Global Education Journal
Vol. 4, 2 (May-August, 2026), pp. 303-309
EISSN: 2986-6529, DOI: https://doi.org/10.59525/gej.1454

Analysis of the Classification of Environmental Conditions of Chili Plants into Healthy and Unhealthy Categories Using the Decision Tree Algorithm

Ermita Sari[1], Noviyanti[2], R. Joko Musridho[3], Muhammad Rezky Ramadhan[4], Vivi Damayanti[5], Isvihani Ramadhan[6], Fajar Delonda[7], M. Irsyadul Fikri[8], Amal Alfarizal[9]

[1-9] Universitas Pahlawan Tuanku Tambusai, Indonesia: ermitasari790@gmail.com

ARTICLE INFO

Keywords:

Classification, Chili Plants, Data Mining, Decision Tree J48, Environmental Conditions.

Article history:

Received 2026-04-14
Revised 2026-05-12
Accepted 2026-06-27

ABSTRACT

Changes in various environmental parameters during the cultivation process can have an impact on plant development and have the potential to reduce the quality of the results obtained. Therefore, a method is needed that is able to identify plant conditions quickly, objectively, and accurately. This study aims to build a classification model of chili plant conditions using the Decision Tree J48 algorithm by utilizing environmental parameter data and soil nutrients as the basis for determining healthy and unhealthy plant categories. This research method uses a quantitative approach with data mining techniques on a public dataset consisting of 2,200 datasets. The variables analyzed included nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, soil pH (ph), and rainfall. The process of forming and testing the model was carried out using the WEKA application with the 10-fold cross validation technique. The results showed that moisture was the most dominant attribute in the formation of decision trees, followed by rainfall, temperature, and soil pH as supporting attributes. The resulting model obtained an accuracy rate of 99.5909%, a Kappa Statistic value of 0.989, a Mean Absolute Error (MAE) of 0.0042, and a Root Mean Squared Error (RMSE) of 0.064. These values indicate that the model has an excellent degree of conformity with a relatively low rate of prediction error. Based on these results, the Decision Tree J48 algorithm has proven to be effective in classifying the condition of chili plants based on environmental parameters and has the potential to be applied as the basis for the development of a decision support system to help the process of monitoring plant conditions more quickly, objectively, and accurately.

Corresponding Author:
Ermita Sari
Universitas Pahlawan Tuanku Tambusai, Indonesia; ermitasari790@gmail.com

1. INTRODUCTION

Technological developments in the agricultural sector have encouraged the implementation of modern agricultural systems (smart farming) as an effort to improve the quality and productivity of crops. The use of this technology provides opportunities to increase the effectiveness of the monitoring and management process of plants so that cultivation activities can be carried out more efficiently and optimally. One of the commodities that has received a lot of attention in the development of modern agriculture is the chili plant because market demand and public demand for this commodity continue to increase. In addition to being used as a food, chili peppers also contain various compounds that provide health benefits. The high demand for chili peppers makes this commodity have strategic economic value and contributes to increasing farmers' income in various regions (Nurfalah et al., 2025).

The optimal growth of chili plants is greatly influenced by the suitability of environmental conditions during the cultivation process, such as temperature, air humidity, rainfall, and soil pH. In addition to these environmental factors, the availability of soil nutrients consisting of nitrogen (N), phosphorus (P), and potassium (K) also plays an important role in supporting vegetative growth, plant development, and crop quality. If the environmental conditions are in the appropriate range, plant growth can take place well. Conversely, less favorable conditions have the potential to inhibit crop development and reduce crop productivity. According to Nurfalah et al. (2025), increased extreme rainfall, high humidity, and unstable temperatures are the main factors that affect the condition of chili plants and lead to a decrease in crop productivity. In addition, Amir et al. (2025), states that environmental parameters have an important influence on plant growth and cultivation success rates. Nurfalah et al. (2025), explaining that extreme weather conditions and high humidity can increase the risk of disturbance to crops and cause a significant decrease in crop yields. Therefore, it is necessary to monitor environmental conditions regularly so that maintenance actions can be carried out more quickly and appropriately.

