ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Application of expert system using certainty factor method to identify diseases in rice plants

Isni Azmi[1], Gunawan Gunawan[2], Sawaviyya Anandianskha[3]
[1,3]Information System, STMIK YMI Tegal, Indonesia
[2]Informatics Engineering, STMIK YMI Tegal, Indonesia

ARTICLE INFO

Article history:

Received Mar 18, 2024
Revised Mar 31, 2024
Accepted Mar 31, 2024

Keywords:

Agricultural Technology;
Certainty Factor Method;
Disease Identification;
Expert System;
Rice Plants.

ABSTRACT

This article explores the application of expert systems using certainty factor methods for disease identification in rice crops, highlighting the importance of information technology integration in agriculture. The study aims to develop a system that allows quick and accurate identification of rice disease, using certainty factor methods that are effective in dealing with data uncertainty. This study used a quantitative approach with a quasi-experimental design. The results indicate an effective system for identifying diseases, with significant implications for supporting farmers and improving food security. Suggestions for future research include system integration with mobile applications and real-time data analysis to improve system accessibility and applicability in modern agricultural practices.

Corresponding Author:

Isni Azmi,
Information System,
STMIK YMI Tegal,
#1 Pendidikan Street, Tegal City, Central Java, 52142, Indonesia
Email: isnyazmi03@gmail.com

1. INTRODUCTION

In today's digital era, integrating information technology in the agricultural sector, especially rice crops, is becoming increasingly important. Rice, as one of the world's main food commodities, faces various challenges, including disease attacks that can significantly reduce the quality and quantity of production (Shahzad et al., 2021). Early and accurate identification of diseases in rice plants is crucial to take appropriate handling measures (Sethy et al., 2020), (Li et al., 2020). However, in many regions, particularly developing countries, access to agricultural experts is limited, and knowledge about identifying rice plant diseases is not always accessible.

These limitations pose problems in the health management of rice crops, which can affect a country's food security (Karthikeyan et al., 2020). Diseases of rice crops not only reduce the quality of yields but can also lead to crop failure, which leads to substantial economic losses for farmers (Shan et al., 2021). Therefore, quick and accurate identification of rice plant diseases becomes very important.

This research was conducted to overcome this problem by developing an expert system that uses the certainty factor method to identify diseases in rice plants (Chhetri et al., 2023). The certainty factor method was chosen because of its ability to handle uncertainty and subjectivity in disease diagnosis (Jiang et al., 2021; Ju et al., 2022; Seoni et al., 2023). This research aims to fill the gap between the need for specialist knowledge in remote areas and provide a widely accessible system.

The use of this expert system can strengthen farmers' capacity to identify rice plant diseases more quickly and accurately through online applications, enabling timely interventions to prevent the spread of diseases (Haider et al., 2021). Thus, this research contributes to the sustainable improvement of rice production and helps strengthen food security (Viana et al., 2022).

The certainty factor method in this expert system is integrated with a knowledge base developed through collaboration between informatics researchers and experts in rice crop agriculture (Hosseini & Ivanov, 2020; Orlando et al., 2020; Talukdar et al., 2022). This is a proposed innovation whereby the system will not only identify the disease but also provide initial treatment recommendations based on the level of certainty of diagnosis (Eslam et al., 2020).

With an expert system that is able to detect disease early and provide recommendations for appropriate action, farmers can reduce crop losses, increase agricultural productivity, and optimize the use of resources efficiently. Apart from that, this research will also contribute to increasing farmers' knowledge about healthy and sustainable agricultural practices, as well as ensuring the availability of sufficient food for the community (Mugniati & Witanti, 2024).

This research will significantly contribute to overcoming the problem of disease identification in rice plants by utilizing information technology. Developing and implementing this expert system can provide practical and efficient solutions to increase rice production and strengthen food security in various regions (Ristaino et al., 2021).

2. RESEARCH METHOD

2.1 Research Design

This study used a quantitative approach with a quasi-experimental design to test the effectiveness of the expert system using the certainty factor method in identifying diseases in rice plants (Béné et al., 2020). This design was chosen because it allowed researchers to observe the effects of interventions (expert systems) on subjects (rice disease data) under conditions that were not fully controlled, mimicking real-world application settings. The absence of control groups and complete randomization reflect limited access to resources and practical conditions on the ground.

