A Comprehensive Crop Recommendation System Integrating Machine Learning and Deep Learning Models

1st Asit Ghosh
Department of Computer Science & Engineering Centurion University of Technology and Management,
Bhubaneswar, India
asitg2003@gmail.com

2nd Sunil Kumar Mohapatra
Department of Computer Science & Engineering Centurion University of Technology and Management,
Bhubaneswar, India
sunil.mohapatra@cutm.ac.in

3rd Pritam Pattanaik
Department of Computer Science & Engineering Centurion University of Technology and Management,
Bhubaneswar, India
pritampattanaik378@gmail.com

4th Prasant Kumar Dash
Department of Computer Science & Engineering
CV Raman Global University,
Bhubaneswar, India
prasantdash@cgu-odisha.ac.in

5th Sujata Chakravarty
Department of Computer Science & Engineering Centurion University of Technology and Management,
Bhubaneswar, India
chakravartys69@gmail.com

2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU) | 979-8-3503-6178-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/IC-CGU58078.2024.10530724

***Abstract*— A crucial development in recent agriculture is the emergence of crop recommendation systems, which provide a unique method for optimizing crop yields and resource use. In this work, we have proposed an advanced "Crop Recommendation System" to help farmers choose the right crop according to different soil and environmental parameters. Our system is trying to improve crop yields and agricultural productivity using machine learning (ML) and deep learning (DL) models. The data used for training and evaluation was collected from the Kaggle dataset repository. It includes the eight most important features such as nitrogen (N), phosphorus (P), potassium (K), rainfall, pH, temperature, humidity, and crop labels representing 22 different crops. This work employs machine learning and deep learning models such as Decision Trees, Random Forests, XGBoost, Support Vector Machine, K-nearest neighbors, Naive Bayes, Artificial Neural Networks, Deep Neural Networks, and Temporal Convolutional Networks. Based on performance evaluation, Random Forest and TCN are the best models, showing 99.2% and 99.9% accuracy, respectively. The performance of other models was also excellent, with accuracy levels ranging from 93.8% to 98.7%. SVM was a little behind, but still performed well with 93.4% accuracy. The project also includes parameter optimization for XGBoost and TCN. The TCN model gives a better result than XGBoost using parameter optimization. The outcome of this research shows that machine learning and deep learning models perform well in crop recommendation systems and the TCN model demonstrates its commitment to accurate and efficient crop recommendations. This research work supports precision agriculture by providing a web interface to the farmers with reliable parameters to improve crop selection based on environment and soil conditions.**

Keywords-Machine Learning, Deep Learning, Parameter Optimization, Recommendation System, Precision Agriculture.

I. INTRODUCTION

The need for precise and efficient crop management practices has become more crucial because of rapid climate change and the increasing demand for agricultural sustainability. This study presents a unique "Crop Recommendation System" to help farmers with the ability to decide the best crop for their field. By using the capabilities of both DL and ML models, this system aims to improve the performance of traditional crop management systems. This work uses a data-driven methodology and applies a carefully curated dataset for crop recommendations. The dataset contains important features like rainfall, pH, temperature, humidity, N, P, K, and crop labels that represent a different variety of 22 crops. The importance of applying ML models including Decision Tree, Random Forest, XGBoost, SVM, KNN, and Naive Bayes as well as DL models including ANN, DNN, and TCN to furnish farmers with better clarity and understanding of agricultural ecosystems [1]. The result shows that Random Forest and TCN achieve the best with an accuracy of 99.2% and 99.9%, respectively. To present the most efficient models, this work investigates the capability of Decision Trees, XGBoost, KNN, Naive Bayes, ANN, and DNN, which show impressive accuracy. SVM comes in last place with an accuracy of 93.4%. This research includes a set of procedures for parameter optimization for XGBoost and TCN. Notably, TCN maintains its power even after development, showing its power and efficiency in processing complex patterns of agricultural data. Additionally, having an easy-to-use website for farmers is also beneficial for this research work. The website allows farmers to simply choose the best crop for their field by entering soil and environmental parameter details [2]. The real-time web interface reduces the generated data into useful advice for farmers. Fig 1 represents the overall workflow of this work.

