



Smart Farming: Potato Disease Diagnosis Using Machine Learning

Kalaivani D

Dept. of Information Science and Engineering, New Horizon College of Engineering, Bangalore, India

Sujay B

Dept. of Information Science and Engineering, New Horizon College of Engineering, Bangalore, India

Nidish Kumar G

Dept. of Information Science and Engineering, New Horizon College of Engineering, Bangalore, India

Tharakeshwar S

Dept. of Information Science and Engineering, New Horizon College of Engineering, Bangalore, India

Monika

Dept. of Information Science and Engineering, New Horizon College of Engineering, Bangalore, India
monikapadmanathan2004@gmail.com

Abstract: Potato cultivation has been very popular for the last half of the century. But potato production gets hampered due to some diseases which are increasing the potato production for the farmers, which in turn is affecting the life of farmers. An automated and immediate disease detection process is required to increase potato production and improve the system. Our goal is to diagnose the potato diseases in the potato and by using the leaf pictures we are going to use the advanced machine learning technology. This paper clarifies about the process and machine learning systems with which how the potato leaf diseases will be identified and classified. Image processing is the best solution for identifying and analyzing the diseases. In this analysis, there is a division for 200 pictures of affected potatoes and potato leaves, which has been taken from plant town base data. To identify the sick and sound leaves, a few models which are ready to be used has to be utilized. When it is tested using 25% of the test data and the rest 75% by the train data, its predicts the accurate range from 99 to 23%.

Keywords: Potato Disease, Image processing, Machine Learning, Agriculture

Nomenclature

Abbreviation	Description
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CV	Computer Vision
DL	Deep Learning
GPU	Graphics Processing Unit
ML	Machine Learning

1. Introduction

Potato the major harvesting crops in country. The production of potatoes are affected by many different kinds of pests and infections. Common diseases such as Early Blight, Late Blight, and Bacterial Wilt significantly reduce both crop productivity and market value if not identified at an early stage. These diseases directly affect crop quality and yield. Hence our potato exports aren't up to our expectations. The farmers of major areas face many hampers on the disease every year. The farmers and businessmen in our country are facing many problems with the diseases in the case of export to countries such as Russia, Indonesia, Malaysia, Sri Lanka, Thailand, etc. This brings us to the after effects of the global pandemic, which saw the increase of prices of many everyday-commodities, among which was the potato. While this can be viewed as an issue to the common man, it has benefited farmers greatly. This increase in price paired with increased overall potato production in India, presents a huge opportunity for farmers. Potatoes have become one among the most profitable crops. On 2024, 513,537 metric tons worth of fresh potatoes/chilled (approximate total export value being 110 million USD) and 172,444 metric tons worth of frozen potatoes (approximate total export value being 200 million USD). Thus, given the importance of potatoes in the global market, it then becomes vital to ensure that the production quality is on the top of facilitating the health and well-being of potato crops. Hence, to aid and ensure that, early disease detection then becomes crucial. Given that a lot of practices and solutions exist for such an issue, it needs to be noted that they are often slow and time consuming, as well as expensive. Traditional disease diagnosis relies heavily on expert visual inspection, which is not always readily available in rural farming regions. A

modern, more efficient solution is then obligated to be present.

This project tackles exactly that. Recent developments in artificial intelligence and computer vision have enabled automated plant disease diagnosis with high efficiency and minimal human intervention. By means of image processing models, said diseases can be diagnosed faster and easier [10]. Using image processing, the disease can be identified by isolating key symptoms that are present on the potato plant or inside the potato, then matching it with certain threshold parameters.

Furthermore, appropriate treatments have been provided as well, with further speeds up the recovery process and recovery rate of the crops. Providing treatment recommendations immediately after disease detection assists farmers in taking timely corrective measures and minimizing crop damage. The model used performs very well on the training data with it having 99% accuracy on sample data. Therefore, the system is sure to accurately and diagnose the diseases.

1.1 Objectives

1.1.1 Create an Automated Detection System

The core targets of the project is to develop and promote a system that intelligently and automatically identifies diseases in potato plants. The automated system is intended to reduce manual inspection effort while providing consistent and reliable disease diagnosis. This can be achieved via image processing techniques. Specifically, by using image processing to analyze the images of potatoes and potato plant's leaves, the system aims to find patterns and symptoms which indicate the presence of certain diseases. Once the model is fully accustomed to images, the system can not only differentiate between infected and healthy leaves, but also specify the exact kind of disease that the plant has, all with minimal human interaction.