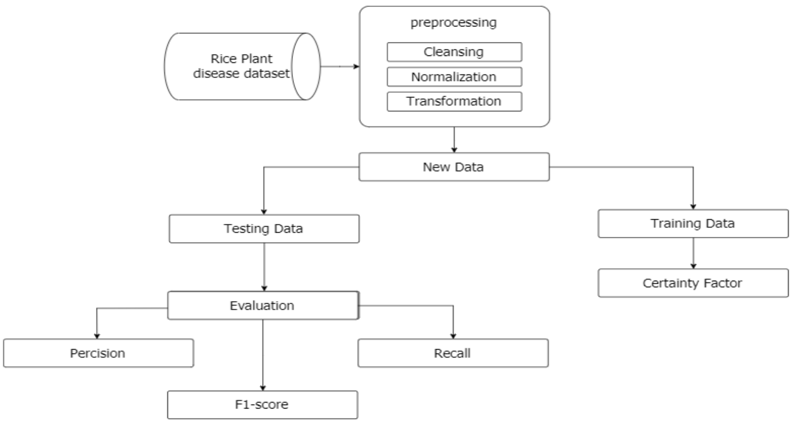

Figure 1. Research flow

Figure 1 describes a structured research flow to develop an expert system for identifying rice plant diseases. The process begins with pre-processing rice plant disease datasets, which involves cleaning the data from discrepancies or loss of value, normalizing to ensure the data is in a consistent format, and transforming the data into an analysis-ready format. The resulting new data is then divided into two parts: training data, which is used to develop a model with certainty factor methods, and test data, which is used to evaluate the model. Model evaluation is done by measuring precision, recall, and F1 scores, key metrics for assessing system performance in correctly identifying diseases. This diagram summarizes the methodical process from initial data processing to final model evaluation.

2.2 Data Collection

The dataset consists of entries that record various disease cases in rice plants, including symptoms, types of diseases, and plant conditions (Patil & Kumar, 2022). Data were collected from several articles that have conducted disease research on rice plants and related literature. Each entry has been annotated with a disease diagnosis confirmed by an agricultural expert, which will be used as ground truth for expert system evaluation (Alqahtani et al., 2023).

2.3 Pra-processing Data

Data pre-processing involves several essential steps to prepare a dataset for analysis, including data cleansing, normalization, and transformation. Data cleansing is done to remove or correct missing or inconsistent entries. Data normalization is necessary to measure variables on the same scale, facilitating analysis. Data transformations, such as category encoding, are performed to convert text data into a format that algorithms can process.

2.4 Research Procedure

The research procedure begins by dividing the dataset into training and testing. Training sets are used to develop expert system models, while test sets are used for evaluation. System experts were trained to model the relationship between symptoms and disease diagnosis using certainty factor methods. After training, the system is tested on a test set to assess its disease-identifying performance. Training Set: Used to develop or 'train' an expert system model. Testing Set: Used to evaluate how well a trained model can predict or classify never-before-seen data. Usually, this division is done randomly with a certain proportion (for example, 70% for training and 30% for testing).

Expert System Training: This stage will use training sets to develop an expert system model. Using certainty factor (CF) methods can model confidence or confidence in the relationship between symptoms and disease diagnosis (Santra et al., 2020). The certainty factor is calculated using the formula Where is the combined certainty factor between hypothesis $CF(H,E) = CF(H) \times CF\ (E|H) CF(H,E) H$ and proof *E* (Santhoshkumar & Dhinesh Babu, 2020). is the initial certainty factor of the hypothesis before looking at the evidence. Is the certainty factor of the proof $CF(H) CF\ (E|H) E$ given by hypothesis *H*. In the of a medical expert system, $CF(H)$ it can be the initial level of confidence that a patient has a particular disease based on prior medical knowledge and $CF\ (E|H)$ it can be the level of confidence that certain symptoms are related to that disease (Hamedan et al., 2020).

