National College of Ireland

A Hybrid Machine Learning Approach for Crop Classification, Yield and Fertilizer Prediction for Sustainable Agriculture

MSc Research Project
MSc. Data Analytics

Lynnet Grace Nakagiri
Student ID: X23217677

School of Computing
National College of Ireland

Supervisor: Harshani Nagahamulla

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Lynnet Grace Nakagiri
Student ID:	x23217677
Programme:	MSc. Data Analytics **Year:** 2024
Module:	MSc. Research Project
Supervisor:	Harshani Nagahamulla
Submission Due Date:	12/12/2024
Project Title:	A Hybrid Machine Learning Approach for Crop Classification, Yield and Fertilizer Prediction for Sustainable Agriculture
Word Count:	7240 **Page Count** 22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Lynnet Grace Nakagiri
Date:	11/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A Hybrid Machine Learning Approach for Crop Classification, Yield and Fertilizer Prediction for Sustainable Agriculture

Lynnet Grace Nakagiri
x23217677

Abstract

The increase in world population has put pressure on food production and this has made it more crucial to make sustainable decision-making in agriculture. This study proposed a hybrid machine learning model to predict crops, yield, and fertilizer requirements. Base models (Random Forest, Gradient Boosting, and LSTM) are integrated with meta-models (SVR, XGBoost, and MLP). MLP meta-model was the best hybrid framework with accuracy of 99.96% and RMSE of 0.0415 for crop classification. RMSE and MAE of 0.1628 and 0.0912 were achieved in yield prediction and RMSE of 0.0294 and R^2 of 0.9990 in fertilizer prediction. These results highlight the hybrid model's capacity to address agricultural challenges and its potential to enhance precision agriculture technologies. This shows that hybrid model is capable of addressing agricultural challenges.

Keywords: machine learning, agriculture, crop classification, yield prediction, fertilizer optimization, hybrid model, precision agriculture.

1 Introduction

Agriculture is vital to the world's economy in poverty eradication, and food security as it signifies the path towards feeding the exponentially increasing population. As the global population grows, there is the need to improve agricultural productivity owing to the rising demand for food (Abdel-Salam, Kumar & Mahajan, 2024). These objectives can only be achieved by proper assessment of crops, yields and the proper utilization of fertilizers, which are key components of sustainable agriculture. The food production sector has been subjected to several challenges that limit its possibility for enhanced performance including instability in the climate, resource constraints, and ineffective use of fertilizers which leads to uncertainty when making informed decisions (Abbas et al., 2020).

The traditional methods are not able to give reliable solutions to these challenges in agriculture like changing weather and different types of crops in regions, hence Machine learning models provides solution in learning intricate patterns on complex data in huge datasets (Agarwal & Tarar, 2021). Ideal agricultural practices will be realized by machine learning models due to its ability to identify patterns. However, Agarwal & Tarar (2021) in his study noted that challenges arise when complicated models in agriculture are used especially when translating outputs to farming practices.

This research, therefore, goes out to investigate how these various approaches of blended learning; ensemble learning, conventional learning, and deep learning approaches, are implemented. The base models to be incorporated in the hybrid framework of this research include Random Forest (RF), Gradient Boosting (GB), and Long Short-Term Memory (LSTM) with Support Vector Regressor (SVR), Extreme Gradient Boosting (XG Boost), and Multi-Layer Perceptron (MLP). These helps further improves the accuracy and reliability of the hybrid model by making use of predictions from the base and meta models. Each model brings unique strengths to the model. The RF model provides robustness in overall performance, XGBoost outperforming others in regression, and LSTM for capturing temporal information of sequential data. By using the base model results, the different types of meta-models, like SVR for non-linear relationships, XGBoost for better regression predictions and MLP for multi-layer architectures and complex interactions make predictions even better. This new approach employs the benefits of the conventional approaches in identifying relationships and patterns while incorporating deep learning for crop and yield prediction alongside optimizing fertilizer prediction.

1.1 Research Question

"How effectively can a hybrid machine learning model predict crop yields and optimize fertilizer requirements to enhance agricultural productivity and sustainability?"

1.2 Research Objectives for Crop Yield and Fertilizer Forecasting

To address the research question effectively, the following key objectives are established.

To design a hybrid machine learning framework that combines base models (Random Forest, Gradient Boosting, and LSTM) with meta-models (SVR, XGBoost, and MLP) to improve crop classification, yield prediction, and fertilizer forecasting

To optimize the performance of the hybrid framework through hyperparameter tuning, including grid search, and assess its impact on key predictive tasks

To evaluate the hybrid model using key performance metrics such as Accuracy, RMSE, MAPE, MAE, and R^2 to determine its effectiveness in agricultural predictions

To compare the hybrid framework's performance with individual base and meta-models to validate its superiority in precision agriculture tasks

1.3 Research Objectives for Crop Yield and Fertilizer Forecasting

The report is organized as follows: the first section is the introduction about the research topic, the challenges faced when using of traditional machine learning models, research question and objectives. Section 2 entail literature review in crop yield production and fertilizer forecasting which highlights key models and techniques. The third section describe methodology used in development of the hybrid machine learning model. The fourth section is design specification. Results of the model's performance using agricultural data set is presented on the section five. In section six, findings and implication on agriculture alongside future research direction is discussed. Finally, section 7 concludes the report by summarizing the contribution and recommendation for future work.

2 Related Work

This section offers a brief review of previous work done to predict agricultural outcomes with the help of machine learning, deep learning, and both. This review is divided into three main subsections: It includes Machine Learning Techniques, Deep Learning Models, and Hybrid Approaches. It also contains an SOWT analysis that gives an evaluation of strengths and weaknesses to assess the areas of research that are lacking in the existing scholarly work.

2.1 Machine Learning Techniques

Accurate predictions have been made in terms of crop classification, yield and fertilizer recommendation with the advent of machine learning techniques, hence improving agricultural production. Random Forest has been widely adopted in agricultural setting due to its ability to handle complex, high-dimensional dataset while avoiding overfitting it. A study conducted by Bondre and Mahagaonkar (2019) showcased the power of random forest which outperformed other algorithms in feature-rich environments by achieving accuracy of 86.35% in soil classification and 97.48% in crop yield prediction. Another study by Attar et al. (2024) which harnessed the power of RF and Multilinear regression (MLR) in fertilizer prediction recorded accuracy of 93.68% in Random Forest which outperformed other conventional methods including MLR. In regions with heterogenous conditions and when subjected to large dataset, Random Forest struggle with scalability which calls for Ensemble methods like Boosting with have been proved to be effective. A study by Waoo and Tiwari (2024) demonstrated that Boosting improves accuracy through iterative refinement by obtaining a prediction accuracy of 97.43% after they applied this technique.

