MIND (Multimedia Artificial Intelligent Networking Database) Journal

MIND Journal | ISSN (p): 2528-0015 | ISSN (e): 2528-0902 | Vol. 6 | No. 2 | Halaman 183 - 193
DOI : https://doi.org/10.26760/mindjournal.v6i2.183-193 Desember 2021

Leveraging MobileNet, InceptionV3, and CropNet to Classify Cassava Plant Disease

GRADY MATTHIAS OKTAVIAN, HANDRI SANTOSO

Information Technology Pradita University
Email: grady.matthias@student.pradita.ac.id

Received 25 Agustus 2021 **|** *Revised* 20 Oktober 2021 **|** *Accepted* 28 November 2021

ABSTRAK

Singkong adalah tanaman yang tumbuh di sub-saharan Africa dan sering dijadikan sumber karbohidrat bagi manusia. Namun, tanaman singkong tersebut memiliki banyak penyakit yang dapat mengancam ketersediaan bahan makanan bagi jutaan orang. Terdapat banyak upaya dan penlitian yang menggunakan kecerdasan buatan dalam bentuk computer vision agar dapat membantu petani mendiagnosa apakah tanaman singkong mereka sehat atau tidak hanya dengan mengambil gambar dari daun tanaman mereka. Pada publikasi ini, penulis melatih tiga jaringan saraf artifisial yang bernama CropNet, MobileNet, dan InceptionV3 untuk dapat mengklasifikasikan gambar-gambar berupa penyakit tanaman singkong. Pembaruan yang dibawa penulis adalah dengan membuat sebuah algoritma gabungan yang mengkombinasikan hasil prediksi dari ketiga jaringan saraf artifisial yang telah dilatih guna mendapatkan hasil prediksi yang lebih akurat. Ternyata, metode penggabungan algoritma ini mampu memberikan nilai akurasi lebih tinggi 6.8% ketimbang nilai rata-rata akurasi dari masing-masing model.

***Kata kunci**: pembelajaran mesin, visi komputer, klasifikasi gambar, jaringan saraf artifisial, kecerdasan buatan, penyakit tanaman*

ABSTRACT

In sub-Saharan Africa, cassava is widely grown and considered to be a large source of carbohydrates for human food. However, the plant is plagued with diseases which can threaten food supply for millions of people. By using computer vision, researchers attempted to create an image classification model that can tell farmers whether the plant is sick or not by taking pictures of their leaves. In this short paper, the author attempts to train three Convolutional Neural Network: CropNet, MobileNet, and InceptionV3 that can classify cassava plant diseases based on visual data. As a novelty, the author creates an ensemble voting classifier that combines the prediction of CropNet, MobileNet, and InceptionV3 to create a better prediction. Turns out, creating an ensemble voting classifier enables us to achieve an accuracy score which is 6.8% higher than the average individual scores of each model.

***Keywords**: machine learning, computer vision, image classification, convolutional neural network, artificial intelligence, plant diseases*

1. INTRODUCTION

Over 40% of Africans rely on cassava as their main source of calories, with over 145 million tons of casava harvested in 2014 **(Gutowski, et al., 2018)**. For small farmers in low-income areas, growing cassava is their main occupation and considers cassava as a food security crop. Cassava is mainly chosen because the plant adapts well to the geographical and climate situation in sub-Saharan Africa (low soil fertility and irregular rainfall patterns). However, cassava farmers see a challenge as the plant is not invulnerable against diseases that might affect their crop harvest results **(Roossinck, 2020)**. These diseases can be diagnosed as certain patterns appear in the cassava leaves. While it is possible to have botanists or experts checking out the plants one by one to see their health, it would be a physically taxing process. Artificial intelligence technology attempts to leverage this by creating a computer vision model which can help to detect cassava diseases in real time. One of the main issues that the modelers have to face is picking the best model architecture so the model can be deployed in a robust platform which eases the usage of said model.

The task which is attempted to be performed is called image classification, in which the input image will be passed through layers in a convolutional neural network. The last layers tend to be dense layers which will finally tell us the probability of the input image belonging to a few predetermined classes **(Krizhevsky, Sutskever, & Hinton, 2012)**. The convolutional neural network architecture is used to process pictures because it allows the network to process each pixel and the relation it has with its surrounding pixels, and thus reduces the size it took to process images compared to classical machine learning frameworks which counts all pixels as feature columns. This method of processing images has led convolutional neural network to gain widespread attention and recognition in the computer vision field of study.

