LONTAR KOMPUTER VOL. 16, NO. 2 AUGUST 2025 **p-ISSN 2088-1541**
DOI : 10.24843/LKJTI.2025.v16.i2.p07 **e-ISSN 2541-5832**
Accredited Sinta 2 by RISTEKDIKTI Decree No. 158/E/KPT/2021

Multi Classification of Strawberry Leaves Using Support Vector Machine (SVM) Method on Smart Greenhouse Plants Based on Internet of Things (IoT)

Osphanie Mentari Primadianti[a1], Agung Surya Wibowo[b2], Muhammad Zimamul Adli[a3], Agung Muhamad Toha [a4]
[a]Electrical Engineering, Universitas Islam Nusantara, Jl. Soekarno-Hatta No.530, Buah Batu, Bandung, West Java, Indonesia
[b]Electrical Engineering, Telkom University, Jl. Telekomunikasi No. 1, Bandung, West Java, Indonesia
[1]osphanie@uninus.ac.id
[3]zimamuladli@uninus.ac.id
[4]agungmuhamadtoha@gmail.com
[2]agungsw@telkomuniversity.ac.id (Corresponding author)

Abstract

Strawberry plants, or Fragaria x ananassa, are shrubs in the Rosaceae (rose) family that produce sweet and scented red fruit. Strawberries are high in vitamin C and other minerals. The benefits of growing strawberries in smart greenhouses are one of the hydroponic farming sectors advances. The construction of a smart greenhouse system that can be monitored and controlled automatically simplifies agricultural research, which formerly relied on traditional farming and wet labs with long study timeframes and high expenses. This innovation makes it easier for researchers to study the impact of the Internet of Things (IoT) on strawberry plant growth by using several sensors in the greenhouse and Artificial Intelligence (AI) to save time and money in optimizing strawberry plant growth. Meanwhile, the Support Vector Machine (SVM) algorithm with a multi-classification category on leaves 3, 4, and 5 achieved Precision: 0.96, Recall: 0.95, F1-Score: 0.95, and Accuracy: 0.95. The accuracy level reaches 95%, implying that machine learning can be used in strawberry cultivation to assist hydroponic farmers.

Keywords: *Support Vector Machine (SVM), Machine Learning, Multi-Classification, Strawberry Plants, Smart Greenhouse*

1. Introduction

As a population, humans have built their civilization on agriculture. This aims to ensure that the ever-increasing demand for food is met responsibly by using both inexpensive and high-quality items. In order to address fundamental food demands, agriculture has been the backbone of society throughout human history. It serves as the basis for civilization and social advancement. In order to meet the growing demand for food, agriculture has evolved from traditional practices that depend on the strength of humans and animals to the technologically advanced modern period.

Since plants grown in greenhouses are shielded from pests and disease [1], they yield higher-quality plants than those grown outdoors, which is why many contemporary farmers have chosen to plant in greenhouses. Often referred to as plant houses, greenhouses serve as containers for plants to grow in accordance with environmental requirements. A greenhouse is a structure with a bubble-like form that is covered with transparent or light-absorbing materials to maximize production and shield plants from erratic weather conditions that could harm their growth.

The strawberry plant is one kind of plant that can be grown in a greenhouse. Because of their great demand and nutritional value, strawberries (Fragaria × ananassa) are one of the most economically valued horticultural crops in the world [2]. Their cultivation is, nevertheless, extremely susceptible to environmental influences and prone to a number of plant diseases, which can drastically lower quality and productivity. One of the most obvious markers of plant health is

leaf condition, and it is essential to identify anomalies in leaves early on to avoid extensive crop damage.

More effective crop health monitoring and management have been made possible by recent developments in smart agriculture and Internet of Things (IoT) technology [3], which integrate sensors, automation, and data analytics. Precision farming can be supported in a smart greenhouse by IoT devices that continuously record environmental data (such as temperature, humidity, and soil moisture) [3] and plant leaf data in Excel format.

In this study, one of the primary challenges lies in the multi-class classification of strawberry plants based on leaf count. Plants with three leaves (class 0) are categorized as young, those with four leaves (class 1) are considered to be in the early maturation stage, and those with five leaves (class 2) are classified as mature. Using information from the greenhouse's temperature panel, temperature sensors, pH levels, nutrients, humidity, and water volume [5-7], specifically in order to get around this, strawberry leaves are increasingly being subjected to machine learning techniques. The most popular supervised learning algorithm among them is Support Vector Machine (SVM), which is renowned for its efficiency in multi-classification tasks and its resilience when dealing with high-dimensional data [8-10].

