2025年臺灣國際科學展覽會
優勝作品專輯

作品編號　200019

參展科別　環境工程

作品名稱　Revolutionizing Potato Agriculture: Harnessing Machine Learning Techniques for Disease Detection and Management

得獎獎項　三等獎

就讀學校　Aitchison College, Lahore
指導教師　Syeda Khadija Bashir
作者姓名　Rafey Ataullah
Mohammad Lajwar Jamal

作者照片

Investigating the Utilization of Emerging (IoT) Sensor Technology in Potato Agriculture for Crop Disease Prediction and Preventive Measures in the Context of Smart Farming and Climate Change

Authors: Rafey Ataullah, Mohammad Lajwar Jamal

Table of Contents

Abstract

Aim: The aim of this study is to make a disease-predicting model trained on data from weather stations and API using machine learning that gives the farmer the ability to predict crop diseases before they set in, allowing them to take timely preventative measures and reduce wastage.

Materials and Methods: In this study the Internet of Things (IoT) sensors throughout agricultural fields of potato crops in Jafferabad, Depalpur Punjab. The sensors collect real-time data on environmental conditions, such as precipitation, air temperature, relative humidity, wind speed, and direction, Dew Point, VPD, and the Delta T values, to identify subtle disease indicators and patterns within the environmental data. Our novel machine-learning program makes use of the data collected by the weather station and analyses them.

Results: Using the data, one predictive statistical method using Python 3.8.0 was created which uses the data from the weather station which can predict diseases before they set in.

1. Introduction

In an era marked by an increasing global population and the ever-growing demand for food, ensuring food security is of paramount importance. Globally, more than one billion people consume potatoes hence making potatoes the third most important food crop in the world. [1] Potatoes are a sustainable crop, as the potato plant grows more food on comparatively less land, less time, and less water than other crops. As a result, the importance of the potato is increasing globally. [2]

In Pakistan, Solanum tuberosum, the potato crop is mostly sown in spring, summer, and autumn in different agro ecological regions ranging from plains to hilly areas. In Punjab, most of the potato crop is cultivated in Kasur, Depalpur, Sahiwal, and Pakpattan. [3] Potato cultivation is less labor and time-intensive from sowing till harvest (less than 90 days) which is considerably less than other major crops. However, despite all these advantages potato productivity in Pakistan is not promising compared to other developing countries. [4]

Potato crop all over the world faces significant challenges, particularly in the form of early and late blight of potatoes, bacterial common scabs, wilt and soft rot, diverse viral infections, and nematodes. These diseases, if left unchecked, can devastate crops and lead to food shortages, impacting communities and economies worldwide. There is a dire need to develop forecast and early warning services for predicting the time of appearance of disease and hence optimizing the use of fungicides etc. [5]

The relationship between the weather and the disease in potato crops is well established. In some diseases strains of bacteria are present in almost all tubers, including seed, but generally do not cause disease and remain dormant unless weather conditions are favorable. [6] [7]

Given the limitations of traditional farming practices and the significant impact of weather on potato health, cost-effective weather forecasts and early disease warnings are crucial. Such technology would empower Pakistan's small-scale potato growers (predominantly cultivating less than 25 hectares) to optimize resource allocation and implement timely disease control measures.

The aim and objective of this study is that by integrating IoT sensors for real-time data collection about environmental conditions, for data analysis and prediction, we can create a sophisticated system that enables early detection of crop diseases. This early warning system will allow farmers and agricultural stakeholders to take proactive measures to prevent the spread of diseases. It will also lead to more efficient and sustainable agricultural practices.

In an era where agriculture is increasingly vulnerable to shifting climate patterns and emerging diseases, this research introduces a novel paradigm by harnessing the power of IoT sensors to collect real-time data from the fields and then convert them into a location-specific forecast model.

2. Materials and Methods

Initially, the Internet of Things (IoT) sensor (Weather Station- Pessl Instruments uMetos 3.3) is strategically positioned in the agricultural fields of potato crops in Jaffarabad, Depalpur, and Kasur Punjab to continuously collect real-time data on key environmental factors such as temperature, humidity, soil moisture, and air quality. We have used two parameters temperature and humidity which are more relevant to potato disease progression (6,7). The Internet of Things (IoT) sensor used was specifically the Pessl Instruments iMetos 3.3 Weather Station. This advanced sensor continuously gathers real-time data on crucial environmental factors, including temperature, humidity, soil moisture, and air quality.

Focusing on the parameters of temperature and humidity, which are particularly relevant to the progression of potato diseases, our research aims to develop a predictive model trained on data from weather stations and API, for anticipating the onset of diseases before they manifest. The core research process involves the seamless integration of the data from the weather station in a predictive statistical model using Python 3.8.0 which can predict the onset of disease before they set in.