1.1.2 Dataset Creation and Utilization

Training the system involves making it "learn" and find patterns among a large number of example images. A diverse dataset containing images captured under different environmental conditions improves the robustness of the trained model. First, images of both healthy and infected leaves are collected. These images then undergo preprocessing. Preprocessing is a step that prepares the raw images into a cleaner image set, such that no errors or false alarms may arise. This includes resizing, cleaning and labeling the images in order to prepare them for training the machine learning model.

1.1.3 Model Implementation and Evaluation

This step involves the usage of various different machine learning algorithms to create our most optimal and appropriate system. Comparative evaluation of different deep learning models helps identify the most suitable architecture for disease classification. Under VGG16 architecture, CNNs (Convolutional Neural Networks) is used as the base algorithm for the machine learning model [4][10].

The model is then trained using the previously created image dataset, and the results are collected and evaluated. Evaluation happens on the basis of model performance and accuracy.

1.1.4 Achieve High Accuracy in Disease Detection

The system has to have a high level of accuracy in order to meet the desired outcome. High classification accuracy ensures reliable diagnosis and increases user confidence in the proposed system. In order to achieve this, the model is fine-tuned, i.e. adjusting and improving data quality, adjusting the setting and using methods like data augmentation. Data augmentation is a process that involves adding more training samples by rotating the pre-existing images, zooming in on the images, and such. By ensuring a high accuracy level, the system aims to aid farmers make faster and better informed decisions.

1.1.5 Early Diagnosis to Minimize Loss

If diseases are detected earlier, farmers can cure or even prevent the onset of these diseases. Early intervention helps reduce the spread of infection to neighbouring plants and improves overall crop health. This reduces crop damage, prevents major losses, and cuts down the need for expensive treatments which is caused due to prolonged effects of said disease(s). The system helps farmers make faster and better decisions, which leads to higher yields and lower additional costs.

2. Literature Review

1. "Detection and differentiation between potato diseases using calibration models trained with non-imaging spectrometry data" by Bienkowski (2019) [1]: This study uses non-imaging spectrometry, specifically spectral values of potato samples, which help identify the

disease of the plant may be going through. The study emphasizes the importance of spectral analysis as a rapid alternative to traditional visual inspection methods. It focuses on trying to find a method of rapid and non-destructive disease detection. They specialized in using spectral reflectance data, i.e. how much light is reflected from a surface as data for their analysis. Using this data, the calibration models which they utilized to differentiate between multiple diseases. This research demonstrated the potential for disease diagnosis in agriculture without the need for high-definition images.

2. **"Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine" by Monzurul Islam (2017) [2]:** This research focuses on disease diagnosis of potato plants using images as input and makes use of Support Vector Machine (SVM). The study demonstrated that machine learning techniques can effectively assist in agricultural disease monitoring. SVM uses a process called image segmentation, wherein the image is divided into multiple parts, and then isolating the most required parts of the image and using that to recognize the features present in said image. Specifically, K-means clustering was used to perform segmentation, a method that groups similar pixels based on colors and intensity. Using this, multiple diseases were able to be successfully classified, often with good accuracy. However, the one downside to this research was that a relatively small dataset was used, in which images were often captured under good lighting and controlled conditions. In comparison, CNN-based methods can scale better with large datasets, handle diverse lighting, and offer end-to-end learning with improved real-world adaptability.
3. **"Detection of Potato Disease Using Image Segmentation and Machine Learning by Md. Adif Iqbal and Kamrul Hasan Talukder (2020) [3]:** This research employed image segmentation and traditional machine learning to detect potato leaf diseases. The study highlights the effectiveness of combining image preprocessing with machine learning classification techniques. It used 450 images from the Plant Village dataset, which may not represent real-world variability. The model's accuracy could be affected by uncontrolled factors like lighting and background noise in field conditions. The system lacks evaluation in actual agricultural environments, limiting its practical relevance. Also, the computational demands of the method are not clearly addressed, raising concerns for low resource settings. The segmentation-based approach may struggle with generalization across disease types. In contrast, CNNs offer end-to-end feature learning, better adaptability to field images, and scalability for real-world deployment.
4. **"Potato Leaf Diseases Detection Using Deep Learning" by Divyansh Tiwari (2020) [4]:** This research is based on applying transfer learning using deep neural networks. Transfer learning has become a widely adopted technique because it reduces training time and improves classification performance. Transfer learning is the process of using a pre-developed machine learning model is reused for other, different yet related task. It involves taking the pre-trained model which has trained on a large dataset, for example thousands of images from online and then fine tuning it to use for a specific task, like detecting diseases in potato plants. Logistic Regression provided the best performance, reaching up to 98% accuracy. It showed the possibility of using transfer learning for real life uses such as potato plant disease detection
5. **Application of Transfer Learning to Detect Potato Disease from Leaf Image by Farabee Islam (2019) [5]:** This research is focused to train a model to identify diseases in potatoes. Multiple state-of-the-art deep learning architectures were investigated to determine the most effective model for disease classification. Multiple different CNN architectures were used as a basis and tested, models like InceptionV3, Xception, VGG19 were used in this testing process. These models were fine tuned on a dataset containing potato leaf images, specifically aiming for two diseases: Early Blight and Late Blight. This system provided a near perfect accuracy of 99.5% when the architecture used was Xception. However, the results weren't useful in real life scenarios as the training data contained mainly images taken under ideal conditions, with perfect lighting. This doesn't represent the field conditions. This along with the fact that no field testing was conducted leads to major drawbacks in the study.
6. **Potato Leaf Diseases Detection and Classification System, Girish Athanikar and Ms. Priti Badar (2016) [6]:** This research was dedicated to find and classify the leaf diseases using image processing. The study demonstrates the early application of machine learning techniques in plant disease diagnosis. They used infected leaf images as input, put said images through segmentation using K-means clustering and extracted features such as color and shape. They then used Back Propagation Neural Network as the machine learning algorithm. This provided results with accuracy of upto 92%. However, this study, like the previous one, lacked any real-world testing and also doesn't specify what type of disease is infecting the plant, but only that the plant was infected. Furthermore, Back Propagation Neural Network, while it provided a high accuracy, is an outdated method by today's standards