As a component of an Internet of Things (IoT)-based smart greenhouse system, the goal of this research is to develop a model that can recognize various issues with strawberry leaves using the SVM approach [11]. It is anticipated that the model will aid in the early identification and categorization of various leaf anomalies, allowing farmers or automated systems to make timely and focused interventions. This study advances precision agriculture and the sustainable production of strawberries in controlled environments by integrating dataset processing, machine learning, and IoT infrastructure.

2. Materials and Methods

This research was conducted at the Smart Greenhouse facility of Universitas Islam Nusantara (UNINUS). The greenhouse is equipped with an integrated Internet of Things (IoT) system that continuously monitors key environmental and plant-related parameters using various sensors. The parameters collected include

- Suhu (Temperature)
- Kadar pH (pH Level)
- Nutrisi (Nutrient concentration)
- Kelembaban (Humidity)
- Volume Air (Water Volume)
- Panel Temperature (Temperature of solar panel/control unit)

Table 1 shows the database structure used in the Smart Greenhouse system for strawberry plant growth classification. The dataset is divided into three classes that represent different growth stages: Class 0 (young, 3 leaves), Class 1 (early maturation, 4 leaves), and Class 2 (mature, 5 leaves). Each class contains six features: temperature, humidity, soil moisture, soil nutrient concentration, light intensity, and CO_2 level. For each class, data were collected from 22 individual field samples, ensuring a balanced and representative dataset for training and evaluating machine learning models.

Table 1. Database Smart Greenhouse

Class	Features	Field Number
0	6	22
1	6	22
2	6	22

In a smart greenhouse setting, the image below [Figure 1] illustrates the process of a multi-class classification system for the conditions of strawberry plant leaves utilizing an Internet of Things-based machine learning approach. The following are some significant stages that make up the diagram:

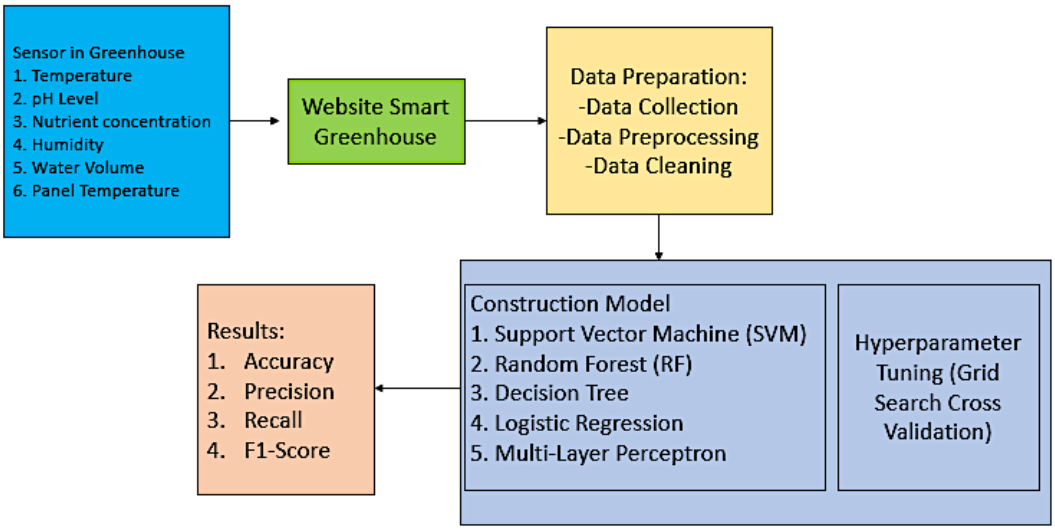

Figure 1. Methods

1. Greenhouse Sensor

The greenhouse system has six different kinds of sensors that gather environmental information that influences strawberry plant growth, including:

a. Temperature: The ambient air temperature of the plant.
b. pH level: The planting medium acidity or alkalinity is indicated.
c. Nutrients: The concentration of nutrient solution.
d. Humidity : The humidity of the greenhouse air.
e. Water Volume: The amount of water that is accessible.
f. Panel Temperature: The sensor or control unit temperature.