2.5 Data Analysis

Data analysis involves using performance metrics, such as Accuracy, precision, recall, and F1-score, to evaluate the effectiveness of expert systems in identifying diseases in rice crops (Ahad et al., 2023). This analysis helps in understanding the strengths and weaknesses of the system in various conditions and types of diseases. Accuracy is the ratio of the number of correct predictions (positive and negative) to the total number of predictions made. Accuracy provides a general idea of system performance but is not always the best metric for systems with an unbalanced distribution of classes (Yağcı, 2022). Precision measures the proportion of optimistic predictions that are correct.

$$\text{Presisi} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (1)$$

Precision is necessary in contexts where the cost of false positives is high, for example, in diagnosing severe plant diseases. The recall measures the proportion of actual positive cases that are correctly identified.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (2)$$

Recall is essential when not identifying an actual positive (disease) has more significant consequences than misidentifying. The F1-Score is the harmonic mean of precision and recall, balancing the two. This is especially useful when you need metrics that consider false positives and negatives.

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi \times Recall} \quad (3)$$

Using these metrics, you can conduct a comprehensive analysis of how your expert system works in diagnosing various diseases in rice crops. For example, high precision indicates that the diagnosis is most likely correct when the system identifies a disease. Meanwhile, high recall suggests the system effectively catches actual disease cases. F1-score balances precision and recall, which is very useful in cases where you need to consider both aspects in balance (Chicco & Jurman, 2020).

2.6 Evaluation

The final evaluation of the expert system involves a comprehensive analysis of the test results, including identifying cases where the system performed well or underperformed. This evaluation aims to identify potential areas for improvement and further system development.

Once the system is trained, it will evaluate it using a test set. This evaluation usually involves calculating performance metrics such as Accuracy, precision, recall, and F1-Score. For example, Accuracy is calculated using formulas (Maxwell et al., 2021).

$$Accuracy = \frac{\text{Number of correct prediction}}{\text{Total number of case}} \quad (4)$$

Where "number of correct predictions" is the number of cases in which the expert system successfully identified the correct diagnosis, and "total number of cases" is the total number of cases in the test set (Jain et al., 2020).

3. RESULTS AND DISCUSSIONS

This research developed an expert system using the certainty factor method for disease identification in rice plants. The system was successfully trained and tested with significant effectiveness based on a dataset that includes various cases of disease with symptoms, disease types, and plant conditions.

Table 1. Disease dataset

Disease Code	Disease Name
P1	Blast
P2	Brown leaf spots
P3	Leaf Blight
P4	Rice Tungro
P5	Rhizoctonia solani
P6	Striped spots

Table 1 is a list of rice plant diseases identified with unique codes P1 to P6, including Blast disease, Brown leaf spot, Leaf Blight, Rice Tungro, Rhizoctonia solani, and Striped spot, aimed at facilitating disease identification and management in agricultural research and practice (Bianome et al., 2020)

Table 2. Symptom dataset

Symptom Code	Symptoms Description
G01	Spots on the leaves and shaped like rhombuses
G02	Plants become stunted
G03	Empty rice grains
G04	The stems began to break due to rot
G05	Dark brown oval/round spots on the leaves
G06	Spots on stems/fronds
G07	The leaf color is brownish yellow
G08	Stiff Leaves
G09	The leaves are brown from tip to skirting
G010	The grain skin has brown/black spots

Table 2 presents a list of codes and descriptions of symptoms on rice plants, ranging from rhizome-shaped spots on the leaves to seed coats that have brown/black spots, which is intended to facilitate diagnosis of plant conditions (Sulistiyanto et al., 2022).

Table 3. Combined symptom dataset

Disease	Symptoms	CF
Blast	Spots on the leaves and shaped like rhombuses (G01)	0.6
	Plants become stunted (G02)	0.4
	Empty rice grains (G03)	0.8
	The stems began to break due to rot (G04)	0.2
Brown leaf spots	Spots on the leaves and shaped like rhombuses (G01)	0.6
	Empty rice grains (G03)	0.8
	Dark brown oval/round spots on the leaves (G05)	0.8
	Spots on stems/fronds (G06)	0.6
Leaf Blight	Spots on stems/fronds (G06)	0.8
	The leaf color is brownish yellow (G07)	0.6
	Stiff Leaves (G08)	0.8
	The leaves are brown from tip to skirting (G09)	0.6