7. **Detection and Classification of Leaf Disease Using Artificial Neural Network**, by **Malvika Ranjan and colleagues (2015) [7]**: This study presented a method for detecting and classifying plant leaf diseases using Artificial Neural Networks (ANN). The work contributed to the growing interest in automating agricultural disease diagnosis through intelligent systems. This methodology included image acquisition, pre-processing, segmentation of infected areas, feature extraction, and classification via ANN to identify specific diseases. The system's strengths lie in its ability to automate, efficiently, and accurately classify leaf diseases, aiding in early diagnosis and reducing manual labor. However, it has drawbacks, such as the necessity for a large, well-labeled dataset for training the ANN and its sensitivity to image quality, which can impact classification accuracy.
8. **Content Based Paddy Leaf Disease Recognition and Remedy Prediction Using Support Vector Machine** by **F. T. Pinki, N. Khatun, and S. M. M. Islam (2017) [8]**: This research introduces a system that identifies paddy leaf diseases and recommends treatments through Support Vector Machine (SVM) classification. The inclusion of remedy prediction makes the system more useful for practical agricultural decision-making. Their approach includes capturing images, enhancing disease spots through pre-processing, extracting features based on color and texture, and classifying the diseases using SVM for accurate identification. The primary benefit of this method is its high classification accuracy and effective remedy suggestions, which support farmers in managing diseases. However, the system has limitations, such as its dependence on high-quality images and difficulties in dealing with complex backgrounds or overlapping symptoms, which could impact detection accuracy.
9. **Application of Support Vector Machine for Detecting Rice Diseases Using Shape and Colour Texture Features** by **Yao Q., Guan Z., Zhou Y., Tang J., Hu Y., and Yang B., (2009) [9]**: The research outlines a method for automatically detecting rice diseases using SVM classification. This work is considered one of the earlier studies demonstrating the applicability of machine learning techniques in agricultural disease diagnosis. This study employs image processing techniques to extract shape and colour texture features from rice leaf images, which are then input into the SVM for classifying and identifying various diseases. The main advantage of this technique is its high accuracy and efficiency in disease detection, which can greatly assist in early diagnosis and crop management. However, it also has drawbacks, including the necessity for a well-curated dataset and potential challenges in generalizing results across different environmental conditions or rice varieties. This method showcases the promise of machine learning in precision agriculture but needs further development for wider application.

3. Methodology

All the proposed model, the CNN algorithm is utilized to know the various kinds of infections, it is having 7 classes of potato sicknesses and 99% accuracy rate as referred to fig. 1.