2. Website Smart Greenhouse

The Smart Greenhouse Website, a web-based monitoring system that serves as both a user interface and a location to integrate data from IoT devices, transmits and displays all sensor data.

3. Data Preparation

Important procedures to transform raw data into data that is ready for analysis are included in this step, including:

a. Data Collection: A web-based system that retrieves data from sensors.
b. Data Preprocessing includes data manipulation, encoding, and normalization.
c. Data Cleaning : Eliminating duplicates, and missing values from data.

4. Model Construction and Training

Upon data preparation, the subsequent phase involves constructing and training several categorization models. Five machine learning algorithms are employed as potential models:

a. Support Vector Machine (SVM)

One way to describe non-linear SVMs is as linear separators in a high-dimensional space [12]. The input space $\mathbb{R}^d$ is mapped using a non-linear mapping Φ; note the space $\mathcal{H}$. Therefore, $\mathcal{H}$ has the same geometrical interpretation. In $\mathcal{H}$, the linear separator f^k parameter w^k of form (1) is never calculated. Specifically (it could have an infinite or enormous dimension) [12]. However, it is referred to as a linear combination of images using pi of the support vectors (with indices in N_s^k for the input data) [12].

$$w^k = \sum_{p \in N_S^k} \alpha_k^p \, y^p \Phi(x^p) \quad (1)$$

Therefore, the normalization factor π_w^k that will be applied in this work will be specified by:

Table 2. The Best Hyperparameters

Model	Range	The Best Parameter
SVM	C = 0.1-100, Kernel=linear/rbf,Gamma=scale/auto	C=10, gamma='scale', kernel= 'rbf'
Random Forest	estimators= 50-200, max_depth= 10-30, min_samples_split = 2-10, min_samples_leaf = 1-4, bootstrap=True/ False	min_samples_leaf=1, min_samples split=5, n_estimators=100
Decision Tree	Criterion=gini/entropy, max_depth=0-15, min_samples_split=2-10, min samples_leaf=1- 4	criterion='gini', max_depth=3, min_samples_leaf=2, min_samples_split=2, random_state=42
LR	multi_class='multinomial', solver='lbfgs', max_iter= 0-200	multi_class='multinomial', solver='lbfgs', max_iter=200
MLP	Hidden_layer_sizes = (50-100),max_iter = 0-1000	hidden_layer_sizes = (64, 32), max_iter=1000, random_state=42

$$\frac{1}{\pi_w^k} = \sum_{p,p' \in N_s^k} \alpha_k^p \alpha_k^{p'} y^p y^{p'} K(x^p).\phi(x^{p'}) \quad (2)$$

$$= \sum_{p,p' \in N_s^k} \alpha_k^p \alpha_k^{p'} y^p y^{p'} K(x^p, x^{p'}) \quad (3)$$

where K is the kernel function that makes it simple to calculate dot products in H [12].

b. Random Forest

Using the idea of ensemble creation of decision trees, Random Forest (RF) is a popular option in the bioinformatics field [13]. It is an effective, interpretable, and non-parametric classification method that offers excellent classification accuracy for a range of applications [13].

c. Decision Tree

Learning a model from a collection of categorized cases that can predict the class of previously unseen instances is known as classification. Two characteristics set hierarchical multi-label classification apart from regular classification: (1) a single example may belong to multiple classes, as in this study; and (2) the classes are hierarchical. Then, the decision tree induction in this system operates similarly to the big margin approaches used for structured output prediction [14].

d. Logistic Regression

When there are several explanatory variables, odds ratios can be obtained using logistic regression. With the exception of the response variable being binary, the process is fairly similar to multiple linear regression. The effect of every variable on the odds ratio of the observed event of interest is the end result [15].

e. Multi-Layer Perceptron

A generalization of a perceptron is a collection of layers of perceptron, where every neuron in one layer is connected to every neuron in the other layers. The architecture or configuration of a multilayer perceptron, including the number of layers and neurons per layer, defines it [16].

Each model undergoes evaluation via hyperparameter tweaking by grid search cross-validation to identify the most effective parameters and achieve optimal predictive outcomes.