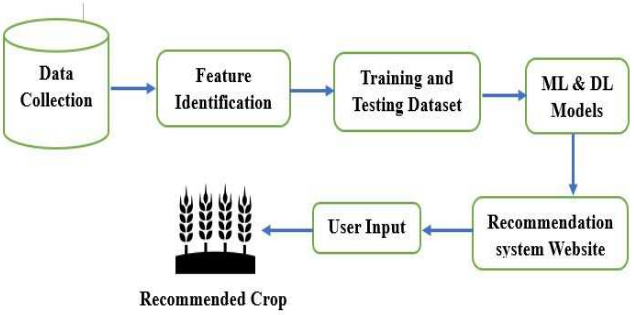


Fig 1: Workflow for an Integrated Crop Recommendation System

Starting with data collection, we identify critical features before creating training and testing datasets. For accurate predictions, ML and DL models are used. The website of the recommendation system facilitates user input by processing data to suggest the best crop for specific conditions. The workflow of the model starts from the data acquisition phase to the crop recommendations phase.

II. RELATED WORK

This section discusses a detailed review of the existing research work performed on crop recommendation systems. Significant contributions from several research articles are noted. Crop prediction techniques like K-Nearest Neighbour and Naive Bayes were first presented by Medar et al. [3]. Using machine learning techniques like Random Forest and XGBoost, Dhruvi Gosai et al. examined data collection with characteristics like pH and humidity where XGBoost outperformed among all [4]. To increase accuracy, Bandara et al. successfully implemented Nave Bayes and SVM in a predetermined region, taking into account farmer input[5]. Different machine learning algorithms such as Naive Bayes, XGboost, and Random Forest are studied by Varun Prakash R et al. for crop prediction[6]. Real-time data was collected to achieve high precision. The authors present a crop recommendation system based on a team approach and a majority decision technique in precision agriculture. The model uses Random Tree, CHAID, K-Nearest Neighbor, and Naive Bayes, yielding an 88% prediction accuracy [7]. The earlier study includes relevant research, such as the use of classification algorithms for yield prediction, the flexible Crop Yield Prediction Framework (XCYPF), and information mining techniques for crop yield estimation in various locations [8]. The research contributes by offering a recommendation system suitable for Indian agriculture, intending to guide farmers in optimal crop selection based on site-specific data. The importance of precision agriculture in India is addressed in [9], along with the common issue that farmers have while choosing the right crops for their soil types. According to [10] the authors review recent research on smart recommender systems for agricultural applications. According to research [11][12], IoT improves crop recommendation by providing real-time environmental data and optimizing agricultural decisions.

A Recommendation system that makes use of a soil database, crop knowledge offered by experts, and criteria like soil characteristics gathered from laboratory tests is suggested as a solution for precision agriculture. It employs SVM and ANN as an ensemble learning model for precise and effective crop recommendations. The idea of using machine learning techniques for crop prediction highlights the significance of increasing agricultural output in India. Precision farming, recommendation systems, ensemble models, SVM, ANN, and Random Forest are the primary techniques to implement this system. This integrates IoT and ML technologies for soil testing, addressing the crucial problem of crop recommendation in Indian agriculture. A complete approach is demonstrated by the implementation of machine learning techniques such as Decision Tree, Naïve Bayes, Support Vector Machine, Logistic Regression, Random Forest, and XGBoost, and the emphasis on soil characteristics including temperature, moisture, pH, and NPK nutrients. Majority Voting, an ensemble technique, is used to produce precise crop predictions based on input from several learners. The significance of the suggested method in achieving high accuracy, particularly with algorithms like Naïve Bayes, Random Forest, and XGBoost, is highlighted by a comparative study with comparable research. The system's capacity to raise agricultural yield and lower climate change-related hazards is consistent with the more general objective of precision.