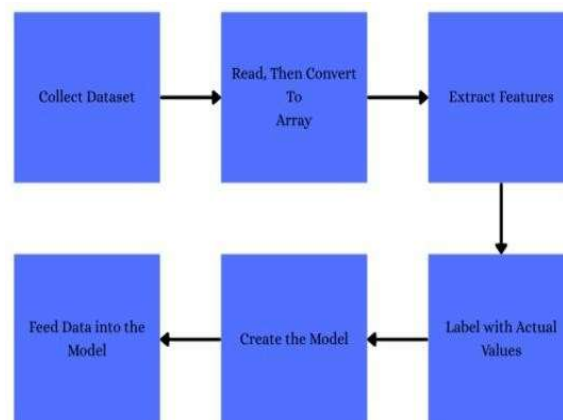


Fig. 1. CNN Workflow for Potato Disease Detection

3.1 Dataset

This information has been collected by the taking pictures directly from the fields of a village. Collecting images directly from agricultural fields enables the dataset to capture natural variations in lighting,

background, and disease severity, making the trained model more suitable for real-world farming environments. Data on about five types of the diseases of potatoes and the potato leaves was collected in Table 1. A total of 200 potato and the leaves images was collected. With the data being split between the leaves and disease. Through this, it has been noticed that diseases are far more contagious in potato leaves than in potatoes themselves.

Table 1. *Class-wise Distribution of Potato Disease Images*

Class	Disease Name	Image (Number)	Train Data	Test Data
01	Early Blight of potato	428	329	100
02	Potato leaf roll virus	394	274	75
03	Hollow heart of potato	221	218	80
04	Scab of potato	154	150	98
05	Soft rot of potato	238	200	68

3.2 Preprocessing

Four types of image processing steps need to be taken. These steps include normalizing the image, changing the color of the image, and identifying the properties, and finally Image processing such as filtering and transformation of the image [10]. Image preprocessing plays a crucial role in reducing unwanted variations and improving the quality of input images before feature extraction. Python's OpenCV Library has been used for this purpose. Features of the OpenCV library include: Read & write images, Capture and save the videos, Image processing such as filtering and transformation, Detection the feature, Video or picture object detection. The picture document is perused using OpenCV. OpenCV represents colors in BGR format. Whereas, in Pillow, colors are represented using the RGB format.

3.3 Model Architect

CNNs analyzes an image piece by piece, focusing on small parts at a time. CNNs are better at recognizing local patterns or similarities rather than matching entire images globally. The convolutional layers automatically learn hierarchical visual representations, beginning with simple edge features and gradually identifying complex disease patterns present in potato leaves. This ability to detect fine-grained differences makes CNNs highly effective for plant disease detection [4][5]. When integrated into mobile apps or smart farming tools, they can assist farmers by analyzing images of crops in real time and identifying potential diseases with high accuracy. By providing accessible and intelligent diagnostic support, such tools can promote informed decision-making, reduce crop loss, and empower agricultural communities ultimately improving productivity and sustainability in the farming sector.

4. Social Benefits

- **Improved Farmer Income:** By enabling early and accurate disease detection, the system helps reduce crop losses and lower production costs, leading to increased yields and higher profits for farmers. Early intervention also minimizes the spread of diseases to healthy crops, thereby improving overall farm productivity.
- **Sustainable Farming Practices:** Accurate diagnosis minimizes the overuse of pesticides and other chemical treatments, supporting environmentally friendly and sustainable agricultural practices. The reduction in unnecessary chemical application also helps preserve soil fertility and protect surrounding ecosystems.
- **Increased Access to Technology:** The project empowers rural and smallholder farmers by providing accessible, user friendly tools for disease detection, fostering the digital transformation of agriculture. Easy-to-use intelligent systems encourage greater adoption of precision agriculture technologies among farming communities.
- **Enhanced Economic Stability:** With improved productivity and lower risk of crop failure, farmers especially small-scale and marginal ones gain better financial security and economic resilience. Improved agricultural productivity also contributes positively to the local and regional economy.
- **Strengthened Food Security:** Increased and more reliable potato production contributes to a stable food supply, supporting national food security goals. Consistent crop production helps meet the growing food demand while reducing dependence on imports.
- **Support for Rural Development:** Technology-driven solutions encourage innovation in rural communities, helping build a more resilient agricultural sector and improving overall quality of life in rural areas. The proposed system also promotes knowledge sharing and digital awareness among farmers through modern agricultural practices.

