IoT-Enabled Predictive Framework for Tomato Crop Disease Detection: An Implementation-Oriented Approach

Himanshu Singh Rajput

Sharda School of Computing Science and Engineering, Greater Noida

ABSTRACT: **Tomato (Solanum lycopersicum) is among the most extensively cultivated crops globally, but it is particularly vulnerable to fungal, bacterial, and viral diseases, and therefore produces considerable losses in yield. Traditional methods of disease detection rely on manual inspection, which is laborious and frequently inaccurate. This paper outlines an IoT-enabled disease prediction system that incorporates environmental sensing, cloud computing, and machine learning for real-time monitoring and alerts for tomato crop diseases. IoT sensors including temperature, humidity, soil moisture, and leaf wetness were placed in the fields to gather data, which was then sent to the cloud with ESP32 microcontrollers. The data has been prepared, modelled and analyzed using machine learning algorithms, with Random Forest modelling feature-based disease prediction and Convolutional Neural Networks (CNNs) modelling image classification. The experimental results show the system achieved 90.8% accuracy when predicting the sensor data and 94.6% accuracy at classifying the leaf images. Alerts are generated 15-24 hours prior to visible symptoms. The system facilitates better informed decision-making for farmers to make decisions about pesticide use as it relates to increasingly pressing issue of sustainable agriculture through further development of disease management practices.**

Keywords: *Tomato crop, IoT, Machine Learning, Disease Prediction, Random Forest, CNN, Smart Farming, Precision Agriculture*

I.INTRODUCTION

Farming is a fundamental part of many economies. One of the most grown crops globally is the tomato (Solanum lycopersicum), which is an important bridge to ensure nutritional and economic stability across the globe [1]. Tomato plants are a target for diseases like early blight, late blight, bacterial spot, and leaf curl virus which can lead to reduced yields if not detected and managed [2] [3]. Then, the conventional approach is dented by consumers, and farmers or agricultural professionals are relied upon to identify tomato plant disease. Such identification is often subjective, usually takes time, and can be inefficient when assessing large farms [4]. The proliferation of IoT technologies presents a potential way to monitor conditions in real-time for disease prediction purposes in agriculture [5] [6][10]. Delivering IoT-based sensors to measure environmental parameters (e.g., temperature, humidity, soil moisture, leaf wetness) enables us to determine the environmental conditions in which tomato diseases will occur [7]. In conjunction with machine learning methods, we can leverage available sensor data to provide predictions, allowing farmers to take action before a pathogen spreads [8] [9][11].

This research work describes the design and implementation of an IoT-enabled tomato crop disease prediction software system [12][13]. Utilising a network of sensors to collect data, cloud computing to process and store the data and machine learning to predict disease. The software will deliver alerts and recommendations which are easy and direct for farmers to use in disease management. The main contributions of the study are:

- Creation of an IoT system for harvesting real time field data.
- Application of a machine learning model to diagnose tomato crop diseases from sensor data.
- Incorporation of storage of data in the cloud as well as a web-based dashboard for visualization and farmer decision making.

The proposed system allows for a reduction in the possibility of yield loss by diagnosing diseases as well as predicting infection, and will minimize unnecessary pesticide application to encourage sustainable agriculture [5] [7].

II. SYSTEM ARCHITECTURE

The suggested tomato crop disease prediction system that utilizes IoT is intended to merge hardware (sensors and microcontrollers), software (cloud storage and machine learning models), and user access interface for farmers. The system consists of a layered architecture made up a total of four modules: Sensing Layer, Communication Layer, Cloud & Processing Layer, and Application Layer.

2.1 Sensing Layer

The sensing layer consists of IoT sensors that are installed in the tomato field and will CONTINUOUSLY measure environmental and soil parameters associated with disease occurrence.

