



Machine Learning–Based Fertilizer Recommendation System using Soil and Climate Parameters

Dr Saravanan G¹, Amshavalli M², Lithan T³, Manoj Kumar M⁴, Rohith Kumar S M⁵

Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College Erode,
Tamil Nadu, India¹

Assistant Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College Erode,
Tamil Nadu, India²

Student, Department of Computer Science and Engineering, Erode Sengunthar Engineering College Erode,
Tamil Nadu, India³

Student, Department of Computer Science and Engineering, Erode Sengunthar Engineering College Erode,
Tamil Nadu, India⁴

Student, Department of Computer Science and Engineering, Erode Sengunthar Engineering College Erode,
Tamil Nadu, India⁵

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03.2026; Published: 28.03.2026.

ABSTRACT: This project creates a smart crop and fertilizer recommendations system that uses soil properties such as NPK values and pH, as well as climate factors such as rainfall and temperature, to support agricultural decision making. The system will use Boruta and Recursive Feature Elimination to identify important features of the environment, while also using Cat Boost-based classification and regression models to predict the best crop type, fertilizer type and amount needed. A system of expert knowledge is also embedded to detail efforts on nutrient guidance and application timing. The solution will be deployed via a Stream lit application, where farmers can upload soil and crop data in a batch format and get interactive data-driven recommendations. By considering environmental inputs, machine learning predictions, and expert knowledge, the system seeks to encourage sustainable farming practices, appropriate fertilizer use, and improved crop productivity.

KEYWORDS: Fertilizer Recommendation System, Machine Learning, Soil and Climate Analysis, Cat Boost

I. INTRODUCTION

Agriculture is critical to food security, and modern agriculture relies even more on data-driven solutions to maximize crop yield. Traditional fertilizer application approaches often lead to over- or under- application of nutrients, leading to less yield, soil degradation, and environmental negative impact. To overcome these issues, we need an intelligent recommendation system that incorporates multiple factors including nutrient composition of soil (NPK and pH), as well as rainfall and temperature. Machine learning models use domain-relevant knowledge to guide farmers with potential crops and the best fertilizers to use. This project will advance the development of data-driven, crop-specific advice for farmers, using new methods of feature selection, predictive modeling, and verified domain information about fertilizers. The solution will improve yield and the utility of resources, while also encouraging more sustainable agricultural practices.

1.1 FERTILIZER RECOMMENDATION SYSTEM

A Fertilizer Recommendation System is a decision- support system aimed at aiding farmers in determining the right type and quantity of fertilizers to apply to their crops, considering soil conditions, crop needs, and environmental conditions. The system analyzes specific parameters, including quantitative and qualitative soil nutrient levels (NPK),



pH, rainfall, and temperature, and ultimately provides recommendations based on these factors to improve nutrient application by using data. More modern systems incorporate machine learning models and predefined expert knowledge databases to make accurate fertilizer predictions that yield higher crop yields, increase cost savings, and reduce environmental impact while supporting sustainable farming practices. This approach is a more efficient method for reducing nutrient wastes, lessening environmental impact, and maintaining soil's health overall.

1.2 MACHINE LEARNING

Machine Learning (ML) is a subfield of artificial intelligence that allows computer systems to identify and learn patterns based on the collected or historical data without being explicitly programmed. In the agriculture domain, ML algorithms will review features found in both historical and real-time data, i.e., conditions of soils, weather, crop yields, and seasons, to find acceptable crops, fertilizer use, and farming practices. Examples of ML methods used include classification, regression, and feature extraction to capture an equally complex relationship between the features defined and the outcome variable. Using the ability to learn via the algorithms, the Fertilizer Recommendation System can provide data-driven recommendations to increase crop productivity while limiting resource wastes, as well as support sustainable agricultural practices in general.

