MEKATRONIKA
e-ISSN: 2637-0883
VOL. 3, ISSUE 2, 1 – 5
DOI: https://doi.org/10.15282/mekatronika.v3i2.7159

ORIGINAL ARTICLE

Time-Series Classification Vegetables in Detecting Growth Rate Using Machine Learning

Ezahan Hilmi Zakaria[1], Mohd Azraai Mohd Razman[1,*], Jessnor Arif Mat Jizat[1], Ismail Mohd Khairuddin[1], Zelina Zaiton Ibrahim[2], Anwar P. P. Abdul Majeed[1] and Suhaimi Puteh[1]

[1]Faculty of Manufacturing and Mechatronics Engineering, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.
[2]Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

ABSTRACT – IoT based innovative irrigation management systems can help in attaining optimum water-resource utilisation in the exactness farming landscape. This paper presents a clustering of unsupervised learning based innovative system to forecast the irrigation requirements of a field using the sensing of a ground parameter such as soil moisture, light intensity, temperature, and humidity. The entire system has been established and deployed. The sensor node data is gained through a serial monitor from Arduino IDE software collected directly and saved using the computer. Orange and MATLAB software is used to apply machine learning for the visualisation, and the decision support system delivers real-time information insights based on the analysis of sensors data. The plants organise either water or non-water includes weather conditions to gain various types of results. kNN reached 100.0%, SVM achieved 99.0% owhile Naïve Bayes achieved 87.40%.

ARTICLE HISTORY
Received: 14th May 2021
Revised: 22nd June 2021
Accepted: 28th July 2021

KEYWORDS
Feature extraction
Machine learning
Classification
Sensor reading
Chilli plant

INTRODUCTION

Vegetables are parts of plants that are always used and consumed by humans or other animals as a daily food in their life. It is commonly used and applied to plants collectively to ensure the part of vegetables is edible to be eaten, including the flowers, fruits, stems, leaves, roots, and seeds [1]. Alternatively, it is applied randomly, often is used for culinary and cultural traditions. Some food is derived from excluding the fruits, flowers, nuts and cereal grains but includes savoury fruits such as tomatoes, and courgettes, flowers such as broccoli, and seeds such as pulses.

To maintain a healthy plantation are a must to produce good quality fruits and vegetables. The problem is occurred on why some plants can become wilt and weak. Each veget and fruit able must be provided with a good amount of water, nutrient, air, light, temperature and space. Their health is affected based on the quantity and condition of each environmental factor to receive the most optimum quality. Also, people nowadays are busy with working that they cannot manage times for watching the plant.

RELATED WORK

Before proceeding with the experiment, it is essential to know the model and method of learning to apply to the experiment's data. Collecting data is the main start before proceeding with the experiment's process [2]. To identify the healthiness and disease of the plant, the machine learning technique [3]–[5] was often applied to gain accuracy and performance. The clustering algorithm, as defined by L.a.b (value from the CIELAB colour scale) and coordinates of x and y-axis of the pixels [6], [7]. The input images are segmented into atomic regions. It defines approximately the number k of superpixels N/k pixels, where N is the number of pixels in the picture. The centre moved to the lowest gradient value over a neighbourhood of 3 x 3 pixels avoiding centroid allocation in edge regions that contain noisy pixels.

The classification is done in the machine learning method as we use the machine learning described from the review, each pixel needed to be assigned one by one to class [8]. However, a single pixel did not contain the information required to make a statement from the class affiliation. Thus the surrounding area is a must for classification. That area can be classified by using Conventional Neural Network (CNN). The example CNN network architecture is given to acknowledge how the category was made to classify. Therefore, research is done to present data extraction using multiple sensors tacclimatizationnts' acclimatisation and classify the condition from clustering the appropriate state.

*CORRESPONDING AUTHOR | Mohd Azraai Mohd Razman | ✉ mohdazraai@ump.edu.my

METHODOLOGY

Plants' Vitality

The progress of plantstartsrts from the seed until it becomes significant, and healthiness depends on the condition for each plant assigned applied in the experiment. Overview of these plants shows that the plants' height and the number of leaves are increasing gradually.