2.2 Deep Learning Models

Agriculture sector has been revolutionized by deep learning models by offering ability to learn temporal patterns and complex spatial from large datasets. In processing multi-sensor data, it has been identified that Long-Term Memory networks (LSTMs) and Convolutional Neural Networks (CNNs) provides the best solution. Sharma et al. (20204) in their study employed hybrid deep learning model that leveraged the power of CNN and Recurrent Neural Network (RNNs) in prediction and achieved an accuracy of 90% with a mean absolute error (MAE) of 2.17% and root mean squared error (RMSE) of 2.94%. an effective strategy used in this study is integration of multi-modal data such as soil composition, crop histories and weather patterns. Oikonomidis et al. (2022), similarly applied hybrid CNN-based models like CNN-LSTM and CNN-DNN, achieving an R^2 of 0.87 and RMSE of 0.266, which highlights the problem of handling complex feature interactions in crop datasets by combining CNN with deep neural networks. Deep learning models despite their accuracy, often requires extensive computational resources which limits deployment in regions with scarce resources.

2.3 Hybrid Approaches

The most effective methods for crop yield prediction have emerged as hybrid approaches that integrate machine learning and deep learning techniques. Ayalew and Lohani (2023)

combined both RNNs and Support Vector Machines (SVM) with LSTMs and further achieving 96% accuracy for crop yield prediction in Ethiopia's Lower Kulfo Watershed. This study highlighted the ability of accounting temporal and spatial variability in agricultural systems by using hybrid models. The strengths of different algorithms are combined by these approaches to improve robustness and accuracy in prediction. Batool et al. (2022) studied and noted that integrating the AquaCrop simulation model together with XGBoost regression, achieves an RMSE of 0.12t/ha for XGBoost and an RMSE of 0.48t/ha for simulation model. These hybrid approaches are useful for integrating diverse sources of data into cohesive framework, data sources such as weather data, soil parameters and simulation models. However, there key challenges that still remain as computational complexity and reliance on datasets of high-quality.

Table 1 : Summary of Reviewed studies on Crop Yield and Fertilizer prediction

Author(s)	**Models Used**	**Results (Metrics)**	**Key Findings**	**Weaknesses**
Bondre & Mahagaonkar (2019)	SVM, Random Forest (RF)	Soil classification accuracy: RF 86.35%, SVM 73.75%;	SVM excels in yield prediction and RF performs better for soil classification	Limited generalizability to diverse regions
Attar et al. (2024)	Random Forest, Multilinear Regression (MLR)	RF Accuracy: 93.68%	RF outperformed MLR and other methods for fertilizer optimization	MLR accuracy was significantly lower
Waoo & Tiwari (2024)	KNN, Boosting	Accuracy: 97.43%	Boosting improved prediction accuracy incrementally	Lacks integration of external variables
Parashar (2024)	Regression Algorithms	crop yield prediction accuracy: 92% and 100% for fertilizer recommendation	Regression algorithms effectively analyze soil attributes	Overfitting risk
Sharma & Rathore (2024)	CNN and RNN (Hybrid Deep Learning)	Accuracy: 90% MAE: 2.17% RMSE: 2.94%	Hybrid models effectively process multi-sensor data	Computational complexity
Oikonomidis et al. (2022)	CNN-DNN, CNN-RNN, CNN-LSTM, CNN-XGBoost	Best: CNN-DNN, RMSE: 0.266, MSE: 0.071, MAE: 0.199, R^2: 0.87	Effective feature extraction via CNNs and CNN-DNN excels in handling complex agricultural datasets	High computational cost for complex hybrid models
Miqdad et al. (2024)	Hybrid LSTM and SVM	MAPE: 1.84%	Combining LSTM and SVM improves rice yield predictions.	Dependency on historical data
Lahari et al. (2024)	Decision Tree, SVM, Random Forest, Neural Networks, Hybrid	Accuracy: 99.5% and Precision :99.58	Combines multiple algorithms for crop prediction accuracy Hence	Limited interpretability

			high accuracy in diverse conditions	
Batool et al. (2022)	AquaCrop Simulation, XGBoost Regression	AquaCrop: MAE 0.45 t/ha, RMSE 0.48 t/ha; XGBoost: MAE 0.093 t/ha, RMSE 0.120 t/ha	Hybrid models outperform standalone models, with XGBoost yielding the best performance	Simulation models require extensive data
Ayalew & Lohani (2023)	SVM, LSTM, RNN	Accuracy: 96%	Hybrid SVM-deep learning improves predictions in resource-limited regions	High computational complexity
Varshitha & Choudhary (2024)	Random Forest with Feature Selection	Accuracy: 98%	Feature selection improves recommendation accuracy significantly	Limited scalability for large datasets
Archana & Saranya (2020)	Voting-based Ensemble Classifier (Naïve Bayes, Random Forest, CHAID)	Accuracy: 92%	Ensemble learning combined multiple models for improved prediction	May not scale well with large datasets
Mukul Kumar et al. (2024)	Ensemble Techniques, CNN-XGBoost	Accuracy: 98%	CNN-XGBoost excels in high-dimensional datasets	Heavy computational cost
Shahhosseini et al. (2020)	Machine Learning Ensembles (Various Base Learners)	RRMSE: 7.8%; Mean Bias Error: -6.06 bushels/acre	Ensemble models outperform individual models in yield forecasting	Limited application to regions beyond study area
Malashin et al. (2024)	Deep Neural Network (DNN) optimized with Genetic Algorithm (GA), Explainable AI (LIME)	$R^2 = 0.92$ for crop yield prediction	GA- optimized DNN improves accuracy in predicting sustainable crop yields	Requires extensive computational resources and high quality datasets

2.4 Summary and Gaps in Literature

As evidenced from the discussed literature above and in Table 1, both the machine learning and deep learning models offer optimism for the agricultural predictions with certain drawbacks. RF and GBM algorithms are strong in variability and explainability, but they are not good at capturing temporal patterns. Models such as LSTM can make better predictions for time series, but they need large amounts of data and computational resources.

There is potential in employing what we may refer to as the ‘blending’ of both machine learning and deep learning models to overcome these challenges. The pros of using more complicated algorithms and high computations are available but the cons such as the need to overcome greater challenges constitutes a bottleneck. Another weakness is a high emphasis

on the model's performance and low regard for its interpretability and scalability required for deploying models practically, especially in environments with limited computational resources and capacities.

To address these gaps, this present research aims at developing an optimized and efficient hybrid machine learning model that combines base models (Gradient Boosting, Random Forest, and LSTM), meta-models (XGBoost, SVR, and MLP), and a final neural network layer.

3 Research Methodology

In this section, the Cross Industry Standard Process for Data Mining (CRISP-DM) (Rahmadi et al., 2023) will be employed in developing the predictive architecture of crop, yield, and fertilizer prediction in this project as seen in the Figure 1 below. The first stage is business understanding while the second phase is the data understanding where we carry out exploratory data analysis (EDA) to find patterns, relationships and insights within the data. The third stage in CRISP-DM model is data preparation and assists in the preparation of data for modeling and includes steps like data cleansing, encoding of categorical features, scaling of continuous features-variables for modeling. The fourth phase is modeling where the predictive models are built using machine learning techniques. The fifth phase is the evaluation phase which is used to assess the performance of the model using metrics like RMSE, MAE and MAPE and R^2 score to determine extent to which the project goals have been achieved. Lastly, deployment which is the last step in this process will not be conducted in this particular project.

