RESILIENCE FARMING USING MACHINE LEARNING FOR PREDICITIVE CROP RECOMMENDATIONS

Sukhvinder Singh Bamber

Computer Science & Engineering, University Institute of Engineering & Technology, Panjab University SSG Regional Centre, Hoshiarpur, Punjab, India.

***Abstract**: Food security and sustainable agricultural development face challenges such as climate-induced environmental stresses, crop yield predictions, and feeding a growing global population. This research proposes an ML-based crop recommendation system to help farmers make the right selection of crops using data-driven insights. The system focuses on the enhancement of productivity, reduction of resource wastage, and resilience improvement against environmental changes. The system uses ML algorithms like Random Forest, Naive Bayes, and XGBoost. It analyzes the characteristics of soil, climate conditions, and nutrient availability. Random Forest and XGBoost are good for handling complex and high-dimensional data to unravel complicated patterns and relationships of parameters. Naive Bayes offers a very simple yet effective baseline model for comparison. These models are evaluated against each other to decide the most accurate and robust one to predict crop. The system provides recommendations for farmers without test reports on soil data and general recommendations based on general types of soil, while detailed soil data leads the system to provide specific crop selection for farmers related to the levels of nutrients like nitrogen (N), phosphorus (P), and potassium (K). Therefore, this two-fold approach makes the system flexible and accessible to varying users from small-scale farmers to large agricultural operations. It combines leading-edge ML techniques in delivering accurate, scalable solutions to modern challenges in agriculture, thereby contributing toward sustainable practices under dynamic environmental conditions.*

***Keywords*:** Machine Learning, Sustainable Agriculture, Crop Recommendation, Crop Prediction, XGBoost, Random Forest, Naïve Bayes.

1. INTRODUCTION

For the past few years, the agriculture sector has faced various shocks, including adverse impacts of climate change, increasingly growing demands of the population, and the use of sustainable farming practices. These issues have led to a less predictable nature of crop yields while posing a blow to food security. Additionally, the global population is growing alarmingly, which also translates to an increasingly higher pressure on food production. The United Nations projections put the world population at approximately 9.7 billion by 2050 [41], so there is a tremendous call for improvement in food supply, too. In looking after all of this, clever and intelligent solutions must be found for agricultural processes. Application solutions based on AI, ML, and IoT are supported by researchers and technologists for sustainable crop production as well as resource management in agriculture. One of the likely approaches is the implementation of an intelligent crop recommendation system to avail proper recommendations of crop choice to farmers concerning the environmental conditions and soil. The systems use advanced machine learning algorithms, big data analytics, and IoT sensors to provide recommendations that can help optimize resources and maximize yields. Crop recommendation systems examine different input parameters like soil properties, climate conditions, market demand, and crop price to identify the most appropriate crops for a region. This paper takes on a new research work in the direction of the development of an automated crop recommendation system that provides farmers with accurate crop suggestions, helping them decide on optimal agricultural productivity. Climate change has affected the predictability of growth patterns, making it challenging for farmers to decide on which crops they have to cultivate. Traditional practices are not workable anymore and become inefficient with extreme weather events and uncertainty in environmental conditions. Crop recommendation systems, using data-driven insights, guide the population in decision-making in crop selection. This is a significant tool in modern agriculture because these systems draw from many factors that include soil health, climate adaptation, resource optimization, yield maximization, economic viability, and sustainability to give precise and practical crop recommendations. This research will draw from other existing works in the field of crop recommendation systems. Previous research proved that models such as regression models, XGBoost, and Random Forest can be applied to achieve high-accuracy predictions in crops. This research is beneficial to have a comparison on the effectiveness of different approaches in crop recommendation. Besides, IoT sensors with big data analytics can monitor in real time and analyze the environmental conditions to boost accuracy in crop recommendations. This way, the system ensures adaptability to any changing conditions, as it always produces dynamic recommendations concerning what climate or soil can bring.

The proposed system addresses some of the most critical aspects of modern agriculture:

Soils have a quality that is one of the pre-basic constituents in determining crop suitability. Our system analyzes different properties regarding soil type, such as pH levels, nutrient content, and texture to suggest crops suitable for the specific conditions of a given field. Based on its understanding of the unique characteristics of the soil, the system can then recommend crops more likely to thrive and thus reduce the risk of low yields and promote soil health. Climate Adaptation: The system shall incorporate climate information, such as temperature, rainfall, and moisture levels, indicating crops likely to survive in the local climate. This enables the farmer to be given crops that will be able to survive in certain conditions allowing him to adapt to challenges posed by climatic changes. Resource Optimization ensures that resource use supports sustainable agriculture. Its recommendation for crops best matched to local conditions limits the use of unnecessary inputs like fertilizers, pesticides, and water on that piece of land. This will reduce farmers' costs, decrease pollution associated with agriculture, and thus support sustainable agriculture. Yield Maximization: It would guide the crops with maximum potential for maximum yields based on past crop performance histories and current environmental conditions. This allows farmers to produce at maximum levels, increasing profitability to achieve bigger food security. Economic Feasibility: The system, in addition to environmental feasibility, would ensure that crop price and market demand would lean towards selected crops to be economically viable for farmers. This consideration helps the farmer make decisions that are lined up with their economic objectives and environmental sustainability. Sustainability: The system encourages sustainable agriculture by promoting the diversification of crops and the conservancy of resources. With time, such practices lead to soil health and resilience, providing a basis for long-term agricultural productivity.

The research paper discusses the development of a crop recommendation system via choice and implementation of appropriate machine learning algorithms, data processing techniques, and metrics used for performance evaluation. Thus, the system is ready to handle large datasets that analyze extensive information on soil and climate conditions, historical crop yields, and market trends. We are ready to prove its relevance as a reliable tool in modern agriculture by evaluating the accuracy and effectiveness of our system to existing approaches. Still, implementing such a system has some challenges. There is another problem with data privacy because soil and climate conditions contain data that should not be exposed. There is also an issue of scalability because the system has to accommodate data to be read from a large number of geographic sites with various environmental settings. Lastly, new technologies introduced in agriculture are a challenge, especially in small-scale farming in which they may not have all the necessary resources and technical skills for implementation. This paper addresses the existing technical and adoption-related challenges by developing solutions like encryption of data for privacy protection, usage of cloud-based infrastructure for scalability, and training programs to aid adoption among farmers. An advanced crop recommendation system for enhanced crop management and raised food security will be introduced in this paper. This crop recommendation system will integrate the most current technologies. It addresses multiple challenges that affect agriculture today, from optimizing resource use to improving productivity and ensuring minimal environmental impact, all these factors add up to economic viability for the farmer. This approach is holistic in using a selection of machine learning models to calculate the best crop to plant based on soil moisture levels, nutrient levels, temperature, and rainfall. The IoT sensors allow for real-time monitoring of the factors, thereby ensuring data collection and prompt insights. Along this line, Big Data analytics will proceed to advance the capabilities of the system in handling vast volumes of agro data, changing it into valuable information leading to the practice of smart farming. This research, therefore, constitutes a very great contribution to the knowledge base for precision agriculture, a strong building block for future advancements in crop recommendation systems. Future studies may look to enhance this system to offer greater predictive precision, such as having easier scalability to support bigger datasets and their applications. Adding remote sensing data, satellite images, etc., can also yield a more accurate level of data on crop health, weather conditions, and soil, leading to more precise suggestions. This would be important to develop these systems so that agriculture might be made more sustainable. It gives farmers the ability to better manage the impacts of climate change, optimize resource use, and respond to the demands of a growing global population. And, ultimately, this work supports efforts toward long-term food security and resilience in agriculture, benefiting both the farmer and society at large.

Paper constitutes literature survey in section 2 for the in-depth study of the work done till date, proposed methodology is described in section 3 which includes: machine learning, methodology implemented, pseudocode, correlation colour bar, variable relationship and implications. In the section 4 the results are discussed and in the section 5 future scope while the paper has been concluded in the section 6 followed by references.

2. LITERATURE SURVEY

Sai Teja M.et al. (2022) discussed a crop recommendation system that enhanced productivity and guided farmers in economic management and decision-making for different crops. The system uses a regression algorithm on various soil parameters and weather conditions. It recommends crops like wheat, rice millets, peas, and green gram with 99% accuracy outperforming ANN with 93% and LSTM with 97% accuracy in this implementation.

Agarwal A. et al. (2023) have comprehensively studied crop recommendations based on soil properties. The study focused on a machine learning algorithm called XGBoost, Random Forest, Decision tree, naïve Bayes, SVM, and Logistic Regression 1. XGBoost, naïve Bayes, and Random Forest stand out with more than 99% accuracy. The analysis was based on soil properties, nutrient content, pH level, etc.

Avanija J. et al. (2024) proposed a hybrid methodology with AntLion optimization and a decision tree algorithm. This method proposed mixing two methodologies to improve crop recommendations. This hybrid approach has increased the accuracy of the decision tree. The model has achieved an accuracy of 98.86% in crop recommendation. Despite hybridization, the model cannot achieve accuracy above 99%, which XGBoost and Random Forest offer.

Sani S. R. et al. (2023) went into the details about the Random Forest Algorithm for crop recommendation due to the high accuracy through the numerous decision trees. The system they proposed has achieved a robust accuracy of 99%. This system targeted farmers, researchers, and policymakers in making informed decisions on crop planning and management. The system can reach various parts of the county as it can be deployed on cloud platforms. They also have visualized and predicted labels and actual labels using scatter plots.

Gorika B. P. et al. (2023) focused on young farmers, selecting suitable crops based on environmental factors. The authors have tried various algorithms like K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), as well as Naïve Bayes (NB). Among them, The Naïve Bayes achieved the highest accuracy at 99% and is followed by RL at 98%. The paper discussed the use of the Google Assistant-based crop recommender platform.

Ramachandra A. C. et al. (2023) developed a system that recommends crops based on factors such as soil nutrients, temperature, humidity, and rainfall. The performance of the model was evaluated using an accuracy score. The model was based on the comparison of the Support Vector Machine (SVM), Random Forest, and Naïve Bayes. The Naïve Bayes has archived the highest accuracy of 99.09%

followed by Random Forest at 99.08%. This paper targets the use of IOT components to consume live data and devise an optimal approach to implement in India and internationally in the future.

Gupta R. et al. (2021) highlighted the integration of IOT and Wireless Sensor Networks in smart agriculture to enhance food production to meet the growing demand of the growing population of the world. The implementation of WSNs such as ZigBee, WiFi, SigFox, and LoRaWAN is used in IOT_WSNs for smart agriculture. The paper discussed the use of IOT-WSNs for smart irrigation, soil moisture monitoring, fertilizer optimization, and pest and disease control. Despite its usefulness scalability, reliability, data privacy, and security remain the biggest challenge in future directions.

Mowla M. N. et al. (2023) proposed a soil fertility mapping architecture-based IOT-based model to assist machine learning-based context-aware fertilizer recommendations. The model's accuracy against the standard soil chemical analysis method compares different machine learning algorithms like Logistic Regression, SVM, KNN, and GNB. The GNB model comes out with 96% and 94% accuracy from training and testing datasets, respectively. The study focused on increasing crop yield and reducing resource waste.

Kulkarni V. et al. (2022) stabilized the connection among various IOT-based smart agriculture technologies, including wireless sensor networks, RFID, GPS, and unmanned aerial vehicles (UAVs). These devices are oriented towards increasing the yield production of crops. The analysis focused on crop monitoring, irrigation management, and greenhouse automation. However, the limitations of technologies, security, energy consumption, and high cost are the biggest obstacles to implementation.

Madhumathi R. et al. (2023) proposed a machine learning approach for predicting soil nutrient and crop recommendations, which allows optimization of agricultural practices. Regression models have been applied, including Boruta, Ridge, and ElasticNet models, aiming to predict the NPK levels of soil with actual values obtained with 91% accuracy using the Boruta model. On the other hand, a crop recommendation system uses a KNN technique to suggest appropriate crops concerning matters such as NPK, pH, temperature, and humidity with almost the same accuracy rate. Future updates may include real-time data sources to enhance recommendations in supporting an efficient and profitable farming practice.

Khan A. A. et al. (2022) described a crop prediction system based on weather and the use of big data analytics. The system uses parameters like temperature, rainfall, humidity, wind speed, and soil types from various datasets. The system used MapReduce to process large datasets efficiently. The system Recommends for the top 3 of, seasonal crops, year-round crops, suitable soil types and seed varieties, and expected weather conditions for optimal yield. The paper includes the future scope of the addition of more factors like soil moisture and cloud cover and incorporating disease warnings and fertilizer recommendations.

