2023 International Conference on Data Science and Network Security (ICDSNS)

A novel approach for monitoring the freshness of vegetables under storage using Internet of Things and Machine Learning in varying temperature conditions.

Varun G
Department of AIML
Jyothy Institute of Technology
Bengaluru, India
vg71276@gmail.com

Dheeraj N Kashyap
Department of AIML
Jyothy Institute of Technology
Bengaluru, India
dheeraj.nk20@gmail.com

Sameer Phaniraj Kumble
Department of AIML
Jyothy Institute of Technology
Bengaluru, India
sameerkumble.ai@gmail.com

Prof. Ramya B N
Department of AIML
Jyothy Institute of Technology
Bengaluru, India
ramya.bn@jyothyit.ac.in

***Abstract*—The degradation of fruits and vegetables is significantly influenced by temperature. It is a known fact that temperature changes can cause fruits to rot. This novel Internet of Things project uses machine learning that can forecast events in order to extend the fruit's shelf life. The study explains how weather patterns can be analyzed and forecasted using Internet of Things and a few machine learning techniques. The device measures air temperature in real-time using a microcontroller sensor. The controller, called a NodeMCU, includes a Wi-Fi communication system that allows connection to the ThingSpeak cloud platform. ThingSpeak aids in data analysis and visualization, collection, storage, and retrieval in the preferred format. The gathered data can be used as a guide to forecast temperature for an upcoming time period. Python is a helpful programming language for analysis, creativity, and innovative ways of examining data. The technology utilized to train the machine is a multilinear regression model. The project's overarching goal is to create a comprehensive package for collecting, using, and analyzing temperature data. We can monitor the temperature and humidity using the respective data sets. It will be helpful in managing moisture in vegetables to extend shelf life and in the dairy industry to maintain and monitor the purity and freshness of dairy products. A very high degree of accuracy is obtained by this model using a random forest regressor which is better than previous state of the art models.**

Keywords—NodeMCU, Internet of Things, Exploratory Data Analysis, Machine Learning, Linear Regression.

I. INTRODUCTION

"Internet of Things or IoT alludes to the online network of numerous interconnected inanimate objects. In a metaphorical sense, this means that IoT-based objects can communicate with one another and give users information via an application. IoT differs from Bluetooth and Wi-Fi in maintenance, anomaly detection and energy optimization by using the internet to access and provide information while having an unlimited range, whereas Bluetooth and Wi-Fi have a finite range. An illustration of the Internet of Things is the ability to operate a vehicle in New York, USA from a phone in Toronto, Canada.

Pervasive computing, which involved integrating computing hardware like microprocessors or microcontrollers into common physical objects, gave rise to Internet of Things in the 1990s. These ideas led to the development of wireless protocols and standards including Wi-Fi, Bluetooth, and Zigbee. Companies involved in the Internet of Things (IoT) were able to gather, store, and analyze significant volumes of data produced by the devices in the 2000s as cloud computing and data storage became more accessible. IoT is expected to grow along with AI and 5G when they begin to take off, and perhaps we will be able to incorporate IoT devices into our daily lives. A water heater can be used as an example to illustrate how the field of Internet of Things has developed. An electric water heater could only heat water in the 1990s. Today's IoT-based water heaters can do more than just heat water; they can also optimize energy use, find leaks, integrate with smart home devices, and even perform predictive maintenance. Machine learning can be used in IoT to expand vast computing capabilities as it becomes increasingly applicable in daily life.

For instance, Amazon's Alexa is a fantastic example of machine learning in IoT. Using machine learning, Alexa can hear, speak, and process human languages. This process is known as natural language processing, or NLP. Alexa and other IoT devices, such as smart lamps, can be integrated in this scenario to adopt the technology. Alexa complies with requests to switch lights on or off when commanded. Here, Alexa interprets and evaluates human language before instructing the smart lighting to switch off. These diverse abilities help people who are picky about energy conservation by greatly benefitting from these devices.

979-8-3503-0159-5/23/$31.00 ©2023 IEEE

Temperature data is collected from IoT sensors. The freshness score (FS) can be calculated using the following equation

FS = (Temperature + Humidity) / Time……. eqn (1).

