Potato Disease Detection Using CNN

B.P Chaitra[1], Sunil Kumar[2], Vipul Sharma[3], Abhishek Kumar[4] and Bhuvan H C[5]
[1,2,3,4,5]*Computer Science and Engineering, P.E.S College of Engineering, Mandya, Karnataka, India*

Abstract— In this study, a Convolutional Neural Network (CNN) is used to classify potato leaf diseases using machine Learning. The proposed method involves preprocessing leaf image data, training a CNN model on this data, and evaluating the model's performance on a test set. Experimental results demonstrate that the CNN model achieves an impressive overall accuracy of 93% in identifying Early Blight, Late Blight, and Healthy potato leaves. This approach presents a reliable and effective solution for disease identification in potatoes, crucial for maintaining food security and reducing agricultural losses. Importantly, the model performs well even in cases of severe infections. This research underscores the potential of machine learning techniques for classifying potato diseases, offering a valuable tool for automated disease management in potato farming.

Keywords— Convolutional neural networks, image processing, machine learning, potato disease detection, agricultural technology.

I. INTRODUCTION

Potato is a crucial global food crop and a major source of income for farmers. However, It is vulnerable to various diseases, especially leaf diseases, which can lead to significant yield and quality losses. Early and accurate disease detection is key for effective management. Traditional diagnosis methods rely on visual inspections, but these are subjective and time-consuming. Recently, there has been increasing interest in using machine learning for automatic disease detection.

Potato (Solanum tuberosum) is the fourth-ranked food crop globally, valued for their adaptability and high carbohydrate content. While they can be stored for a long time, storage diseases are common, affecting both the dining and processing markets. Wherever potatoes are produced, diseases both in the field and during storage can be a limiting factor in the ability to produce them sustainably and profitably.

Figure 1: Potato (Solatium Tuberosum)

In this study, we propose a machine learning method for categorizing potato leaf diseases. Specifically, we explore using convolutional neural networks (CNNs) for automatically detecting and classifying these diseases. Our main aim is to create a precise and effective approach for early disease detection and classification, enabling farmers to respond promptly and control the spread.

The model is designed to detect and identify various diseases affecting potato leaves, leveraging CNNs' ability to identify these diseases more effectively than the human eye. Remarkably, some pre-trained neural network architectures have achieved error rates as low as 3%, surpassing the best 5% error rate of human vision. The human top-5 error rate for large-scale images is 5.1%, higher than that of pre-trained networks.

II. LITERATURE SURVEY

Islam et al. (2017): Proposed a method combining image processing and machine learning to identify diseases in potato plant leaves using the 'Plant Village' database. Their method classified 300 images with a 95% accuracy rate, highlighting the importance of contemporary phenotyping and plant disease detection for food security and sustainable agriculture.

Velmurugan and Renuka Devi (2017): Developed software for automatic detection and classification of plant leaf diseases. They proposed a novel algorithm that achieved 94% accuracy on a database of about 500 plant leaves. The process involved color transformation, green pixel masking, segmentation, and texture statistics computation from SGDM matrices.

Dasgupta et al. (2019): Focused on addressing food scarcity in developing countries, particularly India, by tackling diseases like Early Blight and Late Blight in potato plants using deep learning and transfer learning techniques. The paper also reviewed literature on the fungus Alternaria Solani responsible for Early Blight disease.

Cecilia et al. (2019): Researched the use of image processing and machine learning to detect pests and diseases in blueberry plants. They created their own image database and used various filters and algorithms for image enhancement and noise removal. The study achieved an 84% accuracy rate using deep learning, although it lacked a literature review.

Iqbal and Talukder (2020): This study investigated the use of image processing and machine learning methods to identify and classify diseases in potato leaves. The authors processed 450 images of healthy and diseased potato leaves from the Plant Village database using image segmentation techniques. They employed seven different classifier algorithms to distinguish between healthy and diseased leaves, with the Random Forest classifier achieving the highest accuracy rate of 97%. The study also included a review of prior research utilizing machine learning and image processing for plant disease identification.