Davrazos G. et al. (2023) recommended an IoT-enabled crop recommendation system for smart agriculture using machine learning that allows it to give crop recommendations based on factors such as nutrients in the soil and the climate. In this research, the authors tested various machine learning models on an open-source dataset that contained 22 types of crops and the influencing factors such as nitrogen, phosphorus, potassium, and rainfall. Among the varied machine learning models tested, Random Forest gave results with the highest accuracy regarding suitable crop recommendations. Being that this system integrates IoT data with machine learning, therefore, it guides the farmers in making decisions from crop selection perspectives considering sustainable agriculture objectives and increasing resource efficiency together with the crop yield prediction.

Chhikara S. and Kundu N. (2022) investigated an IoT-enabled crop recommendation system for smart agriculture, using machine learning to recommend crops based on different factors: including soil nutrients and climate. Testing different machine learning models, that is, Random Forest, XGBoost, and the Decision Tree, on an open-source dataset containing 22 crop types and factors- among them are: nitrogen, phosphorus, potassium, and rainfall-the authors tested the system, and identified that the

Random Forest model was the best in giving accuracy prediction results on the suitability of crops. The self-sustained crop prediction system, which considers the data of IoT along with machine learning, would aid the farmers in optimizing their crops based on sustainable agricultural objectives and improve the predictions related to resource efficiency and crop yields.

Padmavathi A. et al. (2024) investigated the efficiency of the XGBoost algorithm in classification and regression on different data sets. A comparison to traditional Gradient Boosting is made. Its multithreaded architecture improves the speed of training, and sparse data handling is particularly beneficial to the accuracy of predictions. The study measures performance on a sample set of Pima Indians Diabetes, Airfoil Self-Noise, and Banknote Authentication datasets that together show that XGBoost in most cases performs better than Gradient Boosting, especially in speed. Although it could not claim the same precision for every dataset, XGBoost's feature importance plots could be used to show which key predictive attributes each set would need. Some future studies recommended the use of real-world data with anomalies and tuning techniques to further optimize the algorithm.

Gosai D. et al. (2021) presented an IoT-based solar-powered agricultural irrigation system that will monitor and control applications remotely. Implemented for a farming community, this is one way that allows farmers to check through a smartphone app the water levels, temperature, and humidity, thereby cutting costs while saving energy. The microcontrollers used are ESP32, Firebase for the real-time storage of data, and a friendly Android app. It is solar powered thus reducing dependence on grid electricity and carbon footprint, hence more sustainable. Future work would include mechanisms to introduce automation algorithms and make a switch to drip irrigation from flood irrigation, hence increasing efficiency.

Santhanam R. et al. (2016) presented a hybrid model development for smart agriculture using IoT and ML technologies. Considering the high level of global food demand these days, methods of traditional farming leave much to be desired concerning resource efficiency. The paper uses the sensor network of the IoT to monitor agricultural parameters such as soil moisture, nutrient presence, and pest presence whereas models of ML can be used to analyze the data gathered. With datasets such as dry beans and types of soil, this paper suggests superior accuracy for crop classification, anomaly detection, and even soil-type classification, thereby making the model apparent with resultant superior predictive accuracy, resource efficiency, and promising scalable and sustainable agriculture.

Habibi M. A. et al. (2022) developed an IoT-based agricultural irrigation system with solar energy support to enable remote monitoring and control. The technology is developed for the Kelompok Tani Tawang Makmur (KTTM) farming community, to deal with some of the exorbitant irrigation costs and labor-heavy irrigation demands in that region by incorporating sensors measuring water level, soil moisture, temperature, and humidity. With the ESP32 microcontroller, the collected data are streamed in real-time to a cloud-based Firebase app to be monitored by people on their mobile phones. It reduces energy consumption and carbon emissions while providing an efficient means for sustainable agriculture: scalable water management.

Monaco S., Bottazzi P. and Altobelli F. (2020) presented some analyses of rice cultivars under drip irrigation in Tuscany, Italy, for optimal water consumption and yield. Twelve rice varieties have been evaluated regarding growth, yield, and water efficiency. It is indicated that, from research done, the highest production through drip irrigation was attained; however, the yields of millet differed. In general, the round rice group adapted well to agro-environmental conditions though the Long B group could not adapt. The techniques suggested are precision agriculture and water management intending to improve water use efficiency by bringing down the high watermark of water consumption in drip irrigation. The Risobiosystems project of the Italian Ministry of Agriculture provided an impetus for development in the drip-irrigated rice area and suggested further areas of research in fine-tuning irrigation scheduling with precision agriculture for better water and crop efficiency.

Roy R. and Aslekar A. (2022) discussed the role of the Internet of Things (IoT) in enhancing farm productivity, focusing on precision agriculture methods that integrate IoT sensors, cloud computing, and data analytics. It highlights the benefits of IoT applications such as automated irrigation systems, weather monitoring, and soil management. IoT-enabled devices provide real-time data on crop conditions, soil health, and climate factors, which allow farmers to optimize water usage, reduce resource wastage, and improve crop yields. The research underlines IoT's potential to transform agriculture in India, addressing challenges of water scarcity, climate variability, and productivity enhancement.

Deshmukh M. et al. (2022) proposed a crop recommendation system that uses the XGBoost algorithm to improve crop productivity based on the parameters of soils and weather and predict the three best crops. Regarding that, information on soil nutrients (N, P, K), pH, temperature, humidity, and rainfall will provide better crop choices with aversions in loss. Moreover, the model can present recommendations on soil fertility and five-day weather forecasting that could be incorporated with such a decision. The performing results of crop prediction using these models are comparable to the other algorithms like Random Forest and Naive Bayes, which also have high accuracy and reliability.

Amin A. B. et al. (2021) discussed a wireless sensor network and a drip irrigation system designed to monitor wheat growth under drought stress. Soil moisture management is managed with ESP32-based systems that need to maintain desirable ranges of moisture in potted wheat plants (40-70%). Drought intensities are simulated with the help of the setup, and environmental parameters like air temperature, light, and humidity are monitored to study the responses of wheat to different phenotypes. Results are considered to affect the reproductive stages of wheat instead of shoot growth, and the system offers a controlled environment to study the genetic and phenotypic responses of wheat to drought through precision agriculture.

Kethineni K. et al. (2024) introduced a web-based agriculture recommendation system, applying deep learning to help farmers select appropriate crops, fertilizers, and pesticides. The system applies SVM and Random Forest machine learning models to figure out crop recommendations, considering soil characteristics and climate as parameters. Fertilizer recommendations apply soil nutrient levels in the form of NPK, while a CNN model helps identify pests by uploading images. The proposed crop optimization system aims to raise the quality and yield of crops, thereby making agriculture more productive and accessible to Indian farmers.

Sharma A. et al. (2021) presented a machine learning-based approach for guiding Indian farmers in crop selection based on soil properties levels of N, P, K, and pH, as well as humidity and rainfall conditions. The paper compares five ML algorithms: Decision Tree, Gaussian Naive Bayes, Logistic Regression, Random Forest, and XGBoost, with the most accurate model implemented within a mobile application. The system takes data from sensors on farms spread all across India and creates a model, which will be accessible without an internet connection via an app for Android, assisting farmers in making the right decisions to enhance productivity.

Iniyan S. et al. (2024) proposed an online system to recommend crops and fertilizers to Indian farmers. Choosing K-Nearest Neighbors, Random Forest, and Linear Regression as the ML classifiers, produces the output based on the properties of the soil, climate, and crops grown previously. The system also interacts with the nutrient variability in soil and climatic conditions to achieve sustainable practice. The model proposed also stresses user access through a website, giving predictions about the improvement in crop yield and profitability for India's widely varied agricultural regions.

Thendral R. and Vinothini M. (2023) recommended crops and fertilizers for improving agricultural productivity by increasing crop yield and decreasing environmental costs. Inputs in the form of characteristics of soils, crop yield data, and fertilizer usage for giving a sharp point-by-point recommendation toward significantly higher crop yield with minimal environmental costs are brought into play through models of CNNs, RNNs, and LSTMs. So far, the model has been confirmed in the aspect of accuracy, nutrient use efficiency, and cost-effectiveness, thus revealing the potentiality of this system in making agriculture sustainable.

Ramya J. and Sankar M. (2023) described a machine learning system that can decide on the best selection of crops for agriculture. The paper focuses on the need to establish increased productivity within agriculture through targeted crop recommendations by evaluating attributes in soil, climate conditions, and nutrient levels. This can be achieved by testing the several supervised models applied in this research, including Random Forest, Decision Tree, SVM, and k-NN, of which the best model proved to be Random Forest. According to this data-driven precision agriculture, there will be a need for educating the respective farmers on what kind of crop would thrive best according to the environmental condition specification.

Kulkarni N. H. et al. (2018) improved crop productivity through a crop recommendation system using ensembling technique in an ensemble-based approach, which enhances the level of accuracy in crop recommendation for better agricultural results. The system proposed here calculates and measures the data for soil properties as well as climate and recommends a suitable crop based on a set of various machine learning algorithms such as Random Forest, Naive Bayes, and Linear SVM with a classification accuracy of 99.91%. Therefore, this system that presents the improved crop selection process would be able to integrate multiple models through majority voting and assist farmers to enable them to make the most appropriate choices, hence helping in the development of the desired productivity.

Raheem M. A. et al. (2023) suggested a system using machine learning (ML) models toward precise crop recommendation based on soil and environmental conditions. The case study discusses the application of Random Forest and XGBoost, amongst other ML algorithms, on processing the soil attributes such as N, P, and K in addition to the weather factors. Techniques applied to improve its performance include data preprocessing, feature engineering, and visualization. The study concludes with Random Forest achieving high accuracy at 98.9%, followed by XGBoost, which is at 98.2%. Future enhancements include GPS-based predictions, disease detection, and fertilizer recommendations for scalable agricultural solutions.

Gayathiri B. et al. (2023) developed an approach for crop yield prediction and fertilizer suggestions via machine learning. In this paper, Random Forest is utilized for crop yield prediction, and fertilizers are suggested based on soil nutrients (NPK) via K-Means clustering. Datasets of crop production and levels of nutrients are employed from Salem district in Tamil Nadu. The yield prediction was highly accurate, with Random Forest; however, the K-Means revealed high effectiveness in fertilizer recommendation clustering. The proposed system aims at improving agricultural productivity and sustainability through crop and fertilizer recommendations to farmers.

Begum S. et al. (2024) based on Random Forest (RF) and K-Nearest Neighbors KNN), introduced a machine learning-based crop recommendation system that assesses soil properties, climate factors and historical data to predict suitable crops, yield, and market prices. The accuracy of the system was 96% using RF and 78% using KNN. A user-friendly web application is created based on Flask to allow recommendations. This can foster greater productivity and efficient resource use by farmers, adaptability according to market demands, and further support sustainable agriculture and food security.

Chowdhury M. et al. (2024) proposed an ICT-based solution for many problems that plague the agriculture sector-the problems being faced by farmers in Bangladesh. The smartphone application is a crop planner, that helps to diagnose crop diseases, access agri-experts for queries, update on weather, facilitate online selling of crops, furnish loan application facilities, allow training approaches, and fertilizer recommendations, all aiming at increasing productivity and profit. Evaluations showed strong farmer acceptance (56–68%) in terms of its applicability and impact. It outperforms existing solutions by integrating comprehensive agricultural support.

Bhola A. and Kumar P. (2024) explored advanced methodologies of smart crop recommendation techniques through machine learning methods. The study takes into consideration the existing models that combine soil parameters with weather conditions and market data in predicting optimal crop yield and profitability. Prior studies show contributions in terms of models; ANN, SVM, RF, and ensemble methods to deal with the challenges involved in precision agriculture. The authors present a new two-phase framework that amalgamates crop filtration based on ANN and yield prediction using RF, which shows a high accuracy of 99.10% with efficient usage of resources. This work marks an important proposition relating to data-driven tools to increase agricultural productivity.

Akkem Y. et al. (2024) discussed a literature review on the XAI integrated with crop recommendation systems, pointing out a growing necessity in the support towards its incorporation to promote trust, transparency, and utility in smart farming. It discusses key XAI methods, LIME, SHAP, and LRP for the interpretation of predictions made by AI models. Applications of XAI in agriculture, counterfactual explanations, and ethical considerations are also discussed. The evaluation of machine learning models and IoT frameworks makes explainability for smart farming instills higher confidence levels in farmers and enhances decision-making processes. The authors propose an XAI-based architecture that brings actionability and understandability to agricultural stakeholders' preferred AI recommendations.