In this paper, the author attempts to explore, fine tune and compare three convolutional neural networks on the cassava disease dataset. The first model is a small CNN called MobileNet, which highlights its light size and processing speed. The second model is InceptionV3 which is the currently latest version of the Inception CNN trained mainly on ImageNet dataset. The third model is a CNN developed by Google and TensorFlow which specifically used to tackle the cassava plant disease problem. In addition to the three separate models, as a novelty, the author creates a new ensemble classifier that combines the predictive power of the three individual models. Creating an ensemble classifier boosts the predictive accuracy of each individual models, and that claim can be proven in our model evaluation result.

2. LITERATURE REVIEW

2.1. Cassava Diseases

Cassava (*Manihot esculenta Crantz*), also named yuca, is a vegetation with tuberous roots who has the scientific family name of *Euphorbiaceae*. Sub-saharan Africa, New Guinea, and South America commonly cultivate Cassava as a primary source of diet. Among the countries in those region, Nigeria, Indonesia, Brazil, and Thailand are countries with the most production of Cassava. Cassava is very rich is starch and mainly used in food, however, it is also used in technological industries of food as key ingredient in the production of starch derivatives **(Zhu, 2015)**.

Cassava serves as the main source of carbohydrate needs for people living in developing regions, such as South America, sub-Saharan Africa, and New Guinea. Cassava is notable for

having important nutrients and starch which are quite resistant. It is also able to withstand harsh conditions, and therefore widely grown in tropical countries. However, Cassava is a plant that is often plagued with a few diseases, which comes from many sources of pathogens – viral, bacterial, and even from smaller organisms like pests and insects. Furthermore, each different pathogen needs a different type of ways to cleanse it. A cassava crop failure spells doom for the continuation and sustainability of food supply in many critical regions of the world, and therefore, novel ways are kept being developed to be able to detect and stop cassava diseases before its widespread damage is too late to be contained. There are four disease condition which are explored in the dataset: Cassava Mosaic Disease (CMD), Cassava Bacterial Blight (CBB), Cassava Greem Mite (CGM), and Cassava Brown Streak Disease (CBSD) (Cock, 2019)

Cassava Mosaic Disease are caused by virus in the genus of Begomovirus. These viruses are circular single-stranded DNA viruses which are transmitted by whiteflies which infect cassava plants. Cassava plants infected by CMD produces a variety of foliar symptoms that include mosaic, mottling, misshapen and twisted leaflets, and an overall reduction in size of leaves and plants. It will produce few or no tubers, which are usually harvested. The CMD is the most severe and most widespread disease in sub-Saharan Africa **(Kumar, 2019)**.

Figure 1. Cassava Mosaic Disease (Kumar, 2019)

The Cassava Bacterial Blight (CBB) is caused by the pathogen Xanthomonas axonopodis. This bacterium is capable of infected most members of the plant genus Manihot. In cassava, symptoms include blight, witting, dieback, and vascular necrosis. In leaves, it leaves an angular necrotic spotting, often with a chlorotic ring encircling the spots **(Thind, 2019)**.

Figure 2. Cassava Bacterial Blight (Thind, 2019)

The Cassava Green Mite (CGM) is a disease caused by the pest Mononychellus tanajoa. The mite pierces and sucks juices from the leaves, causing yellowing, mottling, death and leaf fall. Stems show a 'candle stick' effect with the loss of terminal shoots. To mitigate such mites, the options of introducing its natural enemies and/or cultivating resistant varieties of cassava **(Lebot, 2019)**. If chemical control is used, it might give rise to a resistant population of mites, and the cassava natural quality might be tampered with.

Figure 3. Cassava Green Mite (Lebot, 2019)

The Cassava Brown Streak Disease (CBSD) is a viral infection caused by the RNA virus known as Ipomovirus. This virus gives the cassava severe chlorosis and necrosis, giving them a yellow to brown mottled appearance. Chlorosis may be associated with the veins of leaves, spanning from the mid vein, secondary and tertiary veiins, or in blotches unconnected to veins **(Begomoviruses: Occurrence and Management in Asia and Africa, 2017)**.