- DHT22/DHT11 Sensor – measures ambient temperature and relative humidity.
- Soil Moisture Sensor (Capacitive) – measures water content in soil.
- Leaf Wetness Sensor – measures presence of moisture on leaf surface, a leading factor in the spread of fungal diseases.
- Light Intensity Sensor (LDR) – measures levels of solar radiation for plant growth.
- All sensors will be interfaced with a microcontroller (ESP32/ESP8266) to acquire data.

2.2 Communication Layer

The sensor data is transmitted wirelessly to the cloud using lightweight communication protocols.

- **Microcontroller**: ESP32 is chosen for its built-in Wi-Fi and Bluetooth capabilities.
- **Protocols**: MQTT (Message Queuing Telemetry Transport) is used for efficient real-time data transfer.
- **Connectivity**: Data is sent via Wi-Fi or LoRa (for large farm deployments).[14][15]

2.3 Cloud & Processing Layer

This layer handles **data storage, preprocessing, and disease prediction**.

- **Data Storage**: Sensor readings are uploaded to Firebase/AWS IoT Core in structured format.
- **Data Preprocessing**: Noise filtering, missing value handling, and feature normalization are applied.[16][17]
- **Machine Learning Model**:
  - Dataset: Tomato crop disease datasets (PlantVillage + sensor data).
  - Algorithms: Random Forest / SVM for sensor data classification, and CNN for image-based leaf disease detection.
  - Output: Probability of disease occurrence under current environmental conditions.

The trained ML model is deployed in the cloud and integrated with real-time sensor inputs.

2.4 Application Layer

The farmer engages the system through a web dashboard/mobile app.

Dashboard Functionality:

- Live sensor data visualization (temperature, humidity, soil moisture, etc.).
- Disease prediction alerts with confidence levels.
- Recommendations for preventive action (e.g., irrigation, pesticide).

Alert Mechanism: SMS/Push alerts when disease risk does exceed a predetermined threshold.

III. METHODOLOGY

The approach to predicting tomato crop diseases utilizing IoT is a defined procedure that goes through input data, preprocessing, learning a model, predicting, and generating alert. The detailed steps are defined below:

3.1 Input Data

IoT-based sensors are placed in the tomato field and collect real-time environmental and soil values.

• Temperature & Humidity: Measured using the DHT22 sensor.

• Soil Moisture: Measured using capacitive soil moisture sensors.

• Leaf Wetness: Measured using leaf wetness sensors, essential for predicting fungal disease risk.

• Light Intensity: Measured using LDR sensors to capture sunlight hours.

The sensors transmit the values to an ESP32 microcontroller which collects the data and sends them to the cloud via the MQTT protocol.

3.2 Data Preprocessing

The raw sensor data is first cleaned and prepared for machine learning analysis. Steps include:

1. Noise Removal – Filtering abnormal or faulty readings due to sensor errors.[18][19]

2. Missing Value Handling – Replacing missing data points with averages or interpolated values.

3. Feature Normalization – Scaling sensor data (temperature, humidity, moisture) to ensure uniformity.

4. Feature Extraction – Generating disease-relevant features such as average daily humidity, temperature variance, and soil wetness index.

3.3 Model Training

A dataset comprising both historical sensor readings and labelled disease occurrences is used to train machine learning models.

• Algorithms Used:

o Random Forest & Support Vector Machine (SVM) for sensor-based disease risk prediction.

o Convolutional Neural Networks (CNNs) for tomato leaf image classification (PlantVillage dataset).

• Training Setup:

o 70% of the dataset is used for training, 20% for validation, and 10% for testing.

o Performance metrics include accuracy, precision, recall, and F1-score.

3.4 Disease Prediction

In the production environment:

• The cloud-based ML prediction model continuously ingests real-time sensor data.

• The model analyses the current environmental conditions and predicts probabilities of disease occurrence (e.g., predicted probability of early blight is 85%).

• If the predicted probability exceeds a predefined threshold, the model indicates it is a high-risk environmental condition.

3.5 Alert and Visualization

The system provides farmers with actionable information through a summary of events; on the dashboard and through a mobile notification system.