Although monitoring the environmental conditions of plants has a very important role, its implementation in the field is still mostly done manually. This method takes a relatively longer time and has the potential to result in less accurate decisions. In addition, manual monitoring methods are considered less efficient when applied to agricultural areas that have a wide land coverage. As technology develops, the application of data mining in agriculture opens up opportunities to utilize environmental data in automatically analyzing crop conditions based on several parameters, such as temperature, humidity, soil pH, and rainfall. Amir et al. (2025), explaining that the data preprocessing and classification process has an important role in producing more accurate analytical models in agricultural data processing.

Several previous studies have explained that classification methods can be used to support the analysis of agricultural data. Nurfalah et al. (2025), indicating that the Decision Tree algorithm is effectively used in the process of analyzing agricultural data because it is able to generate easy-to-understand decision rules. In addition, the study also explains that temperature and humidity parameters have a significant influence on plant conditions. The results of the study show that the Decision Tree-based classification method has the potential to be used in identifying plant conditions based on the results of the observed environmental parameters.

Based on the problems that have been explained, this study aims to classify the environmental conditions of chili plants into healthy and unhealthy categories by utilizing the Decision Tree algorithm based on environmental sensor data. The study used a public dataset obtained from Kaggle and processed using the WEKA application to build a classification model based on the parameters of temperature, humidity, soil pH, rainfall, and soil nutrient content. The results obtained are expected to help farmers in identifying plant conditions more quickly and accurately so that they can support decision-making in cultivation activities, such as watering arrangements, environmental condition management, and preventive measures against plant damage.

2. METHODS

The developed model was used to group plant conditions into two categories, namely healthy and unhealthy, based on environmental parameters and soil nutrients contained in the dataset. The selection of the J48 Decision Tree algorithm was based on its ability to generate easy-to-understand classification rules while describing the relationships between parameters that affect plant conditions. In addition, this algorithm has been widely used in various studies in the field of agriculture to support the process of classification and data-driven decision-making (Tariq et al., 2023).

The dataset used in this study is a public dataset obtained from Kaggle and contains information about the environmental conditions of plants and soil nutrient content. The variables used include nitrogen (N), phosphorus (P), potassium (K), temperature, air humidity, soil pH, rainfall, plant labels, and condition attributes that act as classification targets. At the classification stage, the attributes of the condition are divided into two categories, namely healthy and unhealthy. All stages of research, from data processing, model formation, to model performance evaluation, were carried out using the WEKA application. The research stages include data collection, data preprocessing, formation of classification models using the J48 Decision Tree algorithm, data sharing, model evaluation, and interpretation of results. Each stage is carried out sequentially to produce a classification model that is able to identify the condition of chili plants based on the environmental parameters available in the dataset.

The data collection stage was carried out using public datasets obtained from Kaggle as the main source of research. The dataset was chosen because it contains information about environmental parameters and soil nutrients that are in accordance with the purpose of the study, which is to classify the condition of chili plants into healthy and unhealthy categories. The variables used in the formation of the classification model include nitrogen (N), phosphorus (P), potassium (K), temperature, air humidity, soil pH, and rainfall as parameters that represent the environmental conditions of the plant (Amir et al., 2025). On the contrary, these inconsistencies have the potential to affect plant health so that all attributes in the dataset are used as a basis in the classification process (Nurfalah et al., 2025).

In this study, preprocessing includes checking blank data, identifying duplicate data, checking data types on each attribute, and adjusting data formats to suit the processing process in the Weka application. In addition, a relabeling process was carried out on the attributes of the condition so that the classification target consisted of two categories, namely healthy and unhealthy. The dataset that has gone through the preprocessing stage is then used as input data in the process of forming a classification model. The next stage in this study is to build a classification model using the Decision Tree J48 algorithm through the WEKA application. The model was developed by utilizing the attributes of nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, soil pH, and rainfall as predictor variables to determine the category of chili plant conditions, namely healthy or unhealthy. The J48 Decision Tree algorithm forms a classification model in the form of a decision tree based on the attributes that have the best ability to distinguish healthy and unhealthy categories. The resulting model is then used as a basis to analyze the relationship between environmental parameters and the condition of chili plants without changing the characteristics of the data contained in the dataset (Tariq et al., 2023). After the classification model is successfully formed, the next stage is to conduct an evaluation using the 10-fold cross validation method. In this method, the dataset is divided into ten parts (fold), then each part is used alternately as training data and test data until all data gets the same opportunity to be tested. This approach was chosen because it was able to produce a more representative evaluation and reduce the potential for overfitting in the classification model developed (Tariq et al., 2023).