III. METHODOLOGY

The architecture in Fig 2, represents the methodology to design the model, which consists of the most six important phases starting from data collection to model representations. Each of the phases of the proposed model is discussed in the listed subsections.

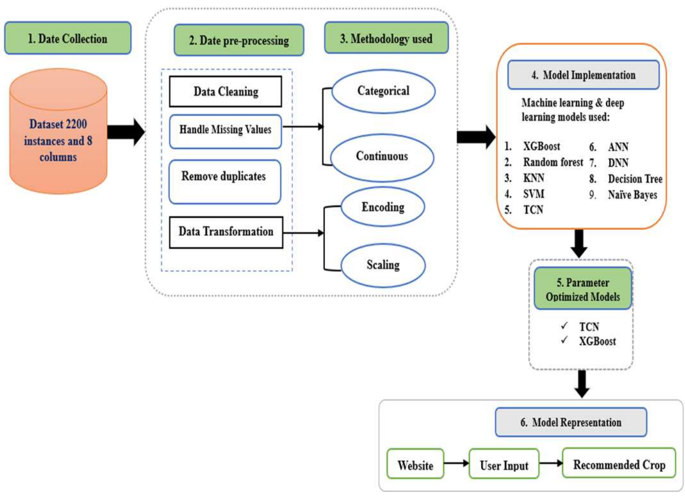


Fig 2: Proposed Architecture for Crop Recommendation System

A. Dataset Description

This project's dataset is the foundation for effective model training and evaluation. It was obtained from Kaggle [13]. It includes essential factors such as N, P, K, rainfall, pH, temperature, humidity, and crop labels, as well as information on 22 different crops. This data acts as the foundation for the models' accurate prediction and recommendation capabilities. TABLE 1 describes the general overview of the dataset.

TABLE 1: OVERVIEW OF DATASET

Dataset Name	Characteristics	Classes	Instances	Crops
Crop Recommendation dataset	Multivariate	8	2200	22

B. Data Preprocessing

Throughout the stage of data preprocessing, we took great care to maintain the integrity of the dataset. To avoid redundancy, redundant features were eliminated carefully. Outliers were addressed and handled properly. Moreover, categorical variables were transformed into numerical patterns using data transformation methods like label encoding to enable smooth model integration. To measure and evaluate the numeric characters, scaling was deployed. These steps strengthened the dataset's robustness, allowing the models to extract meaningful patterns and provide farmers with precise crop recommendations [14].

C. Model Selection

A variety of ML and DL techniques, including random forest, KNN, decision tree, DNN, KNN, Naive Bayes, ANN, XGBoost, SVM, and TCN were employed in the proposed method..

1. Decision Tree

A decision tree is a tree-like model that produces decisions by continuously dividing the dataset according to its attributes. Decision trees are useful in this paper because they simplify the complex relationships between soil and environmental variables, making it simpler to determine the ideal growing conditions for various crops.

2. Random Forest

Random Forest is a group of Decision Trees that improves accuracy by combining predictions from different trees. It does a great job of preventing overfitting and making accurate predictions about what crops to grow based on a wide range of environmental factors. The mathematical equation is given in equation (1).

$$\hat{Y} = \frac{1}{M} \Sigma_{i=1}^{m} f_i(X) \quad (1)$$

3. XGBoost

An optimized gradient boosting algorithm called XGBoost is great at dealing with data that doesn't have linear relationships. Its "boosted tree" method makes predictions more accurate, which makes it very useful for fine-tuning crop suggestions based on the complex dataset [15]. The mathematical equation is given in equation (2).

$$\hat{y}_j = \Sigma_{n-1}^{n} f_n\left(x_j\right) \quad (2)$$

4. KNN

KNN sorts data points into groups based on how close they are to other points in the feature space. As part of this project, KNN helps suggest crops by finding similar areas with similar weather conditions. The mathematical equation is given in equation (3).