III. PROPOSED METHODOLOGY

Data collection, pre-processing of the data, data augmentation, and disease classification make up the main four stages of the methodology as shown in figure 2 :

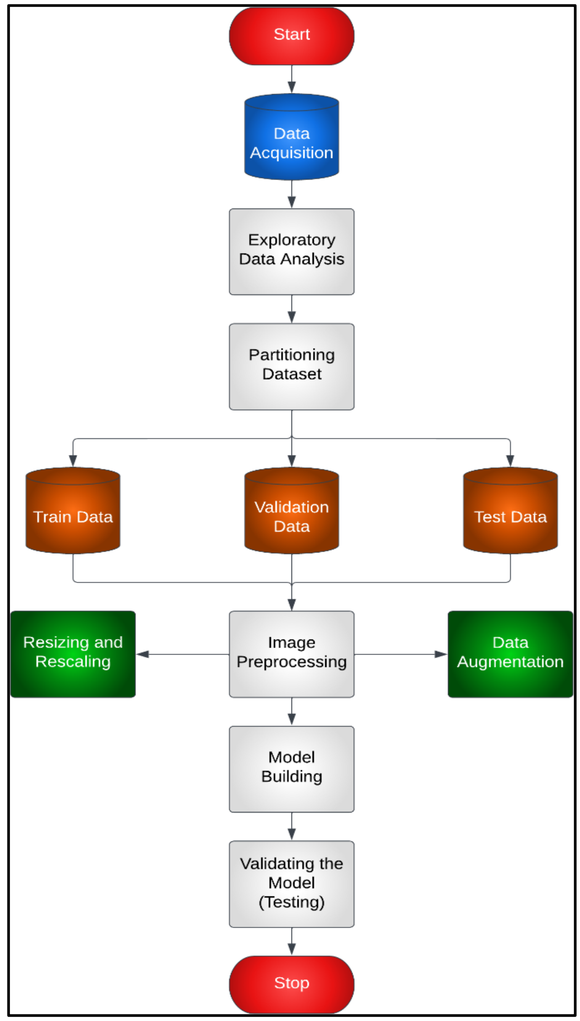

Figure 2: Methodology Flowchart

A. *Data Acquisition and Description*

A range of image sizes and resolutions were obtained from the Arjun Tejaswi-curated 'Plant Village' dataset [1] on Kaggle. There are around 20,600 photos in this collection, representing 15 different plant disease kinds. As seen in Figure 3, this dataset contains 2152 photos of potato plants that are divided into three categories: healthy, early blight, and late blight.

B. *Data Preprocessing*

Three classes comprise the potato-disease dataset. These classes consist of 1000 photographs for the Early-blight class, 152 images for the Healthy class, and 1000 images for the Late-blight class. As a result, 46.4%, 7.06%, and 46.4% of the dataset were split up. The dataset is divided into three parts: training (70%), testing (15%), and validation (15%).

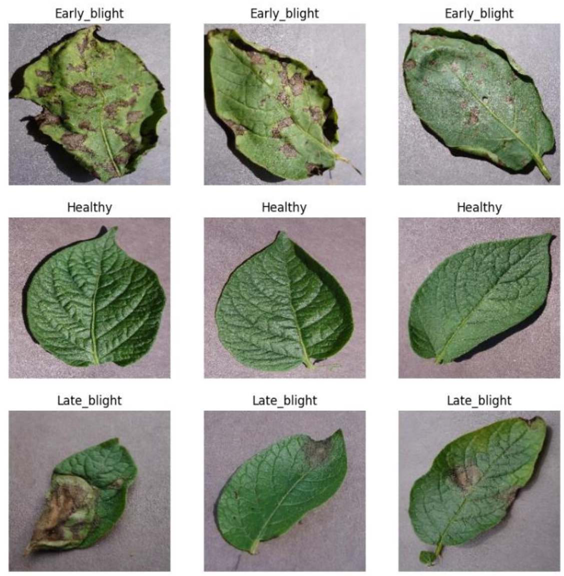

Figure 3: Images of Potato Leaves with diseases.

To make the process of splitting the dataset easier, the photos were separated into batches prior to splitting.

Additionally, several functions were implemented to make better use of the system's resources [2], such as "prefetch," which loads the next batch of photos onto the CPU while the GPU is busy training the images, and "cache," which reduces reading time.

C. Data Augmentation

A large amount of data is required for deep learning (Deep Network), as opposed to the shallow networks utilized in machine learning. Lack of training data and unequal amounts of data for each class are common problems for machine learning and deep learning [3]. Data augmentation is the strategy used to address this problem. The process of altering data without altering its original meaning is known as data augmentation. 5100 datasets are still not enough for this research to operate at its best, thus data augmentation is required.

However, because the leaves have few attributes to take into account, just the resizing and rescaling of the images (1/255), as well as the random rotation of 0.2 degrees, are used for this dataset.