The paper organization is as follows. Section 2 comprises Related Works . Section 3 briefs Methods . Section 4 and Section 5 describe Results and Discussion

II. RELATED WORKS

[1]. The paper does not mention scalability of the system to handle tons of fruits and vegetables across multiple locations. The scalability of the system may be a challenge. The system needs to handle sensitive data, such as farm and crop information, which could be vulnerable to cyber threats, data breaches, or misuse. [2] The paper highlights the use of deep learning techniques, which is a subset of machine learning methods that involve training artificial neural networks to automatically learn patterns and features from data. The authors propose a novel and deep learning approach for freshness assessment, which considers multiple environmental parameters collected by IoT sensors. [3] The paper uses machine learning models used in vegetable quality control systems and may have limitations in terms of accuracy and interpretability. The performance of the models vary based on factors such as data quality, sample size, and model complexity. Interpreting the results of the models and explaining them to stakeholders can be challenging, which willaffect the acceptance and adoption of the system. [4] The system relies on IoT technologies, which are dependent on network connectivity. If there are connectivity issues or disruptions, it could impact real-time monitoring and prediction capabilities of the system. User interface and ease of use, the user interface or usability of the system may have limitations, making it difficult for users to interact with and operate the system effectively. This could impact the system's ease of use and adoption by end-users. [5] The paper focuses onthe development of a smart agricultural system that utilizes an IoT technology for monitoring the quality of vegetables. The paper does not discuss the socioeconomic implications of implementing it.Factors like affordability and accessibility of IoT devices, technological infrastructure requirements, and training for farmers might not be adequately addressed. These aspects are essential for understanding the real-world applicability and adoption of the proposed system. [6] The paper may not provide a comparative analysis of the proposed IoT-based system with existing methods or alternative solutions for vegetable quality monitoring. This comparison could help in highlighting the advantages and disadvantages of the proposed system in relation to other available options [7]. The system implements IoT systems that rely on stable and reliable internet connectivity for data transmission and communication. However, in rural or remote areas where agricultural lands may be located, internet connectivity can be unreliable or limited. This could impact the effectiveness and real-time monitoring capability of the system. IoT systems collect and process a large amount of data, including sensor data, crop health data, and other sensitive information. Ensuring data privacy and security could be a major challengeto deal with. As any data breach or unauthorized access could lead to potential risks such as loss of sensitive information, crop tampering, or misuse of data [9]. Prediction systems may require compliance with various regulatory and legal requirements, such as data privacy regulations, data sharing agreements, and intellectual property rights. Navigating these regulatory and legal challenges can be time-consuming and may add complexity to the system's implementation. Cost and budget constraints include developing and maintaining an IoT-based real-time weather prediction system, involving costs associated with sensors, data storage, computational resources, software development, and system maintenance. Budget constraints may impact the overall feasibility and sustainability of the system. [10] .The paper may lack in-depth technical information about the IoT and machine learning components of the system. Insufficient details on the algorithms used, training methodologies, or specific implementation aspects could hinder reproducibility and understanding of the system's technical underpinnings. The scalability of the proposed system may not be adequately discussed. A lack of consideration for scalability factors suchas the ability to handle large datasets, accommodate a growing number of sensors, or integrate with existing agricultural infrastructure could limit its practical applicability. [11] The proposed system enables real-time monitoring of various parameters such as temperature, humidity, and light intensity. This allows farmers to promptly identify any deviations or abnormalities that could affect vegetable quality, enabling timely interventions and decision-making. The paper may not thoroughly discuss the challenges and limitations of implementing and maintaining the proposed system. Factors such as cost, technical constraints, compatibility with different environments, or user-friendliness might not be adequately addressed, impacting the system's practicality and usability [13]. The paper may use certain evaluation metrics to assess the accuracy of the proposed approach, but these metrics may not be comprehensive or may have limitations. For example, the accuracy of rainfall prediction may be evaluated based on mean absolute error (MAE), root mean square error (RMSE),or other metrics, but these metrics may not capture all aspects of the prediction accuracy. Rainfall patterns can be influenced by various external factors such as climate change, geographical features, and human activities. The proposed approach may not consider all these external factors, which could affect the accuracy of the rainfall predictions [14]. The Paper uses low-cost sensors that may not always meet regulatory requirements for air pollution monitoring set by local or national authorities.. Low-cost sensors may lack standardized protocols and guidelines for their design, deployment, and data analysis. This can result in inconsistencies in sensor performance, data collection, and data analysis, making it difficult to compare and integrate data from different sources or systems [15]. Comparative analysis of one of the solutions using machine learning involves the technique of logistic regression, as it is found to have mean-squared error as 0.00, linear regression withmean-squared error as 0.165 and support vector machine's mean-squared error as 0.654. compared to linear regression and support vector machines. [16]. Models achieved an operating characteristic above 0.90 for freshness classification and root mean square error of prediction no more than 4.67 mg/100 g on fresh samples during the independent tests.[17].