The number of leaves of each plant is taken to see how much the plant has been growing for both places, either outside or in the greenhouse, as in Table 1. The fertiliser is applied to the soil. Each plant is labelled, which is A and B for Outside. Plant A being watered while B is not watered. For C and D that stay in the greenhouse, C is watered while D is non-watered. The result below shows the data taken during the run of the experiment.

Table 1. Number of leaves taken during the experiment

Plant	Number of Leaves			
	4/5/2021	13/5/2021	10/6/2021	28/6/2021
A	15	26	43 (1 chili)	63 (8 chillies)
B	14	22	31(3 chillies)	26 (6 chillies)
C	16	34	37	43 (5 chillies)
D	8	14	20	27

Sensor Read Data Collected

The data acquired is inspected to justify whether the design of the experiment discussed is legitimately suitable and acceptable before progressing through the following process, data pre-processing. The findings here should contribute to research on the plant's growth rate by demonstrating that the methods of experiments in this study could get enough data retrieval concepts for elevating the identification techniques. The data retrieved will be shown in Figure 1 as the data is altered to observe the data better, and Figure 2 depicts the overall flow of the methodology used in this research to acquire the findings.

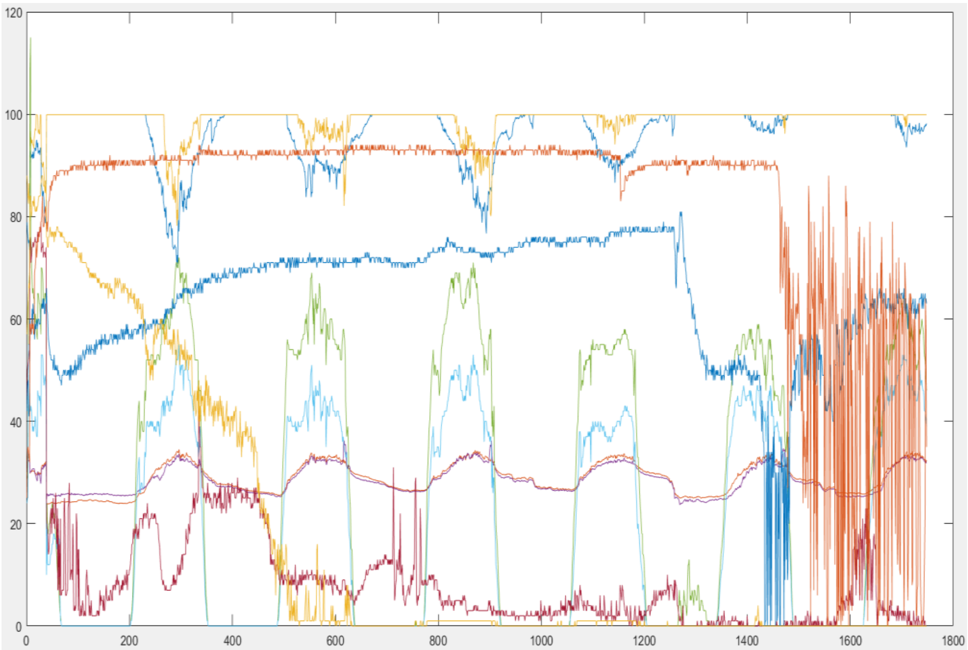

Figure 1. Overall Data Collected

Feature Extraction

Feature extraction is a process that reduces the data by dimensional which the raw dataset is lessened to more controllable groups for processing. This is done after collecting data from the plants. The most crucial characteristic of large datasets is that they contain a vast number of variables. The computing resources process depends on the variables given, as the higher the variable, the higher the processing time need to compute. Feature extraction helps find the convenient feature from large datasets by selecting and combining the variables into components. Consequently, it helps reduce the amount of data effectively. It is an easy process, but it is proficient at explaining the actual dataset with accuracy and originality.