$$\hat{y} = \arg\max_{j\in\{1\dots,p\}} \Sigma_{k=1}^{k} w_{ij}\,[y_i = j] \quad (3)$$

5. Naïve Bayes

Naive Bayes is a probabilistic model that figures out how likely something is to happen given certain conditions. It helps with this project by giving a probabilistic framework for suggesting crops based on the parameters that are given. The mathematical equation is given in equation (4).

$$P(A|\,B) = \frac{P(B|A)P(A)}{P(B)} \quad (4)$$

6. SVM

There is a powerful classifier called SVM that can work with both linear and nonlinear data. In this project, SVM helps find the best decision boundaries in the feature space so that accurate crop suggestions can be made based on many different factors. The mathematical equation is given in equation (5).

$$\hat{y}(x) = \text{sign}(\Sigma_{i=1}^{n} \alpha_i y_i\, K(x_i, x) + b) \quad (5)$$

7. ANN

ANNs copy the neural structure of the human brain and are very good at recognizing complex patterns. ANN helps with this project by learning complicated connections that give us a more complete picture of how environmental factors affect crop choices [16]. The mathematical equation is given in equation (6).

$$y = f(k_1a_1 + k_2a_2 + \dots + k_na_n + b) \quad (6)$$

8. DNN

DNN enhances feature learning as well as conceptualization, which enables the model to better understand the complicated interactions within the data and, thus, provide precise crop recommendations. Equation (7) provides the mathematical equation of DNN.

$$\hat{y}=f_L\,(W_Lf_{L-1}(W_{L-1}\dots f_2(W_2f_1(W_1X+b_1)+b_2)+\dots)+b_L) \quad (7)$$

9. TCN

When data is organized in a specific order, TCN is an effective tool for identifying temporal relationships. In this study, TCN assists in identifying patterns in time-varying environmental parameters, providing us with valuable information for crop recommendation systems. Equation (8) provides the mathematical equation of TCN.

$$y_{i+1} = \sigma(k * y_i + b) \quad (8)$$

D. Training and Validation

An 80/20 split of the dataset was made, with 80% going to train the model and 20% going towards testing. During training, models adjusted their parameters to identify complex patterns in most of the data. The 20% set aside for testing acted as a clean sheet of paper on which we could impartially evaluate the models' ability to generalize to cases that had never been encountered before. By thoroughly verifying the models on real-world conditions that are not presented during training, our technique assures the dependability of our crop recommendation system and develops reliability in their practical value for farmers.

E. Performance Evaluation Metrics

Each crop recommendation system model's accuracy, precision, recall, and F1 score were extensively examined. Precision confirms that suggested crops satisfy the parameters and that the forecast model is correct. Recall evaluates how well the model can recognize every significant crop without making any errors. Precision and recall are balanced to provide a comprehensive performance metric in the F1 score. Accuracy is essentially a measure of prediction accuracy. By outlining each model's advantages and disadvantages, this thorough analysis was done by selecting the most reliable algorithm for precision agriculture.

F. Recommendation System

We developed a farmer-friendly website to make our sophisticated crop recommendation algorithms more accessible. This user-friendly platform lets farmers enter parameters. Following submission, the system evaluates the input data using trained models and immediately suggests crops based on the geographic regions of the farmers. Farmers may improve crop choices and increase agricultural production and sustainability by using this straightforward interface, which provides practical insights [17].

IV. RESULT AND DISCUSSION

Here, we present an in-depth analysis of the density distribution of all essential features for model development. This graphical representation sheds light on the spread and concentration of each variable, providing valuable insights into the dataset's characteristics.