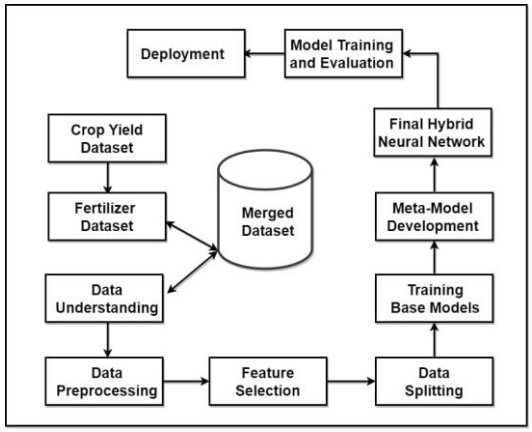

Figure 1: CRISP-DM methodology process applied for this research.

3.1 Data Collection

The primary data used for this project were collected from multiple agricultural datasets that included information on crop yields, environmental conditions, and fertilizer requirements. The two datasets used were:

The crop and yield dataset were obtained from Kaggle and contains 19690 rows and 10 columns. This data is for Indian states from 1997 to 2020 and the columns are extensively described in the Table 2 below with columns.[1]

Table 2: Summary of the Yield and Fertilizer prediction data

Column	Data Type	Description
Crop	Character (Chr)	The name of the crop cultivated.
Crop_Year	Integer (Int)	The year in which the crop was grown.
Season	Character (Chr)	The specific cropping season
State	Character (Chr)	Indian state where the crop was cultivated.
Area	Numeric (Float)	Total land area under cultivation for the specific crop.
Production	Integer (Int)	The quantity of crop production.
Annual_Rainfall	Numeric (Float)	The annual rainfall received in the crop-growing region
Fertilizer	Numeric (Float)	The total amount of fertilizer used for the crop
Pesticide	Numeric (Float)	The total amount of pesticide used for the crop

Table 3: Detailed Summary of the Fertilizer dataset

Column	Data Type	Description
Temperature	Integer (Int)	Value of Temperature rate
Humidity	Integer (Int)	Value of Humidity rate
Soil Type	Character (Chr)	Type of soil where the crop is grown
Crop Type	Character (Chr)	The name of the crop grown.
Nitrogen	Integer (Int)	Rate of Nitrogen rate
Potassium	Integer (Int)	Value of Potassium rate
Phosphorous	Integer (Int)	Value of Phosphorous rate
Fertilizer Name	Character (Chr)	Name of type of fertilizer used

The fertilizer dataset was obtained from Kaggle and contains 99 entries and 9 columns.[2] Each column provides the necessary nutrient information (Nitrogen, Phosphorus, Potassium), moisture content, humidity, and other soil characteristics required to determine appropriate fertilizer use. All columns are described in Table 3 above.

3.2 Data Preprocessing

To ensure the quality of dataset and suitability in model development, data preprocessing was done. Several critical steps were encompassed in the preprocessing workflow with each uniquely designed to enhance compatibility, consistency and integrity.

3.2.1 Standardizing Column Names

The columns in the fertilizer dataset were standardized to ensure compatibility because this was one of the challenges faced during data merging. Changes included renaming attributes such as 'Temperature' to 'Temperature', 'Phosphorous' to 'Phosphorus', Crop Type' to 'Crop' and ensuring all column names were consistent and devoid of whitespace issues.

[1] https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian-states-dataset
[2] https://www.kaggle.com/datasets/gdabhishek/fertilizer-prediction

3.2.2 Merging Datasets

The crop yield data and fertilizer data were merged on key attributes such as Crop, resulting in a unified dataset that combined crop yields with environmental factors and fertilizer requirements. During merging only five crops were considered (Maize, Sugarcane, Tobacco, Wheat, and Barley) which facilitated a more targeted analysis as key crop of interest were focused on in the project.

3.2.3 Descriptive Statistics

The next step involved the use of descriptive statistics on the merged data set to test the nature of the data contained in the variable, mean values, variability, and dispersion or range. This involved the use of descriptive statistics where the mean, standard deviation, minimum, and maximum values were computed and examined to see if there was anything unusual about the data.

3.2.4 Data Cleaning

Missing values were checked in the combined dataset and none were found hence the combined data was used as is without imputation. Duplicate values in the merged dataset were checked and none were found suggesting that the database was unique and of high quality.

During data merging, outliers were identified using Interquartile Range (IQR) statistical method. Abdipourchenarestansofla et al. (2022) stated that IQR method define outlier values beyond 1.5 times the IQR from first and third quartiles. Outliers were retained in this project since their presence did not have any significant impact on the performance of the model.

3.3 Exploratory Data Analysis (EDA)

Exploratory data analysis was done to understand patterns of the data. The analysis included visualizing the distribution of crops, which revealed Sugarcane as the most prevalent crop (as seen in Figure 2) highlighting its significant representation and potential impact on the analysis. Boxplot was plotted which showed the presence of outliers in the data.

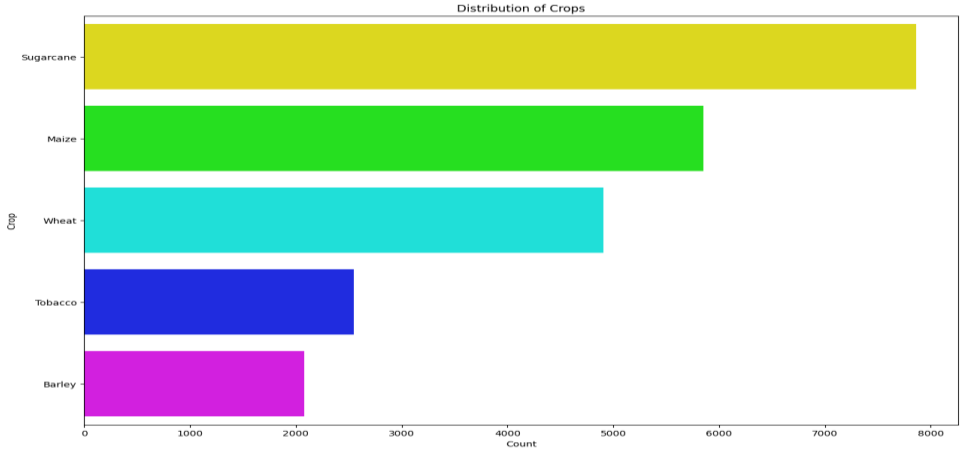

Figure 2: Distribution of Crops

Also, heatmap was used to check for collinearity and correlation between features as seen in Figure 3. These correlations provided useful insights for model building by identifying which features had a more significant impact on the target variables.

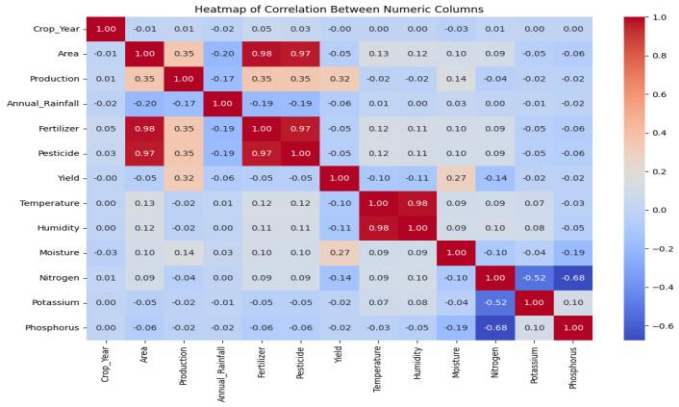

Figure 3: Heat map showing correlation of the features in the merged dataset

3.4 Feature Engineering.

The feature engineering phase involved several steps to prepare the dataset for effective model training and these include:

3.4.1 Label Encoding

Season, State, Crop, Soil Type, and Fertilizer name were the categorical variables that were label encoded in order to transform the categorical variables into numbers. These categorical variables posed a significant challenge, and as such, they were converted into numerical features using label encoding.