Mahale Y. et al. (2024) proposed crop recommendation and forecasting machine learning-based system for Maharashtra, India from 2001 to 2022. This system is a combination of the Random Forest algorithm for crop predictions and also uses the LSTM algorithm for three-month weather forecasts. Optimization of data based on Expectation-Maximization has been done to bring a prediction accuracy of 92%. Crop-specific recommendations as well as a crop calendar have been implemented. The tool supports informed agricultural decision-making because it addresses the diversity of agro-climatic zones in Maharashtra, thus adapting to climate variability. The outcomes enhance the sustainability of farming, with scalability for similar regions.

Sadeh Y. et al. (2024) introduced the VeRsatile Crop Yield Estimator, termed VeRCYe. The APSIM crop model incorporated with satellite data will be used for yield estimation at the field and sub-field scales. Combining LAI from PlanetScope and Sentinel-2 imagery will give this method high spatiotemporal resolution, so accurate yield prediction will be achievable without relying on ground calibration. Applied to 27 Australian wheat fields, VeRCYe demonstrated strong yield estimation ($R^2 = 0.88$) and mapping capabilities, even making early yield forecasts two months ahead of harvest. It can be scaled to adapt to other crops and regions, accounting for climate variability and boosting agricultural productivity.

Yeşilköy S. and Demir I. (2024) suggested the application of AgERA5 reanalysis combined with crop phenology data to predict the yield of winter wheat over five agroclimatic zones in Türkiye using the deep learning MLP model. Results from the model reveal that combining both meteorological and phenological data improves predictions to an extent where even errors less than 10% may be reached. End This then provides scalable, sturdy responses that enable the betterment of food security and mitigation of the impacts of climate change on agriculture.

Jorvekar P. P. et al. (2024) introduced a new approach toward crop yield prediction within an automated crop management system. It applied deep learning models - an optimized Deep Convolutional Neural Network (DCNN) for predicting crop yields through processing in soil and meteorological parameters. The weights of the DCNN are fine-tuned using the SA-WOA to improve the model in terms of prediction accuracy. This model supersedes the traditional state-of-the-art methods by providing low values of computational error with an MAE of 0.247 at a learning rate of 70.

El-Kenawy E. M. et al. (2024) worked for sustainable agriculture and predictive methodology to enhance to enhance potato yield forecasting, which remains a critical pursuit in achieving sustainable agriculture. The paper primarily relies on machine learning models KNN, XGBoost, and gradient boosting, while deep learning approaches such as GNNs, GRUs, and LSTMs are used for comparison.

GNNs excelled in spatial data handling, achieving the lowest MSE (0.02363), while LSTMs and GRUs demonstrated strengths in temporal pattern analysis. These findings emphasize integrating complex data sources, fostering precision agriculture, and enhancing decision-making to support food security.

Pushpalatha R. et al. (2024) brought to light the various methodologies applied for crop yield forecasting and their application using artificial intelligence techniques. The statistical models, process-based models, remote sensing approaches, coupled models, and data-driven techniques paper analyzes the performance evaluation criteria for the above-mentioned methods. The review identifies that the hybrid model, which combines crop simulation models with such ML techniques as CNNs and XGBoost, has had superior overall efficiency in comparison to the standalone models. The review also stresses local and regional variables and recent developments in AI-based crop improvement and genomics.

Sudhamathi T. and Perumal K. (2024) proposed a hybrid DNN-RNN framework combining DNN with RNN to account for complexities arising from temporal and spatial factors in accurate crop yield prediction. The crop data collected are normalized using the method of Z-score. Independent Component Analysis (ICA) is used for feature extraction. Particle Swarm Optimization (PSO) selects the best features. All crop yield datasets experiments are tested by metrics of Mean Absolute Error (MAE) and accuracy. The results obtained with it (that is 97% accuracy, 99% precision, 97% recall, 98% F1-score and 4.57% MAPE and 0.0333 MSE with 0.4173 MAE). The DNN-RNN framework contributes to improving agricultural practices and crop yield prediction through far-reaching support for farmers and other stakeholders.

The literature review, although precluded in that regard, missed major advancements in crop recommendation systems. The proposed research work contributed by incorporating advanced machine-learning algorithms such as Random Forest, Naive Bayes, and XGBoost for carrying out a comparative analysis of their effectiveness. It highlighted dynamic recommendations further using real-time data collection through IoT sensors. Furthermore, the proposed research work ensures sustainability by optimizing resources and reducing environmental impact. Unlike the literature review, it suggested a flexible two-fold recommendation system that accommodated small and large farmers with different needs based on data from available soils.

3. PROPOSED METHODOLOGY

Proposed methodology includes studying and training the various machine learning algorithms according to the proposed work, methodology implemented outlines the various steps followed to execute the proposed work. Pseudocode, correlation colour bar, variable relationship and implications describes how the proposed system works as follows:

3.1 Machine Learning

XGBoost is an algorithm that would be based on gradient boosting decision trees to go exceptionally well on tabular data [42] with accuracy and efficiently picking up complex rules, making it suitable for a scenario where crop viability happens to be a function of multiple soil and environmental factors. The Random Forest classifier constructs an ensemble of decision trees in order not to get overfitted and merge predictions. It is a good option when there is minimal information about the soil as it might pick up associations in categorical variables such as soil type or climate. To make an easier and more efficient representation, the Naive Bayes classifier has also been added; thereby, it can also be used to make prompt predictions when available resources for computation are few or the data is scarce. Compared with this, we expect our chosen comparison to lead to the best possible trade-off in terms of predictive accuracy, computational efficiency, and adaptability of changing data inputs toward data-driven crop recommendations in sustainable agriculture.

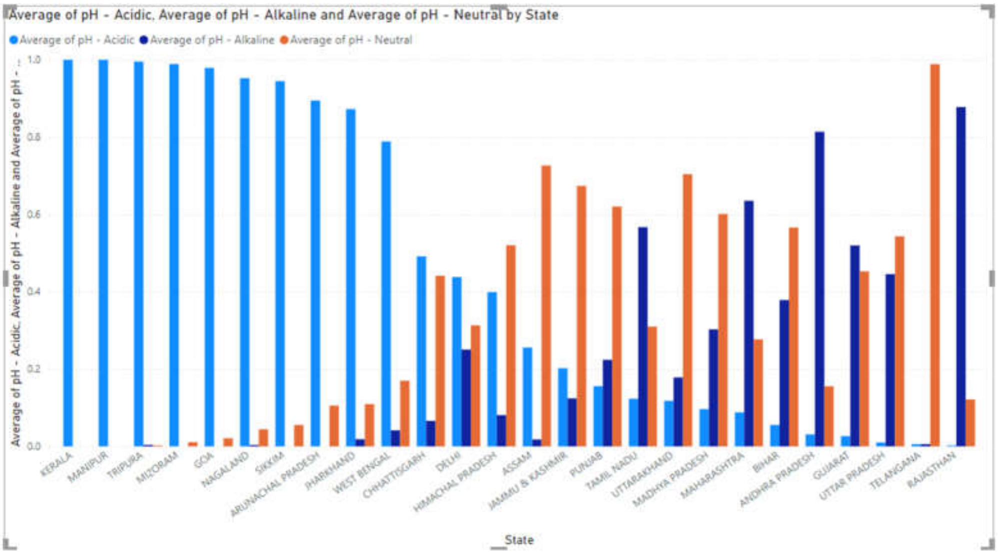

Figure 1: Avg. of pH Acidic/Alkaline/Neutral Vs. States of India.

The above figure 1 depicts, how the average pH [Acidity / Alkaline / Neutral] of soil differs in the states of the country. The pH measures the level of acidity or alkalinity of the substance ranging on the pH scale from 0 to 14. Acidic elements have a pH below 7 while alkaline elements have a pH above 7 and below 14, and 7 is the point of neutrality. Soil microbial activity is a function of the pH of the soil. It is the most important factor influencing crop growth. The bar diagram unequivocally shows that the states Kerala, Manipur, Tripura, Mizoram, Goa, Nagaland, Sikkim, Arunachal Pradesh, Jharkhand, and West Bengal have the highest acidic soil. So, these states might be nutrition-deficient, particularly in calcium and magnesium. The most neutral soils are found in states like Uttarakhand, Assam, J&K, Punjab, Madhya Pradesh, Bihar, Gujarat, and Telangana. Since the pH of the soil is around 7, that makes it the best land for the optimal growth of crops. The above-mentioned states are at a higher potential for yield production. The maximum land with alkaline soil is found in states like Rajasthan, Andhra Pradesh, Maharashtra, Tamil Nadu, Gujarat, and Uttar Pradesh. The excess of calcium carbonate can cause highly alkaline soil, and high pH can lead to deficiencies in iron, manganese, and phosphorus. pH value of soil: The pH value is a decisive factor in crop production and yield. Selection of crops is highly dependent on the pH value of the soil. Most of the crops grow in Neutral pH conditions to produce good growth and yield production. For instance, some plants do well in acidic soils with a pH of 4.5 to 5.5, such as Tea, blueberries, azaleas, and rhododendrons. Lawns require a pH of 5.5 to 6, while roses prefer soils with a neutral pH of 6.5 to 7. Some fruit trees thrive best in alkaline soils, including citrus trees, fig-trees, and pomegranate trees. According to this, the soil of more acidic land tends to be better in states for farming varieties. Because this parameter has its disadvantages, it can be controlled or manipulated through fertilizers. However, it is not recommended to use fertilizer as it permanently damages the soil and renders it barren. It may take years for the environment to replenish the property of the soil. Figure 1 shows the general distribution of soil pH in Indian states. With the help of this graph, a general idea about the crop according to state soil type might be predicated. State vs soil pH average. Provides brief ideas to young farmers on the type of crop that might be cropped in the field with efficient resource management.