5. Evaluation Metrics

The performance of the models is measured by four primary evaluation measures. The model's overall classification accuracy is known as accuracy. where the ratio of accurate predictions to all instances assessed is measured by the accuracy metric [17]. Precision is defined as the ratio of accurate forecasts to all positive forecasts. Whereas precision quantifies the proportion of successfully predicted positive patterns in a positive class out of all predicted patterns [17]. The model's recall is its capacity to identify every true case inside a class. The F1-score measures the balance of performance by taking the harmonic mean of precision and recall.

3. Result and Discussion

The program developed as a result of this work may develop a model that uses strawberry leaves to forecast and categorize hydroponic strawberry plants. To choose the optimal model, hyperparameters must first be chosen. Three variations of the hyperparameter are provided throughout the training process: hyperparameter C, which has a value between 0.1 and 100. Gamma hyperparameter using an auto-value or scale option. RBF and linear are the kernel properties. Following that, SVM will choose the hyperparameter value that will allow the model to train on the 66 provided datasets with the best performance.

We used Grid Search 5-Cross-Validation for hyperparameter adjustment in order to maximize each classification model's performance [18] as shown in Table 2. In order to find the combination that produces the best results during cross-validation, this approach thoroughly searches over a given parameter grid. Below are the best-performing hyperparameters for each algorithm:

a. Support Vector Machine
The parameter grid in the figure defines the search space for hyperparameter tuning of the SVM model using Grid Search with Cross-Validation. It explores a range of values for the regularization parameter C from 0.1 to 100, two kernel types: linear and RBF, and kernel coefficient gamma settings, scale, and auto. By systematically evaluating all possible parameter combinations through cross-validation, this process selects the optimal configuration that delivers the highest model performance.

b. Random Forest
The parameter grid in the figure defines the search space for hyperparameter tuning of the Random Forest (RF) model using Grid Search with Cross-Validation. It tests different numbers of estimators 50-200, max depth 10-30, minimum samples required to split a node/ min samples split 2-10, and minimum samples required at min samples leaf 1- 4. It also evaluates whether to use bootstrap sampling between True or False. By systematically checking all parameter combinations through cross-validation, this process identifies the optimal settings that maximize the model's performance.

c. Decision Tree
The parameter settings define the search space for tuning a Decision Tree model. The criterion parameter tests both Gini and entropy to measure the quality of splits. The max depth is varied from 0 to 15 to control tree complexity. The min samples split ranges from 2 to 10, specifying the minimum samples required to split a node, while min samples leaf ranges from 1 to 4, indicating the minimum samples required in a leaf node. These variations allow systematic evaluation to find the configuration that balances model accuracy and overfitting risk.

d. Logistic Regression
The parameter settings define the configuration for tuning a Logistic Regression model. The multi-class option is set to multinomial to handle multiclass classification directly. The solver is set to lbfgs, an efficient optimizer for multinomial problems. The max iter parameter varies from 0 to 200 to control the maximum number of iterations allowed for the solver to converge, ensuring the model is fully optimized without premature stopping.

d. Multi-Layer Perceptron (MLP)
The parameter settings define the configuration for tuning an MLP (Multi-Layer Perceptron) model. The hidden layer sizes range from 50 to 100 neurons, determining the number of units in the hidden layer and influencing the model's capacity to learn complex patterns. The max iter parameter varies from 0 to 1000, controlling the maximum training iterations to ensure the model has sufficient time to converge to an optimal solution.

3.1 Confusion Matrix

A metric for assessing the precision of machine learning classification utilizing the used algorithm is the confusion matrix. There are two types of it: positive and negative [19].

