# Improving Crop Yield Forecasting with Hybrid CNN-GRU and ARIMA-GRU Models

MSc Research Project
MSc Data Analytics


Sameer Ahamed Sahul Hameed
Student ID: 23326751

School of Computing
National College of Ireland


Supervisor: Prof. John Kelly

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

| | |
|---|---|
| **Student Name:** | ……. SAMEER AHAMED SAHUL HAMEED …… |
| **Student ID:** | …… 23326751 …… |
| **Programme:** | …… MSc Data Analytics ……………………… **Year:** 2024 - 2025 |
| **Module:** | ……… MSc Research Project …………………….……… |
| **Supervisor:** | ……… Prof. JOHN KELLY ………. |
| **Submission Due Date:** | …… 11/08/2025 …… |
| **Project Title:** | Improving Crop Yield Forecasting with Hybrid CNN-GRU and ARIMA-GRU Models |
| **Word Count:** | ……… 7528 ……………… **Page Count**…… 20 …….……. |

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:** ………Sameer Ahamed Sahul Hameed………………………………

**Date:** ………… 11/08/2025 ………………………………………

**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): | |

# Improving Crop Yield Forecasting with Hybrid CNN-GRU and ARIMA-GRU Models

Sameer Ahamed Sahul Hameed
x23326751

## Abstract

Accurate forecasting of crop yields is essential in proper agricultural planning and resource distribution, as well as food security. This study provides a detailed modeling framework in which standard regression methods are combined with deep learning architectures to predict crop yield using environmental, agronomic, and climatological factors. The analysis used a well-organized pipeline, which started with the preprocessing of data and exploratory data analysis, then feature engineering and model development. A wide range of models: Lasso Regression, K-Nearest Neighbors, Decision Tree, Deep Neural Networks (DNN), Long Short-Term Memory (LSTM), CNN-GRU, and a new hybrid model of ARIMA-GRU were trained and tested on the standard regression metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Coefficient of Determination ($R^2$). The findings revealed a definite trend in the performance with the traditional models providing a baseline accuracy and deep learning models providing much better prediction ability. The ARIMA-GRU hybrid model performed best among all of them with an $R^2$ score of 0.986 and the least MAE and MSE and was able to capture not only the linear trends but also the nonlinear trends in the data of crop yield. This work finds that hybrid modeling methods provide a promising way to tackle complex problems in agricultural forecasting and that integrating statistical inference with deep learning has a great potential to assist intelligent and data-driven agricultural systems.

# 1 Introduction

## 1.1 Background

The world of agriculture is becoming more data-oriented, with complex analytics and predictive modeling being used to help make decisions, manage risks and optimize resources. Crop yield forecasting is one of the most important analytical tasks in agriculture. Consistent forecasts of yields can facilitate front-foot action to food security risks, price fluctuations, and climate uncertainty (Pant et al., 2021). In a world of climate variability, population increase, and geopolitical turmoil, there has been a need to enhance the precision and stability of the yield forecasting systems in order to achieve food security and sustainable growth.

Statistical time series analysis has been traditionally used as the basis of crop yield forecasting. Linear relationships and time trend patterns have been modeled via models like the Auto-Regressive Integrated Moving Average (ARIMA) using historical crop and weather data. They are effective when the assumption is stationarity and linearity when future values may be predicted by using a function of the past observations (Elsamae et al., 2021). But, agricultural systems are complicated, non-linear, and subject to a multifaceted variety of factors, such as rainfall, temperature, solar radiation, soil moisture, pest activity, and farming practices. Such complicated dependencies cannot be described using only linear models (Patrick et al., 2024).

The latest updates in deep learning and neural networks architectures have made available more advanced tools to model more complex systems. In particular, RNNs and its variations such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks could be trained to acquire long-range dependencies and temporal dynamics of sequential data. At the same time, Convolutional Neural Networks (CNNs) that were originally designed to learn spatial features on image data have been re-applied to time series by learning temporal patterns over time windows (Venkatnaresh et al., 2024). Nevertheless, each of these models, traditional and modern, is limited in spite of their personal strengths. Deep learning models need large amounts of data and can be sensitive to hyperparameter tuning and are very susceptible to overfitting in low-data regimes. In contrast, statistical models such as ARIMA are unable to learn nonlinear interactions and could be unable to generalize well in the presence of structural changes in climate or crop systems. The observations prompt the necessity of hybrid modeling approaches that enable integrating the advantages of both statistical (interpretability, noise-robustness, seasonality modeling) and deep learning models (flexibility, nonlinear learning, feature representation) (Slater et al., 2022).