1.3 SOIL AND CLIMATE ANALYSIS

In order to ascertain the appropriateness of crops and the proper fertilizer requirements, soil and climate analysis entails assessing important environmental and soil factors. Climate variables like rainfall, temperature, and humidity affect crop growth and production, while soil characteristics like pH, nitrogen, phosphorus, and potassium (NPK levels), and phosphorus and potassium (NPK) levels reveal information about nutrient availability and soil health. The system can forecast crop performance, detect nutrient deficits, and suggest exact fertilizer types and amounts by methodically examining these variables. A machine learning-based recommendation system that incorporates soil and climate analysis allows for data-driven decision-making, maximizing agricultural productivity and encouraging sustainable farming methods.

1.4 CAT BOOST

Cat Boost is a cutting-edge gradient boosting algorithm that works especially well with numerical and categorical data that does not require a lot of preparation. It is reliable and accurate for both classification and regression tasks since it iteratively minimizes mistakes by building an ensemble of decision trees to predict outcomes. The Fertilizer Recommendation System uses Cat Boost models to forecast the appropriate fertilizer type (classification) and the necessary fertilizer quantity (regression) based on crop, soil, and climatic data. Cat Boost is perfect for agricultural datasets because it can effectively handle categorical variables and avoid overfitting, guaranteeing dependable, high-performance predictions for sustainable crop and fertilizer management.

II. LITERATURE REVIEW

In this paper, Alexandros Oikonomidis [1] et.al. proposed the application of deep learning for the problem of crop yield prediction. However, there has been a lack of systematic analysis in previous studies. Hence, the objective of this study is to provide an overview of the state-of-the-art application of deep learning in crop yield prediction. To achieve this, a Systematic Literature Review (SLR) was conducted to identify and analyze the most relevant papers. Out of the 456 relevant studies retrieved, 44 primary studies were selected for further analysis based on selection and quality assessment criteria. The primary studies were thoroughly analyzed and synthesized with regards to key motivations, target crops, algorithms applied, features used, and data sources utilized. The findings revealed that the Convolutional Neural Network (CNN) algorithm is the most commonly used and performs the best in terms of Root Mean Square Error (RMSE).

It is crucial for [2] researchers to acknowledge the challenges in this field before constructing their own models. Meanwhile, practitioners face various obstacles in developing new crop yield prediction models, which are discussed in this SLR paper. These challenges include the selection of model parameters and algorithms, which necessitate a thorough analysis of existing literature. To conduct a comprehensive review, we examined 456 relevant studies. Notably, no systematic literature review has been conducted on the use of Deep Learning in crop yield prediction to date. Oilseed crops are commonly found in fragile agricultural systems that are dependent on weather conditions, particularly in semi-arid regions. Groundnut (peanut) crop, among the various oilseed crops, is highly susceptible to attacks by numerous pests and diseases, more so than many other crops.



Ramesh Babu Palepu [3] et.al. have proposed a system in which agriculture serves as the fundamental function to meet the global demand for food. Particularly in developing countries like India, agriculture is considered the backbone of the economy. The application of data mining techniques in agriculture, specifically in soil analysis, has the potential to revolutionize decision-making processes and enhance crop yields. The analysis of soil plays a crucial role in addressing various agricultural issues. This paper focuses on the role of data mining in soil analysis within the field of agriculture, as well as the work of several authors in this domain. Data mining techniques are highly relevant in the realm of soil analysis. Although data mining is widely used in various sectors and there are numerous off-the-shelf tools and techniques available, the application of data mining in agricultural soil datasets is still a relatively new research field. Nowadays, data mining concepts and techniques are being utilized to solve agricultural problems. The system proposed by V. Rajeswari [4] et.al. highlights the importance of soil in agriculture and aims to predict soil type using data mining classification techniques. The study utilizes JRip, J48, and Naive Bayes algorithms to extract knowledge from soil data and focuses on Red and Black soil types. The findings suggest that the JRip model produces more reliable results and increases the accuracy of soil type prediction. The study also emphasizes the potential of data mining in agriculture for soil classification, wasteland management, and crop and pest management. Other studies have explored various data mining techniques for predicting soil fertility rate, discovering knowledge in the agriculture sector, and estimating crop yield analysis. Efficient methods utilizing data mining can be developed to enhance the accuracy of classification of large soil data sets and address issues in Big Data.

