

# CNN-based Early Blight and Late Blight Disease Detection on Potato Leaves

Susheel George Joseph  
Associate Professor, Kristu Jyoti College  
of Management and Technology  
Changanasery, Kerala, India.  
susheel@kjcmt.ac.in

Mohd. Shaikhul Ashraf  
Assistant Professor, Department of  
Botany  
HKM Govt. Degree College  
Bandipora, Kashmir  
mohdshaikhulashraf@gmail.com

Arun Pratap Srivastava  
Professor  
Department of CSE  
Lloyd Institute of Engineering and  
Technology  
Greater Noida, Uttar Pradesh, India.  
apsvigi@gmail.com

Bhasker Pant  
Department of CSE  
Graphic Era Deemed to be University  
And Graphic Era Hill University  
Dehradun, Uttarakhand, India.  
Bhasker.pant@geu.ac.in

Ajay Rana  
Amity University  
Greater Noida  
Uttar Pradesh, India.  
Ajay\_rana@amity.edu

Ankita Joshi  
Division of Research & Innovation  
Uttaranchal University  
Uttarakhand, India.  
India. Ankitajoshi2010@yahoo.com

**Abstract:** Potatoes are grown commercially in practically every country in the world. Unfortunately, the crop has been affected by a number of different diseases. In order for the gardener to take quick action, they need to have an understanding of the nature of the contamination. They had the notion that if they looked closely at the leaves, they would be able to learn more about the diseases that were plaguing their communities. Many different Convolutional Neural Network (CNN) models and Machine Learning (ML) methodologies have been created in order to provide assistance to farmers in the diagnosis of diseases affecting tomato crops. Deep Learning and Neural Networks are used in the construction of CNN models. This gives CNN models an advantage over other Machine Learning approaches, such as k-NN and Decision Trees. Because it must handle such a wide array of inputs, the notoriously challenging Pre-skilled CNN is notoriously tough to programme. However, it is capable of producing incredible works of art. An outline of a model for a convolutional neural network that is simpler to understand is provided here. It consists of a total of eight hidden levels. The suggested lightweight model beats both state-of-the-art machine learning approaches and pre-trained models in terms of accuracy when applied to the Plant Village dataset, which is available to the general public. The Plant Village dataset has 39 classes, and these classes collectively represent a large number of different plant species. There are ten different diseases that may infect tomato plants, all of which have the potential to inflict damage. While k-NN has the best accuracy (94.9%) among the classic machine learning methods, VGG16 performs exceptionally well among the trained models. After the picture improvement was finished, the images were pre-processed so that the effectiveness of the suggested CNN may be increased. To be more specific, we accomplished this by considering the width of the picture as a random variable and, as a result, altering the brightness of the image correspondingly. On data sets that have nothing to do with Plant Village, the suggested model achieves an outstanding accuracy of 98%.

**Keywords:** *Potato Disease, Machine Learning and Convolution Neural Network (CNN), Image pre-processing.*

## I. INTRODUCTION

Failures in crop production appear to be linked to plants that have developed in an atypical manner [1]. There is a wide variety of plant diseases, each of which has the potential to significantly reduce a yield. In order to appear, plant diseases require particular agricultural plant species and environmental parameters. Sadly, this is the fundamental cause of the worrying development of plant-based infections in the world today. If the farmer is worried about the loss of productivity or the quality of the grain, he or she may choose to use one of the many plant disease control packages that are available for purchase [1]. During the past three to four decades, there has been a significant shift in both the number of plant diseases and the classification systems that are employed for those diseases. If diseases are discovered at an early stage and preventative actions are followed, the grain that is harvested may have a high quality. The recent advancements that have been achieved in visualization have made it feasible to utilize and make use of digital science in the study of plant diseases in a more widespread manner [2]. The proliferation of digital science's applications in real life has been helped by developments such as these.