This research study develops two such hybrid models: (1) a deep learning CNN-GRU model, and (2) an ARIMA-GRU hybrid model, both with the aim to increase fidelity and adaptability of crop yield predictions made based on time series weather and agronomic data.

## 1.2 Problem Statement

Despite the progress achieved in the existing crop yield forecasting models in terms of accessing the historical climate/agronomic data, they still face severe constraints. Statistical methods, although successful in modeling trend and seasonality, are too simplistic in the processes they model and do not capture nonlinear interactions and abrupt changes in the environment. Conversely, data-hungry, computationally demanding, and usually problematic in terms of interpretability and generalizability, purely deep learning-based models are also powerful (Abhishek et al., 2023). The issue, hence, is that the existing single-method models are unable to robustly model linear and nonlinear dependencies in agro-meteorological time series data. Moreover, most of the current methods do not involve rigorous comparison of their models, use limited performance measures or otherwise lack a robust validation protocol that is appropriate to the seasonal fluctuations of the data.

To fill this gap, the study proposes and compares two new hybrid forecasting models: a CNN-GRU model that capitalizes on local and sequential temporal characteristics and an ARIMA-GRU model that combines the statistical and neural learning to improve the residuals. It is believed that these hybrid architectures will perform much better in terms of accuracy, flexibility, and robustness of the forecasting process than traditional and stand alone deep learning approaches.

## 1.3 Research Objectives

The main aim of this research is to enhance the performance and robustness of crop yield prediction with hybrid deep learning models with statistical and neural architecture. Particular goals are:

- To plan and construct a CNN-GRU hybrid deep learning architecture capable of learning local temporal features using convolutional layers and learning sequential dependencies using gated recurrent units.
- To create an ARIMA-GRU hybrid model in which the ARIMA component reflects the linear trend and seasonal patterns, and the GRU component reflects the remainder non-linear elements driven by agro-meteorological factors.
- To thoroughly test and compare the hybrid models with standalone classifical models and deep learning models (Lasso, KNN, Decision Treee, DNN, LSTM) with multiple performance measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and the Coefficient of Determination ($R^2$).
- To use strong model validation procedures such as cross-validation, out-of-sample testing, so that model results are reproducible over time and environmental changes.

In order to show how the proposed hybrid methods can be scaled up and generalized to other crops, other regions and time periods.

## 1.4 Research Question

According to the aims and the context of this research work, the following central research question is formulated:

**How will the use of hybrid deep learning models, namely the CNN-GRU model and the ARIMA-GRU model, enhance crop yield predictions based on time series data compared to existing methods and single deep learning models?**

We suppose that both of the models will target certain weaknesses of the traditional and standalone deep learning models:

- The CNN-GRU model will improve the temporal pattern recognition through localized temporal features extraction (CNN) and long term dependency modeling (GRU).
- The ARIMA-GRU model will be a middle-ground solution as it will capture the deterministic linear patterns with the help of ARIMA and use GRU to represent the more complicated residual patterns.

These models in combination will provide a more accurate, reliable and interpretable yield prediction of crops, particularly when there are climate-induced anomalies and seasonal variations.

## 1.5 Significance of the Study

The proposed study will make theoretical and practical contributions to the sphere of agricultural informatics and predictive analytics.

**Theoretical Contributions:**

- Shows innovative combination of deep learning and traditional statistical models of agricultural forecasting.
- Investigates relative advantages of hybrid neural architectures in modelling real world time series data.
- Offers a replicable methodology which can be extrapolated to other areas with comparable data properties.

**Practical Implications:**

- Helps farmers, policymakers and agribusinesses make better and more timely yield forecasts.
- Improves the predictability of food insecurity early warning systems particularly in the climate-sensitive areas.
- Enables greater resource (fertilizer, irrigation, labor) planning in the presence of anticipated variability of output.

Overall, the hybrid models presented in this paper not only contribute to the academic research in the field of predictive modeling but will also be practically applicable to make agriculture more adaptive, data-driven, and sustainable.

### 1.6 Structure of the Research Study

The following is the organization of the research study:

- **1: Introduction** - Presents the research problem, objectives and significance.
- **2: Literature Review** - Makes an in-depth survey of existing techniques of crop yield prediction, paying special attention to classical and deep learning models, and places proposed hybrid methods in this context.
- **3: Methodology** - Describes the source of the data, the preprocessing, the architectures of the models (CNN-GRU and ARIMA-GRU) and the evaluation process.
- **4: Results and Discussion** - Provides experimental data, comparative studies, and discusses the results with the background of the available literature and model behavior.
- **5: Conclusion and Future Work** - Summarizes the contributions, talks about limitations, and presents directions to future research and possible improvements and applications of the models.

