INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS

ISSN(print): 2643-9840, ISSN(online): 2643-9875

Volume 06 Issue 07 July 2023

DOI: 10.47191/ijmra/v6-i7-50, Impact Factor: 7.022

Page No. 3233-3239

Tomato Fruiting Quality Prediction Using Hydroponics and Machine Learning

Aldrin J. Soriano[1], Cherry G. Pascion[2], Timothy M. Amado[3], Edmon O. Fernandez[4], Nilo M. Arago[5]

[1,2,3,4,5] Technological University of the Philippines, Manila, Philippines

ABSTRACT: The tomato fruiting quality prediction using hydroponics and Machine Learning (ML) focuses on improving tomato quality under a micro-climate setting with the use of various sensors to monitor and analyze the parameters that affect the growth of tomato. This study employed various algorithms such as k-nearest neighbor (KNN), support vector machine (SVM), decision tree, linear regression, and random forest (RF) to find the most appropriate supervised ML algorithm in predicting the tomato fruiting quality. The Random Forest algorithm performs better than the other four ML algorithms at predicting the quality of tomato fruit in the microclimate setup. The RMSE of the Decision Tree is 0.089, the absolute error is 0.040, and the squared correlation is 0.675.

KEYWORDS: Cherry Tomato; Fruiting Quality Prediction; Hydroponics; Machine Learning; Photoperiod

I. INTRODUCTION

Tomatoes are one of the most profitable crops in the Philippines and are widely grown around the world [1]. However, tomato production is characterized by extreme seasonality. In the dry season, there is abundant supply for tomatoes but in the wet season, yield is low, hence, there is limited supply in the market with poor quality of produce that contributes to severe price fluctuations [2]. Cherry tomatoes are packed with vitamins and offers numerous health benefits such as prevention of stroke and prostate cancer [3]. Despite of that, cherry tomato production faces different pest and disease problems, particularly bacterial wilt which reduces yield [4].

Hydroponics is a soil-less way of cultivating crops in water, as one of the potential agricultural approaches. It provides various opportunities in agriculture, particularly in locations coping with issues such as uncontrollable soil deterioration and limited water supplies. Furthermore, this agricultural practice demonstrates excellent results towards an environment-friendly and user-friendly farming and a reliable tool to address food security [5].

Temperature, humidity, water levels, NPK (nitrogen, phosphorus, and potassium) sensors and automatic irrigation are the most important factors in agricultural productivity, growth, and quality [6]. In connection with the advent of big data technologies and high-performance computing, ML has emerged to create new possibilities for data-intensive science in the multidisciplinary Agri-technologies domain [7]. Farm management systems are turning into real-time artificial intelligence enabled programs that provide comprehensive recommendations and insights for farmer decision support and action by using ML to sensor data [7] [8].

Nowadays, plant growth is a complex and dynamic environmentally linked system. Therefore, growth and yield modeling are significant scientific challenge to develop systems that analyze a vast and complex data to make better decisions [9].

ML is the foundation when it comes to artificial intelligence and big data analysis. It provides powerful algorithms that are capable of analyzing, generating patterns, classifying data, and make prediction by itself to perform a specific task like better crop growth [10]. In recent years different ML techniques have been implemented to achieve accurate plant growth, yield, and production prediction for different crops. The most successful techniques are Artificial Neural Networks (ANN), Support Vector Regression (SVR), M5-prime Regression Trees, RF, and KNN [11] [12] [13]. Data intensive approaches allow better decision-making, greater efficiency, and reduced waste while minimizing negative effects for the environment [14].

Compared to laboratory settings, greenhouses can be challenging environments for image analysis, as they are often optimized to maximize crop production thereby imposing restrictions on the possible placement of a camera and thereby its field of view. Further, variation in the colors or brightness of the fruits can be encountered over different plants of the same crop, over time for the same plant, over images of the same plant from different camera positions, etc. [14]. Repeated measurements are difficult because of ongoing work, changing circumstances such as lighting conditions, and an unfriendly atmosphere for electronic equipment.

Most methods for detecting and counting fruits, including tomatoes, have used color space transformations in which the objects of interest stand out, and extraction of features such as shape and texture [14]. In most of these works, the discriminative features were defined by the developers, rather than learnt by algorithms. Computer vision solutions based on hand crafted features may not be able to cope with the level of variability commonly present in greenhouses [15].