Figure 4. Cassava Brown Streak Disease (Begomoviruses: Occurrence and Management in Asia and Africa, 2017)

2.2. Convolutional Neural Network

Convolutional Neural Network (ConvNet/CNN) is a deep learning algorithm which carries input image representation, assign value through learnable weights and biases to different aspects of the images (depths, shades, colors) to extract features and learn distinctive patterns to differentiate visual data stored in images. Other more conventional algorithms use hand-engineered filters, in which every parts of the images are scanned at once, with every aspect of the image considered important. However, Convolutional Neural Network has the ability to extract image features layer by layer, to be able to learn information conveyed within the image, while simultaneously reducing the dimension of the input image **(Iffat Zafar, 2018)**.

Dimensionality reduction is obtained by utilizing a sliding window with a size less than that of the size of the input image. In analogy, the neural network considers a small patch of the complete image one at a time, instead of observing everything all at once. This square patch is called kernel/filters and keeps shifting left to right, and top to bottom, in an attempt to scan the complete image.

Small regression models are trained to detect specific objects in an image. This regression models are tuned by using weights and biased which are randomly initialized at the initiation process of the model, before the first epoch of training has begun. However, as more information are fed into the neural network, the neural network adjusts its weights and biases of each nodes, so the algorithm can achieve better accuracy through approximation.

Overall, the structure of a general convolutional neural network can be seen in the visualization below:

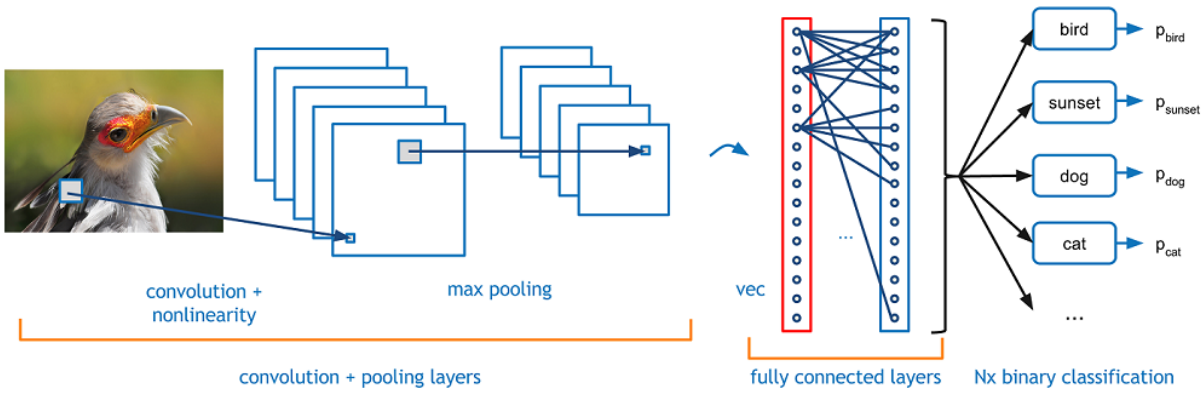

Figure 5. Convolutional Neural Network Diagram (Iffat Zafar, 2018)

There are several layers to consider in a Convolutional Neural Network system. The first layer is called the input layer which serves as the gateway for images data to enter our network to be processed further by the algorithm. Input layer in CNN should contain image data, which are represented by three dimensional matrix that explains the size, shape, and color of our image **(Millstein, 2020)**.

The second layer that serves as the key feature in our convolutional neural network is called the Convolution layer, sometimes called the 'feature extractor layer'. This layer is the primadona of the algorithm, as features of the images are extracted within this layer. Parts

of images are connected to the convolution layer in which it will perform convolution operation which results in the calculation of the dot product between receptive field and the filter/kernel. Result of the operation is a single integer of the output volume. Then, after the convolution layer finishes in reading the whole image, it will then pass the result to the next layer, which is called the pooling layer **(Millstein, 2020)**.

The pooling layer is used to reduce spatial volume of input images after convolution. It is utlizied between two convolution layers. Connecting two convolution layers directly becomes computationally intensive – and this problem is solved by doing a dimensionality reduction first by using a pooling layer. Pooling is done by dissecting the feature matrix into several squares, and taking either the average (mean) of each smaller squares, or taking the maximum value of each squares. The max pooling is more commonly used by advanced algorithms as it tends to converge faster (in the optimization aspect).