## 2 Literature Review

### 2.1 Introduction

Food security management and agricultural planning cannot be complete without crop yield forecasting. It is possible to plan and allocate resources better, regulate the market, and mitigate risks of climatic uncertainties due to the possibility of accurate prediction of yields before harvest. In recent decades, a wide variety of modeling strategies have been used to perform this task, and include both statistical time series modeling and cutting-edge deep learning models. All classes of models have helped in a significant way, but none of them is able to encompass the intricacy of agro-meteorological systems (Badshah et al., 2024; Kowalska et al., 2023). This chapter is a review of the available literature on the topic of crop yield forecasting with emphasis on three main areas namely the traditional statistical methods, deep learning models and the hybrid models. The review places the research in the context of these two bodies of works and explains the reasons why it is necessary to integrate convolutional neural networks (CNNs) with gated recurrent units (GRUs) and combine ARIMA and GRU to enhance the performance of forecasting.

## 2.2 Traditional vs. Machine Learning Methods

In the past, statistical and econometric models, especially those built on linear regressions and seasonal decomposition or models based on time series like ARIMA have been used to make crop yield prediction. The approaches provide interpretability and are appropriate in modeling univariate time series that have stable seasonal patterns. Their fundamental weaknesses however, are that they are linear and stationary, and cannot combine multivariate inputs of various agricultural and environmental sources. The traditional models are usually used in a way that they consider each region or crop separately, thus they have limited ability to make generalizations or transfer learning across geographies. Moreover, they have difficulty incorporating heterogeneous data types including remote sensing indices, soil properties, and satellite-based vegetation measures-data that are becoming more and more critical to real-time spatially conscious forecasting. The shortcomings of the purely statistical methods have contributed to the increasing use of the machine learning methodologies. In contrast to the traditional models, ML algorithms can deal with the high-dimensional input, learn complex nonlinear relationships, and can deal with multi-source data. (Paudel et al. 2021) showed the effectiveness of ML in this regard by creating modular pipeline consisting of weather variables, outputs of crop simulation, remote sensing features, and soil information. Their utilization of such models as ridge regression, support vector regression (SVR), k-nearest neighbors (KNN), and gradient boosting demonstrated better accuracy than the baseline historical averages. Notably, research focused on strict feature engineering in line with crop phenology, and the need to prevent data leakage both of which are essential to real-world forecasting accuracy.

## 2.3 Deep Learning in Time Series Modeling for Yield Forecasting

Early uses of deep learning on crop yield prediction are mixed. As an example, Guo and Xue (2014) compared spatial and temporal neural networks and obtained that spatial neural networks could perform better than time series based models such as NARNN and NARXNN in some cases. Nevertheless, they warned against naive use of neural networks, noting that they were prone to outliers and that they did not generalize well outside training window. The results of their research were that deep learning is to be implemented with sufficient validation methods and where feasible complemented by exogenous data inputs. To some degree, these shortcomings have been addressed with the recent advances in model architecture and training methodology.

One of the most impressive experiments was conducted by (Paudel et al. 2023) who compared LSTM models directly with 1D CNNs and powerful gradient boosted decision tree (GBDT) baseline. They discovered that LSTM models were equal or better than GBDT in a variety of applications, including soft wheat yield prediction in Germany, that recurrent neural networks could learn intricate temporal relationships in raw input sequences. Also, attribution methods showed that deep learning models were able to learn meaningful time-based features, as opposed to merely overfitting to noise. Additional evidence of the strength of deep architectures was given by El-(Kenawy et al. 2024) in complementary work. The fact that they benchmarked conventional ML algorithms such as XGBoost, KNN, and gradient boosting with deep models, MLP, LSTM, GRU, and GNNs revealed that GNN produced the highest R2 scores, whereas GRU and LSTM models produced the lowest MSEs. The results highlighted the ability of deep models to learn spatio-temporal relationships well, automate feature extraction, and also surpass conventional ML algorithms when working in challenging agricultural settings.