Motivated by the above difficulty, the purpose of this study is to predict the tomato fruiting quality with the use of various sensors like pH sensor, humidity sensor, temperature sensor, conductivity sensor, and photosynthetic light sensor. Furthermore, various ML algorithms will be used to determine the cultivation and fruiting quality of tomatoes in a controlled environment setup. It includes growing light to substitute the natural sunlight, peristaltic pump will automatically correct the pH and nutrient content of water solution, ultrasonic mister to increase the humidity, and Peltier cooler to automatically adjust the temperature of nutrient solution. Lastly, determining the significant difference between the prediction and actual yield is vital in this research process to develop a best system that will enhance the current tomato production.

II. METHODOLOGY

A. Research Design

This study used experimental research design which include the definition of objectives and the included variables, planning, designing the process, experimentation procedure, modelling, interpretation of results, and conclusions.

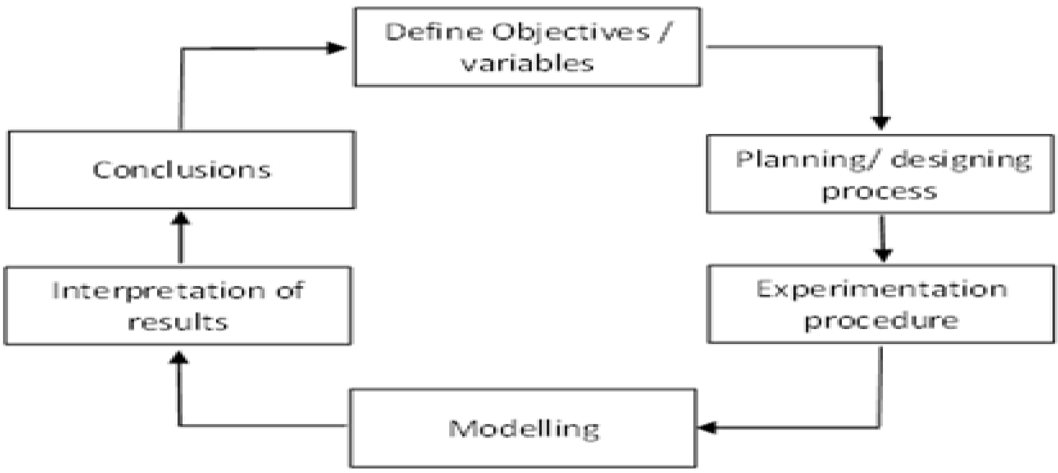

Fig. 1. Experimental Research Model

B. Prototyping Model and System Design

The hardware components of the study are mainly composed of devices and sensors that are used for tomato micro-climate hydroponics system, and an end device that will display the status and quantitative predictions of the ML algorithms. The detailed architecture of automated micro-climate hydroponics system is shown in Figure 2. The system will employ pH, temperature, humidity, salinity, and photosynthetic light level. It is used to gather the parameters measured by the sensors and automate the system using the peristaltic pump for nutrients, an ultrasonic mister for humidity, and artificial light using growing lights.