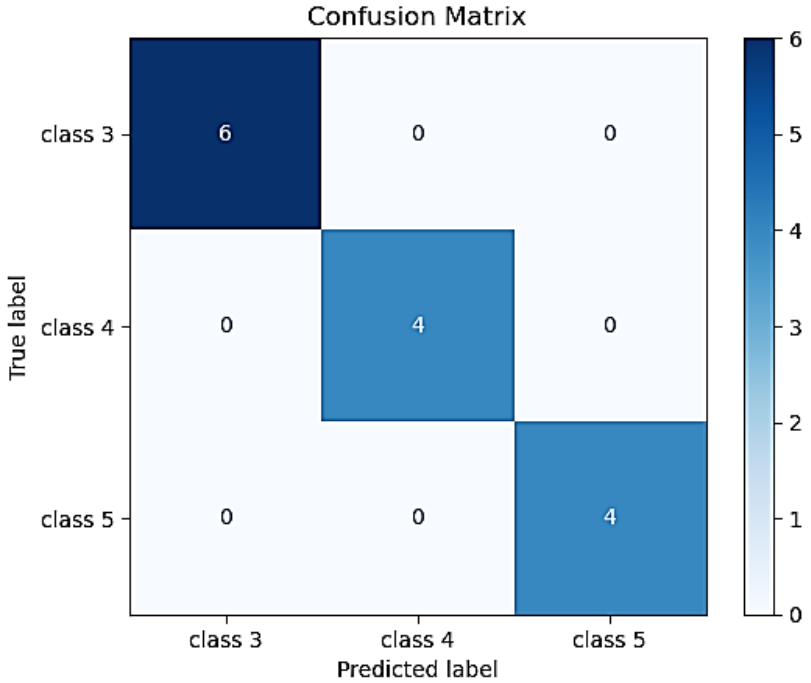

Figure 3. Confusion Matrix

The confusion matrix shows perfect classification performance. All 6 samples of class 3, 4 samples of class 4, and 4 samples of class 5 were correctly predicted with no misclassifications, indicating the model achieved 100% accuracy across all classes. The accuracy, precision, and recall calculations that were performed using the confusion matrix's results are as follows [20-21]:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{F1 Score} = 2 * \frac{precision*recall}{precision+recall} \quad (7)$$

It is possible to formulate the accuracy value of the categorization results in percentage (%) as follows [21]:

$$\text{Accuracy} = \frac{Total\ Prediction\ were\ correct}{Total\ of\ many\ predictions} \text{ x } 100\% \quad (8)$$

3.2 Accuracy Score

Accuracy, precision, recall, and F1-score are the four assessment measures used to compare the performance of five machine learning algorithms in Figure 2 and Table 2.

Table 2. Comparison of Different Methods

Testing	Methods	Precision	Recall	F1-Score	Accuracy
5-Fold Cross-Validation	SVM	0.96	0.95	0.95	0.95
	Random Forest	0.95	0.95	0.95	0.95
	Decision Tree	0.93	0.93	0.93	0.93
	Logistic Regression	0.85	0.85	0.85	0.85
	MLP	0.95	0.95	0.95	0.95