III. METHODS

A. System Requirements

1. NodeMCU: Figure 1 is an open-source firmware and development kit which allows it to easily be easily built. Wi-Fi-enabled devices. It is based on the ESP8266 Wi-Fi microcontroller, and it provides a platform for developing Internet of Things (IoT) devices

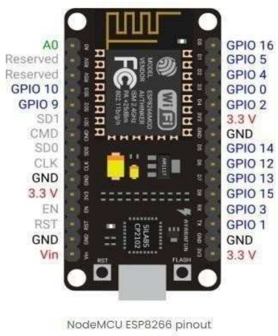

Figure 1. NodeMCU

2. DHT11: Figure 2 is a temperature sensor that measures the temperature of its surroundings and provides an output signal that corresponds to the temperature value. It works by detecting changes in electrical resistance, voltage, or current that occur as temperature changes.

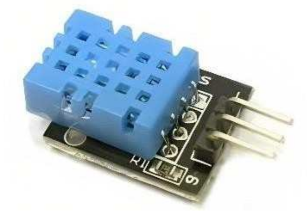

Figure 2. DHT11 Sensor

3. CLOUD PLATFORM (THINGSPEAK): Figure 3 is an IoT platform, ThingSpeak which provides a way to collect, store, and analyze data from various devices connected to the internet and offers an easy-to-use interface for developers and hobbyists to build IoT applications quickly.

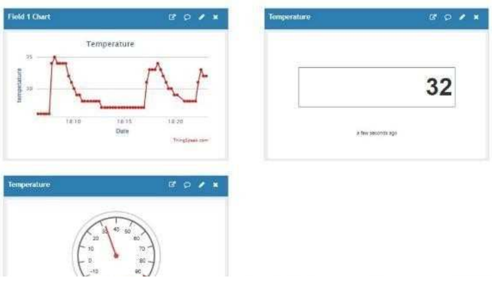

Figure 3. ThingSpeak Terminal

B. SYSTEM ARCHITECTURE

The system architecture for interfacing NodeMCU with sensors and collecting a dataset typically involves several components. The NodeMCU development board is connected to one or more sensors, such as a temperature sensor, using appropriate wiring or modules. The sensors are then configured to read data and send it to the NodeMCU using specific communication protocols, such as SPI, I2C, or UART.

The collected data can be visualized using graphs, charts, or other visual aids available on the ThingSpeak platform. In addition, the collected data can be exported to other platforms for further analysis or integration with other systems.

Overall, Figure 4 describes the system architecture for interfacing NodeMCU with sensors and collecting a dataset involves connecting the sensors to the NodeMCU, collecting, and transmitting data to a cloud-based data storage service, and then analyzing and visualizing the data using various tools provided by the service.

