2, pp. p-1, 2019.
- [24] R. Manavalan, “Towards an intelligent approaches for cotton diseases detection: A review,” *Computers and Electronics in Agriculture*, vol. 1, no. 2, p. 107255, 2022.

Author Profile

Levis Makurumureis a student at Harare Institute of Technology pursuing a Master of Technology Degree in Cloud Computing.