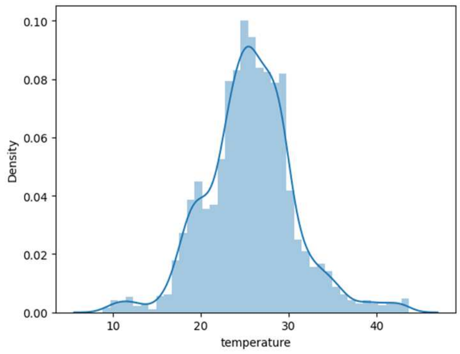


Fig 3: Temperature Density Distribution Plot

Fig 3 represents the density distribution plot for temperature, showing its distribution and concentration in the dataset.

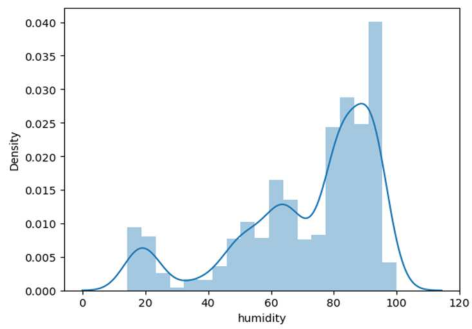


Fig 4: Humidity Density Distribution Plot

Fig 4 shows the actual Humidity Density Distribution Plot, providing a visual representation of the spread and sensitivity of Density percentage in the dataset.

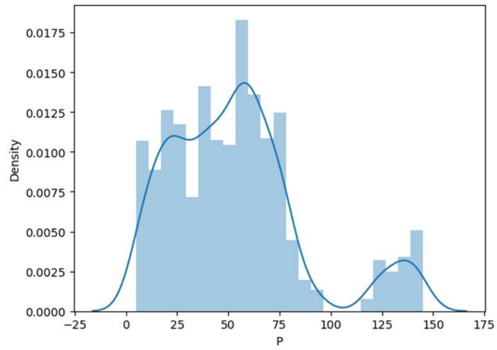


Fig 5: phosphorus Density Distribution Plot

As shown in Fig. 5, the density distribution plot for soil phosphorus shows its distribution and concentration.

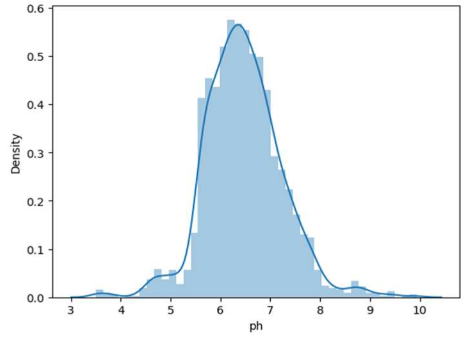


Fig 6: pH Density Distribution Plot

The density distribution plot of pH level in the soil is shown in Fig 6. This plot provides valuable insights into the patterns of level as well as the distribution of this important variable in the dataset.

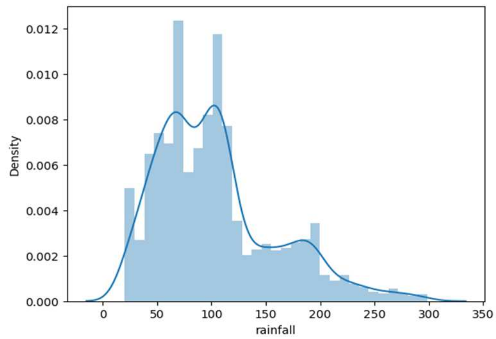


Fig 7: Rainfall Density Distribution Plot

The rainfall volume distribution plot in Fig 7 represents data distribution and concentration. This is crucial for understanding rainfall changes over time.

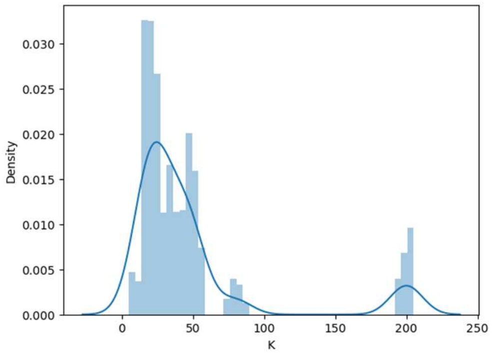


Fig 8: Potassium Density Distribution Plot

The Density Distribution plot of potassium in Figure 8 shows how key soil properties vary and agree across the dataset.