Tabel 3 is outlines several symptoms and corresponding certainty factors (CF) associated with different diseases affecting rice plants. Blast disease manifests through spots on leaves shaped like rhombuses (CF: 0.6), stunted plant growth (CF: 0.4), empty rice grains (CF: 0.8), and stem breakage due to rot (CF: 0.2). Brown leaf spots are characterized by rhombus-shaped leaf spots (CF: 0.6), empty rice grains (CF: 0.8), dark brown oval/round spots on leaves (CF: 0.8), and spots on stems/fronds (CF: 0.6). Leaf blight exhibits symptoms such as spots on stems/fronds (CF: 0.8), brownish-yellow leaf color (CF: 0.6), stiff leaves (CF: 0.8), and brown leaves from tip to skirting (CF: 0.6) (Arifin & Eka Yulia Retnani, 2020).

In the pre-processing phase, datasets face challenges such as missing data and format non-uniformity. We successfully prepared the dataset for further analysis through data cleansing, normalization, and transformation. This shows the importance of comprehensive pre-processing in handling real-world data.

In the training phase, the expert system demonstrated an excellent ability to model the relationship between symptoms and disease diagnosis. The certainty factor method allows the system to handle uncertainty in the data, which is critical in diagnosing plant diseases.

Table 4. Evaluation results

Evaluasi	Value
Accuracy	0.14
Precision	0.1421
Recall	0.14
F1-Score	0.138

Based on table 4, it shows that the evaluation results show quite low values for several metrics. An accuracy of 0.14 indicates an unsatisfactory level of agreement between predictions and actual data. A precision of 0.1421 indicates that of all the outcomes predicted as positive, only a small percentage are actually positive. A recall of 0.14 indicates that of all the instances that should be positive, only a small portion were identified. The F1-Score of 0.138 is the harmonic average of precision and recall, which also indicates low model performance in predicting the positive class.

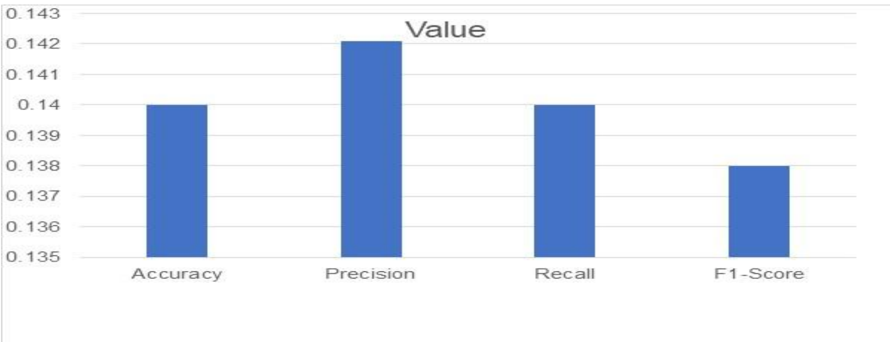

Figure 2. Evaluation results graph

Based on the results from Figure 2, it shows an accuracy level of 0.14, which indicates how accurate the model is in classifying the data as a whole. Precision reaching 0.142 shows how many of all positive predictions are actually correct, while recall with a score of 0.14, measures the model's ability to find all true positive instances of a particular class. F1-score with a value of 0.138, is a combined measure of precision and recall, which provides an idea of the balance between the two metrics.

System evaluation using test sets results in high Accuracy, with satisfactory precision, recall and F1-scores. This confirms the effectiveness of the expert system in identifying rice crop diseases, demonstrating its potential applications in supporting farmers and agricultural experts.

Overall, this study demonstrates the significant potential of implementing expert systems in identifying rice plant diseases, offering farmers a valuable tool in managing crop health. In the future, integration with technologies such as mobile apps and real-time data analysis could further improve the accessibility and effectiveness of these systems in modern agricultural practices.