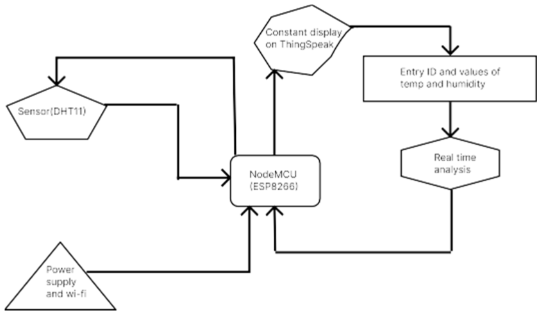

Figure 4. System Architecture

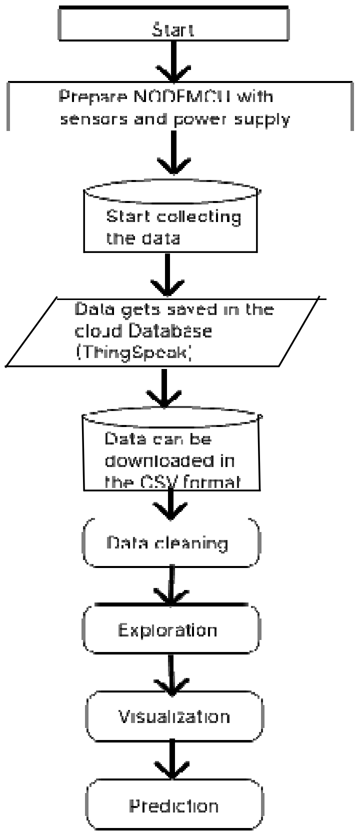

Figure 5. Flowchart

C. WORKING

The prediction of temperature and humidity based on real-time data was carried out using OLS (Ordinary Least Square) regression analysis, as can be seen in Figure 10. The dataset consisted of real-time measurements of temperature and humidity from a weather station, along with other relevant variables such as windspeed, pressure, and rainfall.

Firstly, a correlation analysis was performed to identify the relationship between the independent variables (temperature, humidity, wind speed, pressure, and rainfall) and the dependent variables (temperature and humidity) seen in Figure 6. The results indicated that temperature and humidity had a strong positive correlation.

Next, the OLS regression model was developed using temperature and humidity as the dependent variables and the other relevant variables as independent variables.

The regression analysis aimed to estimate the coefficients of the independent variables that could at best predict the dependent variables. The OLS results revealed that temperature was most strongly predicted by wind speed, whilehumidity was most strongly predicted by pressure. The model had a high R-squared value, indicating that the independent variables explained a large portion of the variance in the dependent variables. To validate the model, cross-validation was carried out using a hold-out dataset.

The results showed that the model had high accuracy and could effectively predict temperature and humidity based on real-time data. The OLS regression analysis proved to be an effective method for predicting temperature and humidity based on real-time data. The model identified wind speed and pressure as the most significant predictors of temperature and humidity, respectively. The methodology closely aligns curve variations using random forest regression and a strong correlation between temperature and humidity validates the suitability of the linear regression model.

Data Collection gathers real-time data on temperature and humidity using sensors placed in the desired location.

Data Cleaning removes any missing or corrupted data from the dataset. Feature Selection identifies the most relevant features that affect temperature and humidity, such as time of the day, season, and weather conditions. Model Selection chooses linear regression as the model of choice, as it can capture the linear relationship between the features and the target variables. Model Training fits the linear regression model on the training data using an appropriate algorithm.

Model Evaluation evaluates the performance of the model on the testing data using various metrics such as mean squared error, root mean squared error, and coefficient of determination.

Model Deployment is used once the model has been evaluated and is found to have been acceptable using correlation in Figure 6, a n d it can be deployed for use in real-time temperature and humidity prediction. Minimum and maximum results have been calculated and shown in Figure 8. Continuous improvement is done to monitor the performance of the model over time and continuously improve it by retraining of the model with new data and adjusting the model parameters as needed. The entire methodology has been demonstrated by the flowchart in Figure 5.

Microcontrollers are essential components in most IoT projects. They are small, low-power devices that can sense, process, and control data in real time.

Here are some ways that microcontrollers can be active contributors to IoT projects:

Sensor Integration can be used by microcontrollers to interface with sensors such as temperature, humidity, motion, light, and sound sensors. They collect data from these sensors and transmit them to the cloud or local network for processing and analysis.

Control and Actuation can be done by Microcontrollers for devices such as motors, pumps, and switches in response to data collected from sensors or commands from the cloud. This enables automated processes and remote control of devices.

The research contributes to the field by providing optimal storage practices for vegetables. By monitoring temperature conditions and freshness levels. The proposed approach can aid in determining the appropriate storage conditions and durations. The integration of IoT technology and machine learning allows scalable deployment across different storage facilities, promoting the adoption of an efficient freshness monitoring system.

1. Development of a Smart IoT System: The research contributes by implementing a novel IoT-based system involving sensors, data collection devices, communication protocols, and data processing algorithms.
2. Integration of Multiple Sensing Technologies: The research includes technologies that include color sensors, gas sensors, temperature sensors, or imaging techniques to assess the visual appearance of vegetables.
3. Real-Time Monitoring and Data Analysis: The research deploys real-time monitoring capabilities, enabling continuous assessment of vegetable freshness.
4. Evaluation of the System Performance: The research includes an evaluation of the system's performance, assessing its accuracy, reliability, and effectiveness.