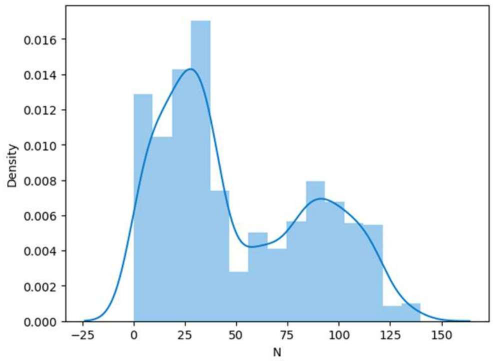


Fig 9: Nitrogen Density Distribution Plot

In Fig. 9, the density distribution plot for soil nitrogen content shows the concentration and spread across the dataset.

We present the findings of our comprehensive evaluation of the crop recommendation system. Each model's performance metrics, which include precision, recall, F1 score, and accuracy, are detailed. The comparison reveals the strengths and weaknesses of various algorithms, guiding the selection of the most effective model for practical implementation in precision agriculture [18]. Fig 10 shows a comparative analysis of model accuracies, revealing differences in performance across different ML and DL models. TCN performs better with 99.60 % accuracy as compared to other models.

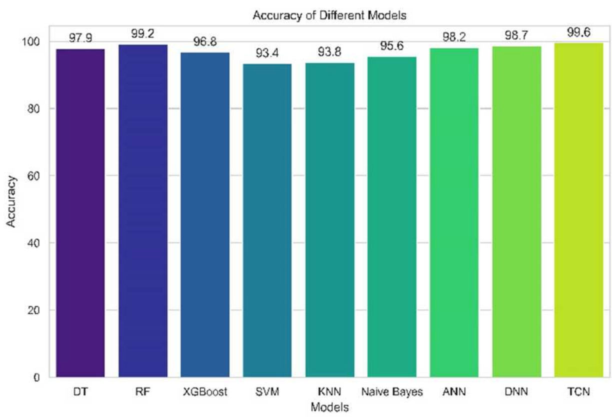


Fig 10: Model Accuracy Comparison Plot

TABLE 2 summarises precision, recall, and F1 score metrics to assess various models' crop recommendation accuracy strengths.

TABLE 2: ANALYSIS OF ML AND DL MODEL EVALUATION MATRICES

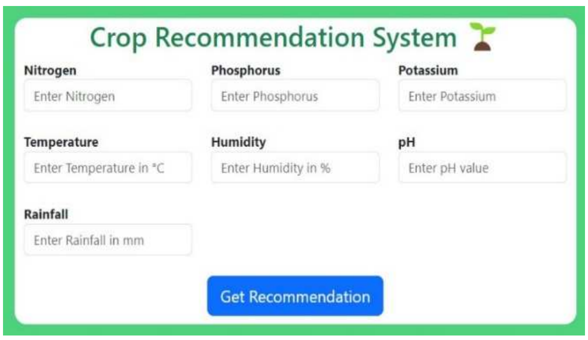
ML/DL Model	Precision	Recall	F1 score
DT	0.94	0.65	0.74
RF	0.85	0.80	0.87
XGBoost	0.84	0.82	0.87
SVM	0.58	0.54	0.56
KNN	0.60	0.46	0.52
NAÏVE BAYES	0.59	0.85	0.43
ANN	0.89	0.89	0.89
DNN	0.72	0.48	0.57
TCN	**0.99**	**0.98**	**0.96**

Parameter optimization enhanced XGBoost's accuracy, highlighting its adaptability. However, TCN showed better accuracy even before optimization because it was built for sequential data. After adaptation, TCN's advantage persisted, highlighting its capacity to identify temporal dependencies. The importance of parameter optimization in directing the preference for TCN in providing detailed and accurate crop recommendations in dynamic agricultural settings is emphasized in this work [19]. The TCN and XGBoost model's performance with parameter optimization is shown in TABLE 3.