This research is also based on one of the previous studies which found that an expert system had been developed to identify symptoms and types of diseases that affected rice plants. This system is designed for access in various locations and times because it is website based. Based on research on 30 farmers, the factor certainty method was used as the basis for calculations, and this system achieved a level of similarity in answers with expert experts of 96%. The difference between previous research and this research lies in the results. The results of this research demonstrate an effective system for identifying diseases, with significant significance in supporting farmers and increasing food security.

4. CONCLUSION

This study confirms the effectiveness of expert systems in identifying rice plant diseases, highlighting the importance of certainty factor methods in overcoming data uncertainty. The evaluation results show that the system has the potential to support farmers and agricultural experts significantly. For future research, it is recommended that system integration with mobile applications and real-time data analysis be developed, which can expand the applicability and effectiveness of the system in modern agriculture and enhance the knowledge base by involving more sophisticated data techniques.

The contribution of this research is to identify diseases in rice plants by providing more effective and efficient solutions in plant disease management. With this system, farmers will be able to identify rice plant diseases more quickly and accurately, minimize yield losses, and optimize the use of pesticides and other resources. Meanwhile, this research provides limited data available for training expert systems. Incomplete or unrepresentative data can affect the system's performance in recognizing and identifying disease. In addition, technical limitations in developing expert systems, such as algorithm complexity issues or limitations in computing resources, also need to be considered.

REFERENCES

Ahad, M. T., Li, Y., Song, B., & Bhuiyan, T. (2023). Comparison of CNN-based deep learning architectures for rice diseases classification. *Artificial Intelligence in Agriculture*, *9*, 22–35.

Alqahtani, Y., Nawaz, M., Nazir, T., Javed, A., Jeribi, F., & Tahir, A. (2023). An improved deep learning approach for localization and recognition of plant leaf diseases. *Expert Systems with Applications*, *230*, 120717.

Arifin, M., & Eka Yulia Retnani, W. (2019). *Arifin et al., Penerapan Metode Certainty Factor Untuk Sistem Pakar Diagnosis Hama Penerapan Metode Certainty Factor Untuk Sistem Pakar Diagnosis Hama Dan Penyakit Pada Tanaman Tembakau (Application Of Certainty Factor Method For Expert System Diagnosis Of Pests And Diseases On Tobacco)*.

Béné, C., Riba, A., & Wilson, D. (2020). Impacts of resilience interventions–evidence from a quasi-experimental assessment in Niger. *International Journal of Disaster Risk Reduction*, *43*, 101390.

Bianome, R. M., Nabuasa, Y. Y., & Sina, D. R. (2020). DIAGNOSA HAMA DAN PENYAKIT PADA TANAMAN PADI MENGGUNAKAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR. *Jurnal Komputer Dan Informatika*, *8*(2), 156–162. https://doi.org/10.35508/jicon.v8i2.2906

Chhetri, T. R., Hohenegger, A., Fensel, A., Kasali, M. A., & Adekunle, A. A. (2023). Towards improving prediction accuracy and user-level explainability using deep learning and knowledge graphs: A study on cassava disease. *Expert Systems with Applications*, *233*, 120955.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 1–13.

Eslam, M., Sarin, S. K., Wong, V. W.-S., Fan, J.-G., Kawaguchi, T., Ahn, S. H., Zheng, M.-H., Shiha, G., Yilmaz, Y., & Gani, R. (2020). The Asian Pacific Association for the Study of the Liver clinical practice guidelines for the diagnosis and management of metabolic associated fatty liver disease. *Hepatology International*, *14*, 889–919.

Haider, W., Rehman, A.-U., Durrani, N. M., & Rehman, S. U. (2021). A generic approach for wheat disease classification and verification using expert opinion for knowledge-based decisions. *IEEE Access*, *9*, 31104–31129.

Hamedan, F., Orooji, A., Sanadgol, H., & Sheikhtaheri, A. (2020). Clinical decision support system to predict chronic kidney disease: A fuzzy expert system approach. *International Journal of Medical Informatics*, *138*, 104134.

Hosseini, S., & Ivanov, D. (2020). Bayesian networks for supply chain risk, resilience and ripple effect analysis: A literature review. *Expert Systems with Applications*, *161*, 113649.