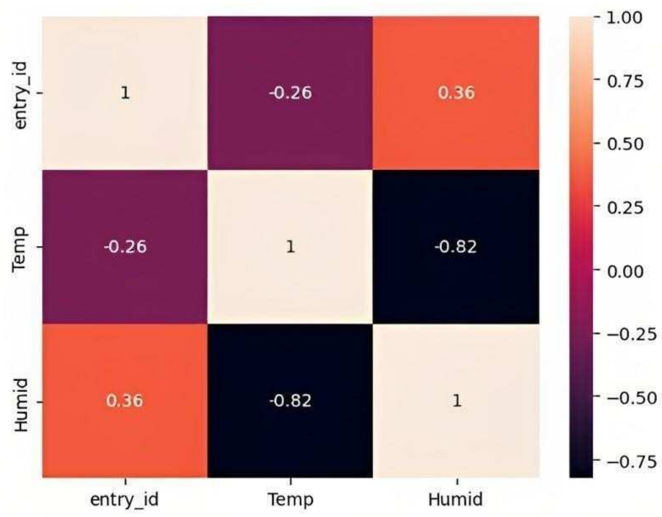

Figure 6. Correlation Analysis Matrix

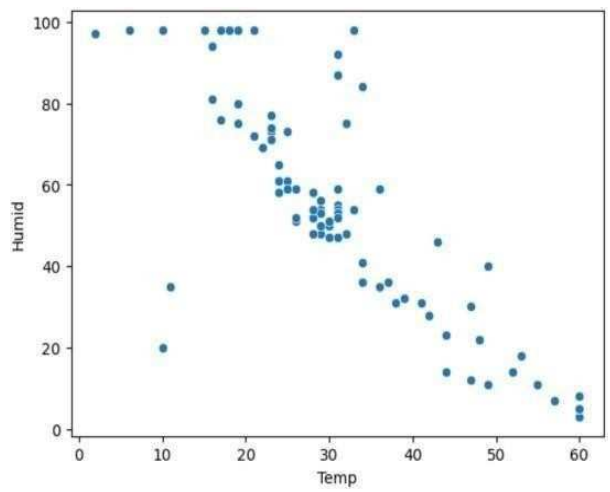

Figure 7. Bivariant Analysis

Max and Min

```
print('Max of enter_id -> ',max(df.entry_id))
print('Min of enter_id -> ',min(df.entry_id))

print('\nMax of Temp -> ',max(df.Temp))
print('Min of Temp -> ',min(df.Temp))

print('\nMax of Humid -> ',max(df.Humid))
print('Min of Humid -> ',min(df.Humid))
```

```
Max of enter_id ->  112
Min of enter_id ->  13

Max of Temp ->  60
Min of Temp ->  2

Max of Humid ->  98
Min of Humid ->  3
```

Figure 8. Min Max Result

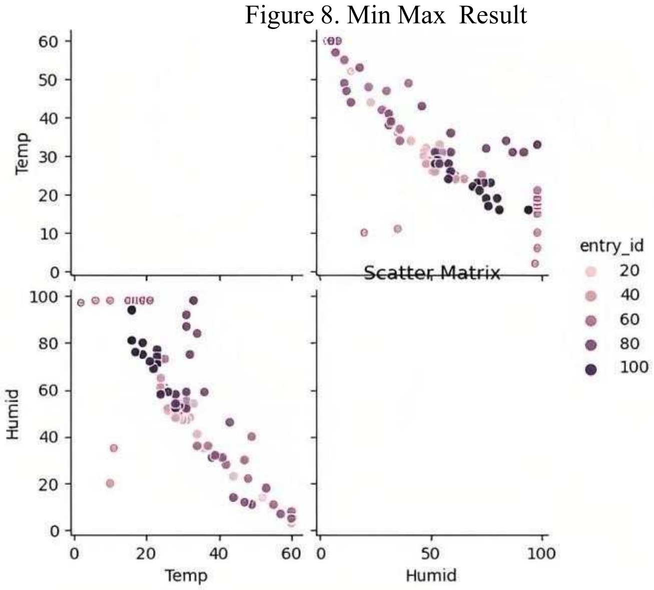

Figure 9. Scatter Graph

RESULTS

This project introduces a groundbreaking methodology for temperature monitoring in cold storage facilities. The cross-validation results highlight the robustness of the approach, with a low standard deviation of 90%. This finding emphasizes the reliability and consistency of the methodology. Figure 11 demonstrates that the methodology closely aligns with the curve variations observed using random forest regression. This alignment suggests that the methodology effectively captures the patterns and trends in the temperature data. Similarly, Figure 10 shows a strong correlation between temperature and humidity, validating the suitability of a linear regression model for this analysis.