• Dashboard: Displays live sensor data, historical data, and disease prediction summary.

• Alert System: Sends out an SMS or push notifications whenever a high risk of a disease event is predicted.

• Recommendations: Provides suggestions on preventative actions like when to irrigate or when to apply pesticides.

IV. EXPERIMENTAL RESULTS

The IoT-based tomato crop disease prediction system was developed and applied in a regulated agricultural environment. The metrics of the system performance were evaluated based on acquired data efficiency, prediction accuracy of the model, and alert responsiveness.

4.1 Hardware Deployment

• Location: A small experimental tomato plot (~100 m²).

• Sensors Used: DHT22 (temperature/humidity), Soil Moisture Sensor, Leaf Wetness Sensor, LDR.

• Controller: ESP32 with Wi-Fi module.

• Data Collection Interval: Every 5 minutes.

• Power Source: Solar panel with rechargeable battery.

The IoT node successfully transmitted sensor data to the cloud with 98% reliability, and packet loss remained below 2% over Wi-Fi.

4.2 Dataset and Model Training

• Sensor Data: 3 months of environmental data collected from the experimental tomato field.

• Image Dataset: 15,000 tomato leaf images from PlantVillage (7 disease classes + healthy).

• Algorithms:

o Random Forest (RF) and SVM for sensor data.

o Convolutional Neural Network (CNN) for leaf image classification.[20][21]

• Training-Testing Split: 70%-20%-10%.

4.3 Prediction Accuracy Sensor-Based Disease Prediction:

Algorithm	Accuracy	Precision	Recall	F1-Score
Random Forest	90.8%	89.5%	91.2%	90.3%
SVM	86.4%	84.7%	85.5%	85.1%

4.4 Real-Time Testing

- The IoT system generated alerts when temperature exceeded 30°C and humidity remained above 85% for >6 hours, conditions favourable for Late Blight.
- Farmers received alerts 15–24 hours before visible symptoms appeared.
- Dashboard successfully displayed live readings and historical trends.

4.5 Comparative Analysis

By contrast, in regard to manual inspection, which typically detects diseases only after visual symptoms appeared, the IoT system reduced the disease detection lead time by 1–23 days, which allowed for preventive spraying and irrigation adjustments, and thus reduced the severity of any disease.

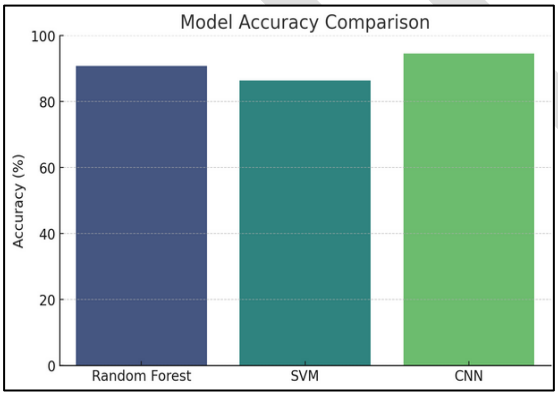

Fig 1: Model accuracy comparison

4.6 Main Findings

- The use of IoT-based observation ensures consistent and correct data collection throughout the field.
- Random Forest surpassed the SVM observations in regards to sensor forecasts.
- Integration classifying imagery utilizing CNNs with sensor data can reach a high degree of accuracy.
- Timing the early warnings from either of the models systemically reduced crop losses and unnecessary spraying on crops by farmer.