p-ISSN 2088-1541
e-ISSN 2541-5832

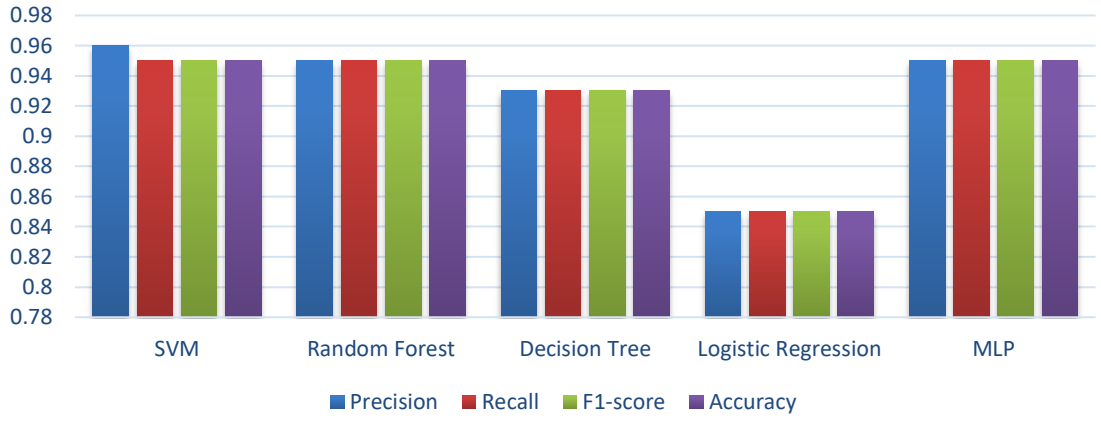

Figure 2. Machine Learning Results

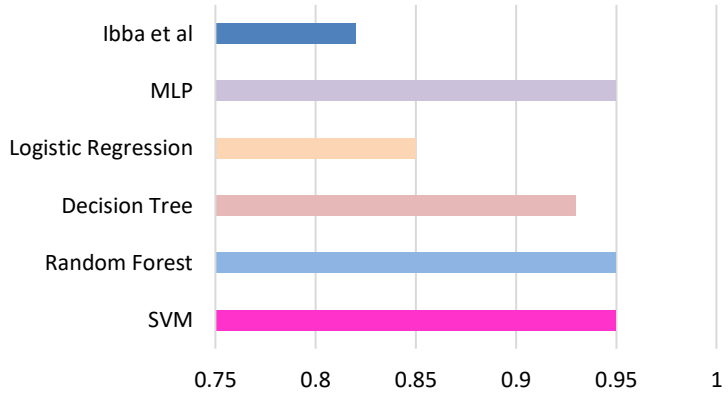

Figure 3. Comparison Results

The comparison results in Table 2, obtained using 5-Fold Cross-Validation, show that the SVM model achieved the highest performance, with a precision of 0.96 and equally high recall, F1-score, and accuracy of 0.95. Random Forest and MLP followed closely, each scoring 0.95 across most metrics. The Decision Tree performed slightly lower at 0.93, while Logistic Regression recorded the lowest scores 0.85 in all metrics. Overall, SVM proved to be the most effective model for this classification task, offering both high precision and balanced performance.

The comparative performance analysis of five classification models such as SVM, Random Forest, Decision Tree, Logistic Regression, and MLP showed that SVM achieved the highest precision 0.96, with Random Forest and MLP tied for second-best 0.95. For recall, Random Forest and MLP shared the top score 0.95, followed closely by SVM 0.95, slightly lower. In terms of F1-score, Random Forest and MLP again led 0.95, with SVM in second place 0.95, marginally lower. Accuracy results mirrored the F1-score pattern, with Random Forest and MLP ranked first 0.95 and SVM second 0.95. The Decision Tree model consistently scored slightly lower 0.93 across all metrics, while Logistic Regression recorded the lowest performance 0.85, indicating reduced effectiveness for the dataset. Overall, SVM excelled in precision, whereas Random Forest and MLP dominated recall, F1-score, and accuracy.

Ibba et al. conducted an experiment on strawberry classification and reported their best performance using the Multi-Layer Perceptron (MLP) method, achieving an F1 Score of 0.82 [22]. In contrast, our study, which applied the Support Vector Machine (SVM) method to a similar classification problem, achieved an F1 Score of 0.95. Since the F1 Score is a balanced metric that considers both precision and recall, the higher value obtained in our research indicates that our model not only made more accurate predictions but also maintained a better balance between

correctly identifying positive cases and avoiding false positives. This improvement suggests that the SVM approach used in our work is more effective than the MLP method applied by Ibba et al., likely due to its ability to handle complex decision boundaries and optimize classification performance in our dataset.

4. Conclusion

Ibba et al. conducted an experiment on strawberry classification and reported their best performance using the Multi-Layer Perceptron (MLP) method, achieving an F1 Score of 0.82. In contrast, our study achieved a substantially higher F1 Score of 0.95 using the Support Vector Machine (SVM) method, with SVM also obtaining the highest precision at 0.96. This demonstrates that our proposed approach outperforms Ibba et al.'s work, offering better accuracy, precision, and recall. Among all the models tested in our research, SVM was the model that performed the best and most consistently across all metrics. Although Random Forest also demonstrated good accuracy and stability, with all metrics around 0.94, it performed marginally worse than SVM, making it suitable for classification tasks requiring high accuracy and low error but still not optimal compared to SVM. Decision Tree models produced metrics hovering around 0.925, delivering decent outcomes but falling short of ensemble methods like Random Forest, and thus are less effective for high-stakes classification problems. MLP in our research achieved results between 0.94 and 0.945, making it comparable to Random Forest and showing promise for handling complex classification tasks; however, it still did not surpass SVM. Finally, Logistic Regression was the weakest performer across all metrics, confirming its limitations in this application. Overall, these results highlight that our proposed SVM based method achieving an F1 Score of 0.95 and precision of 0.96 was superior to Ibba et al.'s MLP approach and other tested models in terms of classification performance.

References

[1] Messelink, G.J., Lambion, J., Janssen, A. and van Rijn, P.C., (2021). Biodiversity in and around greenhouses: Benefits and potential risks for pest management. *Insects*, *12*(10), p.933.