3.4.2 Feature Scaling

In order to implement normalization to the numerical features such as Area, Production, Temperature, Humidity, Nitrogen, and Phosphorus, feature scaling was performed. This operation put the variables on the same scale, which is necessary for gradient-based algorithms and guarantees the correct convergence of neural networks. Also, consistency in the data was assured by standardizing categorical (encoded) and numerical variables which to improve consistency in the dataset. The target variables which include: Yield_scaled, Fertilizer_scaled, Crop, and Fertilizer Name were pre-processed for its usability for yield prediction, fertilizer forecasting, and crop classification.

3.4.3 Feature Selection

In this case, Recursive Feature Elimination also known as RFE was applied with RandomForestRegressor to identify which features are important in predicting the outcome. The Recursive Feature Elimination (RFE) process helped in eradicating all the least significant features and thereby selecting the best 12 features ('Crop_Year_scaled', 'Area_scaled','Production_scaled','Annual_Rainfall_scaled','Fertilizer_scaled','Pesticide_scaled', 'Moisture_scaled', 'Season_scaled', 'State_scaled', 'Crop_scaled', 'Soil_Type_scaled', 'Fertilizer Name_scaled') that had maximum impact on the tasks hence avoiding duplication

of features in the dataset so as to give the best output of the model. This also helped reduce noise and improved model performance by focusing on the most relevant variables.

3.5 Base Models Development

In the crop classification, the Random Forest model performed better than the other models such as SVM and XGBoost with an accuracy of 99.57%, precision of 99.57%, recall of 99.55%, and F1-score of 99.55 %. Another benefit of Random Forest especially for multi-class classification is that it can handle both categorical and numerical features, alongside using bootstrapping to reduce overfitting. Also, due to its capacity to model non-linear relationships, it is appropriate to apply it in complex agricultural data sets.

While assessing yield prediction, XGBoost model was more accurate than models such as SVR and KNN where it gave MSE of 0.0993, RMSE of 0.0414, MAE of 0.0180, R^2 of 0.9985 and MAPE of 3.88%. XGBoost is popular for its gradient boosting structure that actively tunes for ideal performance through error reduction. Compared to other methods, it has powerful regularization techniques and can model non-linear relationships and therefore it is more suitable for regression problems like yield prediction where data can have high variance.

LSTM performed better than other methods such as SVR and KNN in fertilizer prediction with an MAE of 0.0115, RMSE of 0.0228, MAPE, 9.11%, and R^2 of 0.9994. LSTM networks are also the best suited for time series data given their capacity for capturing long-term temporal structures and connections. The other issues like vanishing gradients are also handled well by LSTMs and are useful in learning patterns over extended sequences which in fact is beneficial when analysing factors such as soil moisture, climate and crop rotations when it comes to fertilizer prediction tasks.

4 Design Specification

This design specification involves source data from Kaggle and then Analysis to get the best model for the three categories: crop classification, yield, and fertilizer prediction. Each Analysis involves data cleaning, where missing values, duplicates, and outliers are checked. Exploratory data analysis and data preprocessing using label encoder and scaling are also performed. The best models were picked from each section and data is merged for hybrid modelling. The Figure 4 below illustrates all the steps followed to achieve the hybrid model.

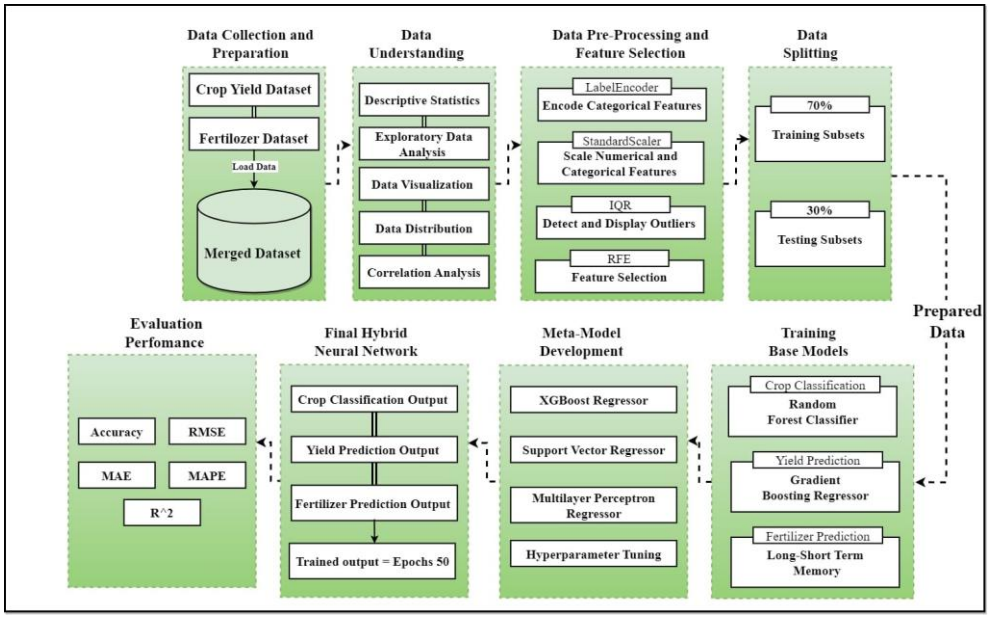

Figure 4: The Process flow design for Crop Classification, Yield and Fertilizer Prediction

4.1 Modelling Technique.

The hybrid model used in this research integrates traditional machine learning, deep learning, and ensemble learning techniques to accomplish three tasks: crop classification, yield prediction and fertilizer prescription.

4.1.1 Base Models Layer.

The base models layer includes the Random Forest classifier model, the Gradient Boosting regressor model and the LSTM neural network model, each of which is trained to perform a specific function on the data on agriculture. The Random Forest model was applied to estimate the type of crop with class probabilities of every potential crop. For the prediction of crop yield, Gradient Boosting Regressor was used. This model is especially suitable for regression problems and provides gradient boosted predictions. The LSTM neural network architecture was then applied to predict the amount of fertilizer needed. LSTMs are particularly good at capturing sequential dependencies in a given dataset and therefore work very well with temporal patterns and modelling of prior conditions that will determine the fertilizer requirement. The selected LSTM network unit consisted of 50 nodes using the ReLU activation function; however, the dropout rate set at 0.3 to prevent over-training of the network model during training

In the implemented model, outputs from the base models: Random Forest for crop classification, Gradient Boosting Regressor for yield prediction, and LSTM for fertilizer prediction were concatenated to form a unified feature set. Random Forest provided crop class probabilities, Gradient Boosting Regressor output a scaled yield prediction, and LSTM predicted fertilizer needs. These outputs were combined into an enriched feature matrix (ensemble_features) and used as input for the meta-models: SVR, MLP, and XGBoost Regressor. This approach allowed the meta-models to leverage the combined insights from all base models, enhancing overall prediction performance.