There is a growing interest among academics working in a variety of fields to investigate the possibilities of computational approaches for the diagnosis of plant issues. This is because it is an excellent approach for monitoring large flower fields and identifying early indicators of sickness on plant leaves. This is owing to the fact that it is an outstanding method. Several points of view have been addressed while thinking about the process of image segmentation. It seems that the K-method collection of criteria is the least effective way for partitioning a picture given how straightforward it is to gather a substantial quantity of data all at once. This study offers a new hybrid technique for image segmentation that can automatically categorise leaf diseases into subcategories. The study uses high-resolution multispectral and stereo images as its primary data sources.

The authors of this essay came up with a plan on how to successfully carry it out. Researchers [3] looked at the potential use of sugar beet leaves. They used a laptop to do picture processing, which led to the development of a novel and more effective method for evaluating the severity of plant diseases. When doing the initial investigation into the leaf area, the Classification criteria were utilised. After that, a Sobel operator was employed to split the contaminated regions into smaller pieces, which made it possible for their respective boundaries to be identified with more ease. To summarize, the number of dots present on a leaf and the pattern in which those dots are distributed may be used to categorise the many illnesses that might affect plants [4].

The algorithms that were created for the purpose of teaching computers can also be used to the process of identifying illnesses in humans. In the past, members of the agricultural research community have conducted experiments on a number of these ideas. K-nearest neighbors and ANN desire trees are the only two of these, however there are a great many more. Support vector machines are one method that has been utilised rather frequently in this industry (SVMs). For instance, in [15], SVMs were utilised in order to explore the outward indications of cotton sickness. After determining the graphical appearance of areas affected with the illness, the authors of [5] demonstrated how to use multi-elegance support vector machines to recognize wheat disease based on its prevalence. This was done after determining the graphical appearance of areas affected with the illness. It's possible that this method may be utilised to diagnose wheat problems (SVMs). After we broke the coding on the leaf, we tested their knowledge of multi-element RBF SVM to see how well they understood it. A professional was brought in to understand the ins and outs of the underlying classifier, and then a variety of statistical methods were used to put it through its paces. A piece of software was designed with the purpose of classifying unhealthy plants into different categories according to the outward manifestations of the illness that they display. In order to differentiate between the various subtypes of the sickness, features that were taken from the centre fringe photos were utilized, as well as Bayes' and SVM. In order to categorise the illnesses, classifiers were utilised. The authors of aimed to solve the difficulty of identifying SVM parameters by using a genetic algorithm (GA) to automatically choose optimum SVM parameters and an orthogonal technique to zero in on the best GA parameters. Together, these two methods were used to narrow in on the best GA parameters. They were attempting to establish boundaries in this manner, and they were making efforts to do so (SVM). In order to recognize the indications and symptoms of nutritional deficiencies, an automated detection system [6] uses a device that is both camera-equipped and equipped with pattern recognition software. This allows the system to identify the signs and symptoms.

As part of this study, we develop an approach for identifying and quantifying the formation of new leaves on

tomato plants that makes use of computers. The entire process may be broken down into its constituent parts, which include preprocessing, segmentation, attribute extraction, and class. Multiple learning-from-experience systems, such as Convolutional Neural Networks (CNN) and Random Forest, were used to sort through images of tomato leaves and determine which ones were diseased and which ones were healthy. This was done in order to determine which tomatoes had healthy leaves and which ones had diseased leaves. The following is a further list of error analysis and repair methods: In order to evaluate the effectiveness of the suggested CNN models, both industry-standard Machine Learning (ML) approaches and pre-trained CNN models were utilised. The data made it very evident that the proposed models were too similar to one another and that they would be better off being considerably distinct from one another. In addition, we found that the suggested CNN model not only achieves accuracy on par with that of the state-of-the-art, but also considerably outperforms pre-trained models and more traditional Machine Learning techniques [7]. This was another finding that we made.

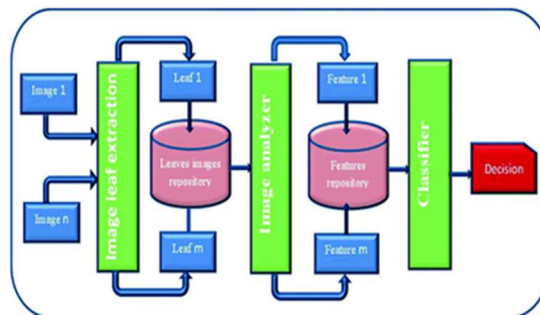


Figure 1. Overall Proposed design system

This paper is discussed as follows, first section and second reveals the introduction about potato leaves and their detection. In the third section, we discuss the proposed CNN model for detecting diseases in Potato leaves. The results of the experiments and their analyses are presented in Section 4, and their interpretation is presented in Section 5.