For comparison, this study also uses the percentage split method with a data division of 80% as training data and 20% as test data. The application of the two evaluation methods aims to compare the performance of the Decision Tree J48 model in classifying the condition of chili plants based on the environmental parameters used in this study. The evaluation was carried out to determine the ability of the Decision Tree J48 model that has been developed in classifying the condition of chili plants into healthy and unhealthy categories. The measurement of model performance in this study uses a confusion matrix with several evaluation indicators, namely accuracy, precision, recall, and F1-score, so that the performance of the model can be analyzed comprehensively. In addition to calculating the evaluation value, this study also observed the attributes that contributed the most to the classification process based on the structure of the resulting decision tree. The results of the evaluation were used to determine the extent to which the model was able to utilize environmental parameters as a basis in the process of classifying the condition of chili plants (Amir et al., 2025).

The last stage is carried out by interpreting the classification model generated by the Decision Tree J48 algorithm. Interpretation is carried out on the structure of the decision tree to understand the classification rules formed from a combination of nitrogen, phosphorus, potassium, temperature, humidity, soil pH, and rainfall parameters. Through this stage, the study not only evaluates the performance of the model, but also identifies the relationships between the parameters used in the classification process. This information is expected to be the basis for supporting the process of monitoring the condition of chili plants more quickly and helping decision-making in data-based cultivation activities (Talaat, 2023).

3. FINDINGS AND DISCUSSION

This section presents the results of the application of the Decision Tree J48 algorithm in classifying the environmental conditions of chili plants based on environmental parameters and soil nutrients. In addition to displaying the results of model formation and performance evaluation, this section also discusses the interpretation of the resulting decision tree and its relationship to previous research.

Decision Tree Classification and Analysis Results

The classification model in this study was built using the Decision Tree J48 algorithm through the WEKA application by applying the 10-fold cross validation method. The dataset used amounted to 2,200 data with predictor variables in the form of nitrogen (N), phosphorus (P), potassium (K), temperature, air humidity, soil pH, and rainfall. Meanwhile, plant conditions are used as a classification target which is grouped into two categories, namely healthy and unhealthy.

The results of the model formation show that the J48 Decision Tree algorithm produces a decision tree with 10 leaves and 19 nodes (size of tree). The structure of the tree that is formed shows that the humidity attribute is chosen as the root node, which shows that this attribute has the highest ability to distinguish the condition of chili plants compared to other attributes. In addition to moisture, the attributes of rainfall, temperature, and soil pH also appear on several branches of the decision tree as supporting parameters in the decision-making process.

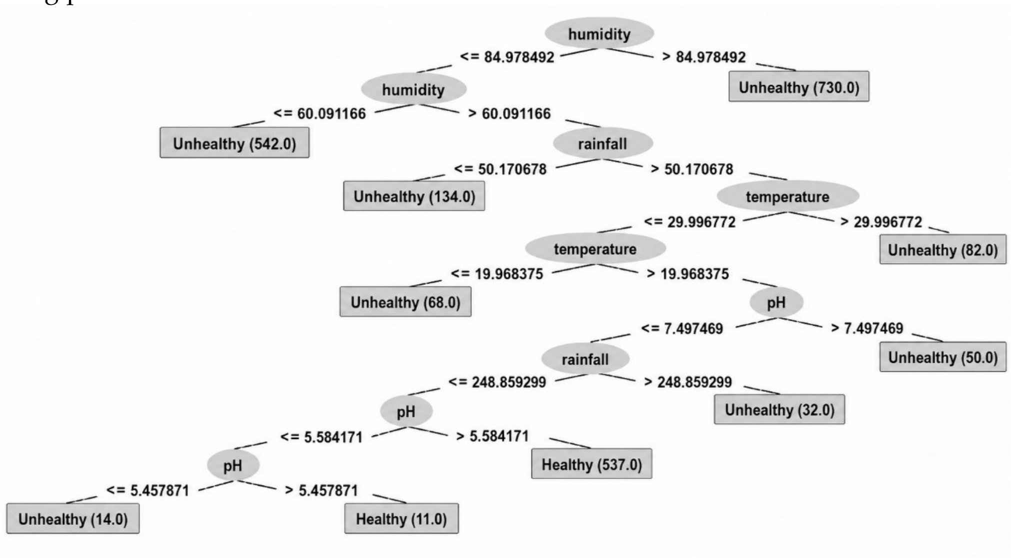

Figure 1. J48 Algorithm Decision Tree

Based on the resulting decision tree structure, the classification process is carried out through a combination of several environmental parameters so that each branch produces a different decision rule. The rule is used to determine whether the condition of the plant is included in the healthy or unhealthy category from the characteristics of the data owned. Thus, the model built not only produces classifications, but also provides an overview of the relationships between environmental parameters that affect the condition of chili plants.