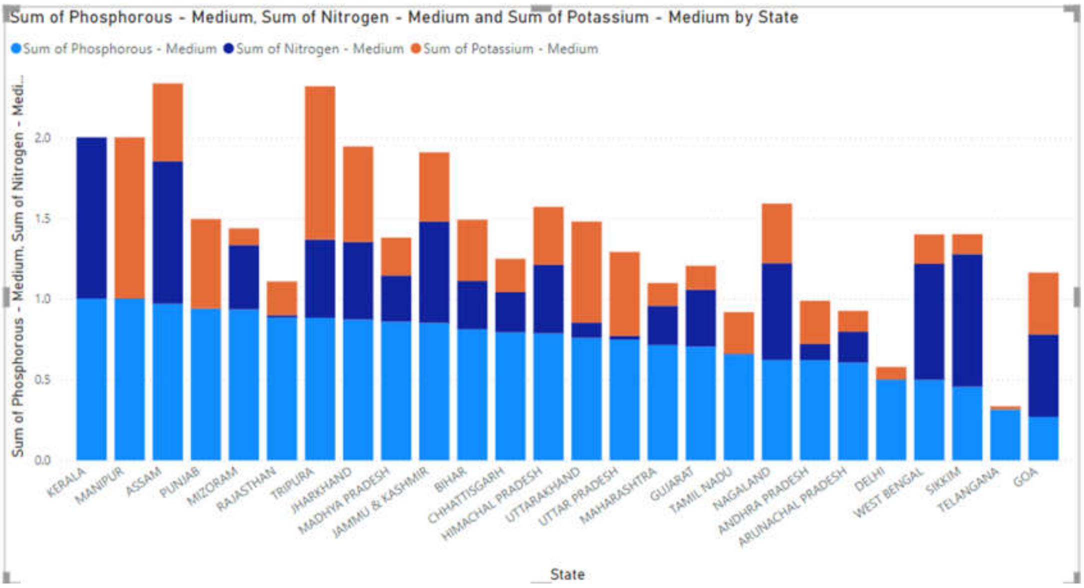

Figure 2: State Vs. Medium of N, P, and K

Figure 2 illustrates the allocation of Nitrogen (N), Phosphorus (P), and Potassium (K) by state and how these nutrients are very essential for crop development and yields. The level amounts of these three nutrients, P, and K-in the soil are considered important indicators of the type of crops that can be raised, as these nutrients are vital to plant health, growth, and resistance to diseases. The major nutrients in the soil include nitrogen, phosphorus, and potassium, and they determine the pattern of the plant's growth and yield as well as their resistance to less-than-favorable conditions. Nitrogen: This nutrient is extremely important for leaf growth and the green color of the plants. Nitrogen constitutes most of the contents of chlorophyll molecules responsible for photosynthesis. Chlorophyll helps convert sunlight into energy for healthy growth. The indications of a lack of nitrogen in plants are a yellowish color and thin stalks which indicate that the plants lack adequate nitrogen levels to maintain healthy and strongly growing appearances. When healthy amounts of nitrogen are present, the leaves will grow green and thick in appearance. Phosphorus is essential for root development, and its availability must be higher at that time when flowers and fruits are developing. It plays a very significant role in early plant development and the formation of healthy roots through which water and nutrients can easily get absorbed. Phosphorus also promotes the flowering and budding of plants. In other words, this nutrient is essential for fruits and seeds to form. Without adequate phosphorus, most likely, a plant will grow stunted without bearing abundant flowers or fruits. Potassium is one of the essential nutrients in the general development of plants and takes part in a large element of the plant's defense mechanisms. It catalyzes several physiological activities, such as water balance, activation of enzymes, and protein synthesis, contributing to the health and tolerance of the plant. Plants grown in potassium-rich soil have greater resistance against stressors like drought and disease. Potassium deficiency usually manifests symptoms at the edge of lower leaves, which can be yellowish or reddish. While lesser sugar content in fruits often results from a potassium deficiency, it brings about some changes in taste and quality. The ratio between these three, which is referred to as the NPK ratio, remains a very essential factor in determining crop yield and affects a farmer's decision when choosing which crop to plant. For instance, if the soil is lacking nitrogen, the plants will appear spindly and washed out. Conversely, if phosphorus is lacking in the soil, it impacts curled leaves and slowed growth. A deficiency in potassium affects the overall health of the plant but particularly impacts the fruits, affecting sweetness; hence, this nutrient is of prime importance to fruit crops. It proves very useful in understanding and addressing soil nutrient imbalances before planting. For example, potassium is short in Kerala states, which may affect crop resilience and quality. Similarly, in Delhi and Telangana states, nitrogen levels are low, which may prevent the plants from growing and increase the leaf size, for instance. However, soil testing must be conducted in the respective fields to collect detailed information about the NPK contents because, although one area might benefit from their nutrient contents, another may not get it because their nutrient levels tend to change dramatically even within the same region.

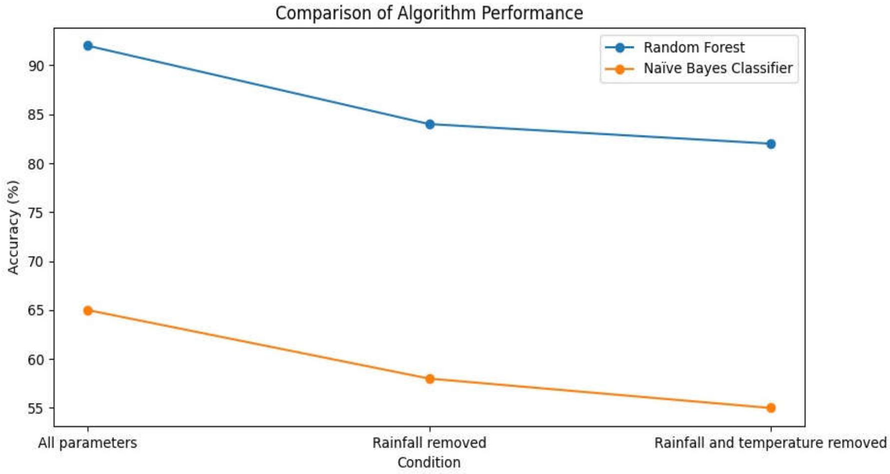

Figure 3: Comparison Graph Between Random Forest and Naïve Bayes Classifier

The "Random Forest Performance" as shown in figure 3 is nothing but a plot illustrating what the removal of key features—rainfall and temperature in this scenario—has on the accuracy of a Random Forest classifier. When all the features are used, the model's accuracy is close to 92%. This is its best-expected performance level. This again is the benchmark against which one evaluates the importance of each feature. When the "Rainfall" feature is removed, accuracy collapses to approximately 86%, indicating rainfall plays a relatively important role in model predictions. Although the model still performs well without the rainfall feature, that 6% reduction in value reflects that it does have some valuable predictive information regarding rainfall. The biggest decline occurs when both rainfall and temperature are removed as features: accuracy falls into the ballpark of around 82%. Such a sharp drop suggests that these two together, at least, are important features that the model relies on for its predictions.

This graph illustrates how these features enhance the model's performance and describe how much reliance is made on rainfall and temperature. This reliance is very possible because Random Forest models depict the aggregation outcomes from several decision trees, so the keys of features enhance the decision-making potential, which has a reliance on the features in importance. The significant amount of loss in accuracy because of the removal of these features underlines their importance. It also points to the necessity of quality data for such crucial features: missing or inaccurate rainfall and temperature data may critically compromise the effectiveness of the model. The above graph suggests that even a few very crucial features removed can cause considerable losses of accuracy; sometimes the weakness of features might be akin to dangers from over-simplification against the benefits of reducing computational costs following feature removal. Thus, there needs to be a balance struck between the model not being too simple or over-parameterized: key predictors cannot be lost in order not to compromise accuracy. The "Random Forest Performance" in figure 3 concludes that rainfall and temperature are necessary features for this model, in both cases, their removal results in large increases in errors. Thoughtful feature selection, appropriate data quality, and knowledge of how features interact with each other all play roles in keeping the model generally effective.

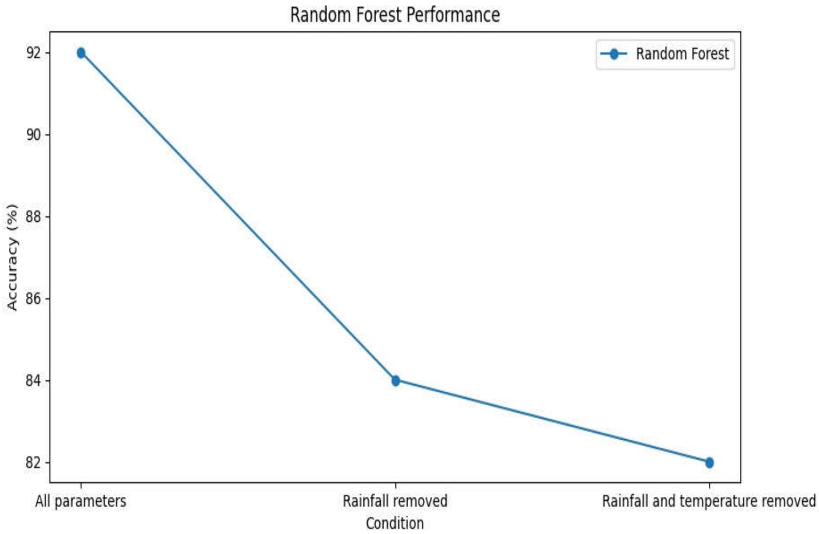

Figure 4: Random Forest Performance Analysis on Rainfall Removal and Both Rainfall and Temp Removal.

The figure 4 depicts "Random Forest Performance" when the key features of rainfall and temperature are removed from the model shows accuracy, whereas the standard accuracy of the model remains at around 92% when all features are included in the model. This provides a good basis for comparison when determining the importance of each feature. Removing the "Rainfall" factor resulted in a loss in accuracy of close to 6%, which shows that rainfall plays a moderately substantial role in the model's predictions. However, the model without it works relatively well, but the drop suggests that rainfall provides valuable predictive information. The largest fall was seen in the case in which both rainfall and temperature factors had been removed, further reducing accuracy to 82%. This sharp drop in performance means that these two features, especially together, are important for the predictive power of the model. The plot below, "Naïve Bayes Classifier Performance," illustrates the effect of leaving out features like rain or temperature on the Naïve Bayes classifier. The classifier was able to hit a baseline accuracy of about 64% given all features, against which we will compare the performance levels. Deleting "Rainfall" from the model drops accuracy down to around 58%, which says rainfall is a moderately significant feature, contributing to the model's predictive accuracy. But when both "Rainfall" and "Temperature" are deleted, accuracy falls to 55%, which says that both are very significant features for the classifier. The fall in accuracy has shown the predictive power that these environmental features provide and may be due to correlation with the target variable. This graph illustrates how the model degrades in performance if key features are removed; it shows how Naïve Bayes, which is usually less sensitive to feature removal, fails without rainfall and temperature data. Such features appear to hold domain information that improves the model's predictive ability. Their aggregate effect, thus, indicates some level of interdependence among them since they uniquely contribute to improving the overall accuracy of the model. This analysis shows how feature selection is imperative to the operation of Naïve Bayes classifiers because it relies on conditional independence assumptions. Selecting those features thought to be most relevant would help achieve greater accuracy and reduce the computational load imposed by such models, where the presence of redundant or irrelevant features may compromise its potential performance. Also, a reduction in accuracy when the significant features are removed poses further questions on data quality. Incomplete or inaccuracy-collected information about the critical features, especially rainfall and temperature, could affect the performance of the model. This leads to an observation that, through exploratory data analysis, some features may require precise and detailed information if the features are critical. The absence of such important features of data could adversely affect prediction reliability, and thus the correctness of data and accuracy become important for proper training of a model. The criticism also touches on the trade-off between the models being complex and simple. Elimination of features may help reduce the model complexity and the computational requirement but may lead to a loss of a critical amount of accuracy through oversimplification at the cost of elimination of very vital features. A balance between simplicity and performance is thus required; simpler models may be computationally efficient but may lose key information essential for the proper making of predictions. So, decisions about what to include in and what to leave out of the model need to be taken with the potential impacts on performance in view.

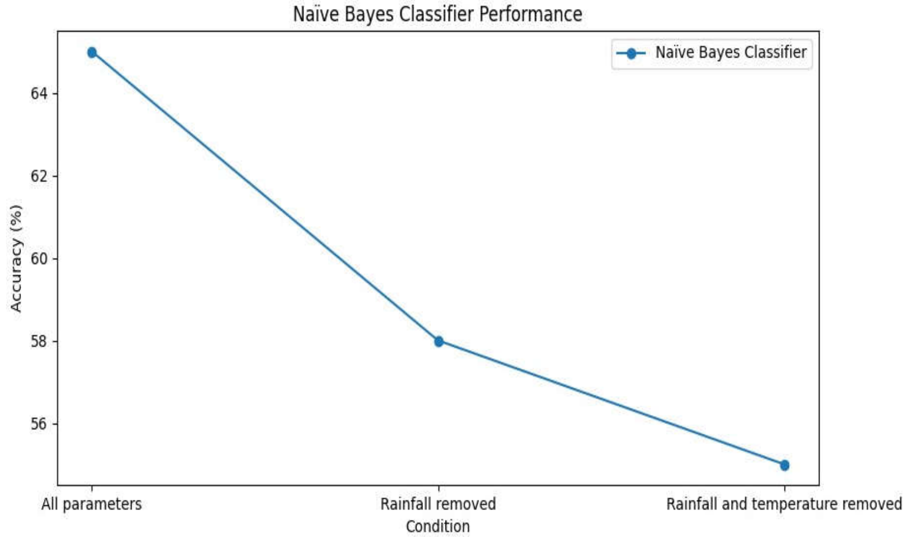

Figure 5: Naïve Bayes Classifier Performance Analysis on Parameter Change (Rainfall & Temp)

The figure 5 shows the performance of a Naïve Bayes Classifier in terms of accuracy concerning the availability of parameters in the data set. The x-axis represents three different conditions: "All parameters," "Rainfall removed," and "Rainfall and temperature removed." Those labels indicate the conditions of which the Naïve Bayes classifier was tested; a demonstration of how model performance was affected in terms of having some of the features removed from the data set. The y-axis represents accuracy as a percentage, showing how much to what extent the classifier performed with each condition. The data points and the trend line indicate a clear relationship between the number of features used in the classifier and its accuracy. In the first condition, with all parameters included, the classifier achieves the highest accuracy, slightly above 64%. This result suggests that the system benefits from Naïve Bayes when all pertinent parameters such as rainfall and temperature and possibly others are included in the model, capturing more information and achieving better classification performance. The presence of all parameters likely enables the classifier to make more informed decisions, suggesting that each parameter contains valuable information that improves accuracy. Moving to the second condition, if rainfall is removed, the accuracy drops significantly to about 58%. This must be a clear demonstration that for the Naïve Bayes classifier concerned, rainfall features play a very important role in this case and its removal significantly lowered the model's effectiveness. Rainfall might be a key variable for the patterns or outcomes that the classifier is trying to predict and, therefore, the lack of rainfall contributes to a reduction in the accuracy of prediction. The loss in terms of 6% in accuracy implies that rainfall accounts for significant information that mainly differentiates one class from another within the dataset. In the third scenario whereby both rainfall and temperature are not accounted for; then, the classifiers' accuracy is reduced to about 56%. This further loss by the model signifies that temperature, like precipitation, is another important variable for the model. Having incorporated both into the exclusion, the classifier then has just very little it can work on, which restricts the capability of the classifier to make the most plausible predictions. The additional loss of around 2% from the preceding scenario then suggests that even though temperature also proves useful, it is not as influential as precipitation within this dataset or given context. However, it is still meaningfully contributing to the performance of the model. In general, this trend of decreasing accuracy from left to right in the graph demonstrates the need for an exhaustive set of parameters in gaining excellent performance by a Naïve Bayes classifier. The highest performance is when all the relevant features are there. Such a classifier relies on the probability distribution of features and has an assumption of conditional independence. The degraded accuracy with further stripping away of features reveals how limited the model is without major predictive parameters. The figure 5 summarizes the sensitivity of performance by the Naïve Bayes classifier to the availability of parameters when all features are available and the best accuracy above 64% for the classifier. Removing rainfall affects performance, dropping accuracy down to around 58%, which establishes that rainfall is the determining factor in this classification problem. Further removal of temperature causes slight incremental drops in accuracy to around 56%, indicating temperature is an important, though relatively less impacting role. This pattern underlines the importance of these environmental factors in whatever predictive task that classifier may be applied to, and it further shows how the Naïve Bayes model depends on the quantity and quality of features for effective classification.