## II. RELATED WORK

Rapid disease detection systems are gaining popularity in a variety of different industries due to the fact that they make it simpler to monitor a large number of crops. This is one of the reasons why. To put it another way, as soon as the leaves of a plant begin to exhibit additional symptoms of a disease, it is possible to diagnose the plant without any further intervention from a human being. This could take place as soon as the first symptoms show up in the patient.

Using high-resolution multispectral and stereo images, the authors of [8] discuss a method for automatically classifying leaf diseases into subtypes. This method can be found in [8]. They examined the leaves of the sugar beets to see whether or not their technique had been effective. Using a method for the processing of digital images on a computer,



better. When data is collected in the form of a picture, we should make sure the analysing model has enough skills to do a good job. Most of the time, more records mean better results. People often use CSV files to keep track of information. In this study, CSV files are used to store 60 characteristics from each picture of one of eight good types of tomato leaf classes. First, photos of torn tomato leaves of different kinds were found on the Internet. To get to the spots that are damaged, the leaf image is cut out and then smoothed. After that, each picture goes through several kinds of photo processing.

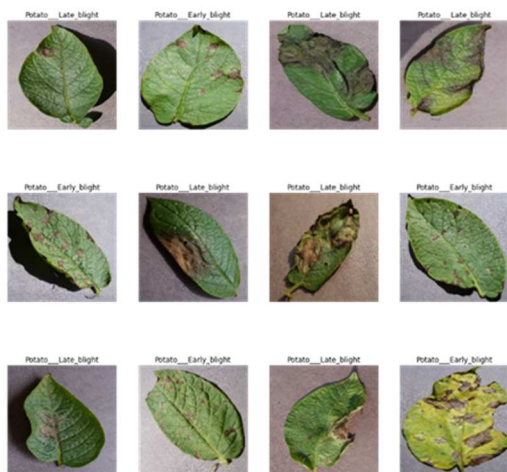


Figure 3. Data Set of Potato Leaves To Diagnosis Disease

- 1) *Image Acquisition:* First, potato leaves are used to make prints (18). In the databases, you can find both healthy and affected potato leaves. The 14000 samples are spread out over eight lines in the dataset. 7 leaflets only have pictures of diseased potato blights, but the other brochure has pictures of healthy ones (16).
- 2) *Image Pre-processing:* This is a key step in the process of making an image bracket. If each image isn't measured (i.e., its size isn't decreased), the processing time will go up during the discovery and bracket stages. Before changing a colour image to an argentine scale image, the size of each potato blight image is cut from 256 by 256 to 128 by 128. (17).



Figure3. Noise reduction in the Potato leaf

- 3) *Segmentation:* In this exploration, we used the most widely used k-suggest clustering method to find interesting areas. We've been using this system

because it's hard to tell the colours apart in splint prints with the naked eye. The L.A.B. creativity space, which is also known as CIELAB and CIE Lab, lets us see how big these differences are. The L-a-b shadeation space is made with the help of the CIE XYZ tristimulus values. The L a b area is made up of the refulgence L or luminance subcaste, the value subcaste a that shows where sedation falls along the red-green axis, and the value subcaste b that shows where sedation falls along the blue-unheroic axis (18).

- 4) *Feature Extraction:* The next goal is to get rid of some important information from the leaf image after it has been cut up and the bad parts of the leaf have been separated. Feature extraction cuts down on the amount of data by measuring certain things about each picture, such as its texture, shading, and shape. In the proposed method, we grade both geometic and histogram skills that will be used to model later. Geometrical features we like to figure out a few geometic skills at this point [19].