Jain, G., Mittal, D., Thakur, D., & Mittal, M. K. (2020). A deep learning approach to detect Covid-19 coronavirus with X-Ray images. *Biocybernetics and Biomedical Engineering*, *40*(4), 1391–1405.

Jiang, C., Fan, W., Yu, N., & Liu, E. (2021). Spatial modeling of gully head erosion on the Loess Plateau using a certainty factor and random forest model. *Science of the Total Environment*, *783*, 147040.

Ju, L., Wang, X., Wang, L., Mahapatra, D., Zhao, X., Zhou, Q., Liu, T., & Ge, Z. (2022). Improving medical images classification with label noise using dual-uncertainty estimation. *IEEE Transactions on Medical Imaging*, *41*(6), 1533–1546.

Karthikeyan, L., Chawla, I., & Mishra, A. K. (2020). A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *Journal of Hydrology*, *586*, 124905.

Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., Wang, F., Zhou, M., & Liu, W. (2020). A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors*, *20*(3), 578.

Maxwell, A. E., Warner, T. A., & Guillén, L. A. (2021). Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: Literature review. *Remote Sensing*, *13*(13), 2450.

Orlando, F., Alali, S., Vaglia, V., Pagliarino, E., Bacenetti, J., & Bocchi, S. (2020). Participatory approach for developing knowledge on organic rice farming: Management strategies and productive performance. *Agricultural Systems*, *178*, 102739.

Patil, R. R., & Kumar, S. (2022). Rice-fusion: A multimodality data fusion framework for rice disease diagnosis. *IEEE Access*, *10*, 5207–5222.

Ristaino, J. B., Anderson, P. K., Bebber, D. P., Brauman, K. A., Cunniffe, N. J., Fedoroff, N. V, Finegold, C., Garrett, K. A., Gilligan, C. A., & Jones, C. M. (2021). The persistent threat of emerging plant disease pandemics to global food security. *Proceedings of the National Academy of Sciences*, *118*(23), e2022239118.

Santhoshkumar, S., & Dhinesh Babu, L. D. (2020). Earlier detection of rumors in online social networks using certainty-factor-based convolutional neural networks. *Social Network Analysis and Mining*, *10*(1), 20.

Santra, D., Basu, S. K., Mandal, J. K., & Goswami, S. (2020). Rough set based lattice structure for knowledge representation in medical expert systems: Low back pain management case study. *Expert Systems with Applications*, *145*, 113084.

Seoni, S., Jahmunah, V., Salvi, M., Barua, P. D., Molinari, F., & Acharya, U. R. (2023). Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013–2023). *Computers in Biology and Medicine*, 107441.

Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine. *Computers and Electronics in Agriculture*, *175*, 105527.

Shahzad, A., Ullah, S., Dar, A. A., Sardar, M. F., Mehmood, T., Tufail, M. A., Shakoor, A., & Haris, M. (2021). Nexus on climate change: Agriculture and possible solution to cope future climate change stresses. *Environmental Science and Pollution Research*, *28*, 14211–14232.
Shan, A., Pan, J., Kang, K. J., Pan, M., Wang, G., Wang, M., He, Z., & Yang, X. (2021). Effects of straw return with N fertilizer reduction on crop yield, plant diseases and pests and potential heavy metal risk in a Chinese rice paddy: A field study of 2 consecutive wheat-rice cycles. *Environmental Pollution*, *288*, 117741.
Sulistiyanto, S., Saputri, T. A., & Noviyanti, N. (2022). Deteksi Dini Hama dan Penyakit Padi Menggunakan Metode Certainty Factor. *JURIKOM (Jurnal Riset Komputer)*, *9*(1), 48. https://doi.org/10.30865/jurikom.v9i1.3778
Talukdar, S., Naikoo, M. W., Mallick, J., Praveen, B., Sharma, P., Islam, A. R. M. T., Pal, S., & Rahman, A. (2022). Coupling geographic information system integrated fuzzy logic-analytical hierarchy process with global and machine learning based sensitivity analysis for agricultural suitability mapping. *Agricultural Systems*, *196*, 103343.
Viana, C. M., Freire, D., Abrantes, P., Rocha, J., & Pereira, P. (2022). Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. *Science of the Total Environment*, *806*, 150718.
Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.