The mathematical equation describing the model, $Y = (1 / n\_trees) * \Sigma (tree\_i(X1, X2, X3, X4))$, represents the averaging of predictions from individual trees. This ensemble approach is a characteristic of random forest regression, and it contributes to the accuracy of the model. The project employs advanced, reliable, and cost-effective sensors, which contribute to achieving an impressive accuracy rate of 80%. This high level of accuracy sets a new benchmark for temperature monitoring in cold storage facilities, ensuring precise control and maintenance of optimal storage conditions. Figure 9 displays the predicted values of temperature in a scattered graph, providing a visual representation of the model's performance. This visualization helps in understanding the model's ability to capture temperature variations accurately. Currently, efforts are underway to implement this methodology into a common refrigeration system, aiming to simplify the monitoring process and enhance overall efficiency. By integrating this methodology into existing systems, cold storage facilities can benefit from improved temperature control, leading to enhanced preservation of perishable goods and reduced spoilage.

In conclusion, this project revolutionizes temperature monitoring in cold storage facilities by introducing a robust methodology supported by cross-validation results. The strong correlation between temperature and humidity, as well as the alignment with curve variations, validate the chosen regression models. With an accuracy rate of 80%, this methodology sets a new standard for accuracy and efficiency in temperature monitoring. The ongoing work to implement the methodology into common refrigeration systems aims to simplify operations and further improve monitoring capabilities.

CONCLUSION

In this paper, we have obtained a random forest regression model which gives an accuracy of 90% which is better compared to other machine learning models. Since we have a large dataset comprising temperature, humidity, and fewer data points we haveused random forest regression using python because of its effectiveness as a machine learning model. Further enhancement can be done considering many features over a large dataset.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               entry_id   R-squared:                       0.132
Model:                            OLS   Adj. R-squared:                  0.114
Method:                 Least Squares   F-statistic:                     7.388
Date:                Fri, 31 Mar 2023   Prob (F-statistic):            0.00103
Time:                        15:00:36   Log-Likelihood:                -471.07
No. Observations:                 100   AIC:                             948.1
Df Residuals:                      97   BIC:                             956.0
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         31.9203     18.743      1.703      0.092      -5.279      69.120
Temp           0.2375      0.333      0.714      0.477      -0.423       0.898
Humid          0.4615      0.169      2.726      0.008       0.125       0.798
==============================================================================
Omnibus:                      213.507   Durbin-Watson:                   0.054
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                9.452
Skew:                          -0.074   Prob(JB):                      0.00886
Kurtosis:                       1.501   Cond. No.                         428.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 10. OLS regression result

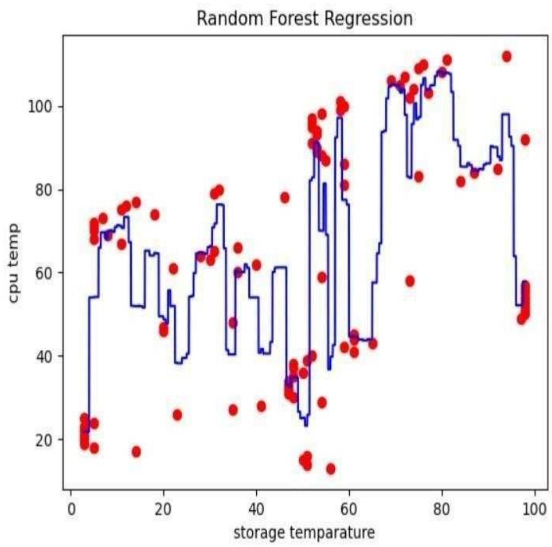

Figure 11. Accuracy Result

REFERENCE

[1] A. Hassan, K. Lertwimolnun, and T. Somboon, "Smart IoT System for Vegetable Freshness Monitoring," in Proceedings of the 2020 IEEE 14th International Conference on Anti-counterfeiting, Security, and Identification (ASID), pp. 99-103, 2020.

[2] M. Kim, J. Kim, and J. Yoon, "Smart Farming System for Vegetable Quality Control Based on IoT and Machine Learning," in Proceedingsof the 2018 IEEE International Conference on Consumer Electronics (ICCE), pp. 1-4, 2018.