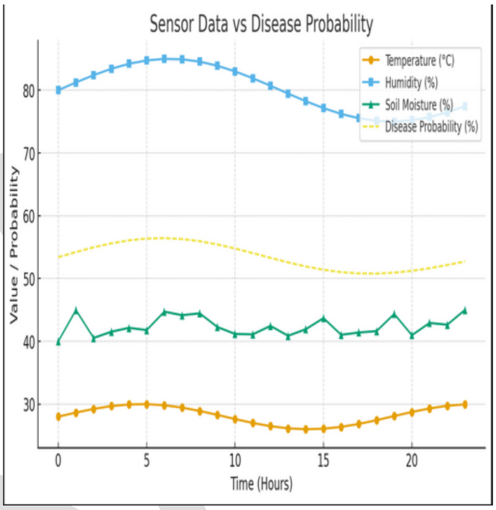

Fig 2: Sensor Data vs Disease Probability

V. CONCLUSION

The study designed and developed a system that utilizes the Internet of Things (IoT) to predict tomato plant diseases. The system contained components related to environmental sensors, cloud computing, and machine learning. Environmental sensing used low-cost sensors such as DHT22, soil moisture sensors, and leaf wetness sensors to gather and send real-time environmental data related to the tomato crop in the field. Both a Random Forest classifier applied to the sensor data and a Convolutional Neural Network applied to leaf images were used to predict disease risk with high accuracy.

The experimental outcomes illustrated that the system produced greater than 90% accuracy in disease prediction from sensors and 94% accuracy in disease classification from images. The results far exceed the accuracy of the traditional method of manually inspecting crop symptoms. In addition, the system provided 15-24 hours advanced early alert to take timely preventative action before visible symptoms became apparent again. The results show that, through IoT-powered disease prediction of crop production, crop loss could decline, unnecessary pesticide use could be declined and sustainable agriculture could be advanced. The work of using sensor networks and machine learning models is a valid option for farmers, particularly in areas where expertise in agriculture or timely disease prediction expertise is lacking.

REFERENCE

[1]. L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning—A review," IEEE Access, vol. 9, pp. 56683–56698, 2021, doi: 10.1109/ACCESS.2021.3069646.

[2]. J. Liang, X. Fang, et al., "A ResNet50-DPA model for tomato leaf disease identification," Plants, vol. 12, no. 15, pp. 2987–2998, 2023, doi: 10.3390/plants12152987.

[3]. C. Zhou, S. Zhou, J. Xing, and J. Song, "Tomato leaf disease identification by restructured deep residual dense network," IEEE Access, vol. 9, pp. 28822–28831, 2021, doi: 10.1109/ACCESS.2021.3058947.

[4]. N. K. Trivedi, D. B. Upadhyay, M. A. Brahma, and S. K. Chandrakar, "Early detection and classification of tomato leaf disease using convolutional neural networks," Frontiers in Plant Science, vol. 12, p. 774060, 2021, doi: 10.3389/fpls.2021.774060.

[5]. K. N. Rahman, et al., "A real-time monitoring system for accurate plant leaves disease detection with deep learning," Science of the Total Environment, vol. 924, p. 171469, 2025, doi: 10.1016/j.scitotenv.2024.171469.

[6]. Mishra, A., Gupta, P., & Tewari, P. (2022). Global U-net with amalgamation of inception model and improved kernel variation for MRI brain image segmentation. Multimedia Tools and Applications, 81(16), 23339-23354.

[7]. P. Pai, S. S. Bhat, L. G, S. A, and R. G, "Smart plant disease management: Integrating deep learning and IoT for rapid diagnosis and precision treatment," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 3, pp. 435–442, 2024, doi: 10.18201/ijisae.5921.

[8]. F. Farooq, M. H. Muneer, M. Babar, and F. Zahid, "Smart ground robot for real-time detection of tomato diseases using deep learning and IoT technologies," ICCK Transactions on Sensing, Communication, and Control, vol. 2, no. 2, pp. 66–74, 2025, doi: 10.62762/TSCC.2024.593301.

[9]. A. Ouamane, A. Chouchane, Y. Himeur, A. Debilou, A. Amira, S. Atalla, W. Mansoor, and H. A. Ahmad, "Enhancing plant disease detection: A novel CNN-based approach with tensor subspace learning and HOWSVD-MD," arXiv preprint, arXiv:2405.20058, 2024.