[2] Kouloumprouka Zacharaki, A., Monaghan, J.M., Bromley, J.R. and Vickers, L.H., (2024). Opportunities and challenges for strawberry cultivation in urban food production systems. *Plants, People, Planet*, *6*(3), pp.611-621.

[3] Rajak, P., Ganguly, A., Adhikary, S. and Bhattacharya, S., (2023). Internet of Things and smart sensors in agriculture: Scopes and challenges. *Journal of Agriculture and Food Research*, *14*, p.100776.

[4] Priyambodo, L., Fuadi, H.L., Nazhifah, N., Huzaimi, I., Prawira, A.B., Saputri, T.E., Afandi, M.A., Nugraha, E.S., Wicaksono, A. and Goran, P.K., (2022). Klasifikasi Kematangan Tanaman Hidroponik Pakcoy Menggunakan Metode SVM. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *6*(1), pp.153-160.

[5] Lavanaya, M. and Parameswari, R., (2018), August. Soil nutrients monitoring for greenhouse yield enhancement using pH value with IoT and wireless sensor network. In *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)* (pp. 547-552). IEEE.

[6] Ahonen, T., Virrankoski, R. and Elmusrati, M., (2008), October. Greenhouse monitoring with wireless sensor network. In *2008 IEEE/ASME International Conference on Mechtronic and Embedded Systems and Applications* (pp. 403-408). IEEE.

[7] Both, A.J., Benjamin, L., Franklin, J., Holroyd, G., Incoll, L.D., Lefsrud, M.G. and Pitkin, G., (2015). Guidelines for measuring and reporting environmental parameters for experiments in greenhouses. *Plant Methods*, *11*, pp.1-18.

[8] Mohamed, E.S., Naqishbandi, T.A., Bukhari, S.A.C., Rauf, I., Sawrikar, V. and Hussain, A., (2023). A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms. *Healthcare Analytics*, *3*, p.100185.

[9] Tang, W., 2024. Application of support vector machine system introducing multiple submodels in data mining. *Systems and Soft Computing*, *6*, p.200096.

[10] Sharma, S., Singh, G. and Sharma, M., (2021). A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans. *Computers in Biology and Medicine*, *134*, p.104450.
[11] Bao, R., Chen, W., Tang, G., Chen, H., Sun, Z., & Chen, F. (2018). Classification of fresh and processed strawberry cultivars based on quality characteristics by using support vector machine and extreme learning machine. *Journal of Berry Research*, *8*(2), 81-94.
[12] Mayoraz, E., & Alpaydin, E. (1999, June). Support vector machines for multi-class classification. In *International work-conference on artificial neural networks* (pp. 833-842). Berlin, Heidelberg: Springer Berlin Heidelberg.
[13] Tripathi, A., Goswami, T., Trivedi, S. K., & Sharma, R. D. (2021). A multi class random forest (MCRF) model for classification of small plant peptides. *International Journal of Information Management Data Insights*, *1*(2), 100029.
[14] Vens, C., Struyf, J., Schietgat, L., Džeroski, S., & Blockeel, H. (2008). Decision trees for hierarchical multi-label classification. *Machine learning*, *73*, 185-214.
[15] Maalouf, M. (2011). Logistic regression in data analysis: an overview. *International Journal of Data Analysis Techniques and Strategies*, *3*(3), 281-299.
[16] Murtagh, F. (1991). Multilayer perceptrons for classification anregression. *Neurocomputing*, *2*(5-6), 183-197.
[17] Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, *5*(2), 1.
[18] Adnan, M., Alarood, A. A. S., Uddin, M. I., & ur Rehman, I. (2022). Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. *PeerJ Computer Science*, *8*, e803.
[19] Naidu, G., Zuva, T., & Sibanda, E. M. (2023, April). A review of evaluation metrics in machine learning algorithms. In *Computer science on-line conference* (pp. 15-25). Cham: Springer International Publishing.
[20]Ting, K. M. (2016). Confusion matrix. In Encyclopedia of machine learning and datamining (pp. 1-1). Springer, Boston, MA.
[21] Luque, A., Carrasco, A., Martín, A., & de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, 216-231
[22] Ibba, P., Tronstad, C., Moscetti, R., Mimmo, T., Cantarella, G., Petti, L., ... & Lugli, P. (2021). Supervised binary classification methods for strawberry ripeness discrimination from bioimpedance data. *Scientific reports*, *11*(1), 11202.