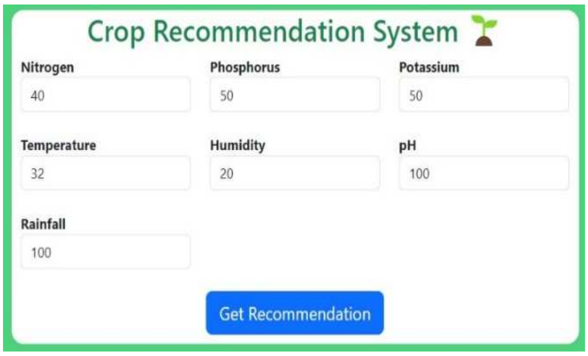
TABLE 3: PERFORMANCE ANALYSIS OF TCN AND XGBOOST MODEL

ML/DL Model	Accuracy	precision	recall	F1 score
TCN	**99.92**	0.99	0.98	0.98
XGBoost	98.92	0.85	0.80	0.87

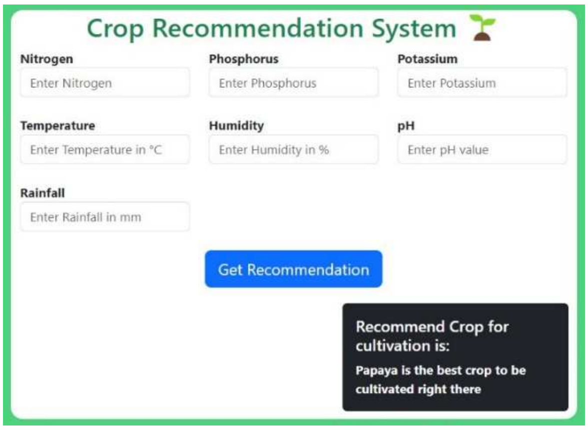
Furthermore, we examine the outcomes of user interaction with the created website, highlighting the system's usefulness and influence on farmers' decision-making procedures. The discussion interprets these findings in the context of the agricultural landscape, providing insights into the system's real-world applicability, potential improvements, and transformative role in modernizing farming practices [20]. Fig 11 (a), (b), and (c) represent the user interface of the crop recommendation system.



(a)



(b)



(c)

Fig 11 (a)(b)(c): Crop Recommendation System for User Interface

CONCLUSION

The Crop Recommendation System, which used a combination of ML and DL models, produced promising results. The comparative analysis demonstrated the accuracy with which models such as TCN predicted optimal crops. The user-friendly website offers farmers a practical interface for receiving personalized recommendations based on input parameters. Precision agriculture advancements have the potential to improve crop yields and sustainability. The project lays the groundwork for future technological integration into agricultural practices, contributing to the ongoing evolution of modern farming methodologies. Large datasets with more features can be used in the future, and ensemble learning models with hyperparameter tuning can be implemented for more accurate predictions.

REFERENCES

[1] Parameswari, Ponnusamy, N. Rajathi, and K. J. Harshanaa. "Machine learning approaches for crop recommendation." In *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, pp. 1-5. IEEE, 2021.

[2] Garanayak, Mamata, Goutam Sahu, Sachi Nandan Mohanty, and Alok Kumar Jagadev. "Agricultural recommendation system for crops using different machine learning regression methods." *International Journal of Agricultural and Environmental Information Systems (IJAEIS)* 12, no. 1 (2021): 1-20.

[3] Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta. "Crop yield prediction using machine learning techniques." In *2019 IEEE 5th international conference for convergence in technology (I2CT)*, pp. 1-5. IEEE, 2019.