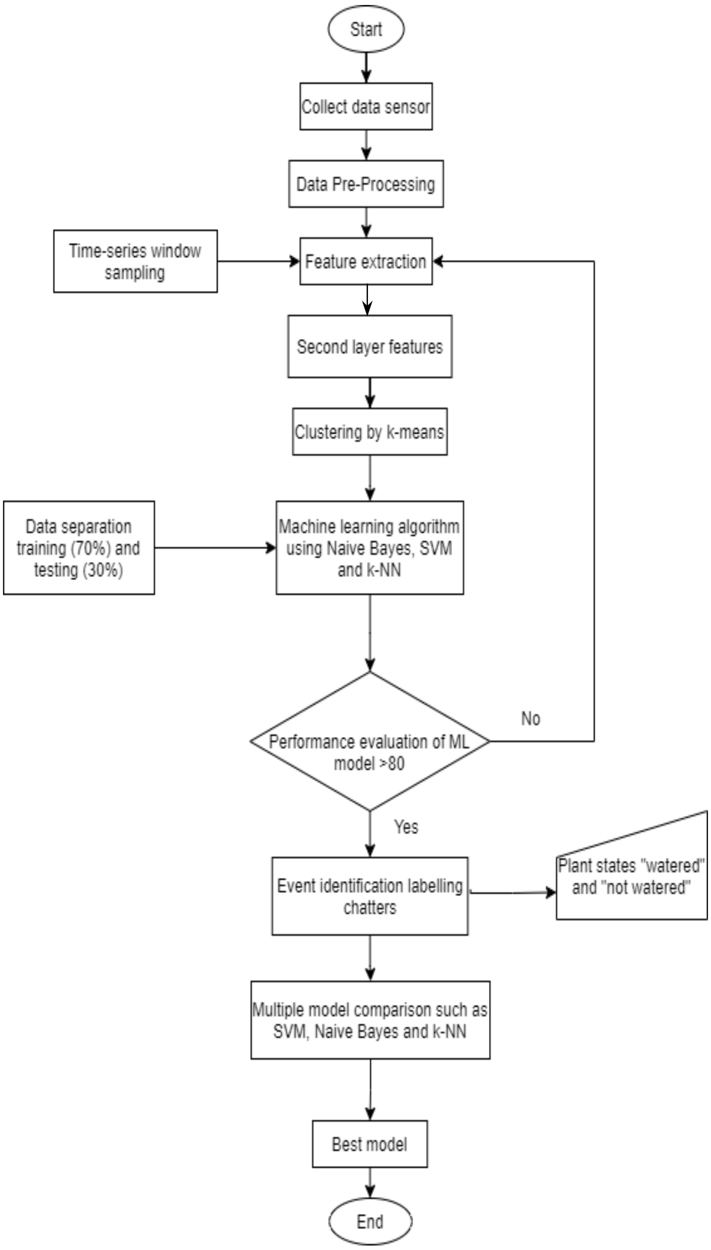

Figure 2. Machine Learning techniques flowchart

Clustering using k-means

The following section is the clustering process of plants growth rate by using k-means. This will show whether the plant needs water or not by observing and clustering the results. This unsupervised algorithm is set to classify the optimum number of clusters where the data is labelled accordingly to the classes.

K-mean is the most straightforward unsupervised learning algorithm that serves for the well-known clustering problem. This learning is used when received unable data (undefined categories and groups of data). K-mean partition is defined in the initial stage. The centre of the cluster is pre-defined. The process is continuously executed if there is no change in cluster classification. The output of this method depends upon the selected cluster centroids.

Event identification

Selecting the plants' instances of growth rate changes is essential as they will be labelled as the responses gained according to the classes clustered previously. The demonstration from the cluster analysis will only provide the best numbers of clusters based on the dataset processed. Before that happens, the classes' description must be changed as the machine only knows how to cluster instead of identifying the types. The time-series data collected are real-time happen where the set of parameters varies in time that can be detonated.