5. Expected Outcome

1. Accurate Disease Detection

The system helps identify potato leaf diseases with high precision, with over 99% accuracy. The trained CNN model automatically extracts disease-specific features from leaf images, enabling reliable classification under different environmental conditions. This means it can reliably distinguish between healthy and diseased leaves, and even classify specific types of diseases. This level of accuracy helps farmers make the right decisions quickly and confidently.

2. Improved Crop Yield

When diseases are detected early, farmers can respond swiftly to treat the plants before the disease spreads further. Timely disease management significantly reduces crop damage and improves the overall health of potato plantations. This prevents major crop damage, ensuring a larger portion of the harvest remains healthy. As a result, potato yields increase, which directly benefits farmers' productivity and income.

3. Cost Reduction for Farmers

Manual inspection of crops and unnecessary use of pesticides are both costly and time-consuming. Automation reduces the dependence on expert agricultural inspectors and minimizes operational expenses. With automated disease detection, farmers can reduce labor costs and use chemicals only when needed, leading to lower overall production costs and higher profit margins.

4. Increased Adoption of Technology in Agriculture

This project introduces advanced tools like machine learning and image processing to rural areas, encouraging farmers to adopt digital technologies. The integration of artificial intelligence into farming practices supports the transition toward smart agriculture. It helps modernize farming practices and promotes a culture of innovation in the agricultural sector.

5. Scalable Solution

Although the system is designed for potato leaves, the same method can be adapted for other crops like tomatoes, rice, or wheat by retraining the model with new datasets. The modular architecture allows the framework to be extended to additional crop disease datasets with minimal modifications. This makes the system scalable and useful for a wider range of crops and farming communities.

6. Enhanced Sustainability

The system helps reduce the overuse of chemical pesticides via early disease detection and support targeted early disease treatment. This targeted treatment approach minimizes environmental pollution while maintaining effective crop protection. This way, the system encourages ecofriendly farming practices. It helps protect the environment, improving soil health and promoting long-term sustainability in agriculture.

6. Result and Analysis

6.1 Training, Testing and the Validation

A sequential CNN model with seven layers is first developed. The dataset is divided into training and testing subsets to ensure unbiased evaluation of the proposed model [1][4]. The Adam optimizer is used to measure the performance error and optimize cross-entropy. Transfer learning is then applied to enhance the model. This is achieved by using built-in Keras modules and Keras Applications that provide pre-trained weights. The deep learning model is modified by removing the final layers and adding additional dense layers to align the pretrained weights with the desired output dimensions. Fine-tuning the network further improves convergence speed and increases the robustness of disease classification.

6.2 Model Performance

Fig. 2 illustrates the distribution of various potato diseases of this study. The distribution demonstrates that the dataset contains multiple disease categories, enabling comprehensive evaluation of the proposed CNN model. The most prevalent is scab of potato at 25%, followed by hollow heart of potato at 22.2%. Potato early blight accounts for 19.4%, while both potato leaf roll virus and soft rot of potato are equally present at 16.7% each. The balanced representation of disease classes improves the learning capability of the CNN and reduces classification bias.