**Pseudo code:**

Image Length:	First, the length of the diseased portion of the splint is determined.
Image Area:	Locate the part of the splint that has become infected, and treat it accordingly.
Image Area Estimate:	Check to see how close the infected area is to the edges and corners of the splint.
Image Perimeter:	Determine which end of the splint has the infection and go to that end.
Euler number:	The following step entails computing the Euler number, also referred to as the Euler Poincare characteristic, of a degraded portion of an image. By examining this value, we will be able to determine the number of components in the picture that are connected to one another.

- 5) *Classification:* CNNs are a type of Multilayer Feed forward Network that are made to work with images [6, 7, 26, 27]. Convolution layers automatically find the parts of an image that will also be fed into the network [20].

- a. *CNN:* CNN's performance is affected by several hyperparameters, as well as the number of iterations, hidden and output nodes, hidden nodes, training algorithm, drop out, response rates, batch size, and so on. In the hyperparameters tuning, the same experiments are done over and over with different hidden layers, epochs, and activation points for

understanding quotes. After fine-tuning, the version's ideal can be reached quickly with high quality.

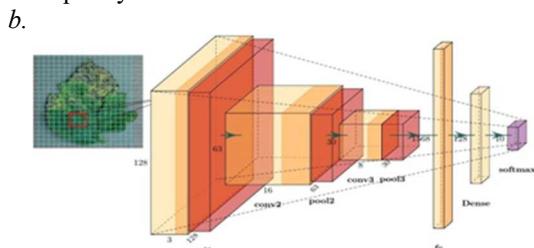


Figure 4. Proposed CNN model [21]

c. *Back propagation in CNN:* The CNN model starts with a weight matrix, and the loss is calculated using equation (16), which is shown below. This is all done through back propagation. The weight values are changed based on how steep the gradient is and how much weight is expected to be lost. To cut down on the common loss, the whole process (both forward and backward propagation) is repeated for a set number of epochs [22].

$$L(a,y) = -(y \log(a) + (1-y) \log(1-a)) \quad (1)$$

Here L is the loss, y is estimated output and a is predicted output for a single take a look at case.

There are many factors that can affect a CNN's performance, including hyperparameters, the training algorithm, drop out, response rates, batch size, and the number of iterations. Repetitive experimentation with various hidden layers, epochs, and activation points is used in hyperparameters tuning to achieve this quote comprehension. After some tweaks, the best possible version can be produced quickly and competently.

#### IV. RESULTS AND DISCUSSION

Using the corresponding equations, the performance of the proposed model was compared to that of other traditional machine learning approaches as well as pre-trained CNN models.

Accuracy is absolutely necessary in order to conduct a comprehensive analysis of the usefulness of a model. When contrasted with the other model, the one in question will provide a more accurate representation of reality. To find a solution, we start by entering numbers into an equation.

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (2)$$

F1-score: It's also known as the harmonic mean of precision and recall, and it's a popular way for comparing the performance of different learning methods. The equation is used to determine the F1-score (3).

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

TABLE II. COMPARISON ANALYSIS FOR EXISTING AND PROPOSED CLASSIFICATIONS

Classifiers	Accuracy	Precision	Recall
KNN	89%	85%	70%
SVM	93%	86%	80%
RF	92.5%	87%	83%
NB	94%	90%	85%
DT	94.6%	92%	90%
Proposed CNN algorithm	97%	95%	96%

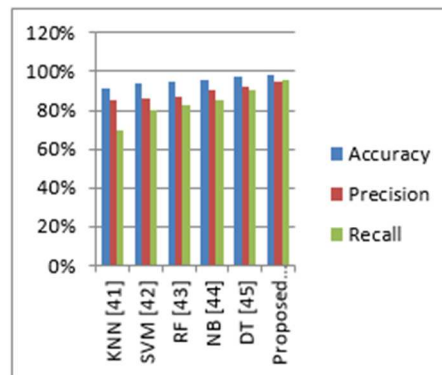


Figure 5. Comparison analysis for existing and proposed classifications

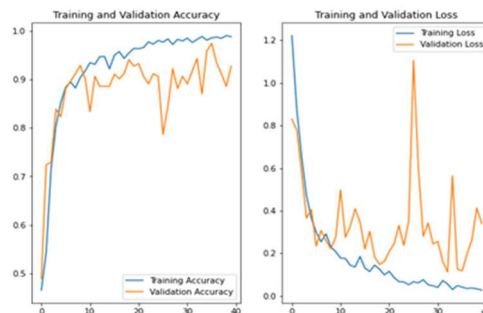


Figure 6. Comparison analysis of training and accuracy (validation vs. loss).