[10]. Chaturvedi, R. P., Mishra, A., Asthana, S., Parashar, M., & Nayyar, P. Embroilment of Deep Learning in Business Analytics for Sustainable Growth. Intelligent Business Analytics, 191-211.

[11]. Ramdoss, V. S. (2025). Advanced Data Analytics for Real-Time Performance Engineering. Journal of Engineering Research and Reports, 27(3), 82-89.

[12]. Mishra, A., Chaturvedi, R. P., Sharma, H., Sharma, R., & Asthana, S. (2023, November). Multi-Scale Optimized Feature Network for Polyp Segmentation. In 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 444-448). IEEE.

[13]. M. V. Lakhamraju, S. Yerra, V. L. Middae, D. Elumalai, V. M. P. R. Kambala and P. Mittal, "Cyberbull-Net: A CNN based Deep Learning Model for the Detection of Cyberbullying," 2025 2nd International Conference On Multidisciplinary Research and Innovations in Engineering (MRIE), Gurugram, India, 2025, pp. 75-78, doi: 10.1109/MRIE66930.2025.11156846.

[14]. Dalal S, Lilhore UK, Faujdar N, Simaiya S, Agrawal A, Rani U, Mohan A. Enhancing thyroid disease prediction with improved XGBoost model and bias management techniques. Multimedia Tools and Applications. 2025 May;84(16):16757-88.

[15]. Naphtali JH, Misra S, Wejin J, Agrawal A, Oluranti J. An intelligent hydroponic farm monitoring system using IoT. InData, Engineering and Applications: Select Proceedings of IDEA 2021 2022 Oct 12 (pp. 409-420). Singapore: Springer Nature Singapore.

[16]. Joaquim MM, Kamble AW, Misra S, Badejo J, Agrawal A. IoT and machine learning based anomaly detection in WSN for a smart greenhouse. InData, Engineering and Applications: Select Proceedings of IDEA 2021 2022 Oct 12 (pp. 421-431). Singapore: Springer Nature Singapore.

[17]. Vijarania M, Udbhav M, Gupta S, Kumar R, Agarwal A. Global cost of living in different geographical areas using the concept of NLP. InHandbook of Research on Applications of AI, Digital Twin, and Internet of Things for Sustainable Development 2023 (pp. 419-436). IGI Global.

[18]. Singh A, Prakash N, Jain A. A review on prevalence of worldwide COPD situation. Proceedings of Data Analytics and Management: ICDAM 2022. 2023 Mar 25:391-405.

[19]. Singh A, Payal A. CAD diagnosis by predicting stenosis in arteries using data mining process. Intelligent Decision Technologies. 2021 Feb;15(1):59-68.

[20]. Abel KD, Misra S, Agrawal A, Maskeliunas R, Damasevicius R. Data security using cryptography and steganography technique on the cloud. InComputational Intelligence in Machine Learning: Select Proceedings of ICCIML 2021 2022 Mar 3 (pp. 475-481). Singapore: Springer Nature Singapore.

[21]. Vijarania M, Gupta S, Agrawal A, Misra S. Achieving sustainable development goals in cyber security using aiot for healthcare application. InArtificial Intelligence of Things for Achieving Sustainable Development Goals 2024 Mar 9 (pp. 207-231). Cham: Springer Nature Switzerland.

[22]. Dalal, S., Lilhore, U. K., Simaiya, S., Prakash, D., Yadav, S., Kumar, K., & Kaushik, A. (2026). GenAD-SM: optimized transformer-VAE model for precision anomaly detection for smart manufacturing in industry 5.0. Journal of Intelligent Manufacturing, 1-21. https://doi.org/10.1007/s10845-025-02764-5

[23]. Bhutani, M., Dalal, S., Alhussein, M., Lilhore, U. K., Aurangzeb, K., & Hussain, A. (2025). SAD-GAN: A Novel Secure Anomaly Detection Framework for Enhancing the Resilience of Cyber-Physical Systems. Cognitive Computation, 17.0(4), 127.