[4] Gosai, Dhruvi, Chintal Raval, Rikin Nayak, Hardik Jayswal, and Axat Patel. "Crop recommendation system using machine learning." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 7, no. 3 (2021): 558-569.

[5] Bandara, Pradeepa S. "Inclination of teachers towards incorporating mobile game based learning into primary education: A Sri Lankan case study." *International Journal of Information Technology and Computer Science (IJITCS)* 10, no. 4 (2018): 66-72.

[6] Prakash, R. Varun, M. Mohamed Abrith, and S. Pandiyarajan. "Machine learning based crop suggestion system." In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1355-1359. IEEE, 2022.

[7] Reddy, Polu Shrikhar, Bollimuntha Amarnath, and M. Sankari. "Study on Machine Learning and Back Propagation for Crop Recommendation System." In 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), pp. 1533-1537. IEEE, 2023.

[8] Manjula, Aakunuri, and G. Narsimha. "XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction." In 2015 IEEE 9th international conference on intelligent systems and control (ISCO), pp. 1-5. IEEE, 2015.

[9] Jain, Sonal, and Dharavath Ramesh. "Machine Learning convergence for weather based crop selection." In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6. IEEE, 2020.

[10] 8. Song, Caixia, and Haoyu Dong. "Application of Intelligent Recommendation for Agricultural Information: A Systematic Literature Review." IEEE Access 9 (2021): 153616-153632.

[11] Mishra, Sushruta, Hrudaya Kumar Tripathy, Brojo Kishore Mishra, and Sunil Kumar Mohapatra. "A succinct analysis of applications and services provided by IoT." In Big Data Management and the Internet of Things for Improved Health Systems, pp. 142-162. IGI Global, 2018.

[12] Mishra, Sushruta, Sunil Kumar Mohapatra, Brojo Kishore Mishra, and Soumya Sahoo. "Analysis of mobile cloud computing: Architecture, applications, challenges, and future perspectives." In Applications of security, mobile, analytic, and cloud (SMAC) technologies for effective information processing and management, pp. 81-104. IGI Global, 2018.

[13] A.Ingle, "Crop Recommendation Dataset", kaggle.com. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset/data (accessed Jan. 19, 2024)

[14] Nischitha, K., Dhanush Vishwakarma, Mahendra N. Ashwini, and M. R. Manjuraju. "Crop prediction using machine learning approaches." *International Journal of Engineering Research & Technology (IJERT)* 9, no. 08 (2020): 23-26.

[15] Ghadge, Rushika, Juilee Kulkarni, Pooja More, Sachee Nene, and R. L. Priya. "Prediction of crop yield using machine learning." *Int. Res. J. Eng. Technol.(IRJET)* 5 (2018): 2237-2239.

[16] Sahoo, Soumya, Sunil Kumar Mohapatra, and Bijayalaxmi Panda. "Classification using extreme learning machine." Compusoft, An International Journal of Advanced Computer Technology 2, no. 12 (2013): 415-421.

[17] Jayaraman, Vaishnavi, Saravanan Parthasarathy, Arun Raj Lakshminarayanan, and S. Sridevi. "Crop Recommendation by Analysing the Soil Nutrients Using Machine Learning Techniques: A Study." In *Computational Intelligence in Data Science: 4th IFIP TC 12 International Conference, ICCIDS 2021, Chennai, India, March 18–20, 2021, Revised Selected Papers 4*, pp. 15-26. Springer International Publishing, 2021.

[18] Chauhan, Gaurav, and Alka Chaudhary. "Crop recommendation system using machine learning algorithms." In *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 109-112. IEEE, 2021.

[19] Gopi, P. S. S., and M. Karthikeyan. "Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model." *Intell. Autom. Soft Comput* 36 (2023): 313-326.

[20] Gupta, A., Nagda, D., Nikhare, P., & Sandbhor, A. (2021). Smart crop prediction using IoT and machine learning. *International Journal of Engineering Research & Technology (IJERT)*, *9*(3).