## 2.4 Convolutional Neural Networks for Time Series Forecasting

Convolutional neural networks (CNNs) were designed to process images, but recent studies have modified CNNs to work with time series. In this regard, 1D CNNs use convolutional filters in the time axis, which explains why they can identify short-term patterns or local features in the data. Such characteristics can be high and sudden temperature, a brief spell of drought, or spotty rainfall bursts, which can affect crop growth stages. CNNs excel in detecting patterns in fixed-size windows and are able to extract features traditional recurrent models may miss. Coupled with RNNs or GRUs, the CNNs become temporal feature extractors with the condensed local patterns passed on to the recurrent units that deal with longer-range interactions. Such a layered solution provides a more sophisticated view of the multi-scale character of time series data. An example is (Gong etl a., 2024) which applied a CNN-LSTM hybrid to enhance short-term weather forecasting. They could learn local variations and time evolution, and their model could outperform standalone LSTM models. Although their study was not directly related to agriculture, their approach can be extremely applicable in crop yield forecasting since short-term weather events play a critical role in the final output (Vallileka et al., 2025). Nevertheless, CNN-based models have not been thoroughly studied in the agricultural yield prediction, especially when combined with GRUs. This gap opens up a possibility to explore the potential of a CNN-GRU model to achieve better performance through the synergetic capabilities of both models to process complex and noisy agro-climatic datasets.

## 2.5 Hybrid Forecasting Models

This concept of integrating statistical models and neural networks is based on the realization that in real world time series data, often linear and nonlinear patterns co-exist. The hybrid models seek to take both by dividing the forecasting process into two complementary steps. Usually, the linear component of the series is forecasted first using a statistical model such as ARIMA. The errors generated by this model are then fitted by a deep learning network which tries to learn the nonlinear trends that are not modelled by the ARIMA. The hybrid ARIMA-GRU model is based on this paradigm. Following the main linear modeled structure using ARIMA, the rest of the residuals are fed into a GRU network. The results of the ARIMA and the GRU models are added to get the final yield forecast. It has been effectively applied in other fields like energy and hydrology. To illustrate, (Waqas et al., 2024) predicted the daily rainfall, and the results were more accurate and robust compared to standalone models using the ARIMA-GRU model. There are also hybrid architectures in deep learning frameworks alone. One such instance is CNN-GRU models, in which CNN layers learn local temporal features of raw time series data, and GRU layers learn the sequential dependencies between those features. (Zhou et al. 2025) proved that this architecture is effective in the prediction of energy demand. Such a model, when used in crop yield forecasting, would be able to learn short-term anomalies (using CNN) and long-term growing season dynamics (using GRU), which could outperform models using only one of the two architectures.

## 2.6 Comparative Analysis and Identified Research Gaps

Table 1: Comparative Analysis of Research Gaps

| Model Type | Strengths | Limitations |
|---|---|---|
| ARIMA | Effective for linear trends and seasonality | Cannot model nonlinearities; poor with exogenous data |
| GRU/LSTM | Good at nonlinear sequential modeling | Prone to overfitting; requires large datasets |
| CNN | Excels at detecting local patterns | Limited in modeling long-term dependencies |
| ARIMA-GRU | Combines trend detection with nonlinear learning | More complex; tuning both components is non-trivial |
| CNN-GRU | Learns both local and sequential temporal features | Computationally expensive; risk of overfitting |

The current literature has identified some of the major gaps, which this study aims at filling. To begin with, although ARIMA and GRU models have been used separately in agricultural forecasting, there is still no indication of the combination of the two models in an ARIMA-GRU hybrid framework to predict crop yield in particular. The majority of hybrid research is on the energy or financial sectors where time series behaviour is not the same as agriculture.

Second, despite the success of CNN-GRU hybrids in other fields, their use with agro-meteorological data is rare. Limited literature exists on how convolutional and recurrent structures may be integrated to deal with the multi-scale, non-stationary, and nonlinear character of agricultural data.

Third, a great number of the existing studies lack a strict comparative framework or validation procedure. Others use only one measure, e.g. mean squared error, and test over short periods of data. This undermines the validity and generalizability of the results especially when it comes to agricultural settings where the seasonal effects are a major variable. Lastly, research that aims at the explainability and interpretability of these models is lacking, and it is necessary to build trust and adoption by stakeholders in agricultural planning.