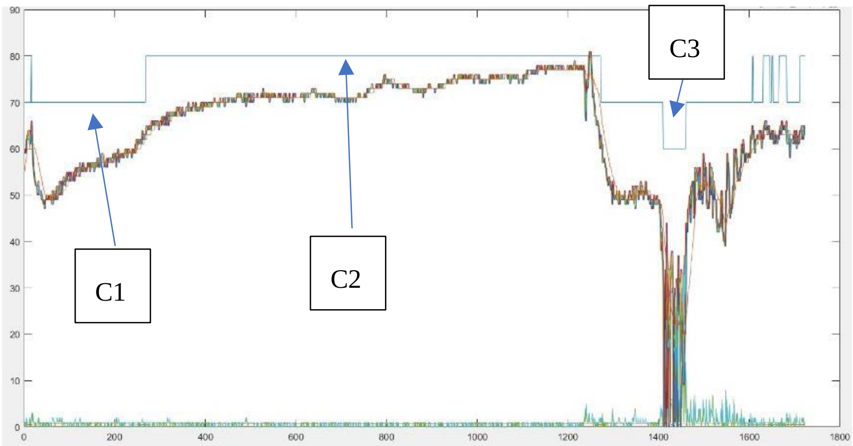

Figure 2. Soil moisture of plant B extracted

Based on Figure 3, the value from 0 to 200 is for the second class (C2), 200 to 1200 is the first class (C1), and 1200 to 1600 is for the third class (C3). For 1600-1800 is an unstable condition as its class is mixed between C1 and C2. The cause of dangerous, it assumes that the plant needs moister to become more stable.

Machine Learning Technique

The process of machine learning is applied to Orange software. The machine learning frame is built for the dataset to run into it, as in Figure 4. It started with inserting the dataset that had been extracted from the sensors. The data went through k-mean clustering and was visualised to see how the clustering happened and the dataset's performance.

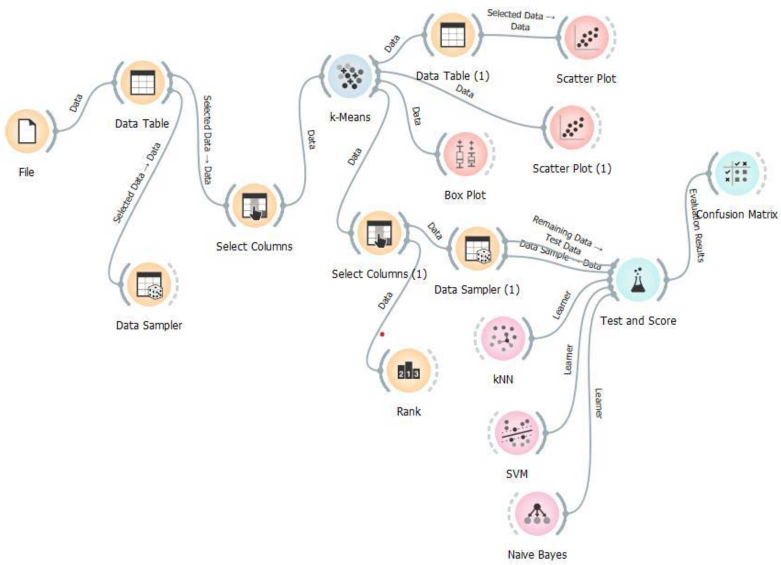

Figure 4. Overall ML Technique Flowchart

The clustered data proceeds with the machine learning model. This is where the classification accuracy result is compared between the models applied in the software. The confusion matrix helps see the detail that happened in classification in the models.

EXPERIMENTAL RESULTS

From the overall of this dataset, we proceed with extracting and splitting the data into train and test. In this experiment, the data for the train is 70%, while for test data is 30%. The information is run into different classification models such as k Nearest Neighbor (k-NN), Support Vector Machine (SVM), and Naïve Bayes. The figure below shows the comparison of each model applied in this experiment.