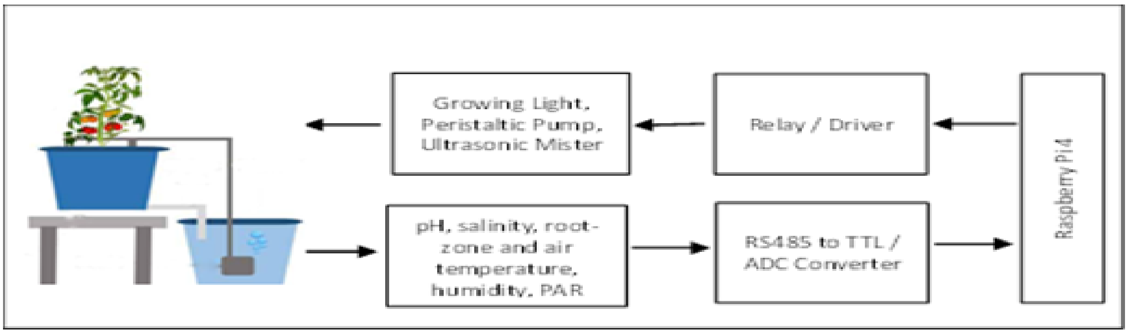

Fig. 2. Architecture of Automatic Micro-climate Hydroponics System

Tomato Fruiting Quality Prediction Using Hydroponics and Machine Learning

In this study, the use of purified water samples for analysis is used, moreover, the use of rainwater or distilled water will be considered for DWC. The value of air temperature will be maintained at 18.3 °C to 32.2 °C and root zone temperature of 23°C to 27°C with a relative humidity that is always higher than 60% [16]. The Nutrient Solution (NS) can be bought premixed based on quantity requirements [17]. The formulation for the NS should be crop targeted and optimal. NS was added for hydroponic tomato crops and the volume of the NS was measured with the help of a peristaltic pump. The amount of nutrients, salts, or impurities, in the water, is referred to the Electric conductivity (EC) limit. For maintaining optimum levels of EC, the researcher set up the EC to levels (1400 ppm to 3500 ppm, note that these values can vary for different crops).

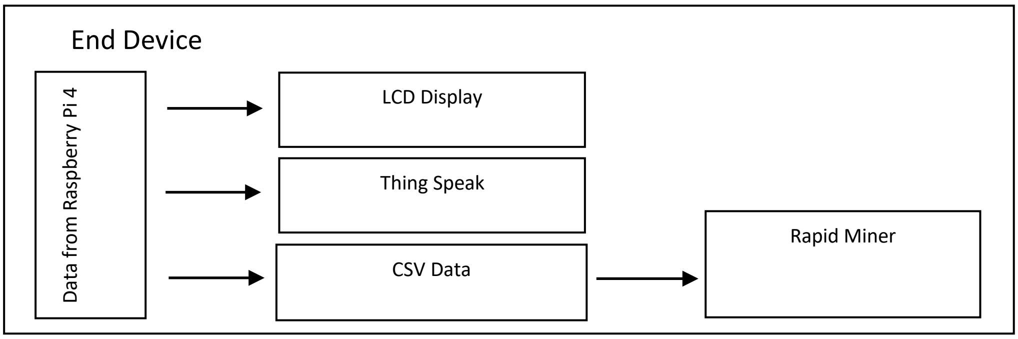

Fig. 3. End device that will display the status and quantitative predictions of the machine learning algorithms

The end-device that will display the status and quantitative predictions of the ML algorithms is shown on Figure 3. This end device is connected to a server that process and store the data from the sensors hydroponic system. Since supervised ML algorithms are used for this study, human intervention will be used to place the quality parameters of the tomato that corresponds to the sensed data.

The evaluation criteria for the quality of tomato fruits are presented in Table I. Meanwhile, the total quantity of the tomato fruits accumulated in the prototype is measured in mm. The total number of harvests and quantity of the harvested tomato fruits will be iterated based on the previous scenario of harvests. Both daily harvest and total harvest are measured in mm.

TABLE I. Tomato Quality Parameters

Parameter / Quality	Average size in mm	Color	Disease
Grade 2	25 mm to 35 mm	Red (Ripe), Light Red, Green	0
Grade 1	<25 mm	Red (Ripe), Light Red, Green	0
Grade 0	Any size	Red (Ripe), Light Red, Green	1

*red - more than 90% of the tomato surface is red in color

*light red - 60% of the tomato surface show pinkish-red or red, provided that not more than 90% of the surface is red