Model Performance Evaluation

After the classification model was formed, an evaluation was carried out using the 10-fold cross validation method to measure the ability of the Decision Tree J48 algorithm in classifying the condition of chili plants. This method was chosen because all data gets the same opportunity to be used as training data and test data so that the evaluation results become more representative. The testing process resulted in a model that successfully classified 2191 data correctly out of a total of 2200 data, while 9 data experienced misclassification. A summary of the evaluation results is presented in Table 1.

Table 1. Decision Tree Model Evaluation Results J48

Evaluation Parameters	Value
Correctly Classified Instances	2191
Incorrectly Classified Instances	9
Accuracy	99,5909%
Kappa Statistic	0,989
Mean Absolute Error	0,0042
Root Mean Squared Error	0,064
Total Instances	2200

Based on Table 1, the Decision Tree J48 model yielded an accuracy rate of 99.5909%, which indicates that almost all of the data was correctly classified. A Kappa Statistical value of 0.989 indicates a very high degree of conformity between the results of the model classification and the actual conditions. In addition, the Mean

Absolute Error (MAE) value of 0.0042 and the Root Mean Squared Error (RMSE) of 0.064 indicate that the prediction error generated is relatively low. These results indicate that the model has excellent classification ability in distinguishing the condition of chili plants.

Discussion

The results of the study indicated that the Decision Tree J48 algorithm obtained an accuracy rate of 99.5909% in classifying the condition of chili plants into healthy and unhealthy categories. The high accuracy value indicates that the environmental attributes and soil nutrients used in this study are able to represent the characteristics of plant conditions well. The combination of nitrogen, phosphorus, potassium, temperature, moisture, soil pH, and precipitation parameters provides sufficient information for the algorithm to recognize classification patterns so that the number of prediction errors generated is relatively small.

From the yield decision tree, the humidity attribute was selected as the root node, which illustrates that this parameter has the greatest influence in distinguishing the condition of chili plants in the dataset used. This condition suggests that changes in humidity levels are the main factor that the algorithm considers in forming classification rules before considering other parameters, such as rainfall, temperature, and soil pH. This suggests that humidity has a better ability to separate data into healthy and unhealthy categories than other attributes. In addition to moisture, the attributes of rainfall, temperature, and soil pH also contribute to shaping the decision rules on some tree branches. From these results, it can be seen that the condition of chili plants is not determined by one parameter alone, but is influenced by a combination of several interrelated environmental factors. This combination allows the algorithm to generate clearer decision boundaries so that the classification process can be carried out more accurately.

The high level of accuracy obtained indicates that the model has succeeded in recognizing the pattern of relationships between parameters contained in the dataset. The number of misclassification errors that only reached 9 data out of a total of 2200 data shows that most of the characteristics of plant conditions can be well studied by the Decision Tree J48 algorithm. In addition to producing high classification performance, the model also generates a relatively simple decision tree structure so that the relationships between parameters can be understood more easily.

The findings of this study demonstrate that the Decision Tree J48 algorithm achieved an exceptionally high classification accuracy of 99.5909%, indicating that the selected environmental and soil nutrient variables effectively captured the characteristics distinguishing healthy and unhealthy chili plants. The decision tree identified humidity as the root node, followed sequentially by rainfall, temperature, and soil pH, suggesting that humidity possessed the greatest information gain in separating the two classes. This result is theoretically consistent with the principles of decision tree learning, where the root attribute is selected because it provides the highest discriminatory power and minimizes classification uncertainty (Tariq et al., 2023). From an agronomic perspective, humidity strongly influences transpiration, pathogen proliferation, and physiological processes in chili plants, making it a biologically meaningful predictor. Rather than relying on a single environmental factor, the model integrated multiple interacting variables to construct decision boundaries, demonstrating that plant health is determined through complex environmental interactions rather than isolated parameters. These findings reinforce the concept of precision agriculture, which emphasizes the integration of multidimensional environmental data to support intelligent crop monitoring and management (Talaat, 2023).