D. Disease Classification using CNN

In artificial intelligence (AI), machine learning (ML) comprises deep learning (DL), also known as deep neural learning or deep neural networks. DL has more layers than ML. Deep learning approaches have improved its state-of-the-art capabilities in numerous disciplines, including as object detection, speech recognition, object categorization, and picture classification [4]. Convolutional Neural Nets are a widely used deep learning option. Based on the state of the leaves, CNN has been utilized in certain studies to identify plant illnesses [5]. Convolutional neural networks, or CNNs, typically consist of a few multi-layer convolutional layers ordered functionally. The stuff on CNN features is of an extremely high caliber. The last output layer in our multiclass categorization process is called Softmax.

In the first convolution (C1), the input picture in various sizes is convolved with 32 filters without any padding, resulting in 32 features with a size of 254x254. These parameters will be taught to the algorithms during the training phase. In every convolution layer, the activation function is a Rectified Linear Unit (reLU). It increases the nonlinearity of the images. The C1 layer is responsible for these traits.

The sample model of CNN can be visualized in fig. 4 as given below:

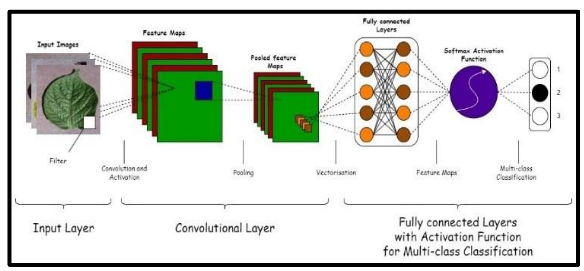

Figure 4: Convolution Neural Network (CNN)

The fifth and final convolution layer is (C5), where 64 feature maps, each measuring 3x3 pixels, are created using 64 filters. The fifth subsampling layer (S5) receives these features and uses them along with maximal pooling and a 2x2 window size. As a result, 36928 shared feature maps measuring 3 by 3 were created. Therefore, no dropout layer is used. A regularization technique called dropout reduces overfitting. Dropout is only used in the model's training phase. In the forward run, dropout prevents a subset of randomly chosen neurons from stimulating downstream neurons and does not update the weights of these neurons in the reverse run. In keeping with the dropout rate, this ignores the neurons during the training phase.

E. Causes and Preventive Measures

Early blight of potato, caused by the fungus Alternaria solani, thrives in warm temperatures (24-29°C) and high humidity. Plants stressed by poor nutrition, particularly nitrogen deficiency, or drought are more susceptible. To prevent early blight, crop rotation with non-host plants, using certified disease-free seeds, and removing infected plant debris are essential cultural practices. Planting resistant or tolerant varieties, ensuring balanced fertilization, particularly nitrogen, and using drip irrigation while watering in the morning can also help.

Late blight of potato, caused by the oomycete Phytophthora infestans, favors cool (15-20°C) and wet conditions with high humidity. Dense canopies and prolonged wet leaf surfaces increase the risk of infection. Preventive measures include crop rotation with non-host crops, destroying volunteer plants and cull piles, and planting certified disease-free seed potatoes. Proper hilling of potatoes during planting and harvesting during dry weather are also important.

IV. RESULTS AND DISCUSSIONS

A. Training Process

The dataset learning experiment made use of the five-layer Convolutional Neural Network architectural model. A 32-batch size, 50 epochs, and a 0.01 learning rate were specified in the study's period specification in order to improve the model's performance. In order to learn how to recognize new photos, the algorithm of the proposed method [6] looks for values in the image collection. The value of each epoch phase needs to match the value of the image that is being learned. The results of the epochs are recorded in order to compute the loss and accuracy values. The accuracy value shows how accurately objects are classified by the system, and the gain in accuracy must be close to or equal to zero. A loss is a sign that the model's value is low. In both plots shown in fig. 5, the findings of the Accuracy and Loss values are displayed. While training, the early and later epochs show more fluctuation values. After the 50th epoch, the training dataset's accuracy, which includes the original images, achieves 93%. The CNN needed 7000 seconds, or approximately 1.9 hours, to complete 50 epochs at an average speed of 104 seconds per epoch.