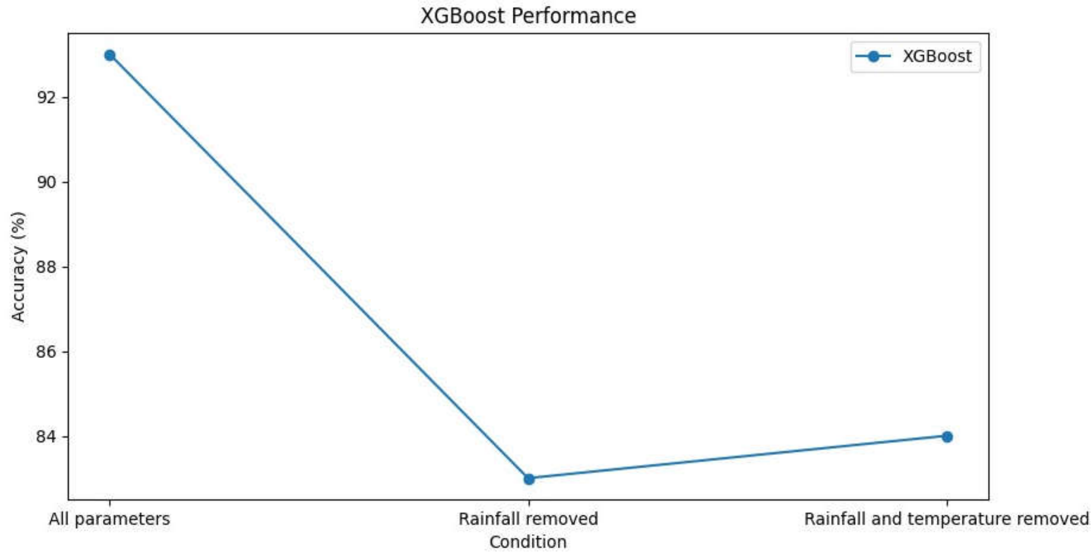

Figure 6: XGBoost Performance Analysis On Parameter Change (Rainfall & Temperature Removed)

Figure 6 entitled "XGBoost Performance" shows the relationship between the accuracy of an XGBoost model with differing conditions where other parameters are either added to or removed from the model. The x-axis has three conditions, namely: "All parameters," "Rainfall removed," and "Rainfall and temperature removed." These conditions relate to the conditions under which the model was tested based on a complete set of parameters, with rainfall removed, and then with both rainfall and temperature removed. The y-axis represents the percentage accuracy of the model and illustrates how the XGBoost classifier performs under each of the conditions. The trend line of this graph and the data points reveal which features influence the accuracy of the model. Under the first condition, "All parameters," meaning the model had the entire data set, the classifier received the highest accuracy and had almost above 92%. Such high accuracy means that when the XGBoost model is given all predictive features-most probably, rainfall, temperature, and maybe more environmental factors-it is optimal. Having all parameters, such a model captures the intricate relationships existing in the data while promoting better discriminative ability and more accurate predictions. XGBoost is a gradient-boosting model, good at obtaining the relationships and interactions among features nonlinearly; therefore, it has comprehensive information translating into high predictive performance. The "Rainfall removed" case is still discernible to have a sharp decline in accuracy with a model performance of approximately 84%. This decline by about 8% in accuracy would indicate that rainfall is a feature that influences things significantly enough for the classifier to need to predict it together with other information that complements this input to make useful predictions. Removing rainfall seems to deprive the XGBoost model of relevant data that it uses to distinguish among the classes or target outcomes involved in this case. A sharp accuracy drop indicates that rainfall is the major factor significantly influencing the classification task, in which it seems that if removed, the performance significantly degrades. In the third condition, "Rainfall and temperature removed," the accuracy stabilizes at about 84.5%. This modest increase in precision suggests that, while temperature elimination does indeed have some effect, it certainly doesn't degrade the performance of the model too much. The temperature should still somehow carry some information of utility, but its impact is much weaker than that of rainfall. The nearly stable accuracy between the second and third conditions suggests that once the rainfall is removed, the temperature itself does not cause another drastic reduction in predictive capability and that this model with the dataset views the information provided by temperature to be less important. This trend in this graph shows that XGBoost is highly sensitive to rainfall as a feature if removed, and there is a sharp fall in accuracy. However, the drop is not as sharp when the feature of temperature is removed; hence, there is an indication that rainfall plays a major role in this classification problem. The model, XGBoost, is known to perform well if key features are present when using structured data, such as those that add informative value to the predictive aspect. This accounts for the loss of accuracy by the absence of rainfall and is like very minimal effects that are attributed to the absence of temperature. In general, the graph suggests that the XGBoost model has a dependency on some key parameters to be able to have high accuracy. With all the parameters being included the model achieves its highest accuracy is over 92%. Precision drops significantly to about 84% when the feature rainfall is excluded. The significance of feature rainfall in this model can be concluded. When both the features, rainfall, and temperature are excluded then the accuracy leaps to 84.5%, which indicates that the temperature has a lesser impact on the model's performance in this context. This pattern illustrates the selective choice of features when selecting XGBoost models. About structured data problems, variables - such as rainfall in this example - that greatly influence the model or enhance its predictiveness, should be included. Other factors, like temperature, could be modestly influential but ought not to be excluded.

3.2 Implemented Methodology

Start: Beginning the workflow.

Import Libraries: In this, all the programming libraries required for the processing and modeling of data are imported.

Load Dataset: In this step, the input dataset is provided, where the model is to be trained and tested

Preprocess Data: In this, cleaning and transforming the data for proper modeling.

Separate Features and Target: In this step, the dataset is separated into features/Input Variables and the target/Output Variable.

Split Data: In this step, data is split into train and test.

Train and Test Models: Initialise, train, and test three different models (Random Forest, Gaussian Naive Bayes, and XGBoost) to predict the crop type and calculate the accuracy.

Compare Accuracy Scores: Determine the model having the highest accuracy.

Take Input: Receive new input data about crop prediction

Predict Crop: Based on input, use the best model to predict the most suitable crop

End

Figure 7 below illustrates a machine learning workflow, and crop prediction based on various classification models. Import the relevant libraries to use when executing functions in the analysis. Load the dataset in the program for further utilization. Load data, prepare it - data cleaning and transformation to better quality and suitability of data for training the model, which may include such steps as missing values handling, feature scaling, and encoding categorical variables. Next, split features-in or independent variables and the target label, dependent variable. You split the data into training and test sets, so you can assess how well the model performs with unseen data.

In the "Train and Evaluate Models" step, three different classifiers are instantiated in total: Random Forest, Gaussian Naive Bayes, and XGBoost. Each is created with parameters that govern the functioning of a classifier. It is an ensemble method where it makes multiple decision trees and averages their output to make predictions, which therefore builds in more robustness and accuracy. Gaussian Naive Bayes is a probabilistic classifier model that assumes Bayes' theorem with normally distributed features in use. XGBoost is an optimized gradient-boosting method for decision trees to improve prediction accuracy. The training dataset is trained by every model, while the test dataset is its target. So, the accuracy of each model is computed using accuracy metrics, which more often depict how good or bad a model performed. It compares the accuracy scores of all three models to pick out which one works the best and then models the crop types with the new input data by the one that shows the highest accuracy. In the final stages, the system inputs new information that, presumably, includes the environmental and soil characteristics that affect crop development and makes predictions based on the best-fit model of which crop is suitable for those conditions. This output concludes the workflow. This diagram will be structured to implement a machine learning system in decision-making for farming, thereby choosing the crop that is the best suited by making precise predictive models.

3.1 Pseudocode

Import Libraries:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import classification_report
from sklearn import metrics, tree
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import GaussianNB
import pickle
import warnings
```

Load Dataset:

```
PATH = 'D:\comp_studies\Research project\Part 1 data\Crop_recommendation.csv'
df = pd.read_csv(PATH)
```

Preprocess Data:

```
le = LabelEncoder()
df['label'] = le.fit_transform(df['label'])
```

Separate Features and Target:

```
features = df[['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall']]
target = df['label']
labels = df['label']
```

Split Data:

```
from sklearn.model_selection import train_test_split
Xtrain, Xtest, Ytrain, Ytest = train_test_split(features, target, test_size=0.2, random_state=2)
```

Train and Evaluate Models:

Random Forest:

```
rf_filename = 'random_forest_model.pkl'
try:
    RF = load_model(rf_filename)
```

```
        except FileNotFoundError:
            RF = RandomForestClassifier(n_estimators=20, random_state=0)
            RF.fit(Xtrain, Ytrain)
            save_model(RF, rf_filename)
        predicted_values = RF.predict(Xtest)
        x = metrics.accuracy_score(Ytest, predicted_values)
        print("Random Forest Classifier's Accuracy is: ", x*100,"%")
```

Gaussian Naive Bayes:

```
        nb_filename = 'naive_bayes_model.pkl'
        try:
            NaiveBayes = load_model(nb_filename)
        except FileNotFoundError:
            NaiveBayes = GaussianNB()
            NaiveBayes.fit(Xtrain, Ytrain)
            save_model(NaiveBayes, nb_filename)
        predicted_values = NaiveBayes.predict(Xtest)
        x = metrics.accuracy_score(Ytest, predicted_values)
        print("Naive Bayes's Accuracy is: ", x*100,"%")
```

XGBoost:

```
        xgb_filename = 'xgboost_model.pkl'
        try:
            XGB = load_model(xgb_filename)
        except FileNotFoundError:
            from xgboost import XGBClassifier
            XGB = XGBClassifier()
            XGB.fit(Xtrain, Ytrain)
            save_model(XGB, xgb_filename)
        predicted_values = XGB.predict(Xtest)
        x = metrics.accuracy_score(Ytest, predicted_values)
        print("XGBoost's Accuracy is: ", x*100," %")
```

Predict Crop Recommendation