The present findings are broadly consistent with previous studies while also providing several important contributions. Nurfalah et al. (2025) reported that humidity and temperature were dominant determinants in predicting chili crop failure using a Decision Tree model, whereas Amir et al. (2025) found that decision tree-based classification effectively identified land suitability for chili cultivation through environmental variables. Similar to those studies, the current research confirms the importance of climatic parameters in determining agricultural outcomes. The remarkably high accuracy suggests that the preprocessing procedure, balanced dataset, and application of 10-fold cross-validation successfully minimized classification bias and enhanced the model's generalization capability. Consequently, this research extends previous literature by demonstrating that integrating environmental and nutrient variables substantially strengthens the predictive performance of interpretable machine learning models in agricultural monitoring (Tariq et al., 2023; Talaat, 2023).

This interpretability is essential for supporting evidence-based management decisions, particularly regarding irrigation scheduling, humidity regulation, soil pH adjustment, and rainfall adaptation strategies. Nevertheless, the outstanding performance reported in this study should also be interpreted cautiously because the model was developed using a single public dataset, which may not fully represent environmental variability across different geographical regions or cultivation systems. Future research should therefore validate the model using larger, multi-location datasets and compare its performance with more advanced ensemble learning algorithms, such as Random Forest and Extreme Gradient Boosting, to determine whether the observed predictive superiority remains consistent under more diverse agricultural conditions. Such investigations would contribute to establishing more robust and transferable intelligent decision-support systems for sustainable precision agriculture (Amir et al., 2025; Talaat, 2023).

So, the results of the study show that the Decision Tree J48 algorithm is able to build an effective classification model in classifying the condition of chili plants based on environmental parameters and soil nutrients. The resulting model not only provides a high level of accuracy, but is also able to show the most influential parameters in the classification process. This information has the potential to be used as a basis for the development of a decision support system to help the process of monitoring the condition of chili plants more quickly, objectively, and data-based.

4. CONCLUSION

This study successfully implemented the Decision Tree J48 algorithm to classify the condition of chili plants into healthy and unhealthy categories based on the parameters of nitrogen (N), phosphorus (P), potassium (K), temperature (temperature), humidity, soil pH (ph), and rainfall (rainfall). The results of the model formation show that moisture attributes play a role as the main parameter in the preparation of the decision tree, while rainfall, temperature, and soil pH are supporting parameters in the classification process. Based on the evaluation using the 10-fold cross validation method, the model produced an accuracy rate of 99.5909%, with a Kappa Statistic value of 0.989, a Mean Absolute Error (MAE) of 0.0042, and a Root Mean Squared Error (RMSE) of 0.064. These values show that the model has a very high degree of conformity with a relatively low rate of prediction error in identifying the condition of chili plants. The findings of this study prove that the Decision Tree J48 algorithm is able to build an effective classification model through the use of the relationship between environmental parameters and soil nutrients so that it can accurately distinguish the condition of chili plants. Therefore, the resulting model has the potential to be developed as the basis for a decision support system to help monitor the condition of chili plants more quickly and objectively based on the results of the classification of environmental parameters, while supporting more effective, efficient, and measurable cultivation management.

REFERENCES

Amir, A., Fachruddin, F., Idris, F., Safriani, M., Saefuddin, R., & Sakti, I. (2025). Perbandingan Model Decision Tree dan Random Forest untuk Penentuan Kesesuaian Lahan Budidaya Cabai dan Terong. 12(5), 695–704. https://doi.org/10.30865/jurikom.v12i5.8672

Amir, A., Fachruddin, F., Idris, F., Safriani, M., Saefuddin, R., & Sakti, I. (2025). *Perbandingan model Decision Tree dan Random Forest untuk penentuan kesesuaian lahan budidaya cabai dan terong*. JURIKOM, 12(5), 695–704. https://doi.org/10.30865/jurikom.v12i5.8672

Chitradurga Manjunath, M., & Palayyan, B. P. (2023). An efficient crop yield prediction framework using hybrid machine learning model. *Revue d'Intelligence Artificielle, 37*(4), 1057–1067. https://doi.org/10.18280/ria.370428

European Journal of Agronomy. (2023). Developing machine learning models for wheat yield prediction using ground-based data, satellite-based actual evapotranspiration and vegetation indices. *European Journal of Agronomy, 146*, 126820. https://doi.org/10.1016/j.eja.2023.126820

