Journal of Scientific and Technological Research, Vol.5 (1), 2023 **Print ISSN: 2309-7167**
https://doi.org/10.59738/jstr.v5i1.23(27-35).kacr3648 **e- ISSN: 2959-7943**

বাংলাদেশ উন্মুক্ত বিশ্ববিদ্যালয়
BANGLADESH OPEN UNIVERSITY

Journal of Scientific and Technological Research

Potato Leaf Disease Prediction: A Machine Learning Perspective

Md Moshiur Rahman[1*], Tamima Yeasmin[1], Md Jahirul Islam[1], Md Dulal Mahmud[2], Md Mahbubur Rahman[3]

[1]School of Science and Technology, Bangladesh Open University, Gazipur-1705, Bangladesh.
[2]Open School, Bangladesh Open University, Gazipur-1705, Bangladesh
[3]Department of Computer Science and Information Technology, Patuakhali Science and Technology University, Patuakhali-8602, Bangladesh

Abstract

Potato leaf disease has mostly two categories; early blight and late blight disease. The disease may be more prevalent in certain weather patterns and have a catastrophic effect on potato crops. In summary, warm, humid weather with frequent rain or heavy dew, temperatures between 15°C and 20°C, and a lack of sunshine are the weather conditions that can cause potato late blight. Drier weather conditions favour early blight, unlike late blight. Warm and dry weather with a lack of rain or irrigation, temperatures between 21°C and 29°C, and high humidity in the morning are the weather conditions that can cause potato early blight. A modified dataset is used for climate-influenced prediction, and the testing accuracy using random forest models is 97%. Analysis of experimental results shows that the suggested potato-leaf disease prediction based on the weather data framework outperforms the outcome of frameworks.

Keywords: Potato, leaf disease, prediction, early blight & late blight.

1. Introduction

Potato late blight is a fungal disease that is caused by Phytophthora infestans. Specific weather conditions can favour the disease, and it can devastate potato crops. Usually, there is a lot of work done on potato leaf disease detection based on image processing using CNN. even though there is also work on weather detection, and this is also based on image processing. This paper tries to predict potato leaf disease (Sharma et al., 2018) based on weather parameters like warm, humid, dry weather, rain status, temperature, sunshine, etc. Previously done potato leaf disease detection can only detect the disease if it has already occurred to the potato leaf. Due to insufficient research in the integration of regional agriculture and weather data, it cannot forecast the disease before it impacts the leaves. Here, the internet collects agro-weather information that effectively helps predict potato leaf diseases. Potato leaf disease prediction based on a single weather parameter like temperature may not be accurate, and hence, real-time disease data is critical for accurate disease prediction and diagnosis. This adds a new dimension to the diagnosis of potato leaf disease. This research is motivated by the urgent need to address the significant threat of potato leaf diseases to agricultural yields. The goal is to proactively manage and mitigate risks by leveraging a climate-influenced prediction model, particularly through advanced data analysis methods like random forest models.

1.1 Research summary

*Corresponding author: Md. Moshiur Rahman (moshiurbou@gmail.com)
Journal homepage: jstr.bousst.edu.bd

This study used a weather dataset that was active from 2006 to the end of 2016. This set of weather data was modified based on our findings. For this approach, we developed a code that consisted of modifying the CSV file and making a new CSV file that merged information about weather and potato leaf diseases. Then the data set is pre-processed for the next processing. The data set is prepared and trained along with a specific label for each classification, which is done in Python through the “Visual Studio Code” platform. After all procedures have been completed, the researcher uses disease prediction methods such as Linear Regression, Decision Tree Regression, Logistic Regression, Decision Tree Classifier, and Random Forest Classifier. Then compared the prediction performance of different regression and classification algorithms on the same weather dataset. Finally, the researcher find the best solution using classification models for predicting potato leaf disease based on weather details.

1.2 Problem Statement

The core problem is the necessity for accurate and prompt prediction of potato leaf diseases, especially early blight and late blight, based on climate conditions. Current frameworks may have limitations, creating a research gap in optimizing predictive models tailored to the climate-influenced aspects of these diseases. The study aims to fill this gap by proposing a modified dataset and utilizing various models to enhance accuracy, providing a more reliable tool for farmers and agricultural stakeholders.