```
l = []
    n = int(input("Enter the value of nitrogen content: "))
    p = int(input("Enter the value of phosphorus content: "))
    k = int(input("Enter the value of potassium content: "))
    temperature = float(input("Enter the value of temperature: "))
    humidity = float(input("Enter the value of humidity: "))
    ph = float(input("Enter the value of pH: "))
    rain = float(input("Enter the value of Rainfall: "))
    data = np.array([[n, p, k, temperature, humidity, ph, rain]])
    prediction = XGB.predict(data)
    l.append(prediction[0])
    print("The crop that should be planted is: ", le.inverse_transform([l[0]])[0])
```

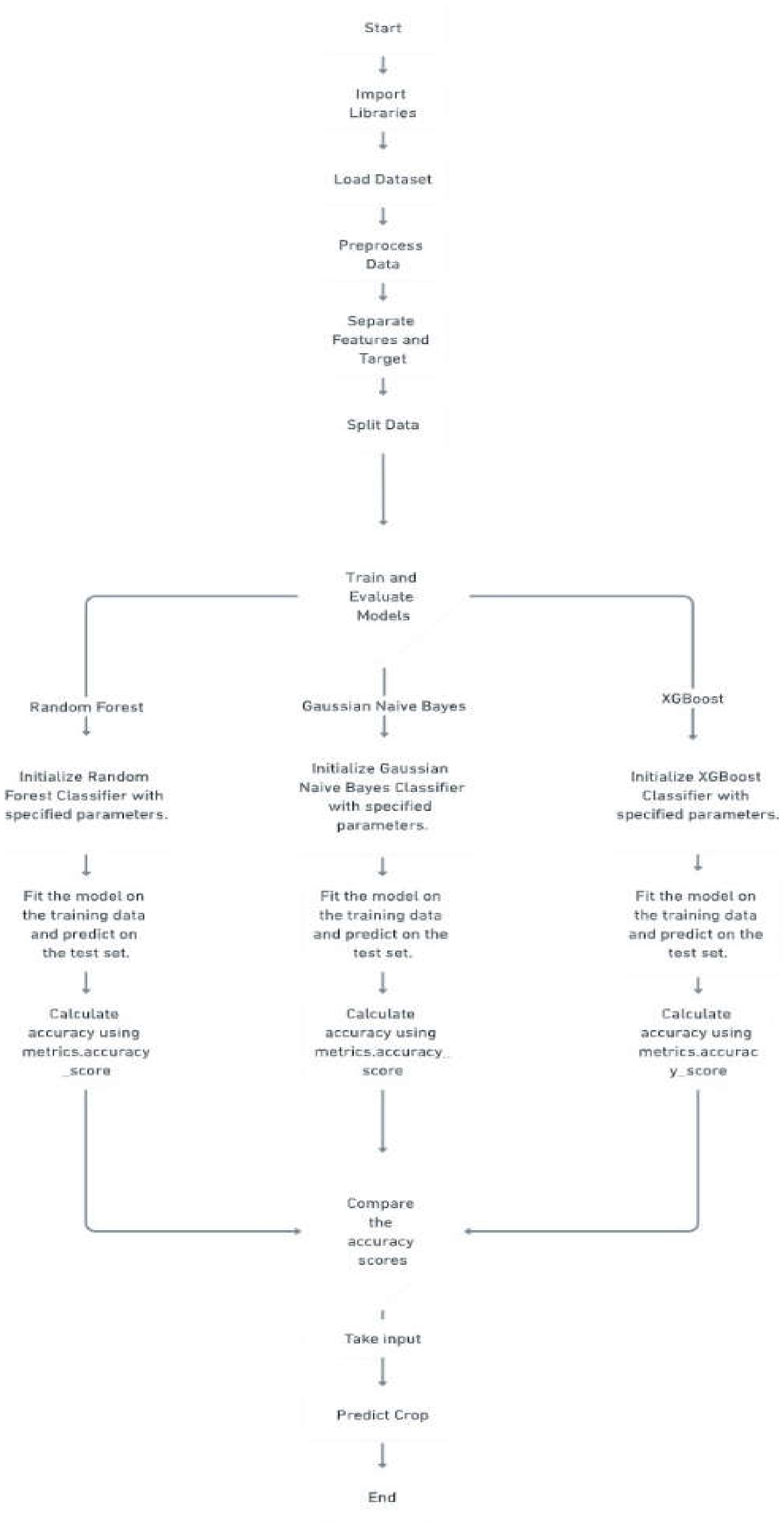

Figure 7: Block Diagram of Proposed Implemented System

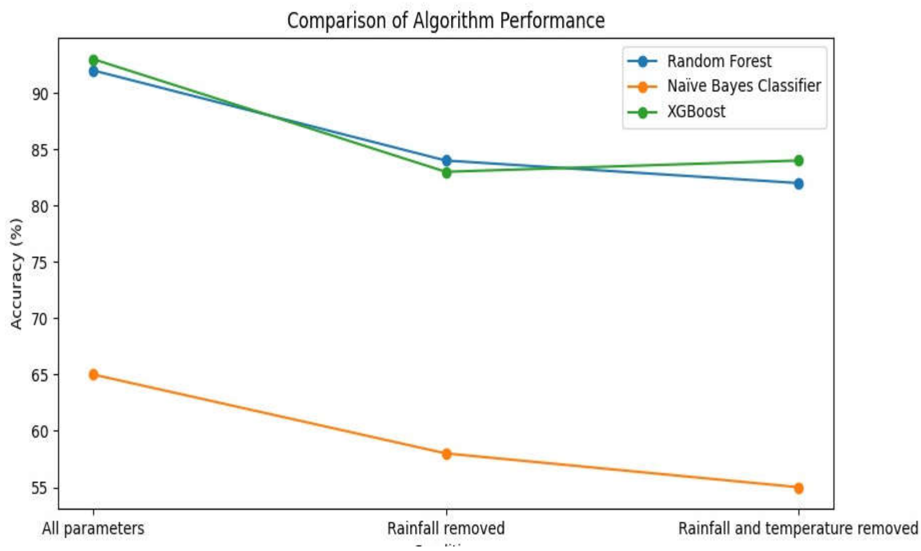

Figure 8: Comparison of Random Forest, Naïve Bayes Classifier & XGBoost Algorithm

Figure 8 compares the accuracy of three machine learning models, which are Random Forest, Naïve Bayes Classifier, and XGBoost, each tested under various conditions concerning included or excluded parameters. The x-axis has three choices in it: "All parameters," "Rainfall removed," and "Rainfall and temperature removed." These are the conditions for testing the model, and the y-axis displays accuracy as a percentage showing which of these conditions each algorithm performed best under. When all parameters are known, then all of these models are correct. XGBoost and Random Forest predict at a very good accuracy of about 92% and 90%, while the Naïve Bayes Classifier starts from about 64%. The skewness shows that the XGBoost and Random Forest models handle the dataset better with all the predictive features because these are pretty good ensemble techniques for revealing complex interactions. As an alternative to this, the Naïve Bayes model, based on conditional independence, does worse with all features because it can't take advantage of feature interactions in the way ensemble methods can. In the second condition, "Rainfall removed," all models see a drop in accuracy, and there's a slight difference by model. XGBoost falls from 92% to 84% with no rain, it's a significant feature. Random Forest also falls less steeply to 86% from 90%. It affects heavily both for XGBoost and Random Forest though Random Forest was a little less sensitive towards it. Naïve Bayes Classifier falls prey to a fall of sorts as its performance crashes down from the supposed 64% to 58%. Rainfall looks reasonable for the Naïve Bayes model, at least less so than the ensemble models do appear to be, based at least in part on perhaps having a slightly smaller total accuracy. In the third case, "Rainfall and temperature removed," the results for each model shift further. Although peaking at about 84.5%, XGBoost stabilizes, implying that removal of temperature adds little value when it does not rain. Compared with this, Random Forest slightly drops to about 85%, which implies there is valuable information in temperature but not as crucial as that from the rain. Random Forest seems more sensitive to multiple feature removal than XGBoost. The Naïve Bayes Classifier sees the accuracy reduced to around 56%, which does start to reveal a cumulative effect from removing rainfall and temperature, both of which were pretty strong for their predictive accuracy. The graph also demonstrated the sensitivity of each algorithm to important key parameters including rainfall and temperature. XGBoost and Random Forest both used all features; but the latter was slightly worse than XGBoost, reaching 100% accuracy. While the ensemble methods were robust, dropping the rainfall and temperature variables significantly reduced the accuracy. Naïve Bayes had a lower baseline in terms of accuracy and dropped much more steeply, so it would seem that overall, its ability to make predictions is relatively fragile to missing information. XGBoost and Random Forest stand out from the others, especially when all the features are included. The most robust one is XGBoost, which shows almost no drop when temperature and rainfall are removed. Random Forest does quite well but is affected more when the features are removed. Naïve Bayes has the lowest baseline and is very susceptible to feature removal, significantly worsening without rainfall and temperature. That stands in sharp contrast to ensemble methods being more robust with missing parameters and Naïve Bayes progressively suffering from lost information. The above graph shows how the accuracy gets affected due to feature selection for both the algorithms used.

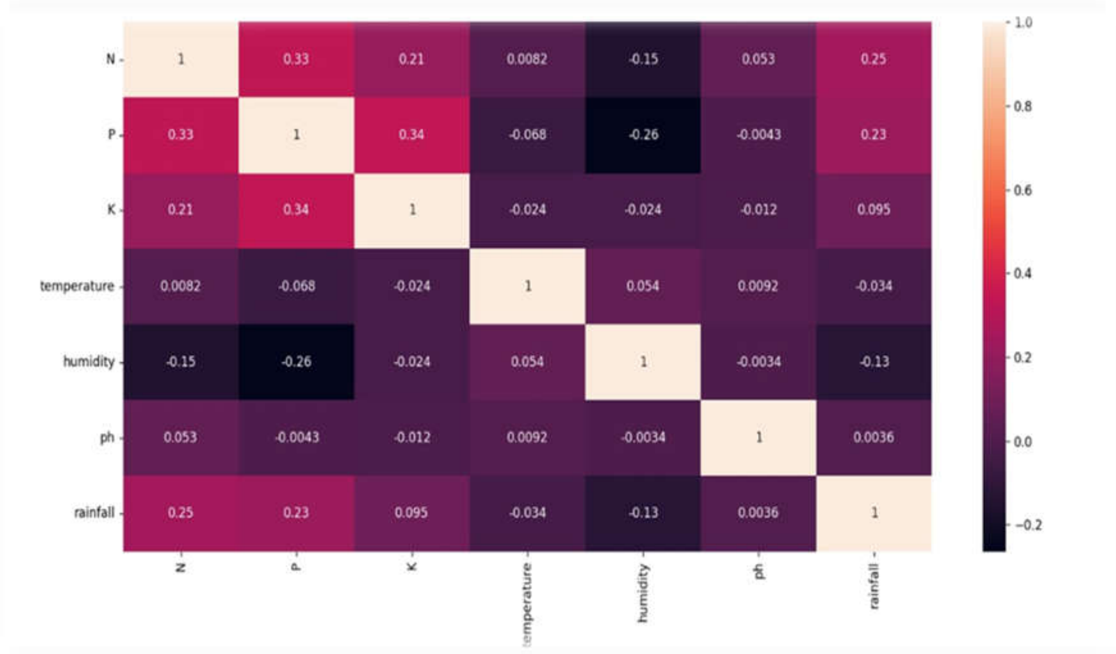

Figure 9: Heatmap Displaying Correlation Matrix.

A heatmap is presented in figure 9, where one can see the correlation matrix that represents the relationship between seven key agricultural and environmental parameters: N-Nitrogen, P-phosphorus, K-potassium, temperature, humidity, pH, and rainfall. The color intensity in a heatmap represents the level of the association; lighter shades represent a higher positive correlation, and a darker color represents a weak or negative correlation. The coefficient value ranges from -1 to +1. This means +1 is a perfect positive correlation, while -1 is a perfect negative correlation. The value is equal to 0, reflecting no linear relationship among the variables.

3.2 Correlation Colour Bar

The color bar in the matrix as shown in the figure 9 ranges from -0.2 to 1.0. The lighter colors show that values have a strong positive correlation. On the other hand, darker shades indicate weak or negative correlations. Some intriguing correlations of modest positive association crop up between Nitrogen and Phosphorus 0.33 and between Phosphorus and Potassium 0.34 These values perform quite well in proving that an increase in one variable goes along with the increase in another. On the contrary, almost zero or negligible correlation emerges in pairs like pH with humidity -0.0034 and temperature with Potassium -0.024 indicating almost no correlation. Some other weak negative correlations are Phosphorus and humidity (-0.26) and Nitrogen and humidity (-0.15), which indicate that higher humidity may be somehow linked to lower nutrient levels of these elements in part.

3.3 Relationship among Variables

Interrelations of the variables give insight into how each might work together in an agricultural or environmental sense:

Nitrogen (N): Nitrogen has a moderate positive correlation with Phosphorus at 0.33, which means that the more Nitrogen is, the more it accompanies the increased levels of Phosphorus. Nitrogen is also weakly positively correlated with Potassium at 0.21 and with rainfall at 0.25, meaning that the variables may rise weakly together. Nitrogen reveals an almost negligible positive correlation with temperature at a value of 0.0082 and a poor level of negative relation with humidity at -0.15, meaning high humidity may be moderately associated with low levels of Nitrogen.

Phosphorus (P): This element has positive correlations with Potassium but at a low moderate level of 0.34. Therefore, an increase in this compound may progress moderately in tandem with the rise in Potassium. Its negative correlation with humidity (-0.26) depicts the fact that lower levels of the Phosphorus compound could also be associated partly with low levels of humidity. Its correlation with temperature is nearly zero (-0.068) and pH (-0.0043) while it has a positive correlation with the rainfall at a weak level (0.23), meaning that the rainfall might sort of correspond to the higher levels of the Phosphorus compound.

Potassium (K): Potassium is associated with relatively minor associations with temperature (-0.024), humidity (-0.024), and pH (-0.012). However, the concentration has an extremely weak positive association with counts of rainfall at 0.095, meaning that the Potassium levels could increase marginally with the occurrence of rainfall.

Temperatures insignificantly relate to other variables with near-to-zero correlations of humidity 0.054, pH 0.0092, and rainfall -0.034. This indicates that changes in temperatures do not significantly affect other variables.

Humidity shows a weak negative correlation with rainfall 0.13 close to near zero with pH at -0.0034. The close-to-nil relationship of pH to rainfall also appears to be close to zero with 0.0036.

3.4 Key Points and Implications

There are two main moderate positive correlations between Nitrogen and Phosphorus (0.33) and between Phosphorus and Potassium (0.34). The latter two may be relatively representative of moderate interdependence wherein these could be partially useful in agricultural purposes where the summation increase of such nutrients influences soil or crop health. Only a few very weak negative correlations emerged, for example with Phosphorus and humidity (-0.26), and Nitrogen with humidity (-0.15). Such results could indicate dilution with humidity, especially for the latter. Putting it simply, very low or even insignificant correlation emerged between most variables, such that apparent interaction among pH and humidity, temperature and Potassium, as well as temperature with other variables, seem minimal. This matrix is a wonderful resource to better understand nutrient and environmental interrelations in farming.

4. Results

The authors have provided some research outcomes for XGBoost and Random Forest. In this regard, test results by the crop recommendation system found that these two models were reliable when the accuracy was at a higher level with or without the removal of an important feature like rainfall and temperature. Even when all the parameters were involved, XGBoost peaked at 92% but tumbled down to as low as only 84.5% when no rainfall and temperature features were involved. The next best was Random Forest, which started at an impressively high 90% accuracy and only showed a moderate reaction to dropping rainfall (up to 86%) and then to rainfall and temperature features, tumbling down to 85%. Comparing it to the Naive Bayes classifier, the loss in accuracy is much more significant from 64% down to 56% when neither rainfall nor temperature is included in the parameters to preserve forecasting accuracy. These results further revealed that although XGBoost and Random Forest performed very well given missing features, Naive Bayes is less robust under similar conditions. In addition, the correlation analysis between the variables of nitrogen, phosphorus, and potassium revealed some modest positive associations, which can be useful for nutrient-specific recommendations for future system updates. The final checks treat the variables of rainfall and temperature as essential to these predictions, such that they drop off very drastically when removed. This shows just how important these variables are to be accurate in crop suitability assessments. Overall, these results indicate that ensemble techniques, such as XGBoost, fit well into dynamic large-scale applications in agriculture when adaptability and predictive accuracy are a prime necessity.