Nabila Chergui [5] et.al. have proposed in their system that recent advancements in Information and Communication Technologies have had a significant impact on various sectors of the global economy. The emergence of Digital Agriculture is a direct result of the widespread availability of digital devices and advancements in artificial intelligence and data science. This new approach to agriculture has introduced innovative processes that enhance productivity and efficiency while also prioritizing environmental sustainability. The utilization of advanced digital devices and data science has enabled the collection and analysis of extensive agricultural datasets, empowering farmers, agronomists, and professionals to gain a deeper understanding of farming tasks and make more informed decisions. This paper presents a comprehensive review of the application of data mining techniques specifically in the context of digital agriculture. The focus of this study is on crop yield management and monitoring, as we delve into the main categories of data mining techniques employed in this domain. Furthermore, we explore a wide range of existing works that utilize data analytics in the field of agriculture.

III. EXISTING SYSTEM

Tamil Nadu's coastal location causes agricultural uncertainty, which lowers output. Although it is impossible to accomplish, greater output should be possible with a larger population and area. Farmers have employed word-of-mouth in previous decades, but the current climate makes it unfeasible. The data used to gain insights into Agri-facts is derived from agricultural elements and metrics. The development of the IT industry has led to various advancements in agriculture sciences that provide farmers with useful agricultural data. It is desirable in the current situation to be intelligent when using modern technical methods in the sector of agriculture. By using the data to create a clear model, machine learning techniques enable us to make predictions. We can handle agricultural problems like crop rotation, crop prediction, water and fertilizer requirements, and protection. Because of the environment's fluctuating climate, an effective method is required to make crop cultivation easier and to assist farmers with their management and production. This could contribute to the improvement of agriculture for future farmers. With the aid of data mining, a farmer can receive a system of recommendations to aid in crop production. Crops are suggested for implementation based on their quantity and climatic characteristics. Data analytics creates a path for the development of practical agricultural database extraction. After analyzing the Crop Dataset, crop recommendations are made based on season and productivity.

IV. PROPOSED SYSTEM

The suggested system is a machine learning-based framework that analyzes soil and environmental factors to suggest appropriate crops and the best way to use fertilizer. To determine the most important parameters, input characteristics including temperature, rainfall, pH, and soil nutrient levels (NPK) are preprocessed using feature selection methods like Boruta and Recursive Feature Elimination. Next, Cat Boost models are used, in which the regressor calculates the necessary fertilizer quantity and the classifier predicts the type of fertilizer that will be best. Furthermore, a knowledge base enhances forecasts by providing specific nutrient requirements and application schedules. Users can submit crop or soil datasets and receive interactive advice, such as crop compatibility and nutrient breakdowns, thanks to the system's Stream illuminated interface. This integrated strategy guarantees better agricultural output, sustainable fertilizer

management, and accurate decision-making.