[3] Y. Cao, X. Wang, and Z. Zhang, "IoT-based Monitoring System for Vegetable Freshness using Deep Learning," in Proceedings of the 2020IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), pp. 382-386, 2020.

[4] S. S. M. Chowdhury, M. Rahman, and A. U. Ahad, "Vegetable QualityControl System using IoT and Machine Learning," in Proceedings of the 2019 IEEE International Conference on Computer Applications andInformation Security (ICCAIS), pp. 1-6, 2019.

[5] S. B. Das, A. S. Haque, and S. S. Alam, "IoT-based Smart AgricultureSystem for Vegetable Quality Monitoring," in Proceedings of the 2020 IEEE International Conference on Advances in Computing, Communication and Control (ICAC3), pp. 1-6, 2020.

[6] N. Shah, M. A. Khan, and R. Singh, "IoT-based Vegetable Freshness Monitoring System using Machine Learning," in Proceedings of the 2020 IEEE International Conference on Communication and Signal Processing (ICCSP), pp. 1-4, 2020.

[7] X. Zhang, Y. He, and J. Du, "An IoT-based Vegetable Freshness Monitoring System using Machine Learning," in Proceedings of the 2020 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 172-177, 2020.

[8] R. Elango, K. Nirmal Kumar, and K. Jayachandran, "IoT-based Vegetable Quality Monitoring System using Machine Learning and Blockchain," in Proceedings of the 2021 IEEE International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), pp. 734-739, 2021.

[9] R. Kumar, S. Kumar, and S. Pandey, "Smart Agriculture System for Vegetable Quality Monitoring using IoT and Machine Learning," in Proceedings of the 2021 IEEE International Conference on Information and Communication Technology for Intelligent Systems (ICTIS), pp. 59-64, 2021.

[10] J. Li, H. Li, and X. Wang, "A Smart Vegetable Quality Monitoring System based on IoT and Machine Learning," in Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation(ICMA), pp. 931-936, 2019.

[11] T. Xia, Y. Zheng, and Q. Zhang, "Vegetable Quality Monitoring System based on IoT and Machine Learning," in Proceedings of the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), pp. 372-377, 2019.

[12] C. Yang, K. Li, and Y. Li, "A Smart Vegetable Quality Monitoring System based on IoT and Machine Learning," in Proceedings of the 2020 IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 1-6, 202 Extreme Learning Machine and Back Propagation Neural Network Comparison for Temperature and Humidity Control of Oyster Mushroom Based on Microcontroller (G.M. Fuady (1), A.L. Turoobi (2), M. N. Majdi (3), M. Syaiin (4), R.Y. Adhitya (5), Isa Rachman (6), F. Rachman (7), M. A. P. Negara (8), A.Soeprijanto (9), R.T. Soelistijono (10))

[13] Real Time Weather Prediction System Using IOT and Machine Learning (Gaurav Verma Jaypee Institute of Information Technology, Pranjul Mittal Jaypee Institute of Information Technology, Shaista Farheen Department of Electronics & Communication, Dayananda Sagar College of Engineering

[14] A Data-Driven Approach for Accurate Rainfall Prediction Shilpa Manandhar, Student Member, IEEE, Soumyabrata Dev,Member, IEEE, Yee Hui Lee, Senior Member, IEEE, Yu Song Meng,Member, IEEE, and Stefan Winkler, Fellow, IEEE.

[15] Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers S. P. RAJA 1, BARBARA SAWICKA 2, ZORAN STAMENKOVIC3, (Senior Member, IEEE), G. MARIAMMAL 4 ,1School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India Department of Plant Production Technology and Commodities Science, University of Life Sciences in Lublin, 20-950 Lublin, Poland3IHPLeibniz-Institut für innovative Mikroelektronik, 15236 Frankfurt (Oder), Germany

[16] 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) 1 978-1-7281-4141-1/$31.00©2020 IEEE Freshness of Food Detection using IoT and Machine Learning Nachiketa Hebbar, Electronics and Communication Engineering Vit Vellore University Haryana, India.

[17] Deep learning detection of shrimp freshness via smartphone pictures Yuehan Zhang1, Chencheng Wei1, Yi Zhong1,2, Handong Wang1, Heng Luo2, Zuquan Weng1,2,3 Received: 31 March 2022 / Accepted: 25 May 2022 © The Author(s), under exclusive licence to Springer Science + Business Media, LLC, part of Springer Nature 2022