```
Enter the value of nitrogen content: 120
Enter the value of phosphorus content: 200
Enter the value of potassium content: 170
Enter the value of temperature: 25
Enter the value of humidity: 78
Enter the value of pH: 8
Enter the value of Rainfall: 200
The crop that should be planted is:  apple
```

Figure 10: Result 1(Prediction of Model)

```
Enter the value of nitrogen content: 56
Enter the value of phosphorus content: 67
Enter the value of potassium content: 89
Enter the value of temperature: 30
Enter the value of humidity: 55
Enter the value of pH: 8
Enter the value of Rainfall: 150
The crop that should be planted is:  chickpea
```

Figure 11: Result 2(Prediction of Model)

```
Enter the value of nitrogen content: 20
Enter the value of phosphorus content: 30
Enter the value of potassium content: 45
Enter the value of temperature: 56
Enter the value of humidity: 40
Enter the value of pH: 4
Enter the value of Rainfall: 200
The crop that should be planted is:  mango
```

Figure 12: Result 3(Prediction of Model)

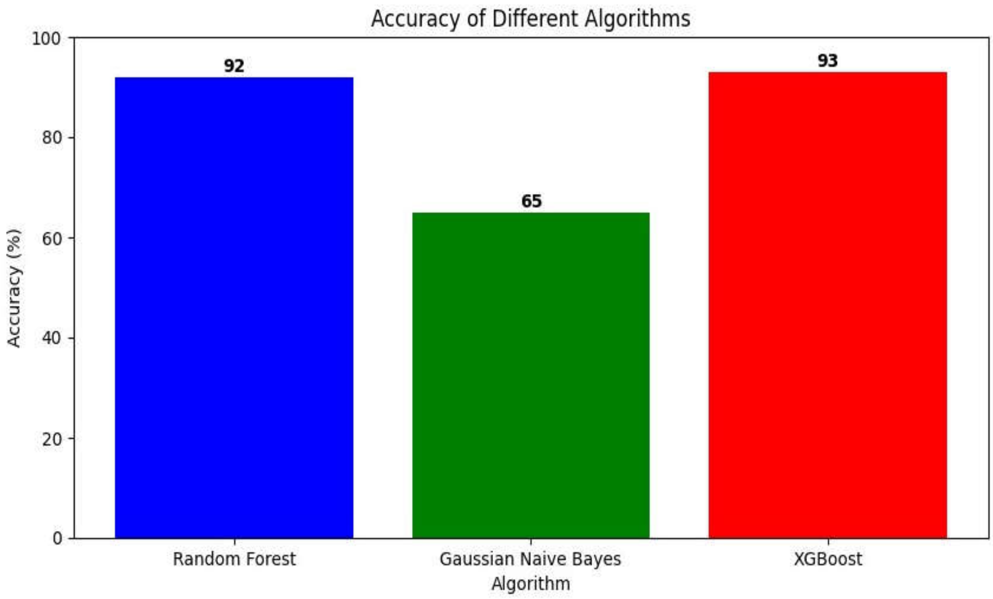

Figure 13: Accuracy Comparison Graph of Random Forest, Gaussian Naïve Bayes and Xgboost.

The bar chart: "Accuracy of Different Algorithms" in figure 13 contrasts the accuracy rate between three crop suitability-predicting machine learning algorithms Random Forest, Gaussian Naive Bayes, and XGBoost. The precision is depicted on the vertical axis as a percentage. It is written above each bar that this is the precision of the algorithm. At 92%, the Random Forest algorithm is too reliable to be used in this application. The lowest accuracy in this case would be that of the Gaussian Naive Bayes which keeps at 65%. In this sense, it would not effectively deal with the complexities of this data set. In terms of accuracy, the highest was achieved through the XGBoost algorithm, which achieved up to 93% accuracy. Thus, it presents the possibility of doing a bit better compared to the Random Forest model and is even stronger in its application were it to be used in producing predictions under this given condition. The accuracy difference appears such that ensemble methods are more robust and suitable for crop recommendation systems compared to Gaussian Naive Bayes since they seem to capture the complex relationships within the data in what seems like a better manner. This graph will focus on the merit of using advanced ensemble methods in boosting higher precision in machine learning-based applications in agriculture, where precision may turn out to be critical in recommending crops.

5. Future Scope

This crop recommendation system offers a basic framework for sustainable agriculture but still allows room for further enhancements. Future work may center on increasing the predictive precision of the system through the provision of data from remote sensing of the satellite imagery of the satellites involved, which would better identify large-scale patterns about precipitation, soil moisture, and crop health. Also, with advanced sensor networks and real-time data from IoT devices, this system can update its recommendations to be applicable in real-time concerning environmental factors when this changes. Therefore, adaptability would be high. Cloud infrastructure will aid this system to scale up in various dimensions for future generations, which will further enable the processing of much larger datasets and areas around the globe with different climates and soil types. Also, as slight changes in soil and environment lead to variances in complex interactions between soil nutrients and productivity in a crop, deep learning techniques can be used in these recommendations for more precision. All these will, in the long run, increase user access through a mobile application interface, supporting various languages, and interactive features which may boost adoption by farmers in many of these regions. A friendly mobile app would also be set up to give periodic updates and notifications, available based on real-time data, to assist quick decisions for farmers. These developments can radically influence the performance of the system in supporting large and smallholder farmers and thereby food security globally.

6. Conclusion

The crop recommendation system, making use of machine learning, is achievable in terms of development and effectiveness, hence capable of assisting farmers in decision-making processes that can enhance the productivity and sustainability of agriculture. Analysis of the comparison of multiple algorithms—that is, Random Forest, Gaussian Naive Bayes, and XGBoost—allowed the conclusion to be drawn regarding which of the three models is more effective in delineating the best crop selection based on environmental and soil parameters. It has been proposed that XGBoost exhibits high accuracy and can predict the best crop choices with a total accuracy rate of 93% followed by Random Forest at 92%, while Gaussian Naive Bayes lags far behind with an accuracy rate of 65%. This means that the choice of a strong algorithm to handle complex agricultural data is paramount, given the intricate dependencies existing between variables such as soil nutrients, rainfall, and temperature. The study highlighted other important features such as rainfall and temperature as keys in determining model predictions. This feature also showed that if these features are removed, the ability of all the models to predict crop suitability went down considerably, thereby establishing once again the weight of such features in ascertaining crop suitability. This will prove helpful for the future upgradation of the system where real-time weather and environmental data can be integrated to potentially boost the adaptability and power of prediction of the system substantially. In addition, relationships found between soil nutrients-nitrogen, phosphorus, and potassium, for instance-provide a basic rationale for targeted recommendations of these nutrients, which would allow farmers to enhance the health and productivity of soils through targeted fertilizer applications. The implications of this research will be profound in farming today as the system to be introduced will be a scalable, technology-based innovation relevant to many farming environments. The system hence caters to different users with recommendations on crop varieties based either on simple types of soils or detailed soil test data. This is a two-stage approach that works both for small as well as for large farmers, who can hence make better use of their resources for productivity despite the inaccessibility of comprehensive soil testing facilities. More added value to this system is its adaptability to sustainable agriculture, thus facilitating optimal use of resources, reducing environmental impacts on ecosystems, and making farmers more responsive to climate change challenges. Looking forward, this paper presents several avenues for future work. Including data from advanced sources, such as satellite imagery and IoT sensor networks, may further refine the capability to predict crops and allow for real-time update responses dynamically to changing conditions. Including a user-friendly mobile application in multiple languages would further make the system more accessible, so that farmers across different regions may access this technology. These upgrades would not only make the crop recommendation system more effective but also more user-friendly and flexible with the intensity of diverse contexts of agriculture across different countries. In conclusion, this work will demonstrate the power of machine learning in turning traditional agriculture into an easier activity for farmers by serving as an efficient tool for crop selection. The high accuracy of the XGBoost and the Random Forest models suggests that those algorithms are well adapted to the complexity of the data from agricultural sources, and their successful use here illustrates the general potential of machine learning in agriculture. With this crop recommendation system, the farmers will be able to obtain better data-driven decisions in achieving long-term goals for food security.

References

[1] Sai Teja M., Sai Preetham T., Sujihelen L., Christy, Jancy S. and Selvan M. P. (May, 2022), "Crop Recommendation and Yield Production using SVM Algorithm", 6th *International Conference on Intelligent Computing and Control Systems (ICICCS 2022)*, Madurai, India, IEEE Xplore, ISBN: 978-1-6654-1035-9, 2022, doi: 10.1109/ICICCS53718.2022.9788274.

[2] Agarwal A., Ahmad S. and Pandey A. (July, 2023), "Crop Recommendation Based on Soil Properties: A Comprehensive Analysis", *14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Delhi, India, DOI: 10.1109/ICCCNT56998.2023.10307999.

[3] Avanija J., Ambati K., Naraganti L., Derangula S. S. and Nashina T. (June, 2024), "Crop Recommendation System using Antlion Optimization and Decision Tree Algorithm", *3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)*, IEEE, Salem, India, DOI: 10.1109/ICAAIC60222.2024.10575152.

[4] Sani S. R., Ummadi S. V. S., Thota S. R., Muthineni N., Swargam V. S. S. and Ravella T. S. (May, 2023), "Crop Recommendation System using Random Forest Algorithm in Machine Learning", *2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2023)*, IEEE, Salem, India, DOI: 10.1109/ICAAIC56838.2023.10141384.

[5] Gorika B. P., ReenaSri S., Dhanushree U. and Sivadhanu K. (Nov, 2023), "Crop Recommendation Systems using Machine Learning Algorithms", *International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, IEEE, B.G., India, DOI: 10.1109/ICRASET59632.2023.10420164.

[6] Ramachandra A. C., Ankitha G.V., Divya I., Vandana P. and Jagadeesh H. S. (July, 2023), "Crop Recommendation using Machine Learning". *International Conference on Data Science and Network Security (ICDSNS)*, IEEE, Tipur, India, pp. 1-5, DOI: 10.1109/ICDSNS58469.2023.10245154.