## 2.7 Summary

This chapter has discussed how the method of crop yield forecasting has evolved over time where traditional statistical models such as ARIMA were replaced by more complex deep learning models such as GRU and CNN. Although each method has its particular benefits, it is also afflicted by crucial shortcomings. ARIMA models can be quite useful at modelling linear and seasonal trends, but are not useful at modelling nonlinear relationships and multivariate inputs. GRUs can learn complex sequential dependencies and fail to capture short-term anomalies. CNNs are good at learning localized temporal patterns, but in isolation they are not good at modeling sequence dynamics. Hybrid models, especially the ARIMA-GRU and CNN-GRU models, come out as the potential solution that combines the best of both worlds. These hybrids have not been well exploited in the agricultural sector despite their potential. The present research study is an attempt to fill this gap and develop, execute, and test CNN-GRU and ARIMA-GRU models to achieve better results in crop yield forecasting.

# 3 Methodology

## 3.1 Introduction

This chapter provides a description of the methodological approach followed in predicting crop yield based on a combination of the conventional regression methods and the state of the art deep learning models. The methodology was systematic and it started with data preprocessing and exploratory data analysis (EDA), feature engineering and transformation, model building and testing, and finally the results were explained using detailed visualizations. The main goal was to find the best predictive model to estimate crop yield using the links between the environmental, agronomic, and climatological variables.

## 3.2 Data Collection and Preparation

The data employed in this research was in a systematic form (CSV) and it had a number of variables which might affect the crop yield. These characteristics were a mixture of numerical and categorical data like rainfall, temperature, days to harvest, type of crop, type of soil and agricultural activities like use of fertilizers and irrigation as each observation variables are **10M rows**. Yield_tons_per_hectare was the variable of interest being the dependent variable, it was the outcome variable of the yield as the features are demonstrated in Table 2.

Initial inspection of the dataset was the first step in data handling to know the structure of the dataset and its content. Functions were used to obtain the overview of data, its types, and existence of missing or abnormal values. In the process, it was observed that some of the categorical variables needed encoding in order to be applicable in model training. As a remedy to this, label encoding was used on columns like Region, Soil_Type, Crop, and Weather_Condition, and converted to numeric form but with ordinal consistency where possible.

Table 2: Attributes Feature Variables of Crop Yield Forecasting Dataset

| Feature | Type | Description |
|---|---|---|
| Region | Categorical | Geographic zone of cultivation |
| Soil_Type | Categorical | Soil classification (Loamy, Sandy, etc.) |
| Crop | Categorical | Type of crop grown |
| Rainfall_mm | Numerical | Seasonal rainfall in millimeters |
| Temperature_Celsius | Numerical | Average seasonal temperature |
| Fertilizer_Used | Boolean | Binary indicator for fertilizer use |
| Irrigation_Used | Boolean | Binary indicator for irrigation |
| Weather_Condition | Categorical | General climatic conditions (Sunny, Rainy, etc.) |
| Days_to_Harvest | Numerical | Crop maturation period in days |
| Yield_tons_per_hectare | Numerical | Target variable - actual crop yield |

Variables that were Boolean such as Fertilizer_Used and Irrigation_Used were transformed into binary integers to ensure that there was consistency throughout the dataset. Outliers in numerical data of Temperature_Celsius, Rainfall_mm, Days_to_Harvest, and Yield_tons_per_hectare were particularly dealt with. The outliers were determined by a standard deviation approach, which would mark the values that were beyond two standard deviations away. Nevertheless, to maintain possible meaningful variations, only evidently inappropriate data like negative yields or zero rainfall in rainy areas were omitted. The output of the preprocessing step was a clean and coherent dataset ready to be used in model training with the richness and diversity that is required to train a robust model.

## 3.3 Exploratory Data Analysis (EDA)

After the preprocessing of the dataset, an extensive exploratory data analysis was carried out to identify any underlying pattern, trend, and relationship between the variables. The step was not only necessary to comprehend the structure and distribution of the data but also to inform the selection of the models and feature engineering choices. EDA was dependent on visualizations. The correlation between rainfall and crop yield was tested using scatter plots and this correlation was found to be extremely different based on the type of crop.