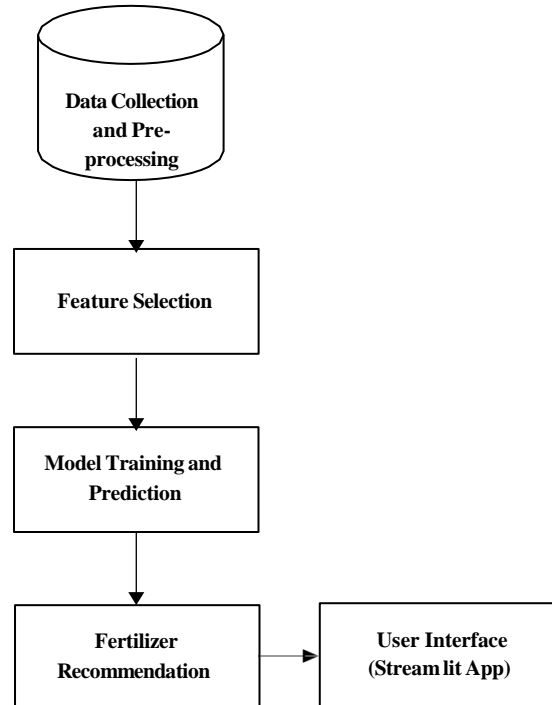


Figure.1. Block diagram

V. MODULE DESCRIPTION

DATA COLLECTION AND PREPROCESSING

This module is in charge of collecting data on crop productivity, season-specific crop information, rainfall, temperature, pH value, nitrogen, phosphorus, and potassium (NPK levels), and other soil characteristics. After being gathered, the raw data is cleaned by standardizing formats, addressing missing numbers, and eliminating discrepancies. Label encoding transforms categorical data, such as crop names, states, districts, and seasons, into numerical form, and scaling procedures normalize numerical features. Preprocessing guarantees that the dataset is accurate, consistent, and prepared for additional feature selection and modeling.

FEATURE SELECTION

The feature selection module finds the most important factors that have a big impact on fertilizer recommendations. To choose the most important characteristics and eliminate the unnecessary or less significant ones, sophisticated methods like Recursive Feature Elimination (RFE) and Boruta (based on Random Forest importance scores) are required. By lowering dimensionality and concentrating exclusively on significant attributes, this procedure enhances the machine learning models' accuracy and efficiency. The model prevents overfitting and improves generalization as a result.

MODEL TRAINING AND PREDICTION

This module uses machine learning algorithms to create predictive models for fertilizer recommendations. A Cat Boost Regressor is used to forecast the amount of fertilizer needed, and a Cat Boost Classifier is taught to identify the kind of fertilizer appropriate for the specified crop and soil conditions. Accuracy for classification tasks and R2 score for regression tasks are used to assess the models' performance after they have been trained and tested on processed datasets. For real-time predictions, the application then saves and re-uses the trained models.

FERTILIZER RECOMMENDATION

Here, the system combines domain knowledge from files with machine learning predictions. These files provide suggested fertilizers, application schedules, and crop-specific nutrient requirements (nitrogen, phosphorus, and potassium). The module creates thorough suggestions that include the type of fertilizer, the ideal dosage per hectare, and the best time to apply it by combining expert-verified knowledge with predictive outputs. This guarantees that the



advice given is grounded on statistics and will be beneficial to farmers in real-world situations.

USER INTERFACE (STREAM LIT APP)

Through a Stream-lit web application, the last module offers an intuitive platform. Bulk uploads of CSV files with crop, soil, and climate data are possible. The input is processed by the system, which also forecasts the need for fertilizer and interactively displays the findings. Each crop's recommended fertilizer kinds, amounts, and comprehensive nutrient tables are available to farmers and researchers. Even non-technical people may utilize the interface because to its scalability, ease of navigation, and support for tabular outputs. Algorithm details

Step 1: Preparing the Data

Gather climate variables (temperature, rainfall) and soil characteristics (pH, NPK levels).

Recognize numerical and category features. Cat Boost does not require a lot of preprocessing to handle categorical features.

Step 2: Selection of Features

To choose the most crucial environmental features for precise prediction, apply techniques like Boruta and Recursive Feature Elimination (RFE).

Step 3: Setup

For every data point, begin with a base prediction, which is often the class probabilities for classification or the average goal value for regression.

Step 4 : Ordered Boosting

Use ordered boosting to build trees in a sequential fashion. This minimizes overfitting by limiting the current tree's training to previous predictions rather than future ones.

Step 5: Building Symmetric Trees

Build symmetric (balanced) decision trees in which every level uses the same feature to divide every node. This guarantees stable forecasts and quicker training