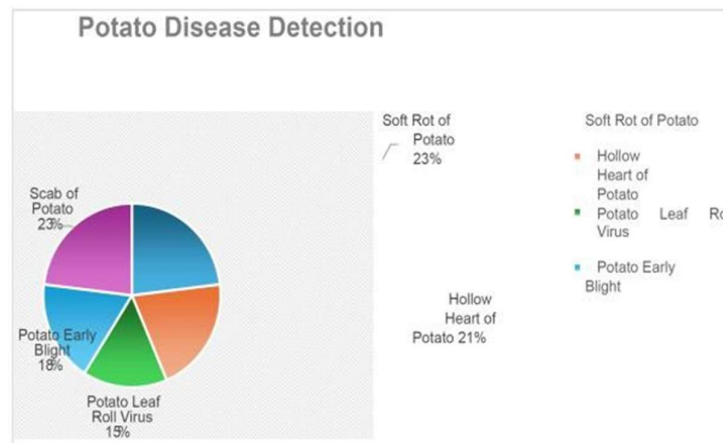


Fig. 2. Class Distribution of Potato Disease

7. Conclusion

This thesis focuses on disease detection in potatoes using machine learning techniques, specifically Convolutional Neural Networks (CNNs). The proposed framework demonstrates that deep learning can serve as an effective decision-support tool for modern precision agriculture. In a country like India, wherein the agricultural sector plays a major role in the economy, the prosperity and well-being of the staple crops becomes vital. Hence, disease detection and diagnosis becomes a key aspect and plays a vital role, as it helps identify diseases faster, which when left unchecked becomes a potential risk factor and can lead to major crop failures. The usage of CNNs is proven to be extremely effective in this field of disease detection using image processing of potato and potato plant's leaf. Within this research, a CNN model has been trained on a dataset which contains images of healthy and infected potatoes and potato leaves. When tested, it provided a generalization accuracy of 99%. This high accuracy indicated it's potential for real-world applications in day-to-day farmers' lives. It could greatly affect agricultural practices and the agricultural industry as a whole positively, especially for farmers with limited education and technical skills, lacking resources and a lack of necessary infrastructure in place. Currently, many farmers in rural areas not only find it difficult, but also often rely on traditional, less effective means of disease detection and management. Often times, a misdiagnosis leads to overuse of pesticides and insecticides, or even the use of wrong pesticides all together. This not only further damages the crops and delays the recovery of crops, but also leads to environmental pollution and degradation by means of these harmful pesticides mixing with nearby water bodies, as well as seeping deep underground and mixing with the group water, contaminating available freshwater sources. This also leads to further degradation of soil. This project is intended to bridge this disparity of knowledge by providing an easy and reliable method of disease diagnosis, which helps farmers not only protect their crops, but also enhance produce quantity and quality. Going forward, this project is planned to be further expanded and diversified. An android application is planned to be developed which is capable of not only finding diseases in potato plants, but multiple other plant species as well, ensuring that the application can be used by farmers of any kind and any scale. The application will make of the CNNs model, similar to the current system, to analyze and diagnose diseases in multiple plant species and provide solutions in accord to the type of plants. The accuracy of the system and the application is expected to increase vastly with addition of images of various different crop diseases present in different crop species. The app will serve as a platform wherein farmers can swiftly diagnose any diseases and treat their crops in a timely manner. Overall, this system has the potential to transform the agriculture sector as a whole, transforming traditional farming practices by incorporating machine learning tools and practices. Overtime, the system's aim is to be a user-friendly platform that aids farmers, improving their produce quantity and quality by ensuring early diagnosis of diseases, which in turn reduces crop failures and ensures that farming-practices as a whole are conducted in a environment friendly manner.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

- [1] Bienkowski, D., Aitkenhead, M. J., Lees, A. K., Gallagher, C., and Neilson, R. "Detection and differentiation between potato (*Solanum tuberosum*) diseases using calibration models trained with non-imaging spectrometry data" *Computers and Electronics in Agriculture*, vol. 167, pp. 105056, 2019.
- [2] Islam, M., Dinh, A., Wahid, K., and Bhowmik, P. "Detection of potato diseases using image segmentation and multiclass support vector machine. In 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE), pp. 1-4, 2017.
- [3] Iqbal, M. A., and Talukder, K. H. "Detection of potato disease using image segmentation and machine learning," In 2020 international conference on wireless communications signal processing and networking (WiSPNET), pp. 43-47, 2020.
- [4] Tiwari, D., Ashish, M., Gangwar, N., Sharma, A., Patel, S., and Bhardwaj, S. "Potato leaf diseases detection using deep learning." In 2020 4th international conference on intelligent computing and control systems (ICICCS) (pp. 461-466, 2020.
- [5] Islam, F., Hoq, M. N., and Rahman, C. M. "Application of transfer learning to detect potato disease from leaf image." In 2019 IEEE International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON), pp. 127-130. 2019.
- [6] Athanikar, G., and Badar, P. "Potato leaf diseases detection and classification system." *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 2, pp. 76-88. 2016.
- [7] Ranjan, M., Weginwar, M. R., Joshi, N., and Ingole, A. B. "Detection and classification of leaf disease using artificial neural network." *International journal of technical research and applications*, vol. 3, no. 3, pp. 331-333, 2015.
- [8] Pinki, F. T., Khatun, N., and Islam, S. M. "Content based paddy leaf disease recognition and remedy prediction using support vector machine." In 2017 20th international conference of computer and information technology (ICCIT), pp. 1-5, 2017.
- [9] Yao, Q., Guan, Z., Zhou, Y., Tang, J., Hu, Y., and Yang, B. "Application of support vector machine for detecting rice diseases using shape and color texture features." In 2009 international conference on engineering computation, 79-83, 2009.
- [10] Kumari, C. U., Prasad, S. J., and Mounika, G. "Leaf disease detection: feature extraction with K-means clustering and classification with ANN." In 2019 3rd international conference on computing methodologies and communication (ICCMC), pp. 1095-1098, 2019.