*green - surface of tomato is completely green in color

C. Establishment of a Predictive Model for Tomato Fruiting Quality Prediction

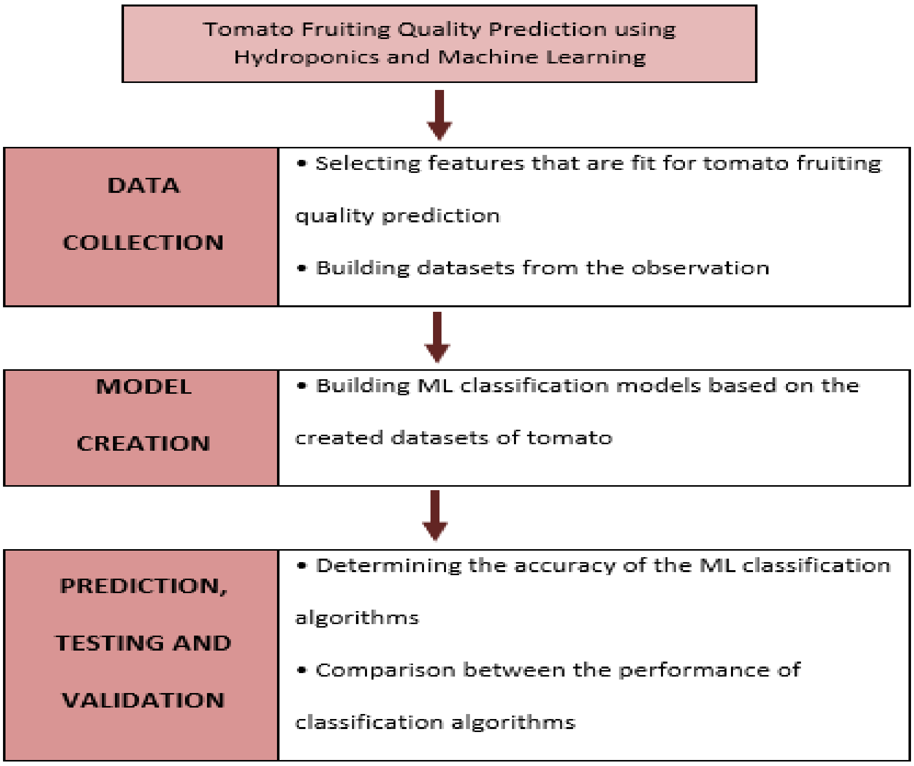

Fig. 4. Predictive Model for Tomato Fruiting Process Flow

- Data Collection.

The process of creating a predictive model for tomato fruiting quality prediction begins with data collection. The requirement analysis for the sensor will be refined in detail. These established guidelines will be used in the design of the system. Through the experiments, the researcher aims to adjust the parameter based on the production of tomato in normal settings to determine the best variables, labels and method needed in establish data sets for the ML

- Model Creation

Observations will be conducted to carry out the creation of datasets. As the values of the selected parameters are obtained using the sensors, human intervention will be utilized to input corresponding labels for the model. The labels will be marked based on the direct observation of the researcher. After collecting enough number of datasets, ML classification models will be created.

- Prediction, Testing, and Validation

The performance of the tested ML algorithms in this study and the comparison of the quantitative prediction is observed using the descriptive research methodology.

III. RESULTS AND DISCUSSION

The overall status of the design project was evaluated based on its functionality in accordance with the set evaluation process. The presented data below discusses the summary of the microclimate environment, and quality performance of the GUI and ThingSpeak to print the values. The measures used to evaluate the performance of five regression Machine Learning algorithms and comparisons are also shown.

Fig. 5. Prototype of the Microclimate Setup

A. Performance of the ML Models for Tomato Fruiting Quality Prediction

Table II shows the performance of the five regression ML on training and validation datasets based on Root Mean Square Error (RMSE), absolute error and squared correlation. Both Random Forest model and Decision Tree algorithm reveals lower RMSE and absolute error, and higher squared correlation in predicting the features used in describing the quality of Tomatoes in 10-fold cross-validation. Among all the quality parameters, the performance of the DT in predicting the quality of the Tomato in the micro-climate setup in terms of the three given metrics is better. However, the performance of SVR model in predicting the quality exhibits larger values on error metrics and smaller value on squared correlation.

All regression machine learning algorithms reveal that DT algorithm provides the smallest values of RMSE and absolute error, as well as the highest squared correlation compared to the other models.