First Image of batch to predict:  
 Actual label: Potato\_\_Early\_blight  
 Predicted label: Potato\_\_Early\_blight



Figure 7. Predicted output as Early Blight

#### V. CONCLUSION

A CNN-based model was constructed specifically for the purpose of this investigation in order to identify the disease

that was harming tomato plants. In the CNN-based definitely architecture that has been presented, there are a total of three convolution layers and one max pooling layer. Within this architecture, each layer makes use of its own distinct grouping of filters. This study made advantage of the information on tomato leaves that was found in the Plant Village dataset. A visually appealing representation of each of the nine types of contamination included in the collection may be found here. Because the photographs that were displayed in the classroom did not all begin in the same manner, it was necessary to use data augmentation techniques in order to standardize them. Evaluations indicate that the version that was evaluated has a level of accuracy while checking the courses that ranges between 76 and 100 percent. In addition to that, testing has revealed that the model is correct 99% of the time.

## REFERENCES

- [1] Srivastava, A., Singh, A., Joseph, S.G., Borole, Y.D., Singh, H.K., WSN-IoT Clustering for Secure Data Transmission in E-Health Sector using Green Computing Strategy, 2021 9th International Conference on Cyber and IT Service Management, CITSM 2021, 2021
- [2] Shrivastava, A.; Ranga, J.; Narayana, V.N.S.L.; Chiranjivi; Borole, Y.D., Green Energy Powered Charging Infrastructure for Hybrid EVs, 2021 9th International Conference on Cyber and IT Service Management, CITSM 2021, DOI: 10.1109/CITSM52892.2021.9589027.
- [3] X. Xie, Y. Ma, B. Liu, J. He, S. Li, and H. Wang, "A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks," *Frontiers of Plant Science*, vol. 11, 2020.
- [4] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Computers and Electronics in Agriculture*, vol. 127, pp. 418–424, 2016.
- [5] M. Brahimi, K. Boukhalifa, and A. Moussaoui, "Deep learning for tomato diseases: classification and symptoms visualization," *Applied Artificial Intelligence*, vol. 31, 2017.
- [6] E. Fujita, Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic investigation on a robust and practical plant diagnostic system," in *Proceedings of the 2016 15th IEEE International Conference Mechanical Learning Application ICMLA*, vol. 2017, Anaheim, CA, USA, December 2016.
- [7] S. Ghosal, D. Blystone, A. K. Singh, and B. Ganapathysubramanian, "An explainable deep machine vision framework for plant stress phenotyping," *Proceedings of the National Academy of Sciences*, vol. 115, pp. 4613–4618, 2018.
- [8] U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudiger, J. T. Heng, and U. R. Acharya, "Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images," *Information Sciences*, vol. 441, pp. 41–49, 2018. S. Yadav, N. Sengar, A. Singh, A. Singh, and M. K. Dutta, "Identification of disease using deep learning and evaluation of bacteriosis in peach leaf," *Ecological Informatics*, vol. 61, Article ID 101247, 2021.
- [9] A. K. Rangarajan, R. Purushothaman, and A. Ramesh, "Tomato crop disease classification using pre-trained deep learning algorithm," *Procedia Computer Science*, vol. 133, pp. 1040–1047, 2018.
- [10] N. Zhu, X. Liu, Z. Liu et al., "Deep learning for smart agriculture: concepts, tools, applications, and opportunities," *International Journal of Agricultural and Biological Engineering*, vol. 11, no. 4, pp. 32–44, 2018.
- [11] V. Sravan, K. Swaraj, K. Meenakshi, and P. Kora, "A deep learning based crop disease classification using transfer learning," *Materials Today Proceedings*, vol. 6, 2021.