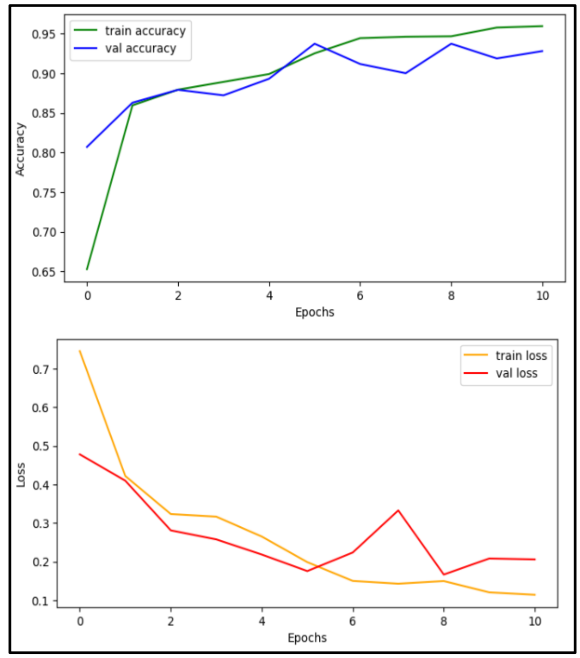

Figure 5: Accuracy and Loss recorded while Training the Dataset

B. Testing Process

After completing the training procedure using the generated dataset, We tested the dataset for different results. The testing loss was reported to be 2.48% and testing accuracy was 93%.

Also, the model trained and tested is capable of signifying the confidence up to what extent it is able to classify or predict the disease. The three results are shown below:

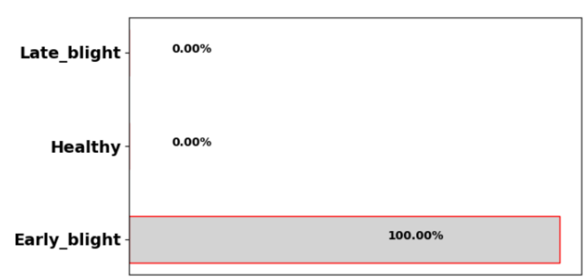

Fig 6.1: Early Blight

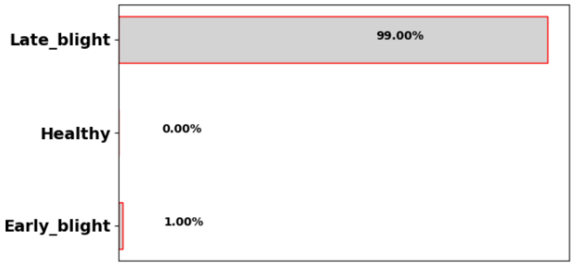

Fig 6.2: Late Blight

Fig 6.3: Healthy

V. CONCLUSION

In this study, we proposed a CNN (Convolutional Neural Network) deep learning approach for classifying potato leaf diseases. The suggested approach consists of pre-processing the leaf images, training a CNN model on the pre-processed data, and evaluating the model's performance on a test set. The results of the trial demonstrated that the CNN model [7] performed exceptionally well in the classification of Early Blight and Late Blight, two distinct potato leaf diseases, with an overall accuracy of 93%.

To sum up, this work shows the promise of deep learning-based methods for managing and diagnosing plant diseases. The proposed method enables early identification of potato leaf diseases, which can significantly improve crop protection and global food security. The development of more efficient and sustainable agricultural techniques is made possible by the extension of this strategy to different crops and environmental circumstances.

VI. REFERENCES

[1] 1 A. Tejaswi, "Plant Village Dataset", Kaggle, version 1, Mar 2020, Accessed Mar. 15, 2024. Available: https://www.kaggle.com/datasets/arjuntejaswi/plant-village

[2] Wikipedia, "Training, validation, and test data sets," *Wikipedia*, Jun. 07, 2019. Accessed: Mar. 26, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Training,_validation,_and_test_data_sets

[3] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, Sep. 2016, doi: https://doi.org/10.3389/fpls.2016.01419

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature Review Articles*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: https://doi.org/10.1038/nature14539

[5] E. Fujita, Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic Investigation on a Robust and Practical Plant Diagnostic System," *2016 IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2016,doi: https://doi.org/10.1109/ICMLA.2016.0178

[6] R. A. Sholihati, I. A. Sulistijono, A. Risnumawan, and E. Kusumawati, "Potato Leaf Disease Classification Using Deep Learning Approach," *2020 International Electronics Symposium (IES)*, Sep. 2020, doi: https://doi.org/10.1109/ies50839.2020.9231784

[7] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases" IJCA, vol. 17(1), pp. 31-38, March 2011, IEEE2010.