Step 6: Leaf Value Calculation

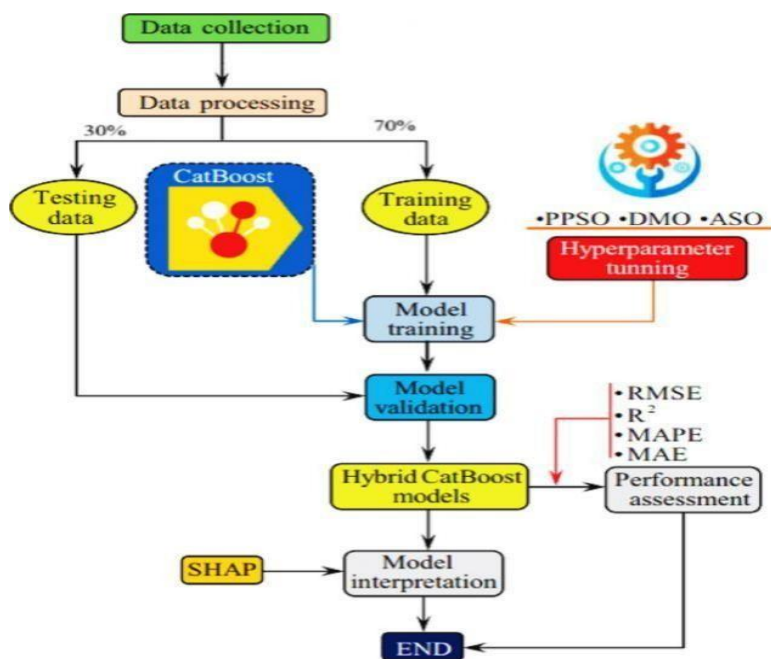
Calculate the leaf values for each tree that minimize the loss function (e.g., mean squared error for regression, cross-entropy for classification).

Step 7: Revising Forecasts

Add the output of the new tree to the ensemble of earlier trees to update the predictions.

Step 8 : Iterative Boosting

Up until the predetermined number of trees or convergence criteria is satisfied, keep constructing trees and updating predictions.



Workflow of Cat boost



VI. RESULT ANALYSIS

Test datasets with soil, climatic, and crop factors were used to evaluate the system and gauge how well the regression and classification models performed. While the Cat Boost Regressor exhibited great R² values in estimating the necessary fertilizer quantity, the Cat Boost Classifier showed high accuracy in determining the appropriate fertilizer type. By generating crop-specific fertilizer recommendations that were in line with domain knowledge from the database, batch predictions made through the Stream Light interface demonstrated that the system could manage numerous entries effectively. For every crop, comprehensive nutrition tables were created, offering precise information on the amounts of nitrogen, phosphorous, and potassium needed as well as when to apply them. All things considered, the outcomes demonstrate how well the system works to provide precise, useful, and data-driven suggestions for sustainable crop and fertilizer management.

Evaluation Metrics

Precision

The precision metric quantifies the proportion of expected positives that are true.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

Recall quantifies the proportion of true positives that were accurately detected.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Accuracy

Accuracy gauges how accurate the model is overall across all classes.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

F1-score

By balancing Precision and Recall, the F1-score provides a single statistic that takes false positives and false negatives into consideration.

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

algorithm	precision	recall	accuracy	F1 score
Cat boost	96.87	97.49	97.07	97.15
RF	86.76	87.88	89.28	87.32
NB	85.29	87.67	85.9	86.47

Table.1. Comparison table

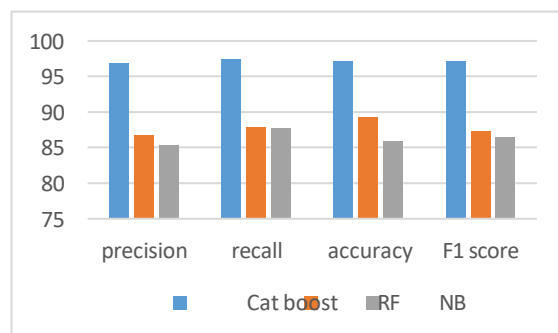


Figure.2. Comparison graph

Cat Boost, Random Forest (RF), and Naive Bayes (NB) are some of the machine learning algorithms that were used to assess the crop and fertilizer recommendation system's performance. With a precision of 96.87%, recall of 97.49%, accuracy of 97.07%, and an F1 score of 97.15%, Cat Boost outperformed the others, showcasing its exceptional capacity to precisely categorize and forecast crop and fertilizer recommendations. With an accuracy of 89.28% and an F1 score of 87.32%, Random Forest performed rather well, whereas Naive Bayes was marginally less effective with an accuracy of 85.9% and an F1 score of 86.47%. According to these findings, Cat Boost is the best model for this system and offers incredibly accurate forecasts for making sustainable farming decisions.