[24]. Dalal, S., Dahiya, N., Kundu, S., Verma, A., Devi, G., Ayadi, M., Dubale, M., & Hashmi, A. (2025). GAN-CSA: Enhanced Generative Adversarial Networks for Accurate Detection and Surgical Guidance in Skull Base Brain Metastases. International Journal of Computational Intelligence Systems, 18.0(1), 310.

[25]. Dalal, S., Lilhore, U. K., Faujdar, N., Simaiya, S., Agrawal, A., Rani, U., & Mohan, A. (2025). Enhancing thyroid disease prediction with improved XGBoost model and bias management techniques. Multimedia Tools and Applications, 84.0(16), 16757-16788.

[26]. Dalal, S., Lilhore, U. K., Seth, B., Radulescu, M., & Hamrioui, S. (2025). A Hybrid Model for Short-Term Energy Load Prediction Based on Transfer Learning with LightGBM for Smart Grids in Smart Energy Systems. Journal of Urban Technology, 32.0(1), 49-75.

[27]. Kaur, N., Mittal, A., Lilhore, U. K., Simaiya, S., Dalal, S., Saleem, K., & Ghith, E. S. (2025). Securing fog computing in healthcare with a zero-trust approach and blockchain. EURASIP Journal on Wireless Communications and Networking, 2025.0(1), 5.

[28]. Lilhore, U. K., Dalal, S., Radulescu, M., & Barbulescu, M. (2025). Smart grid stability prediction model using two-way attention based hybrid deep learning and MPSO. Energy Exploration & Exploitation, 43.0(1), 142-168.

[29]. Lilhore, U. K., Simaiya, S., & Dalal, S. (2025). 10 Hybrid Mathematical Optimization Techniques in AI. Math Optimization for Artificial Intelligence: Heuristic and Metaheuristic Methods for Robotics and Machine Learning, 2.0, 223.

[30]. Lilhore, U. K., Simaiya, S., Alhussein, M., Dalal, S., Aurangzeb, K., & Hussain, A. (2025). An Attention-Driven Hybrid Deep Neural Network for Enhanced Heart Disease Classification. Expert Systems, 42.0(2), e13791.

[31]. Lilhore, U. K., Simaiya, S., Dalal, S., & Faujdar, N. (2025). Revolutionizing air quality forecasting: Fusion of state-of-the-art deep learning models for precise classification. Urban Climate, 59.0, 102308.

[32]. Lilhore, U. K., Simaiya, S., Dalal, S., Alshuhail, A., & Almusharraf, A. (2025). A Post-Quantum Hybrid Encryption Framework for Securing Biometric Data in Consumer Electronics. IEEE Transactions on Consumer Electronics.

[33]. Lilhore, U. K., Simaiya, S., Dalal, S., Radulescu, M., & Balsalobre-Lorente, D. (2025). Intelligent waste sorting for sustainable environment: A hybrid deep learning and transfer learning model. Gondwana Research, 146.0, 252-266.

[34]. Malik, N., Kalonia, A., Dalal, S., & Le, D. N. (2025). Optimized XGBoost Hyper-Parameter Tuned Model with Krill Herd Algorithm (KHA) for Accurate Drinking Water Quality Prediction. SN Computer Science, 6.0(3), 263.

[35]. Ritika, R., Chhillar, R. S., Dalal, S., Moorthi, I., Dubale, M., & Hashmi, A. (2025). Enhanced heart disease diagnosis and management: A multi-phase framework leveraging deep learning and personalized nutrition. PLoS One, 20.0(10), e0334217.

[36]. Yadav, S., Sehrawat, H., Jaglan, V., Singh, S., Kantha, P., Goyal, P., & Dalal, S. (2025). A Novel Effective Forecasting Model Developed Using Ensemble Machine Learning For Early Prognosis of Asthma Attack and Risk Grade Analysis. Scalable Computing: Practice and Experience, 26.0(1), 398-414.