After the visual data is passed through multiple iterations of convolution and pooling layers, the data is finally passed on through fully connected layer which involves weights, biases and neurons. Fully connected layers connects all neurons in one layer to neurons in another layer. Because the images has gone through several convolutions, the computational load in this layer has been reduced, while information about the image is still highly maintained. Last but not least, the data is then passed thorugh a softmax / logistic layer of CNN. A softmax layer is used if the task is multi-class classification (such as the benchmark data that we use in this research), while a sigmoid / logistic layer is used for binary classification.

2.3. MobileNet

MobileNet is a convolutional neural network, which serves as a deep learning algorithm that is primarily used to classify images. MobileNet is the result of a deep learning research in computer vision which attempts to come up with models that can be run in embedded systems **(Vasilev, 2019)**. In order to reduce the number of parameters, MobileNet introduced depth-wise convolutions. The second iteration of MobileNet implements a system known as "Inverted Residual Block" to help improving the performance of the model. The main goal of MobileNet is to ensure the architecture of the network is streamlined and balanced in terms of latency and accuracy. Current MobileNet architecture is on its third iteration, in which it manages to have a 3.2% improvement in accuracy on ImageNet classification compared to MobileNet V2, while reducing latency by 20% **(Wang, et al., 2020)**.

The secret of MobileNet's lightweight yet still competitive performance is in its special pooling layer which implements 'Lite R-ASPP' (Lite Reduced Atreus Spatial Pyramid). This system deploys the global-average pooling in similar manner compared to Squeeze-and-Excitation module, in which the model uses a large pooling kernel with large stride to save computational power, and only one 1x1 convolution in the model. MobileNet V3 applies atrous convolution to the last block of its neural network to extract denser features, and add a skip connection from low-level features to capture detailed information **(Andrew Howard, 2019)**.

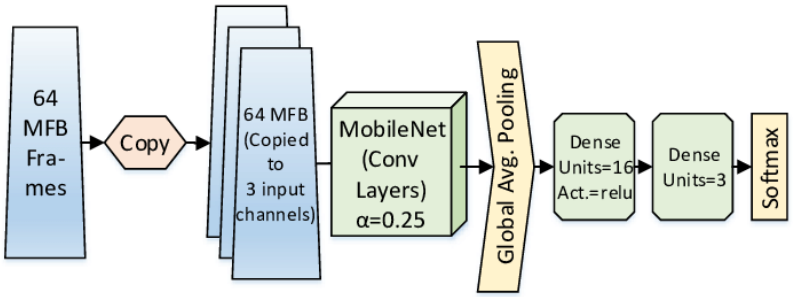

Figure 6. A network architecture of the MobileNet CNN (Andrew Howard, 2019)

In a separate study done by Pan, MobileNet is used to do image recognition and classification to detect welding defects. According to the author of this study, the MobileNet is lightweight, having less parameters to train, and is easily implemented in mobile devices in order to serve a real-time service of image classification **(H. Pan, 2020)**.

2.3. InceptionV3

InceptionV3 is the third version of Google's Inception Convolutional Neural Network. InceptionV3 is trained using the ImageNet dataset on 1000 classes of images, in a total of 1 million training images. Before Inception network was created, CNNs just stack convolution layers deeper and deeper, hoping to get better performance. This does not always work, and deeper networks tend to overfit. It is also computationally intensive. To alleviate this issue, Inception network have filters with multiple sizes that operate on the same level. The network would get wider rather than deeper. Furthermore, the middle part of the network is also equipped with auxiliary classifiers in which softmax is applied to the outputs of two inception models, to compute an auxiliary loss over the same labels. The total loss function is a weighted sum of the auxiliary loss and the real loss. In further improvements to the network, filter banks in the module were expanded to remove the representational bottleneck. In the third version, RMSProp is the default optimizer, and batch normalizations are done for auxiliary classifiers **(C. Szegedy, 2016)**.