4.1.2 Meta-model Layer.

In this layer, different types of meta-models have been used and experimented such as XGBoost (Extreme Gradient Boosting), SVR (Support Vector Regressor), and MLP (Multi-Layer Perceptron). XGBoost was used as the meta-model because it is highly resistant to overfitting, highly scalable, and capable of picking up intricate patterns in the data. The SVR was deemed appropriate because it applies a kernel-based method that may be particularly beneficial in determining the most suitable hyperplane to fit datasets in high dimensions. The MLP is a feedforward neural network, which is known as one of the best types of networks in finding complex relations due to the multi-layer structure.

4.1.3 Meta-model Layer.

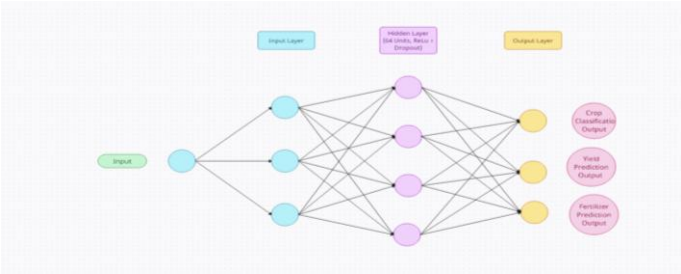

Figure 5: The Final Neural Network architecture for Crop, Yield and fertilizer Prediction.

The final layer in the neural network architecture as seen in Figure 5 is an MLP which accepts the enriched feature set from the base models and the meta-model. This was the final neural network used for making comprehensive predictions of crop classification, Yield, and fertilizer recommendation in one single model. This final stage included the following: Input Layer concatenated outputs from the base models and the meta-model became the input to this layer. The next layer to the network was a hidden dense layer with the aim of improving the merged features and identification of non-linear associations. This layer was composed of 64 units and used ReLU as the activation function while Dropout of 0.3 was also applied to control for overfitting. The output layer consisted of nodes representing each of the three tasks crop classification, yield and fertilizer prediction. The dense neural network was trained with 50 epochs, each epoch having the batch size of 32, the model weights were tuned using the Adam optimizer. It incorporates all these models into one framework for stability and accurate multiple task predictions. These different models in combination together ensured that every kind of learning approach for the estimation of crop type, yield, and fertilizer was taken into account and a very robust manner of generating predictions could be given.

4.1.4 Model Evaluation.

The Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) were used to assess the hybrid model's performance. Each metric provided a different perspective on prediction accuracy, capturing both absolute and related errors. These metrics as seen in table 4 below will be used for crop yield and fertilizer prediction.

Table 4: Evaluation metrics for the Crop classification, Yield and fertilizer prediction

Metric	Formula	What it measures
Root Mean Squared Error (RMSE)	$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2}$	This metric measures the square root of average squared errors between actual and predicted values, making it sensitive to large deviations.
Mean Absolute Error (MAE)	$MAE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\lvert yi - \hat{y}i\rvert}$	This metric calculates the average of absolute differences between actual and political values.
Mean Absolute Percentage Error (MAPE)	$MAPE = \sqrt{\frac{100}{n}\sum_{i=1}^{n}\lvert\frac{yi - \hat{y}i}{yi}\rvert}$	MAPE expresses errors as percentages, offering a normalized measure of accuracy that is easy to interpret.
R^2 score	$r2 = 1 - \frac{RSS}{TSS}$	It shows how a healthy model predicts actual data by indicating the proportion of variance in the target variable that is predictable from the independent variables.
Precision	TP / (TP + FP)	It measures the model's accuracy in predicting positive cases.
Recall	TP / (TP + FN)	It measures the model's ability to capture all actual positive cases.
F1 Score	2 * (Precision * Recall) / (Precision + Recall)	It provides a single measure of a model's accuracy by combining precision and recall.

5 Design Specification

At every level of development and implementation, there are a number of steps and methods involved in creating a machine learning model. Therefore, a thorough strategy is necessary to guarantee the model's correct functioning in the real-world application. In this section we explore the implementation that resulted in the hybrid model crop yield and fertilizer prediction model.

5.1 Tools Used

Jupyter Notebook is used for Analysis with Python as the programming language. Python libraries such as matplotlib, seaborn, NumPy, pandas, sklearn, and TensorFlow are used for data reading, preprocessing, EDA, and modelling. TensorFlow and Keras were used to build and train neural networks, specifically Long Short-Term Memory (LSTM) models and the final multi-layer perceptron.

5.2 Hyperparameter Tuning

When optimizing the performance of the model, hyperparameter is a crucial approach that must be taken. Schratz et al. (2019) conducted a study and concluded that in order to minimize bias, parameter tuning was critical in maintaining the performance of the classifier's consistency. Classification and regression meta-models such as Random Forest, Gradient Boosting Regressor, SVR, MLP, and XGBoost Regressor were tuned using the grid search method which systematically tests a range of combinations in order to identify the hyperparameters with the best performances as presented in Table 5.

- Tuning the Support Vector Regression involved adjusting additional factors such as C, epsilon, kernel and gamma in a bid to better suit the model's fitting strength while

at the same time not compromising on the generality of the model. The tuned parameters used for tuning were C: For the Sigma, Epsilon, Kernel, and Gamma options, specific values have been selected as shown below; Sigma: [0.1, 1, 10], Epsilon: [0.01, 0.1, 0.2], Kernel: [Linear, RBF], and Gamma: ['Scale'].

- To find implicit relationships and prevent overfitting, **MLP** optimization used a hyperparameter tuner to discover the best combinations of hidden_layer_sizes, activation functions, alpha, and learning_rate. The parameters used for tuning were hidden_layer_sizes: [(64,), (128,), (64, 32), (128, 64)], Activation: ['relu', 'tanh'], Alpha: [0.0001, 0.001, 0.01], and Learning_rate: ['adaptive']
- **XGBoost Regressor** tuning uses n_estimators, learning_rate, max_depth, subsample, colsample_bytree and reg_alpha and reg_lambda to reduce the model complexity by strongly regulating the interactions between features and decreasing the likelihood of over-fitting. The parameters used for tuning were N_estimators: [100, 300, 500], [0.01, 0.1], [3, 5, 7], [0.8, 1.0, 1.0], [0.8, 0.6, 1.0], [0, 0.1, 1], and [1, 2].

This approach guaranteed equal fit of all models with the data and reduced the probability of overfitting the models.

Table 4: Detailed Summary Table of the Hyperparameter tuning done for the models.