1.3 Objectives

The study's objectives are diverse and designed to holistically address the challenges posed by potato leaf diseases, particularly early blight and late blight. Firstly, the research endeavors to enhance our comprehension of the environmental conditions influencing the prevalence of these diseases. Through a thorough investigation, the aim is to identify key factors contributing to the manifestation of potato leaf diseases. Secondly, the study seeks to develop and deploy an advanced climate-influenced prediction model, utilizing a modified dataset to elevate the accuracy of disease predictions. Thirdly, the research strives to evaluate and optimize the effectiveness of various models, assessing their performance in predicting diseases under various weather conditions. Additionally, the objectives encompass addressing existing research gaps by proposing a model that tailors predictions to the climate-influenced aspects of potato leaf diseases. The study also includes a comparative analysis to showcase the superiority of the proposed model over existing frameworks. Moreover, the overarching objective is to empower farmers and agricultural stakeholders with a reliable tool for anticipating, managing, and mitigating the risks associated with potato leaf diseases. Ultimately, this research contributes to the broader goal of food security by refining disease management strategies through the implementation of advanced prediction models.

1.4 Challenges

This study's first task in writing this essay was to gather combined weather and potato leaf disease status data and then to preprocess the data to detect diseases. Because the challenge is to find the correct dataset for our project. Unfortunately, researchers have searched a lot, but there are very few datasets available on any authenticated platform, not even on any open-source platform. After analyzing various news, paper-cutting, and weather details and collecting data for our project. This study only focuses on the work of potato leaf disease prediction based on weather details (Shelar et al., 2022) so that can help our farmers. The next challenge was making perfect use of the dataset using selected regression and classifier models and making the output visible. If we did not use the dataset correctly, the researcher would not find the correct result and our challenge was making the confusion matrix visible.

1.5 Background study

Both late blight and early blight are common and dangerous illnesses (Singh et al., 2018). Both are available anywhere potatoes are planted. The phrases "early" and "late" refer to the relative timing of the onset of each disease, even if they can both happen at the same time.

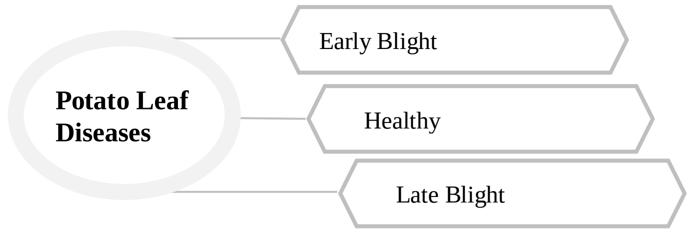

Figure 1: Types of Potato Leaf Disease

Potato early blight is caused by the fungus Alternaria solani, which can also infect tomatoes, other plants in the potato family, and some varieties of mustard. Young, actively growing plants are rarely affected by the disease, often called target spots. It initially appears on older leaves. High humidity levels and warm temperatures encourage the growth of early blight. A dangerous disease called late blight of potatoes is brought on by Phytophthora infestans. It affects potatoes, tomatoes, and occasionally eggplant and other members of the potato family. Late blight is the most dangerous potato disease. In the 1830s, reports of it initially appeared in Europe and the US. It is well-known for igniting the Irish Potato Famine in the 1840s, which resulted in 1.5 million emigrants and a million starvations. The first fungicide was identified in the 1880s, but late blight remained a terrible issue until then. In recent years, it has come up again as an issue. If conditions are right, it can kill plants in two weeks and favours chilly, damp weather (Daniya & Vigneshwari, 2019). The undersides of leaves develop fluffy, white fungal growth when the weather is damp.

1.6 Related works

Plant leaf disease identification and categorization is a challenging process. The following Table 1 shows the list of techniques used in related works in recent.

Table 1: The List of techniques used in related works.

Year	Author	Method	Application area
2023	Gupta and Shah	CNN and IoT	Potato leaf disease identification and mitigation (Gupta & Shah, 2023)
2022	Li et al.	Mask R-CNN, VGG16, ResNet50, InceptionV3, UNet, PSPNet, DeepLabV3+	Potato leaf disease segmentation and detection in complex backgrounds (Li et al., 2022)
2017	Kamble et al.	K-means clustering and SVM	Potato leaf disease classification (IEEE Staff, 2017)
2017	Trimi Neha Tete et al.	Neural network, K-means, and thresholding	Identification of disease from potato, apple, and mango leaves (Tete & Kamlu, 2017)
2015	Prakash M. Mainkar et al.	K-means clustering, grey-level Co-occurrence matrix, and Backpropagation neural network	Identification of disease from potato, tomato, and cotton leaves (Chouhan et al., 2018)

Many researchers have segmented the sick potato leaf using both traditional and soft computing methods. We try to provide a brief overview of some of the soft computing methods used to finish this work in this part. The most popular method for determining and classifying the disease in an image of a potato leaf that has been infected is K-means clustering. Because of its capacity for learning and training, CNN has also been utilized for this mission. The number of soft computing methods utilized to identify potato leaf disease is displayed in Table 1. Each source's application area, author(s), method(s) utilized, and year of publication are displayed in the table. Regretfully, there isn't any prior research on weather-related potato leaf disease.

2. Dataset

To help machine-learning models test and validate the potato leaf disease prediction system, a thorough and trustworthy dataset is needed. Therefore, this study has used weather history details. It is a dataset that was collected from 2006 to 2016. This set of weather data has 96453 weather data, but no dataset combines weather and potato leaf disease. But for our project, we need a dataset that combines weather details and corresponding weather that can cause early blight or late blight of potato leaf disease. For this purpose, we modified the dataset based on collected knowledge and different resources to ensure that the weather is a reason for early blight or late blight. An hourly/daily overview of the Szeged, Hungary region is included in the CSV file for the years 2006–2016. Hourly response data are given for time, precipitation type, temperature, humidity, wind direction, wind speed, visibility, cloud cover, and pressure. The researchers have studied as much as possible and found out that there are different values for both diseases and collected a lot of data from different sources about the potato leaf disease and mentioned it below to ensure that there is no dataset in Kaggle or other resources, even though we have tried to collect data from BARI (Bangladesh Agricultural Research Institute), but there is no proper dataset; hence, the researcher has tried to collect data from previous history and tried to modify our dataset with proper modification. The collected conditions are given below:

In a nutshell, potato blight is caused by weather conditions. The fungal disease Phytophthora infestans causes potato blight, also known as late blight. This disease, which is favoured by certain weather conditions, can be devastating to potato crops (Sharma et al., 2018). Warm, humid weather with frequent rain or heavy dew, temperatures between 15°C and 20°C, and a lack of sunshine are the weather conditions that can cause potato blight. In short, weather conditions cause potato early blight- Potato early blight is a fungal disease caused by Alternaria solani. Drier weather conditions favour early blight, unlike late blight. In summary, warm and dry weather with a lack of rain or irrigation, temperatures between 21°C and 29°C, and high humidity in the morning are the weather conditions that can cause potato early blight. The prediction of potato leaf disease here are two new columns related to potato leaf disease, and they are disease name and disease in numerical value (Sladojevic et al., 2016).

Table 2: CSV dataset details based on weather details

SL	Class	Samples	Training data	Test data
0	Healthy	93699	74955	18741
1	Late Blight	712	1623	421
2	Early Blight	2042	584	129
Total		96454	77162	19291

2.1 Methodology

It is crucial to choose features for categorization that specify the class to which the new data belongs. From this point forward, all of the weather parameters were collected from the dataset, and all parameters were included as features in the dataset. It did not require a data cleaning and transformation procedure for our project datasets since we employed a balanced dataset. The flowchart of this work's process (Figure 2) shows the first phase, which begins with the combination of late blight and early blight disease with healthy classification files from a dataset publicly available on Kaggle.com. Next, the researcher used their own Python script that was created in a Visual Studio code environment with support for Jupyter Notebooks to extract the static features. The dataset, which was prepared and saved as a data frame in a Comma-Separated Values (CSV) file for use in the training procedure, solely included weather information and illness characteristics. Lastly, to identify the proper metrics for assessment, we cross-validate each classifier model. (Mohan and others, 2020).

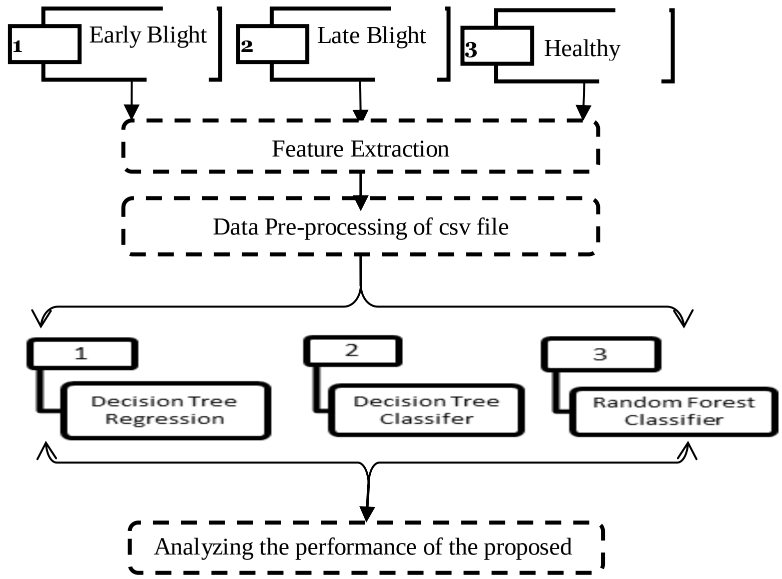

Figure 2: Potato leaf prediction phases

3. Experimental Result

This study begins by training the regression and classifier methods on the dataset (Tilva et al., 2013), followed by testing and evaluation.

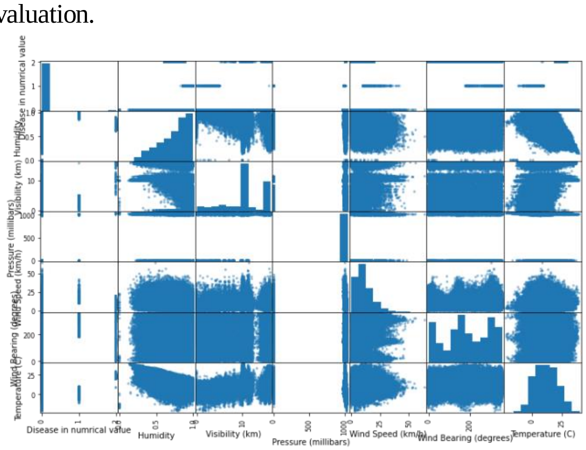

Figure 3: Dataset feature extraction

Considering that all methods are regarded as supervised learning, the researcher decided to combine logistic and linear regression with Random Forest, Decision Tree Regression, and Decision Tree Classifier to create a practical model for disease prediction. They can assist with tasks involving both regression and classification, and they are all easy to use and put into practice. The outcomes of our classification models, including the f1-score, accuracy, precision, recall, and confusion matrix, are now displayed as follows in Table 3.

Table 3: An output of the classification model

	precision	recall	f1-score	support
Healty	0.973126	0.997532	0.985178	14992.000000
Early Blight	0.575000	0.190083	0.285714	121.000000
Late Blight	0.200000	0.015625	0.028986	320.000000
accuracy	0.970842	0.970842	0.970842	0.970842
macro avg	0.582709	0.401080	0.433293	15433.000000
weighted avg	0.953974	0.970842	0.959867	15433.000000

The confusion matrix of the random forest algorithm and CNN is given below:

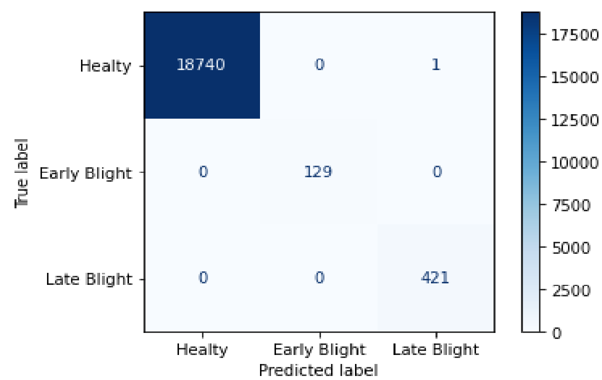

Figure 4: Confusion matrix output of random forest

This study compared the numbers. Figure 5 (b) compared. Given that all of the methods can detect a sizable number of occurrences, the figure validates the quality of the dataset.

```
Linear Regression -- shows error
Scores:  [0.01138387 0.00804961 0.02276847 0.01609974 0.01800006]
Mean:  0.015260349483266789
Standard Deviation:  0.005130387241347213

Decision Tree Regressor -- shows error
Scores:  [0.01138387 0.00804961 0.02276847 0.01609974 0.01800006]
Mean:  0.015260349483266789
Standard Deviation:  0.005130387241347213

Logistic Regression -- shows accuracy
Scores:  [0.97161926 0.97168405 0.97174702 0.97129342 0.97077501]
Mean:  0.9714237519011004
Standard Deviation:  0.0003600918598215322

Decision Tree Classifier -- shows accuracy
Scores:  [0.9897622  0.99041016 0.99040954 0.99034474 0.99079834]
Mean:  0.9903449951742103
Standard Deviation:  0.0003327424221015569

Random Forest Classifier -- shows accuracy
Scores:  [0.9999352 0.9999352 0.9998056 0.9998056 0.9996112]
Mean:  0.9998185605183052
Standard Deviation:  0.00011878277148323273
```

Figure 5(a): Algorithm performance measurement

Figure 5(b): Performance Comparison Bar Chart

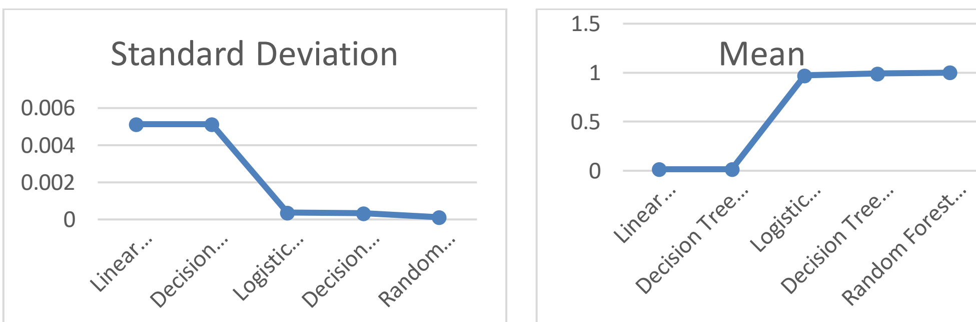

Figure 6: Standard deviation and mean

With a score of 98% and the highest score for accurate predictions—the Random Forest classifier predicted 99% of the dataset correctly—it demonstrated the best degree of accuracy among the classifiers. Furthermore, the Decision Tree classifier correctly predicted 97% of the disease-affected data, achieving the second level of precision. Nevertheless, during testing, the Random Forest Classifier identified 97% of the whole dataset, achieving the highest level of recall.

4. Conclusion

The proposed prediction was performed using a customized dataset, and the analysis was successful. Using this, it was a good idea to predict the disease rather than detect it. The top classification techniques using the current data set are Random Forest and Decision Tree classifiers, which properly categorize the disease with an accuracy of over 95%. This could be a helpful screening method for spotting diseases of the potato leaf. The set of features found in the weather set is the best-performing feature set; however, a suitable dataset is not provided here. Herewith, the researcher has successfully developed a framework for predicting the disease before it is detected. This type of framework will help our farmer or agriculture section.

Limitations

A smart tool is proposed here that can predict potato leaf disease using weather data. So far, we have used weather data that we have built using local sources of information. The researcher do not have any authentic sources of weather information. So, the main limitation of the project is authentic weather data, which is required to correctly predict potato leaf disease in real life. Since the severity of plant diseases varies over time, the models should be enhanced or updated to enable them to identify and categorize illnesses throughout their whole cycle.

Future Directions

The prediction of potato leaf disease still poses several difficulties and significant problems that demand further research and creative solutions. By training the model with authenticated data, we can achieve further improvements. Additionally, researchers may update the model and include new kinds (Kour & Arora, 2019) of plant diseases like tomato, rice, etc. Additional Natural Language Processing may be applied to illnesses to provide suggestions, cures, medications, and sprays. However, the real environment needs to be taken into account for a realistic scenario. It would be more cost-effective to prevent the needless use of fungicides, pesticides, or herbicides if a more effective method of identifying diseased plant patches were developed.

References

Asif, M. K. R., Rahman, M. A., & Hena, M. H. (2020). CNN-based disease detection approach on potato leaves. *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020*, 428–432. https://doi.org/10.1109/ICISS49785.2020.9316021

Chouhan, S. S., Kaul, A., Singh, U. P., & Jain, S. (2018). 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*, *6*(i), 8852–8863. https://doi.org/10.1109/ACCESS.2018.2800685

Daniya, T., & Vigneshwari, S. (2019). A review on machine learning techniques for rice plant disease detection in agricultural research. *International Journal of Advanced Science and Technology*, *28*(13), 49–62.

Gupta, H. K., & Shah, H. R. (2023). Deep Learning-Based Approach to Identify the Potato Leaf Disease and Help in Mitigation Using IOT. *SN Computer Science*, *4*(4), 333. https://doi.org/10.1007/s42979-023-01758-5

IEEE Staff. (2017). *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*. 8–11.

Kour, V. P., & Arora, S. (2019). Fruit Disease Detection Using Rule-Based Classification. In *Advances in Intelligent Systems and Computing* (Vol. 851, Issue March). Springer Singapore. https://doi.org/10.1007/978-981-13-2414-7_28

Li, X., Zhou, Y., Liu, J., Wang, L., Zhang, J., & Fan, X. (2022). The Detection Method of Potato Foliage Diseases in Complex Background Based on Instance Segmentation and Semantic Segmentation. *Frontiers in Plant Science*, *13*(July), 1–13. https://doi.org/10.3389/fpls.2022.899754

Mohan, A., Srivastava, K., Malhotra, G., & Khan, N. U. (2020). Plant disease detection using clustering-based segmentation and neural networks. *PDGC 2020 - 2020 6th International Conference on Parallel, Distributed and Grid Computing*, 506–510. https://doi.org/10.1109/PDGC50313.2020.9315856

Rajan, P. (2016). A Survey on Different Image Processing Techniques for Pest Identification and Plant Disease Detection. *IJCSN International Journal of Computer Science and Network*, *5*(1), 2277–5420. www.IJCSN.org

Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants*, *8*(11), 32–34. https://doi.org/10.3390/plants8110468

Sharma, P., Singh, B. K., & Singh, R. P. (2018). Prediction of Potato Late Blight Disease Based Upon Weather Parameters Using Artificial Neural Network Approach. *2018 9th International Conference on Computing, Communication, and Networking Technologies, ICCCNT 2018*, 1–13. https://doi.org/10.1109/ICCCNT.2018.8494024

Shelar, N., Shinde, S., Sawant, S., Dhumal, S., & Fakir, K. (2022). Plant Disease Detection Using CNN. *ITM Web of Conferences*, *44*(4), 03049. https://doi.org/10.1051/itmconf/20224403049

Singh, B. K., Singh, R. P., Bisen, T., & Kharayat, S. (2018). Disease Manifestation Prediction from Weather Data Using Extreme Learning Machine. *Proceedings - 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages, IoT-SIU 2018*, 1–6. https://doi.org/10.1109/IoT-SIU.2018.8519908

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Computational Intelligence and Neuroscience*, *2016*. https://doi.org/10.1155/2016/3289801

Tete, T. N., & Kamlu, S. (2017). Plant Disease Detection Using Different Algorithms. *Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering*, *10*, 103–106. https://doi.org/10.15439/2017r24

Tilva, V., Patel, J., & Bhatt, C. (2013). Weather-based plant disease forecasting using fuzzy logic. *2013 Nirma University International Conference on Engineering, NUiCONE 2013*. https://doi.org/10.1109/NUiCONE.2013.6780173