- [12] M. B. Tahir, M. A. Khan, K. Javed et al., "Recognition of apple leaf diseases using deep Learning and variances-controlled features reduction," *Microprocessors and Microsystems*, vol. 141, Article ID 104027, pp. 1–24, 2021.
- [13] J. Shin, Y. K. Chang, B. Heung, T. Nguyen-Quang, G. W. Price, and A. Al-Mallahi, "A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves," *Computers and Electronics in Agriculture*, vol. 183, Article ID 106042, 2021.
- [14] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, "Deep neural networks with transfer learning in millet crop images," *Computers in Industry*, vol. 108, pp. 115–120, 2019.
- [15] K. Murali Krishna; Amit Jain; Hardeep Singh Kang; Mithra Venkatesan; Anurag Shrivastava; Suresh Kumar Singh; Muhammad Arif, Development of the Broadband Multilayer Absorption Materials with Genetic Algorithm up to 8 GHz Frequency, Security and Communication Networks, 2022-02-17, DOI: 10.1155/2022/4400412
- [16] Anurag Shrivastava; Chinmaya Kumar Nayak; R. Dilip; Soumya Ranjan Samal; Sandeep Rout; Shaikh Mohd Ashfaq, Automatic robotic system design and development for vertical hydroponic farming using IoT and big data analysis, Materials Today: Proceedings, 2021-07, DOI: 10.1016/j.matpr.2021.07.294
- [17] Anurag Shrivastava; Rajneesh Sharma; Mohit Kumar Saxena; V. Shanmugasundaram; Moti Lal Rinawa; Ankit, Solar energy capacity assessment and performance evaluation of a standalone PV system using PVSYST, Materials Today: Proceedings, 2021-07, DOI: 10.1016/j.matpr.2021.07.258A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," *Frontiers of Plant Science*, vol. 8, pp. 1–7, 2017.
- [18] L. X. Boa Sorte, C. Toledo Ferraz, and F. Fambrini, "Coffee leaf disease recognition based on deep learning and texture attributes," *Procedia Computer Science*, vol. 159, pp. 135–144, 2019.
- [19] S. Ramesh and D. Vydeki, "Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm," *Information Processing in Agriculture*, vol. 7, pp. 249–260, 2020.
- [20] G. A. B. Jayme, "Factors influencing the use of deep learning for plant disease recognition," *Biosystems Engineering*, vol. 172, pp. 84–91, 2018.
- [21] A. Z. da Costa, H. E. H. Figueroa, and J. A. Fracarolli, "Computer vision based detection of external defects on tomatoes using deep learning," *Biosystems Engineering*, vol. 190, pp. 131–144, 2020.
- [22] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.
- [23] A. Fuentes, S. Yoon, S. Kim, and D. Park, "A Robust deep learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, p. 2022, 2017.
- [24] S. Pandiyan, M. Ashwin, R. Manikandan, K. M. Karthick Raghunath, and G. R. Anantha Raman, "Heterogeneous Internet of things organization predictive analysis platform for apple leaf diseases recognition," *Computer Communications*, vol. 154, pp. 99–110, 2020.
- [25] S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain, "Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: an automatic approach towards plant pathology," *IEEE Access*, vol. 6, pp. 8852–8863, 2018.
- [26] D. Oppenheim and G. Shani, "Potato disease classification using convolution neural networks," *Adv. Anim. Biosci.*, vol. 8, p. 244, 2017.
- [27] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers of Plant Science*, vol. 7, pp. 1–10, 2016.
- [28] J. Boulent, S. Foucher, J. Théau, and P.-L. St-Charles, "Convolutional neural networks for the automatic identification of plant diseases," *Frontiers of Plant Science*, vol. 10, p. 941, 2019.
- [29] S. Qiao, Q. Wang, J. Zhang, and Z. Pei, "Detection and classification of early decay on blueberry based on improved deep residual 3D convolutional neural network in hyperspectral images," *Scientific Programming*, vol. 2020, no. 4, 12 pages, 2020.