Model	**Configuration**	**Hyperparameters**
Random Forest Parameters	Default	random_state=42, default n_estimators=100.
XGBoost Parameters	Default	random_state=42, learning rate 0.1, max_depth=3
LSTM Parameters	Default	1 LSTM layer with 50 units, 'relu' activation, dropout of 0.3, optimizer='adam', loss='mse' , epochs=10, batch_size=32
SVR MetaModel	Default	None
	Basic Hyperparameters	kernel='rbf', C=1.0, epsilon=0.1
	Grid Search Optimized	kernel='linear', C=1, epsilon=0.1, gamma='scale'
XGBoost MetaModel	Default	None
	Basic Hyperparameter Tuning	n_estimators=200, learning_rate=0.1, max_depth=4, random_state=42
	Grid Search Optimized	n_estimators=500, learning_rate=0.1, max_depth=5, subsample=1.0, colsample_bytree=1.0, reg_alpha=0.1, reg_lambda=2, random_state=42
MLP MetaModel	Default	None
	Basic Hyperparameters	hidden_layer_sizes=(64, 32), max_iter=1000, random_state=42
	Grid Search Optimized	hidden_layer_sizes=(128, 64), activation='relu', solver='adam', alpha=0.01, learning_rate='adaptive', max_iter=1000
Final Neural Network	Basic Hyperparameter Tuning	Input layer: Combined outputs from the Meta models Dense layer: 64 units, ReLU activation, Dropout layer (0.3), Output layer: 3 units (Crop Classification, Yield and Fertilizer Prediction) optimizer='adam', loss='mse', epochs=50, batch_size=32

5.3 Model Architecture and Training

In this study, the hybrid model comprises of three layers that is the base-model layer, meta-model layer and the final neural network layer. For the Training Process, the dataset was split it into 70% of the data for training and 30% of the data for testing using the Sklearn libraries. The base models were then trained in parallel with the training data of the dataset as discussed above.

6 Evaluation

A critical step in the machine learning process is evaluation, which aims to determine the model's efficiency and make sure its functionality matches the desired design. This procedure guarantees accurate prediction of crop type, yield, and fertilizer needs by applying suitable assessment criteria and carefully examining the model's performance.

6.1 Experiment / Case Study 1

This analysis established that the Default, Basic, and Grid Search Optimized models were different in task-specific variations in performance as seen in Table 6. All the configurations in crop classification showed relatively similar results, with Default SVR possessing the best accuracy, error rates, and efficiency, recording an accuracy of 1.000, RMSE of 0.0000, MAE of 0.0000, and coefficient of determination of 1.000. The Grid search optimized configuration, showed a slightly lesser accuracy of 99.96% and RMSE of 0.0207, essentially implying that tuning of the hyperparameters had only a minimal influence on crop classification.

Table 5: Evaluation metrics for the SVR Meta-model

Metrics	**Default SVR**	**Basic Hyperparameters**	**Grid Search Optimized Hyperparameters**
Crop Classification Metrics			
Accuracy	1.0000	1.0000	0.9996
RMSE	0.0000	0.0000	0.0207
MAE	0.0000	0.0000	0.0004
MAPE	0.00%	0.00%	0.05%
R^2	1.0000	1.0000	0.9997
Yield Prediction Metrics			
RMSE	0.1642	0.1574	0.1628
MAE	0.0885	0.0873	0.0926
MAPE	20.3553%	20.2267%	21.6354%
R^2	0.9765	0.9784	0.9769
Fertilizer Prediction Metrics			
RMSE	0.0590	0.0366	0.0386
MAE	0.0307	0.0277	0.0251
MAPE	21.59%	29.61%	30.31%
R^2	0.9960	0.9985	0.9983

In yield prediction, Default SVR performed well with an RMSE of 0.1642 and MAPE of 20.35% better than Grid Search Optimized SVR which had an RMSE of 0.1628 and MAPE of 21.63%. From the above analysis, it can be deduced that tuning has provided minimal returns for the yield prediction in terms of error metrics.

In the case of fertilizer, the best configuration was Grid Search Optimized SVR, which yielded an RMSE of 0.0386, MAE of 0.0251, and R^2 of 0.9983. These measures demonstrate that hyperparameter tuning is useful for models tasks involving complex relationships and variability since it boosts the enhanced configurations.

6.2 Experiment / Case Study 2

This study investigated how the XGBoost meta model performed in Default, Basic Hyperparameter Tuning, and Grid Search Optimized configurations and established that performance differ according to task type as seen in Table 7.

For crop classification, Default XGBoost gave us 99.97 % accuracy, RMSE of 0.0169, and MAE of 0.0003. Slightly increased RMSE = 0.0207 and MAE = 0.0004 were observed for Grid Search Optimized configuration as compared to Basic configuration which confirmed that hyperparameter tuning did not offer much improvement.

For yield prediction, Default XGBoost had an RMSE of 0.1552, MAE of 0.0874, and MAPE of 20.36%, while Grid Search Optimized had RMSE of 0.1583 and MAPE of 20.95%. The increase patterns also indicate that tuning had a fairly minor effect for this task, which aligns with the general aggression of default values for regression.

Regarding the fertilizer prediction, the Grid Search Optimized XGBoost was statistically significantly superior to the Default and Basic configurations in terms of all the performance indicators, including the lowest RMSE (0.0322), MAE (0.0251), and MAPE (18.70%), and the highest R^2 (0.9988). These results further support the notion that optimizing hyperparameters is crucial particularly when working with data exhibiting complexity and variability in their interactions.

Table 6: Evaluation metrics for the XGBoost Meta-Model

Metrics	**XGBoost Default**	**XGBoost Basic Hyperparameter Tuning**	**XGBoost Grid Search Optimized Hyperparameters**
Crop Classification Metrics			
Accuracy	0.9997	0.9996	0.9996
RMSE	0.0169	0.0207	0.0207
MAE	0.0003	0.0004	0.0004
MAPE (%)	0.03%	0.05%	0.05%
R^2	0.9998	0.9997	0.9997
Yield Prediction Metrics			
RMSE	0.1552	0.1621	0.1583
MAE	0.0874	0.0928	0.0914
MAPE (%)	20.36%	21.54%	20.95%
R^2	0.9790	0.9771	0.9782
Fertilizer Prediction Metrics			
RMSE	0.0573	0.0574	0.0322

MAE	0.0296	0.0329	0.0251
MAPE (%)	33.38%	24.44%	18.70%
R^2	0.9962	0.9962	0.9988

Overall, XGBoost's default configuration is effective for simpler tasks like crop classification and yield prediction, while tuning is crucial for enhancing performance in more demanding tasks like fertilizer forecasting.

6.3 Experiment / Case Study 3

The MLP meta model exhibited task-specific variations in performance across Default, Basic Hyperparameter Tuning, and Grid Search Optimized configurations as illustrated in Table 8.