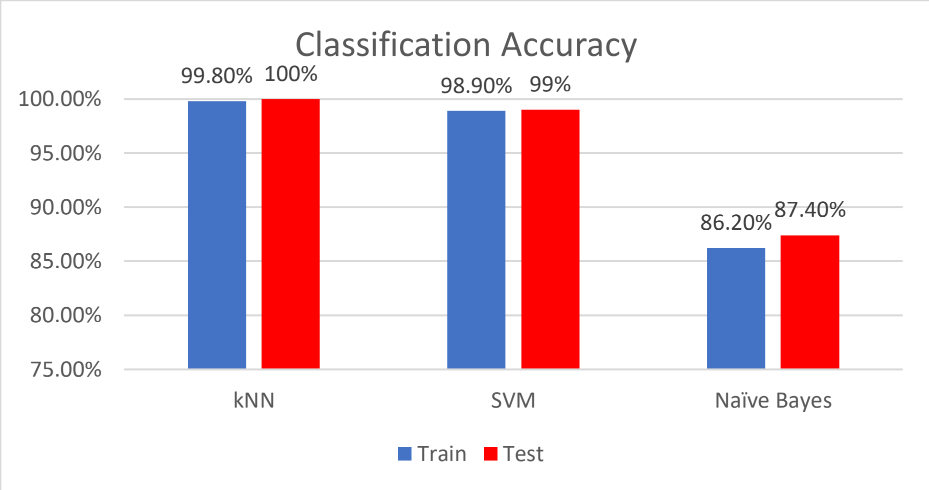

Figure 5 Classification accuracy result of the proposed model

Based on Figure 5, We can observe and compare the detail between different proposed models. k-NN attained the best classification accuracy performance for both train and test datasets while SVM is the second best and Naïve Bayes is the lowest. k-NN got 99.80% on training and 100% on testing, respectively. SVM received 98.90% training and 99% testing, while Naïve Bayes achieved 86.20% training and 87.40% testing. SVM is almost close to reaching the accuracy on k-NN.

CONCLUSION

Based on the objectives of this project, it is shown that overall goals have been accomplished. The first objective is to develop the monitoring system and collect the data from sensors read. It was managed using the Arduino microcontroller to read the sensor data and store it on the computer. It is difficult for long term data reading as it can cause issues and unexpected errors during the experiment. Thus, the time series is not maintained well, but it is still helpful because the defect is about 20% of the overall data.

Next, the classes can be identified after feature extracting and data clustering into each category, respectively. A window sample was applied to resize the data, and the feature pulled combined the data size from 5 minutes, 10 minutes, 15 minutes, 1 hour, and 2 hours, creating 34 components extracted. K-mean clustering is used in this project to cluster the data that has been removed from the sensors. Then the result is visualised in box plot and scatter plot to see the impact of clustered data.

While the last objective is to formulate the machine learning models and classification accuracy performances to predict the health of the plants on a time-series based, it was accomplished by using the software to simulate the different machine learning models to see the classification result. The best model depends on the highest possible classification accuracy percentage. The best model is chosen for its excellent prediction for the healthy of the plants.

ACKNOWLEDGEMENT

The authors would like to acknowledge Universiti Malaysia Pahang for funding this study under the Research Grant (RDU200332).

REFERENCES

[1] S. Puteh, N. F. M. Rodzali, M. A. M. Razman, Z. Z. Ibrahim, M. N. A. Shapiee, and M. A. M. Razman, “Features Extraction of Capsicum Frutescens (C.F) NDVI Values using Image Processing,” *MEKATRONIKA*, vol. 2, no. 1, pp. 38–46, Jun. 2020.

[2] S. Puteh, N. F. M. Rodzali, A. P. P. Abdul Majeed, I. M. Khairuddin, Z. Z. Ibrahim, and M. A. M. Razman, “Classification of Capsicum Frutescens Health Condition Through Features Extraction from NDVI Values Using Image Processing,” *Lect. Notes Mech. Eng.*, pp. 414–423, 2021.

[3] Y. Guo *et al.*, “Integrated phenology and climate in rice yields prediction using machine learning methods,” *Ecol. Indic.*, vol. 120, no. August 2020, p. 106935, 2021.

[4] E. Y. Obsie, H. Qu, and F. Drummond, “Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms,” *Comput. Electron. Agric.*, vol. 178, no. May, p. 105778, 2020.

[5] A. Chaudhary, R. Thakur, S. Kolhe, and R. Kamal, “A particle swarm optimisation based ensemble for vegetable crop disease recognition,” *Comput. Electron. Agric.*, vol. 178, no. August, p. 105747, 2020.

[6] T. C. Venkateswarulu *et al.*, *Screening of ipomoea tuba leaf extract for identification of bioactive compounds and evaluation of its in vitro antiproliferative activity against MCF-7 and HeLa cells*, vol. 58, no. 1. 2020.

[7] E. C. Tetila *et al.*, “Detection and classification of soybean pests using deep learning with UAV images,” *Comput. Electron. Agric.*, vol. 179, no. October, 2020.

[8] F. J. Knoll, V. Czymmek, L. O. Harders, and S. Hussmann, “Real-time classification of weeds in organic carrot production using deep learning algorithms,” *Comput. Electron. Agric.*, vol. 167, no. November, p. 105097, 2019.