[7] Gupta R., Sharma A. K., Garg O., Modi K., Kasim S. and Baharum Z. (2021), "WB-CPI: Weather-Based Crop Prediction in India Using Big Data Analytics", *IEEE Access*, Vol. 9, pp. 137869-137876. DOI: 10.1109/ACCESS.2021.3117247.

[8] Mowla M. N., Mowla N., Shah A. F. M., Rabie K. M. and Shongwe T. (2023), "Internet of Things and Wireless Sensor Networks for Smart Agriculture Applications: A Survey", *IEEE Access*, Vol. 11, pp. 145813-145841. DOI: 10.1109/ACCESS.2023.3346299.

[9] Kulkarni V., Prakasha G., Sridevi N. and Venkatagiri J. (May, 2022), "Soil Classification, Crop Selection and Prediction of Fertilizer Based on Soil Series", *International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)*, IEEE, Bengaluru, India. DOI: 10.1109/IIHC55949.2022.10059636.

[10] Madhumathi R., Ashok Kumar K., Chandra Kumar N., Dhyanesh A. and Arumuganathan T. (Aug, 2023), "Soil Nutrient Prediction and Crop Recommendation System", *International Conference on Circuit Power and Computing Technologies (ICCPCT)*, IEEE, Kollam, India, pp. 976-981, DOI: 10.1109/ICCPCT58313.2023.10245685.

[11] Khan A. A., Faheem M., Bashir R. N., Wechtaisong C. and Abbas M. Z. (2022), "Internet of Things (IoT) Assisted Context Aware Fertilizer Recommendation", *IEEE Access*, Vol. 10, pp. 129505-129517, 2022. DOI: 10.1109/ACCESS.2022.3228160.

[12] Davrazos G., Panagiotakopoulos T., Kotsiantis S. and Kameas A. (July, 2023), "IoT-Enabled Crop Recommendation in Smart Agriculture Using Machine Learning", *14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, IEEE, Volos, Greece, DOI: 10.1109/IISA59645.2023.10345924.

[13] Chhikara S. and Kundu N. (Nov, 2022), "Machine Learning-based Smart Crop Recommender and Yield Predictor", *International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, IEEE, Greater Noida, India, pp. 474-478. DOI: 10.1109/ICCCIS56430.2022.10037678.

[14] Padmavathi A., Gupta A. and Parkash K. B. S. (2024), "Crop Recommendation and Yield Prediction Using Machine Learning Based Approaches", *5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, IEEE, Jamshedpur, India, DOI: 10.1109/ICRTCST61793.2024.10578531.

[15] Gosai D., Raval C., Nayak R., Jayswal H. and Patel A. (2021), "Crop Recommendation System using Machine Learning", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, ISSN: 2456-3307, Vol. 7, No. 3, pp. 554-569, DOI: https://doi.org/10.32628/CSEIT2173129.

[16] Santhanam R., Uzir N., Raman S. and Banerjee S. (2016), "Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets", *International Journal of Control Theory and Applications*, ISSN: 0974-5572, Vol. 9, No. 40, pp. 651-662.

[17] Habibi M. A., Kusumawardana A., Gumilar L., Putra B. A., Adjie R. S. and Izhar, C. A. A. (Oct, 2022), "Implementation of Internet of Things in Agriculture Irrigation System for Remote Monitoring and Controlling with Solar Energy", *FORTEI - International Conference on Electrical Engineering (FORTEI - ICEE),* IEEE, Riau, Indonesia, DOI: 10.1109/FORTEI-ICEE57243.2022.9972917.

[18] Monaco S., Bottazzi P. and Altobelli F. (Nov, 2020), "Performance of Different Rice Varieties Under Drip Irrigation", *International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, IEEE, Trento, Italy, DOI: 10.1109/MetroAgriFor50201.2020.9277628.

[19] Roy R. and Aslekar A. (March, 2022), "IoT in Farm Productivity Enhancement", *International Conference on Decision Aid Sciences and Applications (DASA)*. IEEE, Chiangrai, Thailand, DOI: 10.1109/DASA54658.2022.9765273.

[20] Deshmukh M., Jaiswar A., Joshi O. and Shedge R. (2022), “Farming Assistance for Soil Fertility Improvement and Crop Prediction using XGBoost”, *International Conference on Automation, Computing and Communication (ICACC - 2022)*, Vol. *44*, DOI: https://doi.org/10.1051/itmconf/20224403022.

[21] Amin A. B., Dubois G. O., Thurel S., Danyluk J., Boukadoum M. and Diallo A. B. (May, 2021), “Wireless Sensor Network and Irrigation System to Monitor Wheat Growth under Drought Stress”, *International Symposium on Circuits and Systems (ISCAS)*. IEEE, Daegu, Korea, Print ISSN: 2158-1525, DOI: 10.1109/ISCAS51556.2021.9401545.

[22] Kethineni K., Mekala S. H., Kodali M., Kota V. V. and Jampani J. P. (July, 2024), “A Web-Based Agriculture Recommendation System using Deep Learning for Crops, Fertilizers, and Pesticides”*, International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST)*, IEEE, Vijayawada, India, DOI: 10.1109/ICCIGST60741.2024.10717535.

[23] Sharma A., Bhargava M. and Khanna A. V. (Oct, 2021), “AI-Farm: A crop recommendation system”, *10th International Conference on Advances in Computing and Communications (ICACC)*, IEEE, Kochi, Kakkanad, India, pp. 19-24, DOI: 10.1109/ICACC-202152719.2021.9708104.

[24] Iniyan S., Raja M. S., Srinivasan R. Santhanakrishnan C. and Srinivasan A. (Dec, 2023), “Crop and Fertilizer Recommendation System Applying Machine Learning Classifiers”, *International Conference on Emerging Research in Computational Science (ICERCS)*, IEEE, Coimbatore, India, DOI: 10.1109/ICERCS57948.2023.10433972.

[25] Thendral R. and Vinothini M. (Mar, 2023), "Crop and Fertilizer Recommendation to Improve Crop Yield Using Deep Learning", 2nd *International Conference on Electronics and Renewable Systems (ICEARS-2023),* IEEE, Tuticorin, India, DOI: 10.1109/ICEARS56392.2023.10085054.

[26] Ramya J. and Sankar M. (Sep, 2023), "Crop Recommendation Using Supervised Learning Techniques", 4th *International Conference on Smart Electronics and Communication (ICOSEC-2023),* IEEE, Trichy, India, DOI: 10.1109/ICOSEC58147.2023.10276199.

[27] Kulkarni N. H., Srinivasan G. N., Sagar B. M. and Cauvery N. K. (Dec, 2018), "Improving Crop Productivity Through a Crop Recommendation System Using Ensembling Technique" *3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, IEEE, Bengaluru, India, DOI: DOI: 10.1109/CSITSS.2018.8768790.

[28] Raheem M. A., Hussain M. S. and Ahmed S. A. (Apr, 2023), “Machine Learning Based Crop Recommendation on a Cloud”, *8th International Conference for Convergence in Technology (I2CT)*, IEEE, Lonavla, India, DOI: 10.1109/I2CT57861.2023.10126185.

[29] Gayathiri B., Brindha P., Karthika I., Saranya E., Rajeshkumar G. and Kanna P. R. (July, 2023), “Machine Learning-Based Crop Suitability Prediction and Fertilizer Recommendation System”, *4th International Conference on Electronics and Sustainable Communication Systems (ICESC-2023)*, IEEE, Coimbatore, India, DOI: 10.1109/ICESC57686.2023.10193542.

[30] Begum S., Gadagkar A. V. and Rameesa K. (July, 2024), “Precision Agriculture Revolution: Enhancing Crop Recommendations with Machine Learning Algorithms for Optimal Yield and Environmental Sustainability”, *2nd International Conference on Advances in Information Technology (ICAIT),* IEEE, Chikkamangaluru, India, DOI: 10.1109/ICAIT61638.2024.10690764.

[31] Chowdhury M., Rahman M. O. and Alam S. (June, 2024), “Proprietor: A Farmer Assistance Smartphone Application with Crop Planner, Crop Disease Help, Agri-Expert Search and Crop

Suggestion Features", *15th International Conference on Computing, Communication and Networking Technologies (ICCCNT),* IEEE, Kamand, India. DOI: 10.1109/ICCCNT61001.2024.10725364.

[32] Bhola A. and Kumar P. (2024), "Farm-Level Smart Crop Recommendation Framework Using Machine Learning", *Annals of Data Science*, SPRINGER NATURE, Vol. 12, pp. 117 – 140, DOI: 10.1007/s40745-024-00534-3.

[33] Akkem Y., Biswas S. K. and Varanasi A. (2024), "Streamlit-Based Enhancing Crop Recommendation Systems with Advanced Explainable Artificial Intelligence for Smart Farming", *Neural Computing and Applications*, SPRINGER NATURE, Vol. 36, pp. 20011–20025, DOI: 10.1007/s00521-024-10208-z.

[34] Mahale Y., Khan N., Kulkarni, K., Wagle S. A., Pareek P., Kotecha K., Choudhury T. and Sharma A. (2024), "Crop Recommendation and Forecasting System for Maharashtra using Machine Learning with LSTM: A Novel Expectation-Maximization Technique", *Discover Sustainability*, SPRINGER NATURE, Vol. 5, No. 134, DOI: 10.1007/s43621-024-00292-5.

[35] Sadeh Y., Zhu X., Dunkerley D., Walker J. P., Chen Y. and Chenu K. (2024), "Versatile Crop Yield Estimator", *Agronomy for Sustainable Development,* SPRINGER NATURE, Vol. 44, No. 42, DOI: 10.1007/s13593-024-00974-4.

[36] Yeşilköy S. and Demir I. (2024), "Crop Yield Prediction Based on Reanalysis and Crop Phenology Data in the Agroclimatic Zones", *Theoretical and Applied Climatology*, SPRINGER NATURE, Vol. 155, pp. 7035–7048, DOI: 10.1007/s00704-024-05046-x.

[37] Jorvekar P. P., Wagh S. K. and Prasad J. R. (2024), "Crop Yield Predictive Modeling using Optimized Deep Convolutional Neural Network: An Automated Crop Management System", *Multimedia Tools and Applications,* SPRINGER NATURE, Vol. 83, pp. 40295-40322, DOI: https://doi.org/10.1007/s11042-023-16754-3.

[38] El-Kenawy E. M., Alhussan A. A., Khodadadi N., Mirjalili S. and Eid M. M. (2024), "Predicting Potato Crop Yield with Machine Learning and Deep Learning for Sustainable Agriculture", *Potato Research,* SPRINGER NATURE, Vol. 68, pp. 759 - 792, DOI: https://doi.org/10.1007/s11540-024-09753-w.

[39] Pushpalatha R., Roshni T., Gangadharan B. and Kutty G. (2024), "Computer-Aided Crop Yield Forecasting Techniques - Systematic Review Highlighting the Application of AI", *Environmental Modeling & Assessment*, SPRINGER NATURE, Vol. 29, pp. 1095–1110, DOI: https://doi.org/10.1007/s10666-024-09978-6.

[40] Sudhamathi T. and Perumal K. (2024), "A Novel Hybrid DNN-RNN Framework for Precise Crop Yield Prediction", *International Journal of System Assurance Engineering and Management*, SPRINGER NATURE, DOI: https://doi.org/10.1007/s13198-024-02577-4.

[41] United Nations. (n.d.). *Population*. United Nations. Retrieved September 18, 2024, from https://www.un.org/en/global-issues/population.

[42] Ingle, A. (n.d.). *Crop Recommendation Dataset* [Data set]. Kaggle. Retrieved September 20, 2024, from https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset.