A Survey on AgriForecast a Machine Learning Based Crop Price Prediction and Advisory System for Rural Farmers

R. Nithya[1], N. Narmadhavarshini[2], K. Rajeswari[3]
[1]*Associate Professor, Department of Computer Science and Engineering Vivekanandha College of Engineering for Women, Tamilnadu, India*
[2]*PG Scholar, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India*
[3] *Assistant Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India*

Abstract: Agriculture is the main source of income for many rural farmers, but they often face problems in predicting crop prices and choosing the right crops to grow. *AgriForecast* is a machine learning–based system designed to help farmers make better decisions by predicting future crop prices and giving simple farming advice. The system uses data such as past crop prices, weather reports, soil conditions, and market demand to make accurate predictions. Machine learning algorithms like Random Forest, LSTM, and XGBoost are used to analyze this data and suggest the best crops that can give higher profits. The main goal of *AgriForecast* is to support uneducated and rural farmers who may not have access to smartphones or the internet. Therefore, the system can also work through offline methods like IVR or SMS, where farmers can receive advice in their local language. This project helps farmers reduce risks, increase productivity, and make smart farming choices without depending on middlemen. Overall, *AgriForecast* helps improve farmers' income, promotes sustainable agriculture, and shows how technology can support rural development in an easy and practical way.

Key Words: Machine Learning, Crop Price Prediction, Agriculture, Rural Farmers, Advisory System, LSTM, Random Forest, XGBoost, Sustainable Farming, IVR System, Smart Agriculture, Market Forecasting.

I. INTRODUCTION

Agriculture is the backbone of most developing countries, providing employment and food security for millions of people. In rural areas, farmers depend mainly on agriculture for their livelihood. However, they often face many challenges such as unpredictable weather, fluctuating market prices, poor soil management, and lack of proper guidance. Due to these factors, farmers sometimes make wrong decisions about which crops to grow, when to sell them, or how to manage their resources efficiently. This leads to economic losses and reduced productivity. In today's world, technological advancements like Artificial Intelligence (AI) and Machine Learning (ML) are changing the face of agriculture. These technologies can analyze large amounts of data, find patterns, and make accurate predictions. By using these tools, farmers can receive timely information that helps them plan better. Machine learning models can predict crop prices, estimate yield, and suggest suitable crops depending on soil quality, rainfall, and temperature. The proposed system, AgriForecast, focuses on developing a crop price prediction and advisory system that can help rural farmers make informed decisions.

A special feature of AgriForecast is that it supports offline access. Many rural farmers are uneducated or live in areas with limited internet connectivity. To help them, the system provides advice through IVR (Interactive Voice Response) and SMS services in local languages. This ensures that farmers can easily understand and follow the suggestions without needing a smartphone or technical knowledge.

Overall, AgriForecast bridges the gap between modern technology and rural farming needs. It empowers farmers with knowledge, helps them manage risks, and improves their income through data-driven decision-making. By promoting sustainable farming practices and better market awareness, this project aims to create a positive impact on rural development and food security.

II. LITERATURE REVIEW

BRIDGING THE GAP IN SMART AGRICULTURE: MACHINE LEARNING FOR ACCESSIBLE CROP YIELD AND PRICE PREDICTION

A review of ten research papers shows that machine learning (ML) and deep learning (DL) techniques are transforming modern agriculture. Most studies focused on crop yield prediction and recommendation systems using algorithms Random Forest, Decision Tree, SVM, ANN, LSTM, and XGBoost, analyzing factors such as soil nutrients, rainfall, temperature, and fertilizer use. Hybrid and optimization approaches, including Ant lion and Crayfish Optimization, enhanced accuracy and reduced computation time. Ensemble and neural network models achieved high precision, with accuracy ranging from 90% to 96%. Despite their performance, deep learning models require large datasets and advanced resources, limiting their use for small-scale or rural farmers. Existing systems mostly rely on online platforms, reducing accessibility in remote areas. The literature indicates strong potential for ML in smart farming but highlights a need for offline, user-friendly, low-cost advisory tools. AgriForecast addresses this gap by offering accurate crop and price predictions with simple voice-based advisories for rural communities.

[1] Paruchuri Saiteja A Comprehensive Study on Multi-Factor Crop Yield Prediction This study focuses on analyzing multiple environmental and soil-related factors to improve the accuracy of crop yield prediction. The researchers used a combination of machine learning algorithms such as Random Forest, Regression, and Support Vector Machine (SVM) to process datasets containing variables like soil pH, temperature, rainfall, fertilizer usage, and previous yield records. The main goal was to determine how different parameters collectively influence yield outcomes rather than depending on a single factor. The dataset was pre-processed to remove missing values, and correlation-based feature selection was applied to improve model efficiency. The Random Forest algorithm gave the best performance with an accuracy of about 92%, outperforming other traditional regression models. This study concluded that considering multiple real-world agricultural factors helps in generating more reliable and practical predictions. The overall conclusion highlights that integrating soil and climate data with machine learning can significantly help farmers in planning, reducing risk, and improving overall productivity.

[2] Kriti Priya Shah, A Comprehensive Analysis of Machine Learning Algorithms for Suitable Crop Prediction in Agriculture This paper presents a comparative analysis of different machine learning models for predicting the most suitable crop for cultivation in a given region. The authors used algorithms such as Decision Tree, K-Nearest Neighbor (KNN), and Naïve Bayes to analyze soil nutrients (NPK values), rainfall, and temperature data. The dataset was collected from agricultural repositories and weather stations, then normalized to remove outliers. Decision Tree and Random Forest achieved higher accuracy and interpretability than KNN and Naïve Bayes. The system recommended crops like rice, maize, or pulses based on local environmental and soil parameters. The accuracy of the model was reported to be around 89%. The study concluded that machine learning can guide farmers to choose the best crop for a particular region, reducing wastage and improving output.

[3] J. Avanija, Crop Recommendation System Using Ant Lion Optimization and Decision Tree Algorithm. This Research Introduced A Hybrid Approach Combining Antlion Optimization Algorithm (Alo) With A Decision Tree Model to Recommend Suitable Crops to Farmers. The Alo Was Used to Tune the Hyperparameters of the Decision Tree, Improving Classification Accuracy and Reducing Overloading. The Model Analyzed Variables Like Soil Moisture, Ph, Rainfall, And Temperature to Recommend Profitable Crops Based on Environmental Suitability. The Proposed Hybrid Model Outperformed Traditional Models Like Random Forest and KNN by Optimizing the Feature Weights and Improving Decision Accuracy. Experimental Results Showed An Impressive Accuracy Rate Of About 94%. The Overall Conclusion Was That Integrating Optimization Algorithms With Machine Learning

Can Significantly Improve System Performance And Ensure Better Predictions Even With Limited Data. This Model Is Efficient, Lightweight, And Suitable For Developing Decision-Support Tools For Farmers In Rural Areas.

[4] Nebojsa Bacanin **,** Crop Yield Forecasting Based on Echo State Network Tuned by Crayfish Optimization Algorithm In this paper, the authors proposed a novel model that combines an Echo State Neural Network (ESN) with a Crayfish Optimization Algorithm (COA) to predict crop yield more accurately. The ESN helps capture complex nonlinear relationships between climate variables and yield output, while COA fine-tunes the network parameters for better convergence and minimal error. The dataset included historical data on crop yields, temperature, rainfall, and fertilizer application. The hybrid model achieved superior results compared to conventional neural networks, with an average accuracy of 95%**.** The study concluded that nature-inspired optimization techniques like COA can effectively enhance deep learning models. It also emphasized that such models can assist agricultural planners and policymakers in predicting production trends and managing resources efficiently.

[5] S Sridevi, Crop Yield Prediction Using Machine Learning Techniques Based on Soil and Environment Traits – A Survey This survey reviewed multiple research papers that used machine learning models such as LSTM, SVM, CNN, and Random Forest to predict crop yield based on soil and environmental characteristics. The authors collected data from government agricultural databases, including information on soil type, pH, rainfall, and nutrient composition. The review found that time-series models like LSTM performed better for dynamic climate conditions, while Random Forest worked well for structured, static data. The general accuracy range was between 85% and 90%**,** depending on dataset size and quality. The overall conclusion emphasized that accurate data collection and preprocessing play a more significant role than the algorithm itself. The study suggested combining multiple models and improving real-time data availability for practical on-field implementation.

[6] Deeksha Tripathi, Design of a Precise Ensemble Expert System for Crop Yield Prediction Using Machine Learning Analytics This study introduces an ensemble expert system that combines multiple machine learning algorithms to achieve higher prediction accuracy in crop yield estimation. The proposed system integrates Decision Tree, Random Forest, Gradient Boosting, and XGBoost models to take advantage of each algorithm's strengths and reduce individual weaknesses. The dataset used contained details about soil composition, rainfall, fertilizer amount, temperature, and previous crop yields. Ensemble learning was applied through weighted voting and stacking techniques to improve stability and robustness. The model underwent normalization, feature selection, and cross-validation to ensure efficiency. Among all models tested, the ensemble model produced the most reliable and consistent results, achieving a remarkable accuracy of 96%**.** The study concludes that using an ensemble approach improves the generalization capability of crop yield prediction systems. This ensures better adaptability across different crop types and regions. Overall, the research proved that integrating multiple models enhances performance, reduces bias, and provides more realistic and actionable predictions for agricultural decision-making.

[7] S.M.Mahedy Hasan , Predictive Analytics in Agriculture: Unraveling the Determinants of Crop Yield with Machine Learning This paper explores how predictive analytics can be applied in agriculture to identify the most important factors affecting crop yield. The authors used feature selection techniques and machine learning models such as Random Forest, SVM, and Gradient Boosting to analyze variables like soil nutrients, rainfall, humidity, temperature, and fertilizer type. The dataset was preprocessed using standardization and normalization to remove noise. Correlation analysis was used to determine which parameters had the greatest influence on crop yield. The model achieved 91% prediction accuracy when using Random Forest due to its robustness and interpretability. The study concluded that predictive analytics helps in understanding the interrelationship between soil health, weather, and yield outcomes. By identifying the key factors, farmers can prioritize resources effectively and avoid unnecessary costs. Overall, the research proved that machine learning not

only predicts yield but also provides insights into which environmental and management variables are most significant for improving productivity. The data was standardized and normalized to reduce noise, and correlation analysis was performed to identify the most influential factors. The Random Forest model achieved a prediction accuracy of 91%, demonstrating robustness and interpretability. The study emphasizes that predictive analytics not only forecasts yield but also provides insights into which environmental and management variables significantly affect productivity.

[8] R. Kanagaraj, Predictive Classification Model of Crop Yield Data Using Artificial Neural Network (ANN)This research focused on developing a predictive classification model using Artificial NeuralNetworks to estimate crop yield categories based on historical agricultural data. The dataset included variables such as soil type, nutrient content, rainfall, humidity, and average temperature. The ANN model consisted of multiple hidden layers with sigmoid activation functions, trained using a backpropagation algorithm. The data was split into training and testing sets, with optimization applied using Adam optimizer. The model effectively learned the nonlinear relationships between input parameters and yield outcomes. The ANN model achieved an accuracy of 93%**,** outperforming classical algorithms like Decision Tree and Linear Regression. The overall conclusion highlighted that neural networks can handle large and complex agricultural datasets, offering better generalization even in uncertain weather conditions. The study recommended the integration of ANN-based models into smart agriculture systems for better forecasting and planning, especially in precision farming.

[9] P.Josephin Shermila, Optimization of Agriculture Using Data Science and Machine Learning This paper focused on how data science techniques combined with machine learning algorithms can optimize agricultural activities from crop selection to yield estimation. The researchers collected real-time data on weather, soil nutrients, water levels, and crop health through IoT devices and sensors. Algorithms like Linear Regression, Random Forest, and SVM were used to analyze and forecast crop yield and soil fertility levels. Visualization and statistical analysis tools were used to understand the data trends. The model achieved an accuracy of around 90% in yield prediction. The study concluded that data-driven agricultural systems can significantly enhance productivity, reduce wastage of resources, and promote sustainable farming. The authors emphasized that future agricultural success will rely on integrating machine learning with IoT, GIS, and remote sensing technologies to make real- time decisions and recommendations for farmers. Algorithms such as Linear Regression, Random Forest, and SVM analyzed this data to predict crop yield and soil fertility. Visualization and statistical analysis highlighted trends and patterns, with the model achieving an approximate accuracy of 90% in yield prediction. The study concludes that integrating machine learning with data-driven approaches can improve productivity, reduce resource wastage, and support sustainable farming practices. It emphasizes the potential of combining machine learning with IoT, GIS, and remote sensing technologies for real-time decision-making.

[10] Ramu Boyedi, From Data to Harvest Through Machine Learning-Based Crop Yield Forecasting This research introduced a time-series-based crop yield forecasting system that uses historical yield data and weather conditions to predict future outcomes. The authors employed LSTM (Long Short-Term Memory) networks, which are highly effective for time-dependent data. The model processed variables such as temperature, rainfall, soil pH, and humidity to forecast the next season's yield. The LSTM model was compared with traditional models like ARIMA and Prophet, showing superior performance. The system was trained using normalized and sequential datasets collected from agricultural departments. The model achieved an accuracy of approximately 92%**,** outperforming most baseline models. The conclusion emphasized that using deep learning for time-series forecasting helps capture seasonal variations and long-term dependencies in data. The study also suggested that combining LSTM with real-time market data could support decision-making in both production and pricing. Overall, this model demonstrates the potential of machine learning to transform agricultural forecasting into a precise and intelligent system. Overall, the research demonstrates how machine learning can transform agricultural forecasting into an intelligent, data-driven system.

III. PERFORMANCE METRICS

Model	Dataset(s)	Accuracy	False Positive Rate (FPR)	Optimization
Random Forest, SVM, XGBoost	Kaggle	85 %	Medium	Hyper parameter tuning and cross-validation to improve model performance.
Decision Tree, KNN, ANN	UCI Machine Learning Repository	82 %-	Medium	Parameter optimization and feature selection to reduce errors.
Decision Tree + Antlion Optimization	Research Gate	88 %	Low	Ant lion Optimization to select best features and tune decision tree.
Echo State Network	Kaggle	86%	Medium	Crayfish Optimization for tuning network parameters efficiently.
SVM, Random Forest, ANN	UCI Repository	83%	Medium	Hyper parameter tuning and data preprocessing for better predictions.
Ensemble Models (RF + ANN)	Kaggle	89%	Low	Weighted ensemble method and parameter optimization to enhance accuracy.
XGBoost, Gradient Boosting	Research Gate	87%	Low	Feature selection and hyper parameter tuning for optimal prediction.
ANN	Kaggle	84%	Medium	Backpropagation optimization with learning rate adjustment.
Random Forest, Decision Tree	UCI Repository	85%	Medium	Genetic algorithm and hyperparameter tuning to improve model.
LSTM, RF	Kaggle	88%	Low	Bayesian optimization to fine-tune model parameters for high accuracy.

IV. ANALYSIS

The literature survey reveals that most existing research in agriculture focuses on crop yield prediction and crop recommendation systems using machine learning algorithms such as Random Forest, Decision Tree, SVM, and LSTM. Many studies have combined environmental factors like rainfall, soil nutrients, and temperature to achieve accurate results. Advanced hybrid models such as Antlion Optimization**,** Crayfish Optimization**,** and ensemb methods have shown higher accuracy and improved performance, reaching up to 95–96% in some cases. However, most systems rely on high-quality datasets and internet-based applications, which are difficult to use in rural areas. Although deep learning models provide better precision, they require complex computational resources. The analysis highlights that existing systems lack price prediction, offline accessibility, and farmer-friendly interfaces**,** making them unsuitable for uneducated or remote farmers. Therefore, there is a strong need for a simplified, data-driven system that supports both crop price forecasting and practical advisory services.

V. CONCLUSION

The literature review concludes that the integration of machine learning and data science in agriculture has significantly improved decision-making for farmers. Studies have shown that predictive models can accurately estimate yield, recommend suitable crops, and analyze environmental and soil conditions. Techniques such as LSTM, Random Forest, and ensemble learning have consistently demonstrated high accuracy levels between 90% and 96%, proving their effectiveness in modeling agricultural data. However, a major limitation across existing systems is the lack of accessibility for rural and small-scale farmers who have limited internet connectivity and technical knowledge. Most current models are web- or app-based, requiring smartphones and literacy, which many farmers in developing regions lack. Furthermore, price prediction — a critical factor for income stability — is largely ignored in prior works. Hence, the overall conclusion is that there is a strong need for an offline, easy-to-use, and voice-assisted advisory system that combines crop price prediction and recommendation using efficient machine learning techniques. The proposed *AgriForecast* system fills this research gap by providing accurate, accessible, and real-time advisory support to empower rural farmers and promote sustainable agriculture.

REFERENCE

[1] Paruchuri Saiteja A Comprehensive Study on Multi-Factor Crop Yield Prediction, https://ieeexplore.ieee.org/document/10677737/

[2] Kriti Priya Shah, A Comprehensive Analysis of Machine Learning Algorithms for Suitable Crop Prediction in Agriculture, https://ieeexplore.ieee.org/document/10533902/

[3] J. Avanija, Crop Recommendation System Using Ant Lion Optimization and Decision Tree Algorithmhttps://ieeexplore.ieee.org/document/10575152/

[4] Nebojsa Bacanin, Crop Yield Forecasting Based on Echo State Network Tuned by Crayfish Optimization Algorithm, https://ieeexplore.ieee.org/document/10649266/

[5] S Sridevi, Crop Yield Prediction Using Machine Learning Techniques Based on Soil and Environment Traits – A Survey, https://ieeexplore.ieee.org/document/10493587/

[6] Deeksha Tripathi, Design of a Precise Ensemble Expert System for Crop Yield Prediction Using Machine Learning Analytics https://doi.org/10.1002/for.3183

[7] S.M.Mahedy Hasan, Predictive Analytics in Agriculture: Unraveling the Determinants of Crop Yield with Machine Learning, https://ieeexplore.ieee.org/document/10440997/

[8] R. Kanagaraj, Predictive Classification Model of Crop Yield Data Using Artificial Neural Network, https://ieeexplore.ieee.org/document/10220791/

[9] P.Josephin Shermila, Optimization of Agriculture Using Data Science and Machine Learning, https://ieeexplore.ieee.org/document/10673122/

[10] Ramu Boyed, From Data to Harvest Through Machine Learning-Based Crop YieldForecasting, https://ieeexplore.ieee.org/document/10507915/

[11] Viji Parthasarathy, A Comprehensive Survey of IoT and Machine Learning Innovations in Agriculture.https://doi.org/10.47392/IRJASH.2024.032

[12] Nagaveni B. Nimbalt, Machine Learning Techniques for Crop Yield Prediction Vol. 19 No. 4 (2024) https://orcid.org/0009-0002-6401-8632

[13] Sathiyamoorthy.E, A Systematic Study About The Crop Yield Prediction with Machine Learning Techniques, Volume 10 - Issue 1 (2023) https://doi.org/10.53555/sfs.v10i1.489

[14] Aravind T Review of Machine Learning Models for Crop Yield Prediction http://dx.doi.org/10.4108/eai.7-12-2021.2314568q

[15] Ravi Ray Chaudhary, A critical analysis of crop management using Machine Learning towards smart and precise farming. Vol. 12 No. 5 (2024) https://doi.org/10.62110/sciencein.jist.2024.v12.809