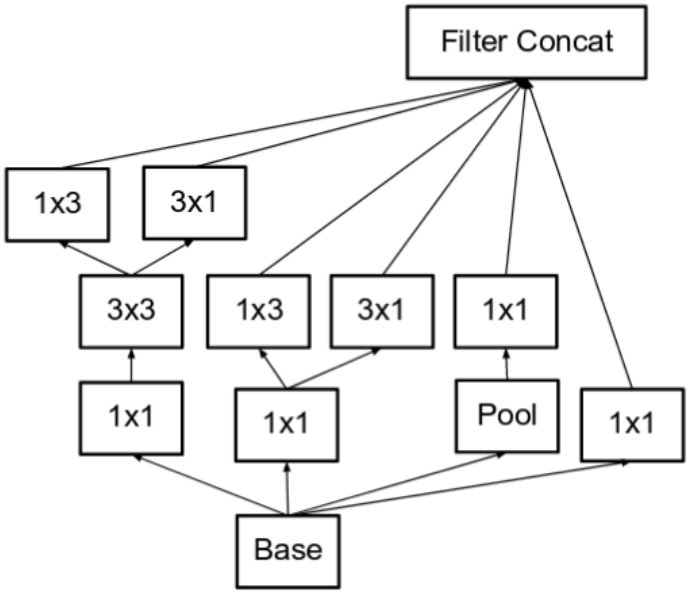

Figure 7. A model architecture of the Inception CNN (C. Szegedy, 2016)

2.4. CropNet

CropNet is a specific CNN which is only trained on Cassava plants, and has 6 output classes: bacterial blight, brown streak disease, green mite, mosaic disease, healthy, or unknown **(Amanda Ramcharan, 2017)**. Unlike MobileNet and InceptionV3, this model is a highly specialized model which is only designed for cassava disease detection, and passing on other images will result for them to be classified as 'unknown'. On paper, this should make the CropNet having the best accuracy compared to our previous two models. CropNet is made by Google and TensorFlow just to classify images of Cassava plants, and it will always give prediction in the form of one of six possible output classes that has been described earlier. If we input an entirely unrelated image into the neural network, it will definitely classify it as 'unknown', so it cannot be used to classify images other than cassava plants.

2.5. Ensemble Voting Classifier

One method to increase the accuracy of machine learning algorithm prediction is to combine the prediction of multiple neural networks to create a better prediction. According to Witten, when we have multiple unbiased models that have comparable accuracy, we can combine outputs of each model by voting **(I. H. Witten, 2016)**. Another research on Ensemble Voting Classifier has been also conducted by Atik Mahabub in 2020, in which he attemps to improve the robustness and accuracy of a fake news detection model. This ensemble voting classifier is done on text data. According to his implementation, by taking three machine learning models and creating an esemble network, the accuracy, ROC score, precision, recall, and f1-score are better than just using a single model. This research supports the notion that creating an ensemble network may help in improving the predictive accuracy of multiple algorithms **(Mahabub, 2020)**. However, in this time, we implement it on image data instead of text.

Since each model outputs a class as their prediction output, voting is done by selecting the label that the majority of the model predicts as the final prediction. For example, if two models predict that a particular image is from class '1', while the third model predicts that the image is from class '2', then our final prediction will be class '1' (the majority). This is done to decrease the chance of a mistake in prediction, since the individual models attempt to patch each other's wrong predictions. During the case where three models have a different prediction, the prediction from CropNet is prioritized.

3. RESULT AND DISCUSSION

The dataset that is used to fine tune and compare these three models come from Kaggle and was hosted on TensorFlow Datasets. The digital images data was provided by Makarere University AI Lab **(Makerere University AI Lab, 2021)**. The dataset name is 'cassava', and consists of leaf images for the cassava plant depicting healthy and four (4) disease conditions; Cassava Mosaic Disease (CMD), Cassava Bacterial Blight (CBB), Cassava Greem Mite (CGM) and Cassava Brown Streak Disease (CBSD). Dataset consists of a total of 9430 labelled images. The 9430 labelled images are split into a training set (5656), a test set (1885) and a validation set (1889). Our models are trained for 20 epochs and the final result is then tested on validation dataset to ensure there is no information leakage.

After our individual models are trained, we take the predicted output from the validation dataset and create an ensemble voting classifiers. We then compare the ensemble prediction result with all three individual models. Here are the evaluation result of the training and validation process:

Table 1. Model Performance Comparison Table

Model	Size (MB)	Trainable Parameters	Validation Accuracy
MobileNet	3.75	1 032 782	82.01%
InceptionV3	77	21 780 646	81.22%
CropNet	15	15 564 502	87.42%
Ensemble Voting Classifier	95.75	38 377 930	89.12%

Since the ensemble voting classifier is a combination of our individual models, its size is the sum of all three of our individual models, and in order to arrive at this model, we need to train all trainable parameters from the three individual models. As we can see from the result, creating an ensemble classifier result in the highest accuracy, although, its size and trainable parameters are also the highest (since it's a combination of three models).

4. CONCLUSION

Based on our training result, it is clear that CropNet is the most accurate individual model for cassava plant disease detection. However, we can see that more generalized models, MobileNet and InceptionV3 have good results as well. This is an important finding, because this means that fine tuning a general CNN can lead to relatively comparable results to specialized models. Another important finding is that the MobileNet is a very efficient CNN – its size is the smallest and is up to 25 times smaller than InceptionV3, however it can achieve a slightly better accuracy than the bigger InceptionV3 model. This is partly because the training and testing dataset consist of relatively homogenous images (they have same dimension, same visual style, and the important object is always in the center). Perhaps, if the task is to classify more difficult images that comes in different resolution, and are less homogenous, then the complexity of the InceptionV3 can be utilized to its maximum potential. In addition to the individual models, it is also noted that creating an ensemble voting classifier, which is made by selecting the prediction label that occurs the most among three individual models, result in a higher accuracy compared to the accuracy of individual models. However, this high accuracy also has its drawback – since we have to train 3 models beforehand, it is not the fastest model to use, and we have to spend time training three neural networks first before arriving at the final prediction result.

REFERENCES

Amanda Ramcharan, K. B. (2017). Deep Learning for Image Based Cassava Disease Detection. *Frontiers in Plant Science*, 1852.

Andrew Howard, M. S.-C. (2019). Searching for MobileNetV3. Retrieved from https://arxiv.org/pdf/1905.02244.pdf

Begomoviruses: Occurrence and Management in Asia and Africa. (2017). Singapore: Springer Singapore.

C. Szegedy, V. V. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 2818-2826).

Cock, J. H. (2019). *Cassava: New Potential For A Neglected Crop.* United States: CRC Press.

Gutowski, A., Wohlmuth, K., Hassan, N. M., Alabi, R. A., Nour, S. S., & Knedlik, T. (2018). *Science, Technology and Innovation Policies for Inclusive Growth in Africa.* Austria: Lit Verlag.

H. Pan, Z. P. (2020). A New Image Recognition and Classification Method Combining Transfer Learning Algorithm and MobileNet Model for Welding Defects. *IEEE Access*, 119951-119960.

I. H. Witten, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques.* Netherlands: Elsevier Science.

Iffat Zafar, G. T. (2018). *Hands-On Convolutional Neural Networks with TensorFlow: Solve Computer Vision Problems with Modeling in TensorFlow and Python.* United Kingdom: Packt Publishing.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Informatics Processing Systems 25* (pp. 1097-1105). Massachusetts: NIPS'12.

Kumar, R. V. (2019). *Geminiviruses: Impact, Challenges and Approaches.* Germany: Springer International Publishing.

Lebot, V. (2019). *Tropical Roots and Tuber Crops: Casava, Sweet Potato, Yams and Aroids.* United Kingdom: CABI.

Mahabub, A. (2020). A Robust Technique of Fake News Detection using Ensemble Voting Classifier and Comparison with Other Classifiers. *Springer Applied Science*, 525.

Makerere University AI Lab. (2021, 11 11). *Cassava Leaf Disease Classification*. Retrieved from Kaggle: https://www.kaggle.com/c/cassava-leaf-disease-classification

Millstein, F. (2020). *Convolutional Neural Networks In Python: Beginner's Guide To Convolutional Neural Networks In Python.* CreateSpace Independent Publishing Platform.

Roossinck, M. (2020). *Virus: 101 Incredible Microbes from Coronavirus to Zika.* United Kingdom: Ivy Press.

Thind, B. S. (2019). *Phytopathogenic Bacteria and Plant Diseases.* United Kingdom: CRC Press.

Vasilev, I. (2019). *Advanced Deep Learning with Python: Design and Implement Advanced Next-generation AI Solutions Using TensorFlow and PyTorch.* United Kingdom: Packt Publishing.

Wang, W., Li, Y., Zou, T., Wang, X., You, J., & Luo, Y. (2020). A Novel Image Classification Approach via Dense-MobileNet Models. *Mobile Information Systems*.

Zhu, F. (2015). Composition, Structure, Physicochemical Properties, and Modifications of Cassava Starch. *Carbohydr Polym.* doi:10.1016/j.carbpol.2014.10.063