To assess the practical applicability of this predictive model, extensive field trials will be conducted in real-world farming environments in the months of December 2024 and January 2025. These trials will further accuracy in real-time.

2.1 Data Observations

Crop health data is being systematically collected. The data set employed in this study was sourced directly from across agricultural fields. The weather station is continuously monitoring various environmental factors, including temperature, humidity, soil moisture, and other relevant parameters. The most critical weather conditions (input features) for the appearance of potato crop diseases in Punjab and how they affected the initial inoculum and disease development were tabulated by conducting a literature review of the relevant studies. The most important weather variables that impacted disease progression were identified to be temperature (minimum and maximum), relative humidity, and precipitation.

The dataset employed in this study was sourced directly from the weather station in the form of a CSV file. For analysis, only the average temperature and average humidity columns were utilized due to their substantial relevance in studying disease-spreading dynamics. The dataset underwent labeling based on the rules obtained through a literature review and was subsequently divided into two subsets: the training data, constituting 80%, and the test data, constituting 20%. This partitioning aimed to comprehensively assess the model's performance, ensuring robust evaluation using independent test data.

The temperature data recorded by the weather station and the data recorded by the OpenMeteo API are mostly in accordance with each other. The key controlling variable is humidity.

2.2 Algorithm Design

The machine learning algorithm for predicting early and late blight in potato crops was designed to integrate advanced data processing techniques with agricultural domain knowledge. The foundation of the algorithm relied on leveraging historical weather data and disease classifications to establish patterns and predict the likelihood of disease occurrence under varying environmental conditions. Data was acquired from the OpenWeatherMap API, which provided key weather features such as daily temperature, humidity, and precipitation forecasts. These raw features were further processed to include additional variables such as the day of the week, month of the year, and averages of temperature and humidity, which served to capture temporal and seasonal effects on disease prevalence.

The preprocessing pipeline included encoding the target variable, crop disease type, using LabelEncoder to convert categorical disease labels (e.g., "Early Blight," "Late Blight," "Healthy") into numerical values suitable for model training. An 80-20 train-test split was employed to ensure reliable performance evaluation. After evaluating several models, a decision tree regressor was chosen for its interpretability, computational efficiency, and suitability for the problem's data distribution. The model was fine-tuned to maximize its predictive performance, taking into account the specific environmental thresholds linked to crop diseases. For instance, "Late Blight" was associated with prolonged periods of high humidity and moderate temperatures, while "Early Blight" was linked to higher temperature conditions.

The algorithm was structured to incorporate real-time weather data, enabling dynamic predictions of disease risks up to seven days in advance. A custom thresholding system was implemented to translate the model's output into actionable disease classifications. This decision-making process was informed by agronomic expertise and aligned with established disease progression models. The final system was deployed as an automated pipeline capable of fetching live data from the

OpenMeteo API, processing the inputs, and predicting the likelihood of disease occurrence. The seamless integration of data acquisition, preprocessing, model prediction, and disease classification underscores the practical applicability of the algorithm in providing early warnings for farmers and agricultural stakeholders.

A mobile application was also built using React Native. The application takes a location input from the user and calls the OpenMeteo API to fetch future forecasts. Future weather condition predictions are compared with the disease determinants in Fig. 1.3 based on users' location. The purpose of the application is to apply this research meaningfully and assist farmers. Moving forward we will also be aiming to use multinomial regression or similar models to calculate the percentage likeliness of disease as well as use a paid API to increase the prediction period from 5 days to 30 days. The prediction region to other potato-producing countries shall be increased as well. Moreover, the mobile application will be made available on platforms like the Google Play Store and the Apple App Store.

Train the Model

```
[ ] from sklearn.metrics import confusion_matrix, classification_report
    import seaborn as sns
    import numpy as np

    model = DecisionTreeRegressor(max_depth=5, random_state=42)
    model.fit(X_train, y_train)

    y_pred_encoded = model.predict(X_test)

    y_pred_encoded = np.rint(y_pred_encoded).astype(int)


    y_pred = label_encoder.inverse_transform(y_pred_encoded)
    y_true = label_encoder.inverse_transform(y_test)

    conf_matrix = confusion_matrix(y_true, y_pred, labels=label_encoder.classes_)
    conf_report = classification_report(y_true, y_pred, target_names=label_encoder.classes_)


    print("Classification Report:")
    print(conf_report)


    plt.figure(figsize=(8, 6))
    sns.heatmap(conf_matrix, annot=True, fmt="d", cmap="Blues", xticklabels=label_encoder.classes_, yticklabels=label_encoder.classes_)
    plt.title("Confusion Matrix")
    plt.xlabel("Predicted Labels")
    plt.ylabel("True Labels")
    plt.show()
```