TABLE II. Comparison of the Performance of Five Regression Machine Learning Algorithms in Fruiting Quality of Tomato

Machine Learning Algorithm	RMSE	Absolute Error	Squared Correlation
Linear Regression	0.415	0.267	0.548
Support Vector Machine	1.019	0.881	0.198
k-Nearest Neighbor	0.278	0.154	0.618
Random Forest	0.192	0.118	0.696
Decision Tree	0.089	0.040	0.675

Figure 6 shows the performance of the five regression ML on 10-fold cross-validation datasets based on Root Mean Square Error (RMSE), absolute error, and squared correlation. It is clear from Figure VI that the Decision Tree Model algorithm performs better than the other four ML algorithms at predicting the quality of tomato fruit in a particular microclimate setup. The RMSE of the Decision Tree Model technique is 0.089 the absolute error is 0.04, and the squared correlation is 0.675. The SVM method has the lowest squared correlations among the five models. The squared correlations of the four machine learning, on the other hand, are very high. However, with 0.675 squared correlation, Decision Tree performs well followed by Random Forest in forecasting quality.

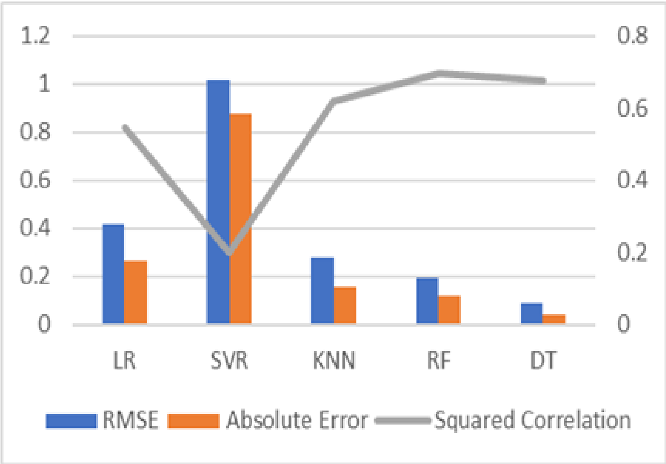

Fig. 6. Comparison of ML Algorithms in Fruiting Quality of Tomato

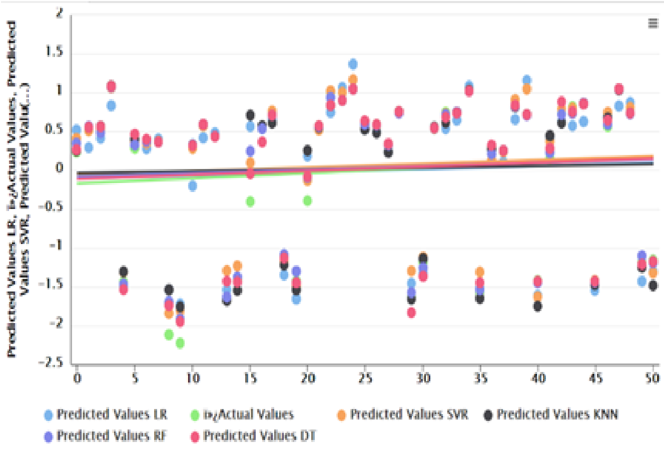

Fig 7. Actual and Predicted Values for Quality Parameters

Meanwhile, in terms of actual prediction, among all parameters, the average of the predicted values obtained using RF and DT algorithms is closest to the average of the actual values quality parameters see Figure 7.

IV. CONCLUSION AND RECOMMENDATIONS

A. Conclusion

1. From the dataset collected, average humidity measures 63.76 %, the average water temperature is 25.0°C, pH value of 5.5, average PAR value is 402 micromole/m2s, average air temperature is 25.0°C, and average TDS value of 2086.12 ppm. The mentioned combination of controlled environment parameters yields continuous growth in the tomato and first harvest happens in Day 59. The total harvest obtained in 8 hours photoperiod is 283.4g, in 12 hours photoperiod is 1087.7g, and in 16 hours photoperiod is 1994.1g.
2. The squared correlation of the quality indicators 0.675 implying that the Decision Tree model used in predicting tomato fruiting quality indicators fits the data well.
3. Linear regression and k-NN gives fair performance on all parameters while SVM performs poorly in predicting quality of tomato, RF performs fairly in predicting all of the parameters.

B. Recommendations

Generally, the study showed the success of tomato fruiting quality prediction using hydroponics and machine learning. The problems and concerns discussed in this research were solved with accuracy and reliability. The following are the recommendations based on the result of findings and conclusions.

1. More datasets must be created to get more precise results in quality prediction.
2. Based on the collected review of related studies, the micro-control environment's temperature, humidity, PH, and TDS were declared constants, while the photoperiod was treated as a variable. Data cleaning, also known as feature selection, was used to eliminate unnecessary data from the data sets. It is advised to have more than three cycles in the future research as these factors are important in the investigation process of its significant relation to predict the tomato fruiting quality.

ACKNOWLEDGMENT

We sincerely appreciate everyone who offered to assist in carrying out this research. Your unwavering assistance and inspiration were truly crucial to the success of this endeavor. And finally, praise be to the name of the Almighty God, the source of all wisdom, power, and unwavering love.

REFERENCES

1) A. L. G. Gorme, Z. C. Gonzaga, O. B. Capuno, R. J. C, S. McDougall, A. D. Goldwater and G. S. Rogers, "Growth And Yield of Tomato (Solanum Lycopercicum L.) as influenced by different soil organic amendments and types of cultivation," Annals of Tropical Research 39 (Supplement B), pp. 116 - 128, 2017.

2) J. Ellamar, "Off-season tomato production: A new technology in tarlac province of Philippines," Acta Horticulturae, vol. 1086, pp. 261-267, 2015.
3) WebMD, "Health Benefits of Cherry Tomatoes," 2021.
4) D. S. Mapa, "Major Vegetables and Rootcrops Quarterly Bulletin," Philippine Statistics Authority, vol. 14, no. 4, p. 10, 2020.
5) P. Sambo, C. Nicoletto, A. Giro, Y. Pii, F. Valentinuzzi, T. Mimmo, P. Lugli, G. Orzes, F. Mazzetto, S. Astolfi, R. Terzano and S. Cesco, "Hydroponic Solutions for Soilless Production Systems: Issues and Opportunities in a Smart Agriculture Perspective," Frontiers in Plant Science, vol. 10, no. 923, 2019.
6) A. C.S., S. S. Suhas G, R. K. Dingra and M. K. Dingra, "Modern Agriculture using Wireless Sensor Network," International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 4, 2020.
7) K. G. Liakos, P. Busato, D. Moshou and S. Pearson, "Machine Learning in Agriculture: A Review," Sensors, vol. 18, p. 2674, 2018.
8) KDnuggets, "Machine Learning in Agriculture: Applications and Techniques," 2019.
9) B. Alhnaity, S. Pearson, G. Leontidis and S. Kollias, "Using Deep Learning to Predict Plant Growth and Yield in Greenhouse Environments," arXiv, 2019.
10) Z. Hong, "A data-driven approach to soil moisture collection and prediction using a wireless sensor network and machine learning techniques".
11) A. Chlingaryan, S. Sukkarieh and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," Computers and Electronics in Agriculture, vol. 151, pp. 61-69, 2018.
12) M. Garanayak, G. Sahu, S. N. Mohanty and A. K. Jagadev, "Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods," International Journal of Agricultural and Environmental Information Systems (IJAEIS), p. 20, 2021.
13) B. Koerner, “Is it better for the environment to wear leather or pleather,” Slate Magazine, vol.3, no. 1, pp. 4–10, 04-Dec-2007. [Online]. Available: https://slate.com/technology/2007/12/is-it-better-for-the-environment-to-wear-leather-or-pleather.html. [Accessed: 04-Dec-2020].
14) "An Introductory Guide to Machine Learning," Apro-software.com, [Online]. Available: https://apro-software.com/machine-learning-agriculture/. [Accessed 6 January 2022].
15) Y. Zhao, L. Gong, Y. Huang and C. Liu, "A review of key techniques of vision-based control for harvesting robot," Computers and Electronics in Agriculture, vol. 127, pp. 311-323, 2016.
16) R. Shamshiri, J. W. Jones, K. R. Thorp, D. Ahmad, H. C. Man and S. Taheri, "Review of optimum temperature, humidity, and vapour pressure deficit Review of optimum temperature, humidity, and vapour pressure deficit," International Agrophysics, vol. 32, pp. 287-302, 2018.
17) T. Fertilizer, "MASTERBLEND Premium Fertilizers," Masterblend International-Tyler Enterprises, 2022. [Online]. Available: https://www.masterblend.com/4-18-38-tomato-formula/. [Accessed 5 May 2022].