Table 7: Evaluation metrics for the MLP Meta-Model

Metrics	**MLP Default**	**MLP Basic Hyperparameter Tuning**	**MLP Grid Search Optimized Hyperparameters**
Crop Classification Metrics			
Accuracy	0.9996	0.9996	0.9996
RMSE	0.0207	0.0415	0.0415
MAE	0.0004	0.0009	0.0009
MAPE (%)	0.05%	0.09%	0.09%
R^2	0.9997	0.9989	0.9989
Yield Prediction Metrics			
RMSE	0.1772	0.1623	0.1628
MAE	0.0932	0.0892	0.0912
MAPE (%)	20.26%	19.91%	21.23%
R^2	0.9726	0.9770	0.9769
Fertilizer Prediction Metrics			
RMSE	0.0459	0.0298	0.0294
MAE	0.0274	0.0210	0.0225
MAPE (%)	30.73%	25.96%	18.66%
R^2	0.9976	0.9990	0.9990

In the crop classification assignment, all models offered comparable results with the Grid Search optimized model offering a 99.96% accuracy, 0.0415 RMSE and 0.09% MAPE. From the confusion matrix, it is evident that the Grid Search Optimized configuration has very low misclassification level, which proves its accuracy and suitability for classifying samples

When it comes to yield prediction, Grid Search Optimized MLP was slightly better than the Default with the RMSE being 0.1628, MAE of 0.0912 and MAPE of 21.23%. Although these differences are small, the fact that the model had this performance suggests that it is well-suited for regression tasks, even on complex data

Out of all the models, the Grid Search Optimized model was the best model for fertilizer prediction that had the lowest RMSE of 0.0294, MAE of 0.0225, and MAPE of 18.66 % with the highest R^2 of 0.9990. One of the key aspects that can be derived from the plot of predictions vs. actual for the fertilizer prediction is that the model shows a near perfect fit for the various patterns present in the fertilizer data.

These results confirm that Grid Search Optimized MLP Meta Model is the most preferred due to its relatively lower errors and the highest R^2, especially in a regression analysis. Although the hyperparameter tuning did not improve the classification performance much, the regression tasks such as fertilizer prediction benefited a lot from this tuning. From these findings, it is clear that there is room for further improvement when it comes to hyperparameter optimization in context of the MLP Meta Models.

The evaluation visualizations show the effectiveness of the MLP model on different tasks. The Figure 6 shows the confusion matrix which makes it very clear how well it has classified the crops for classes like Sugarcane and Maize with high accuracy while misclassifications were fewer for Wheat and Barley classes. Furthermore, the Predicted vs Actual Fertilizer Scatter plot also shown in Figure 6 defines an excellent correlation between the predicted and actual values of fertilizer since most points lie on or very close to the line named “ideal fit” representing well-estimated fertilizers.

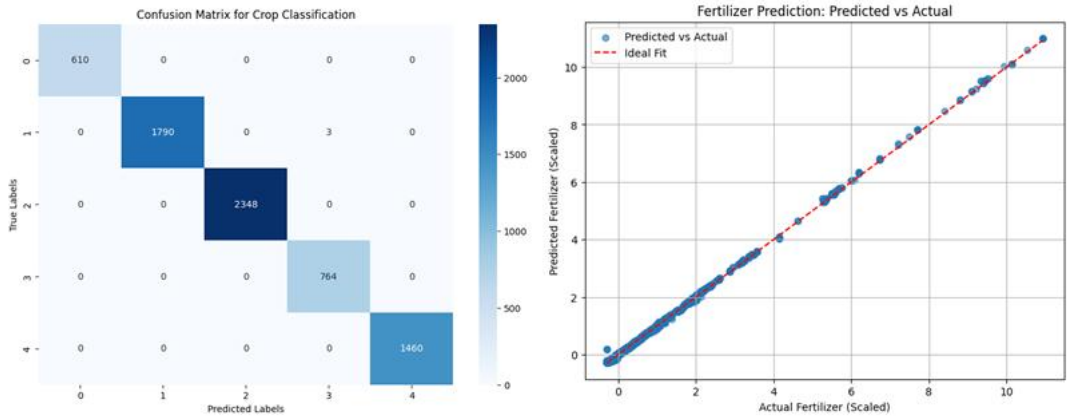

Figure 6: Confusion matrix for the Crop classification and Fertilizer prediction for MLP Meta-Model

6.4 Discussion

As seen the model was trained and tested with default hyperparameters for SVR achieving an RMSE of 0.1642 and MAPE of 20.35% showed that it was effective in predicting the yield. However, in prediction of fertilizer, the results were slightly lower with RMSE of 0.0590 and MAPE of 21.59%. Hyperparameter tuning slightly increased the performance; nevertheless, the grid search parameter optimization further enhanced the outcome of the fertilizer, with an RMSE of 0.0386 and R^2 of 0.9983. Poor performance of SVR in relation to other benchmarks are attributed to its high fluctuations with respect to data variance and inability to capture non-linearity in utmost cases. While SVR was sufficient for basic operations, it failed to provide the best solutions when it came to more complex predictions (Rashid et al., 2021).

The performance of the XGBoost meta model remained relatively stable across almost all analyses, specifically in the fertilizer prediction where, Grid Search Optimized achieved RMSE of 0.0322 and MAPE of 18.70%. For yield prediction, it achieved an RMSE of 0.1583 and MAPE of 20.95%, which are ideal for regression-based models. Nonetheless, in crop classification specifically, its performance was not as commendable as that of the MLP model, achieving only 99.96% accuracy. Such outcomes verify the effectiveness of XGBoost with tasks involving regression analysis (Sitienei et al., 2023), especially when

hyperparameters are appropriately tuned. However, since XGBoost is built around the decision tree, its capacity to capture complex interactions is more limited than in the case of neural networks; therefore, it simply cannot outperform every task.

Finally, in terms of the comparative evaluation, the MLP meta model turned out to give the highest performance and the best results for all of the tasks. For crop classification, it provided the output with an accuracy of 99.96% along with a very low level of confusion as seen in the confusion matrix. Also, in yield prediction, the Grid Search Optimized configuration yielded the lowest RMSE of 0.1628, and the lowest MAE of 0.0912, with R^2 of 0.9776, meaning the model was producing predictions closer to the actual yield. For fertilizer prediction, MLP recorded the lowest RMSE of 0.0294, MAE of 0.0225, and MAPE of 18.66%. The difference between the predicted vs. actual values is even more evident when plotted on a graph implying a perfect fit for fertilizer data. These findings reaffirm that MLP is capable of learning and mimic non-linear and intricate dependencies, which makes it the most flexible and suitable model in this experiment. MLP performs better than other models specifically in carrying out complicated tasks such as fertilizer prediction (Archana & Saranya, 2020). Nevertheless, MLP is not without its limitations, especially, the high computational costs need during the training of the network especially when the hyperparameters are involved

6.4.1 Integration of Results and Theoretical Implications.

Supporting models like Random Forest, Gradient Boosting, LSTM, and meta-learning models (SVR, XGBoost, MLP) are also used in this research. Among them, the MLP meta model showed the greatest overall potential because it provided the best results across tasks. The findings affirm the effectiveness of using mixed models and neural networks in considering higher-order relations and enhanced prediction capability (van Klompenburg et al., 2020). Previously done research is supported by these study that meta-learning and neural network models are applicable in large agricultural dataset datasets.

6.4.2 Implications and Limitations.

The results also highlight the real-world applicability of integrating multiple ML algorithms in crop identification, yield and fertilizer prediction. The dominance of the MLP meta model underscores the appropriateness of using neural networks for tasks that involve non-linear patterns. These insights can inform practitioners in designing decision support systems to optimize resource management which enhances farming sustainability. However, since the data set is specific to a region the results may not be fully generalizable to other locations or weather conditions. The challenge with training and tuning these models such as MLP is that it is expensive in terms of computational complexity especially in resource constrained settings. While hyperparameter tuning has enhanced performance, other sophisticated optimization methods should be examined further to enhance efficiency.