VII. CONCLUSION

The Crop Fertilizer Recommendation System effectively combines expert knowledge and machine learning algorithms to give farmers and researchers precise, data-driven advice. The algorithm forecasts appropriate fertilizer types and quantities as well as the best times to apply them by evaluating crop data, soil parameters (such as pH and NPK levels), and meteorological variables (such as temperature and rainfall). The Stream lit-based interface guarantees user-friendliness by facilitating interactive visualization of nutrient suggestions and batch processing of several records. All things considered, the approach increases agricultural output, encourages sustainable fertilizer management, minimizes resource waste, and facilitates well-informed decision-making, showcasing the potential of fusing data analytics with subject-matter knowledge in contemporary farming methods.

VIII. FUTURE WORK

Future improvements to the system could include adding more environmental and agronomic factors, like soil moisture, solar radiation, the prevalence of pests and diseases, and irrigation schedules, to offer more thorough crop and fertilizer recommendations. Farmers may be able to receive dynamic, location-specific recommendations through integration with satellite imaging and Internet of Things-based real-time soil sensors. Using bigger, more varied datasets can help machine learning models become even more accurate and more regionally generalized. The system can be made a comprehensive decision-support tool for sustainable agriculture by adding crop rotation plans, organic fertilizer substitutes, and predictive alarms for nutrient deficits or climate-related hazards.

REFERENCES

1. Shreya S. Bhanose and Kalyani A. Bogawar (2020) published a research paper titled "Crop And Yield Prediction Model" in the International Journal of Advance Scientific Research and Engineering Trends, Volume 1, Issue 1, April 2016.
2. Tripathy, A. K., et al. (2021) conducted a study on data mining and wireless sensor network for agriculture pest/disease predictions. Their findings were presented in the Information and Communication Technologies (WICT), 2011 World Congress on, published by IEEE.
3. C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
4. C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2



5. C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
6. S.Tamilselvi, R.Prakash, C.Nagarajan, "Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller" Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
7. S.Tamilselvi, R.Prakash, C.Nagarajan, "Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance" Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epr.2025.112428
8. S.Thirunavukkarasu, C. Nagarajan, 2024, "Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
9. C. Nagarajan, M.Madheswaran and D.Ramasubramanian- 'Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model'- Acta Electrotechnica et Informatica Journal , Vol.13 (2), pp.18-31, April-June.2013, DOI: 10.2478/aeei-2013-0025.
10. C.Nagarajan and M.Madheswaran - 'DSP Based Fuzzy Controller for Series Parallel Resonant converter'- Springer, Frontiers of Electrical and Electronic Engineering, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
11. C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.
12. [10] C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
13. Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: <https://doi.org/10.1590/1517-7076-RMAT-2024-0721>
14. M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
15. Ramesh Babu Palepu (2021) conducted an analysis of agricultural soils using data mining techniques. The research was published in the International Journal of Engineering Science and Computing, Volume 7, Issue No. 10, October.
16. Rajeswari and K. Arunesh (2020) conducted a study on analyzing soil data using data mining classification techniques. The research was published in the Indian Journal of Science and Technology, Volume 9, May.
17. A.Swarupa Rani (2020) published a research paper titled "The Impact of Data Analytics in Crop Management based on Weather Conditions" in the International Journal of Engineering Technology Science and Research, Volume 4, Issue 5, May.
18. Pritam Bose and Nikola K. Kasabov (2020), "Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series", IEEE Transactions On Geoscience And Remote Sensing.
19. Priyanka P. Chandak (2021), "Smart Farming System Using Data Mining", International Journal of Applied Engineering Research, Volume 12, Number 11.
20. Vikas Kumar, Vishal Dave (2021), "KrishiMantra: Agricultural Recommendation System", Proceedings of the 3rd ACM Symposium on Computing for Development, January.
21. Savae Latu (2021), "Sustainable Development: The Role of Geographic Information Systems and Visualization," The Electronic Journal on Information Systems in Developing Countries, EJISDC 38, 5, 1-17.
22. Nasrin Fathima.G (2020), "Agriculture Crop Pattern Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, May.
- 23.