Iniyan, S., Varma, V. A., & Naidu, C. T. (2023). Crop yield prediction using machine learning techniques. *Advances in Engineering Software, 175*, 103326. https://doi.org/10.1016/j.advengsoft.2022.103326

Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop yield prediction using machine learning and deep learning techniques. *Procedia Computer Science, 218*, 406–417. https://doi.org/10.1016/j.procs.2023.01.023

Kumar, S., et al. (2024). Crop yield prediction using multi-attribute weighted tree-based support vector machine. *Measurement: Sensors, 31*, 101002. https://doi.org/10.1016/j.measen.2023.101002

Mohan, R. N. V. J., Rayanoothala, P. S., & Sree, R. P. (2025). Next-gen agriculture: Integrating AI and XAI for precision crop yield predictions. *Frontiers in Plant Science, 15*. https://doi.org/10.3389/fpls.2024.1451607

Morales, A., & Villalobos, F. J. (2023). Using machine learning for crop yield prediction in the past or the future. *Frontiers in Plant Science, 14*, 1128388. https://doi.org/10.3389/fpls.2023.1128388

Moussaid, A., El Fkihi, S., Zennayi, Y., et al. (2022). Machine learning applied to tree crop yield prediction using field data and satellite imagery. *Informatics, 9*(4), 80. https://doi.org/10.3390/informatics9040080

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing, 14*(9), 1990. https://doi.org/10.3390/rs14091990

Nurfalah, K., Firmansyah, E., Sutara, B., & Sofiyan, Y. (2025). Prediksi Risiko Gagal Panen Cabai Rawit Merah Menggunakan Algoritma Decision Tree. 5(2), 111–120.

Nurfalah, K., Firmansyah, E., Sutara, B., & Sofiyan, Y. (2025). Prediksi risiko gagal panen cabai rawit merah menggunakan algoritma Decision Tree. *Jurnal Agritech*, 5(2), 111–120.

Pankaj, Bharti, P. K., & Kumar, B. (2023). Crop yield prediction using machine learning: A review of recent approaches. *International Journal of Computer Applications, 185*(24), 27–32. https://doi.org/10.5120/ijca2023922994

Pavani, S., & Beulet, A. S. (2024). Improved precision crop yield prediction using weighted-feature hybrid SVM. *IETE Journal of Research, 70*(3). https://doi.org/10.1080/03772063.2023.2192000

Ramadhan, M. R., & Noviyanti. (2025). Penerapan algoritma Decision Tree pada klasifikasi data pertanian berbasis sensor lingkungan. *Jurnal Informatika Pertanian, 9*(1), 45–56.

Shams, M. Y., Gamel, S. A., & Talaat, F. M. (2024). Enhancing crop recommendation systems with explainable artificial intelligence: A study on agricultural decision-making. *Neural Computing and Applications, 36*, 5695–5714. https://doi.org/10.1007/s00521-023-09391-2

Sharma, V., Honkavaara, E., Hayden, M., & Kant, S. (2024). UAV remote sensing phenotyping of wheat collection for response to water stress and yield prediction using machine learning. *Plant Stress, 12*, 100464. https://doi.org/10.1016/j.stress.2024.100464

Talaat, F. M. (2023). Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. Neural Computing and Applications, 35(23), 17281–17292. https://doi.org/10.1007/s00521-023-08619-5

Talaat, F. M. (2023). Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. *Neural Computing and Applications, 35*(23), 17281–17292. https://doi.org/10.1007/s00521-023-08619-5

Tariq, A., Yan, J., Gagnon, A. S., & Khan, M. R. (2023). Geo-spatial Information Science Mapping of cropland , cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest. Geo-Spatial Information Science, 26(3), 302–320. https://doi.org/10.1080/10095020.2022.2100287

Tariq, A., Yan, J., Gagnon, A. S., & Khan, M. R. (2023). Mapping cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest. *Geo-spatial Information Science, 26*(3), 302–320. https://doi.org/10.1080/10095020.2022.2100287

Tripathi, A., Rathore, B. S., & Singh, D. (2023). Survey paper on agricultural dataset for improving crop yield prediction using machine learning algorithms. *International Journal of Computer Applications, 184*(46), 28–34. https://doi.org/10.5120/ijca2023922571