7 Conclusion and Future Work

From this study we can conclude that machine learning methods allow for the improved and efficient prediction of crop types, yield, and fertilization which are all important factors in effective farming practices. Training base models including Random Forest, Gradient Boosting, and LSTM in conjunction with meta-models like SVR, XGBoost, and MLP helped in improving performances across all the predictive classes. MLP meta-model was the most accurate in crop classification with an accuracy of 99.96% with a lowest RMSE of 0.0294 and MAPE of 18.66% in the fertilizer prediction process after hyper-parameter tuning. Overall, it was observed that MLP was learning the data well and was capable of capturing nonlinear relationships, which making it the best and most flexible model. Moreover, the results from the study revealed that XGBoost performed better in regression tasks, especially in the case of fertilizer prediction, which proved that it can effectively work with non-linear data dependencies.

Although the hybrid model was proved to have high efficiency in this paper, some issues were also found, such as the complexity of computation due to usage of several base and meta-models. Handling these issues in future research such as enhancing model design to avoid high computational complexity while still obtaining precise probability estimates will be appropriate. Expanding the dataset to cover other regions, crops, and weather conditions will improve models in terms of generalizability and versatile.

In order to capture spatial correlations more effectively in crop and yield data, future work could explore advanced architectures such as Convolutional Neural Networks (CNNs). The efficiency and performance of the models can be enhanced furthermore with the help of automated optimization methods such as Bayesian optimization or AutoML for superior tuning. These methods will enhance scalability and applicability of the model in real-life farming environments. Such techniques will improve the ability to scale and apply the model within real-world farming practices.

In general, these study findings assist in providing valuable data that aids in the improvement of superior agricultural practices. From the results obtained from this study, the various stakeholders in the farming industry will be in a position to enhance productivity through the use of hybrid machine learning models. Such outcomes provide some guidance towards the further development of more innovations practices in farming especially by harnessing the potential of data in farming.

References

Abbas, F., Afzaal, H., Farooque, A.A. and Tang, S. (2020). Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms. Agronomy, 10(7), p.1046.

Agarwal, S. and Tarar, S. (2021). A HYBRID APPROACH FOR CROP YIELD PREDICTION USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS. Journal of Physics: Conference Series, 1714, p.012012.

Anbananthen, K.S.M., Subbiah, S., Chelliah, D., Sivakumar, P., Somasundaram, V., Velshankar, K.H. and Khan, M.K.A.Ahamed. (2021). An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms. *F1000Research*, 10, p.1143.

Archana kumaravel (2020). *Crop yield prediction, forecasting and fertilizer recommendation using Data mining algorithm*. [online] International Journal of Computer Science Engineering (IJCSE).

Ashima Kukkar, Rajni Mohana, Sharma, A., Mallik, S. and Shah, M.A. (2024). AgroAdvisor: Crop Yield Prediction, Crop and Fertilizer Recommendation System using Random Forest with Gradient Boosting and DeepFM for Precise Agriculture. *Research Square (Research Square)*.

Attar, N.M., Shah, S.K., Vijayatai Hukare, Vidya Kumbhar and Singh, T.P. (2024). Optimizing Fertilizer Usage using Machine Learning Techniques. [online] doi:https://doi.org/10.1109/mitadtsocicon60330.2024.10575396

Ayalew, A.T. and Lohani, T.K. (2023). Prediction of Crop Yield by Support Vector Machine Coupled with Deep Learning Algorithm Procedures in Lower Kulfo Watershed of Ethiopia. *Journal of Engineering*.

Batool, D., Shahbaz, M., Shahzad Asif, H., Shaukat, K., Alam, T.M., Hameed, I.A., Ramzan, Z., Waheed, A., Aljuaid, H. and Luo, S. (2022). A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning. *Plants*, 11(15), p.1925.

Bondre, D.A. and Mahagaonkar, S. (2019). PREDICTION OF CROP YIELD AND FERTILIZER RECOMMENDATION USING MACHINE LEARNING ALGORITHMS. *International Journal of Engineering Applied Sciences and Technology*, 04(05), pp.371–376.

Jaladi Lahari, Vadlamudi Lavanya and K.B. ManikandanAssistant (2024). Crop Prediction to Aid Farmers using Hybrid Model. *2024 8th International Conference on Inventive Systems and Control (ICISC)*, [online] pp.705–710.

Kumar, M., Saini, P., Kumar, V. and Jain, R. (2024). Digital Cropping: A Comparative Analysis of Hybrid CNN-XGBoost With Stacking & Ensemble Techniques. *2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP)*

Lendy Rahmadi, None Hadiyanto, Ridwan Sanjaya and Arif Prambayun (2023). Crop Prediction Using Machine Learning with CRISP-DM Approach. *Lecture notes in networks and systems*, pp.399–421

M. Abdipourchenarestansofla and Piepho, H.P. (2022). *Yield Observation Outlier Detection with Unsupervised Machine learning in harvest machines*.

Mahmoud Abdel-salam, Kumar, N. and Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. *Neural Computing and Applications*

Malashin, I., Vadim Tynchenko, Andrei Gantimurov, Vladimir Nelyub, Aleksei Borodulin and Yadviga Tynchenko (2024). Predicting Sustainable Crop Yields: Deep Learning and Explainable AI Tools. Sustainability, [online] 16(21), pp.9437–9437.

Miqdad, A., Prima Kristalina and Aries Pratiarso (2024). Hybrid LSTM and SVM Method Rice Yield Prediction in Densely Populated Areas. *2024 International Electronics Symposium (IES)*, [online] pp.510–514

N, V.D. and Choudhary, S. (2024). The new machine learning feature selection method used in fertilizer recommendation. *Bulletin of Electrical Engineering and Informatics*, [online] 13(5), pp.3455–3462.

Oikonomidis, A., Catal, C. and Kassahun, A. (2022). Hybrid Deep Learning-based Models for Crop Yield Prediction. *Applied Artificial Intelligence*, pp.1–18.

Parashar, K. (2024). Crop Prediction and Fertilizer Recommendation Using Regression Algorithms. *International Journal for Research in Applied Science and Engineering Technology*, 12(3), pp.2270–2273

Rashid, M., Bari, B.S., Yusup, Y., Kamaruddin, M.A. and Khan, N. (2021). A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction. *IEEE Access*, 9, pp.63406–63439

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J. and Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406, pp.109–120

Shahhosseini, M., Hu, G. and Archontoulis, S.V. (2020). Forecasting Corn Yield With Machine Learning Ensembles. *Frontiers in Plant Science*, 11

Sharma, A.K. and Rathore, A.S. (2024). Design and Implementation of a Cloud-Based Smart Agriculture System for Crop Yield Prediction using a Hybrid Deep Learning Algorithm. *Current Agriculture Research Journal*, [online] 12(2), pp.714–725.

Sitienei, M., Otieno, A. and Ayubu Anapapa (2023). An Application of K-Nearest-Neighbor Regression in Maize Yield Prediction. *Asian Journal of Probability and Statistics*, 24(4), pp.1–10

van Klompenburg, T., Kassahun, A. and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, p.105709

Waoo, A.A. and Tiwari, M. (2024). OPTIMIZING CROP YIELD AND FERTILITY MANAGEMENT USING KNN AND ML. *ShodhKosh Journal of Visual and Performing Arts*, [online] 5(5), pp.380–385.