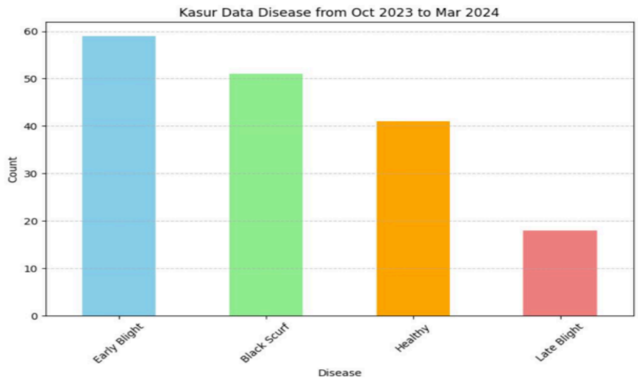
Figure 1.0 Python code to train the model

2.3 Results

Results

The performance evaluation of the machine learning algorithm demonstrated its effectiveness in accurately predicting early blight, late blight, and healthy conditions in potato crops. The model achieved an overall accuracy of 94%, indicating a strong ability to classify crop health and disease states based on weather conditions. A detailed analysis of the classification metrics revealed that the model exhibited a precision of 1.00 for early blight, 0.90 for healthy crops, and 1.00 for late blight.

Figure 2.0



Recall values were 0.95 for early blight, 1.00 for healthy crops, and 0.65 for late blight, reflecting its strengths in identifying early blight and healthy conditions while highlighting a relative limitation in detecting late blight. The F1 scores, which provide a balanced measure of precision and recall, were 0.97 for early blight, 0.95 for healthy crops, and 0.79 for late blight.

The confusion matrix provided further insights into the model's predictive behavior. Of the 164 test cases, the model correctly classified 54 out of 57 instances of early blight, 87 out of 87 instances of healthy crops, and 13 out of 20 instances of late blight. While the model performed exceptionally well in detecting early blight and healthy crops, the lower recall for late blight (0.65) was attributed to a small number of misclassifications, where seven cases of late blight were incorrectly classified as healthy. These results suggest that late blight, which often occurs under complex and overlapping environmental conditions, requires further refinement of the algorithm to enhance sensitivity to its unique climatic indicators.

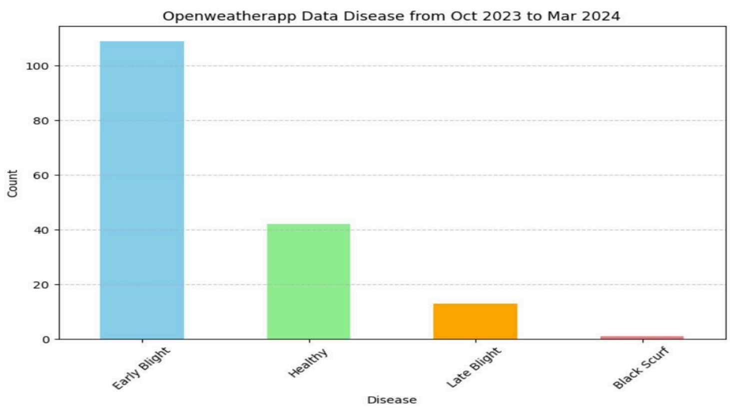


Figure 2.1

Despite these challenges, the high overall accuracy and robust performance in identifying early blight and healthy conditions validate the practical applicability of the algorithm in real-world agricultural scenarios. The weighted average F1-score of 0.94 underscores the model's reliability in delivering actionable predictions. These results demonstrate the algorithm's potential as a valuable tool for farmers and agricultural stakeholders, enabling them to implement preemptive measures to mitigate the risk of disease outbreaks and optimize crop yields.

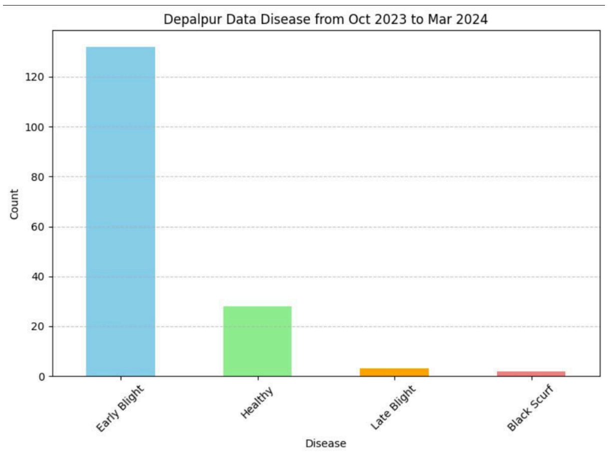
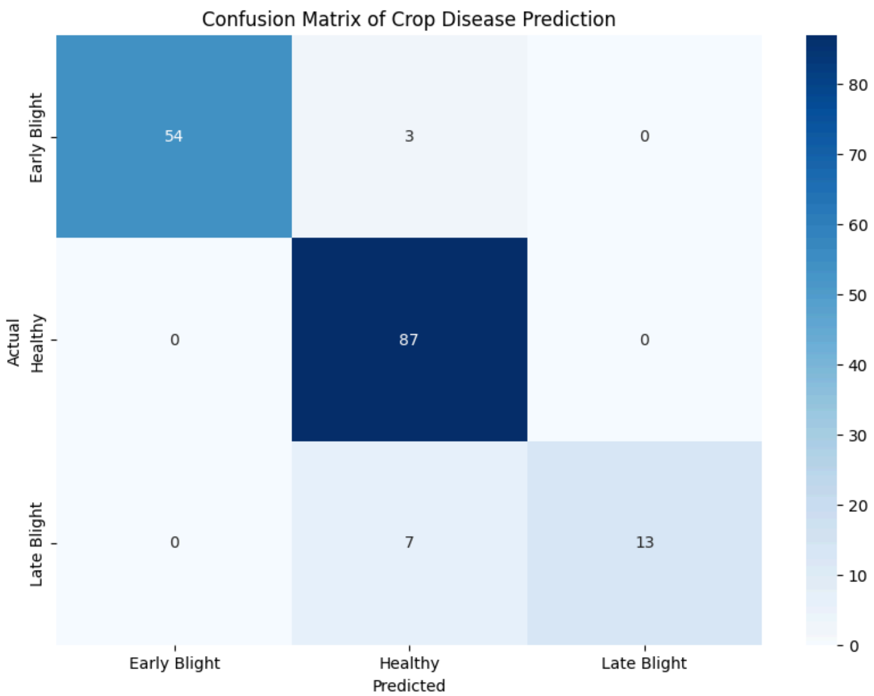


Figure 2.2

Figure 3.0: Confusion Matrix of Test Data



Classification Report:

	precision	recall	f1-score	support
Early Blight	1.00	0.95	0.97	57
Healthy	0.90	1.00	0.95	87
Late Blight	1.00	0.65	0.79	20
accuracy			0.94	164
macro avg	0.97	0.87	0.90	164
weighted avg	0.95	0.94	0.94	164

Figure 4.0: Classification Report of Test Data

2.4 Limitations and Concerns

Real-time testing is in process right now and this will only be possible once the disease sets in which usually starts in mid-December and lasts till February. This predictive model makes use of the real-time data from a weather station, so it is applicable to a radius of 30 km around the weather station. This can be considered a limitation as well as a strength. The strength is that it will be based on real-time data from a specific area and the limitation will be that it will require a weather station for every 30km circle. The predictive model has been trained only on limited data of three years and can only predict the three most common diseases of potato crops. Additionally, Its validity has been compared to only one cycle of harvesting from October 2023 to February 2024. Furthermore, its accuracy was much less when the OpenMeteo API was used to come up with a prediction system for areas where weather stations are not installed.

Bibliography

[1] F. a. A. O. o. t. U. Nations, "FAO," FAOSTAT, 2013.

[2] A. Devaux, J. Goffart and P. Kromann, "The Potato of the Future: Opportunities and Challenges in Sustainable Agri-food Systems.," *Potato Res.,* vol. 64, p. 681–720, 2021.

[3] Government of Punjab, "Ayuub Agricultural Research Institution," [Online]. Available: https://aari.punjab.gov.pk. [Accessed 11 November 2023].

[4] A. Majeed and Z. Muhammad, "potato production in Pakistan: Challenges and prospective management strategies – A review," *Pakistan Journal of Botany,* vol. 50, 2018.

[5] W. A. DAR, F. A. PARRY, BHAT and B. A., "Potato late blight disease prediction using meteorological parameters in Northern Himalayas of India," *Journal of Agrometeorology,* vol. 23, no. 3, pp. 310-315, 2021.

[6] M. Fahim, M. Hassanein, and M. Helmy, "Relationships between climatic conditions and potato late blight epidemic in Egypt during winter seasons 1999-2001," *Applied Ecology and Environmental Research,* vol. 1 , no. 10.15666/aeer/01159172., 2003.

[7] A. K. D. SINGHA, S. K. BHATTACHARYYA, B. RAHMAN, R. Helim, R. R. SAHARIA, B. AJANTA, and P. Sarmah, "EFFECTS OF WEATHER PARAMETER ON LATE BLIGHT DISEASE OF POTATO IN SALCHAPRA-I VILLAGE OF CACHAR DISTRICT," *JOURNAL OF INTERNATIONAL ACADEMIC RESEARCH FOR MULTIDISCIPLINARY,* vol. 2, no. 9, 2014.

【評語】200019

In this study, the Internet of Things (IoT) sensors throughout agricultural fields of potato crops was developed. The sensors collected real-time data on precipitation, air temperature, relative humidity, and wind speed. By using these data and machine-learning program, potato diseases can be detected and predicted.