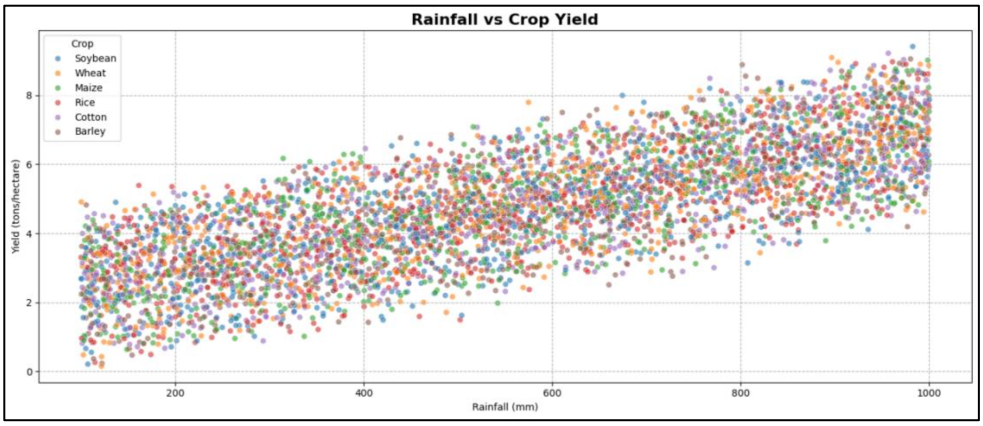


Figure 1: Distribution Scatter Plot between Rainfall and type or Crops

Box plots provided the information on the effect of various soil types and irrigation methods on yield results.

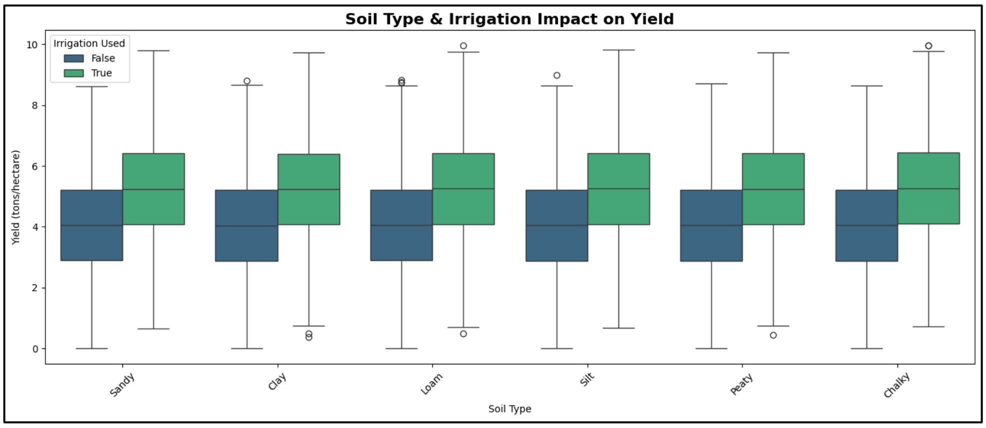


Figure 2: Distribution relationship between the Soil type and Irrigation Impact on Yield

Heatmap was created to see the distribution of average yield of different crops and regions and provided some geographical patterns that might affect productivity in agriculture.

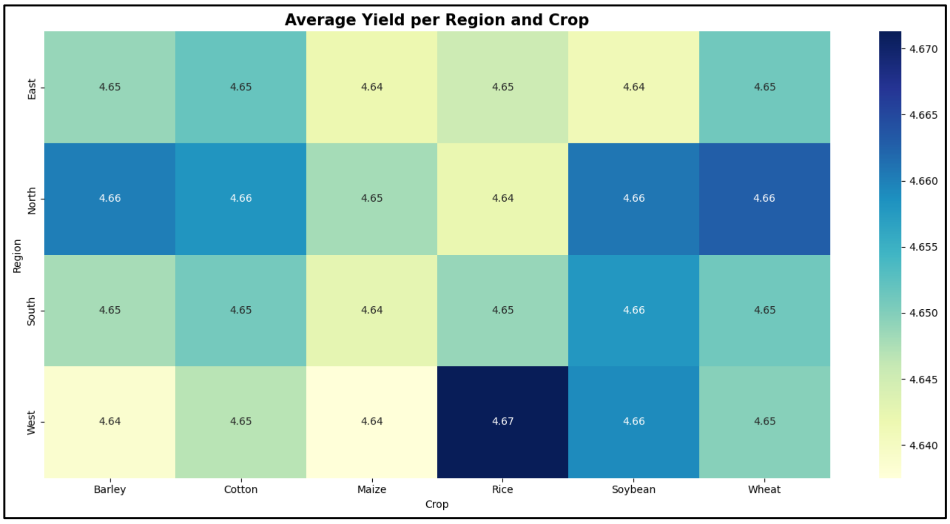

Figure 3: Average Yield per Region and Crop

Moreover, bar plots were used to show how the application of fertilizers and irrigation mechanisms led to differences in yield among the types of crops.

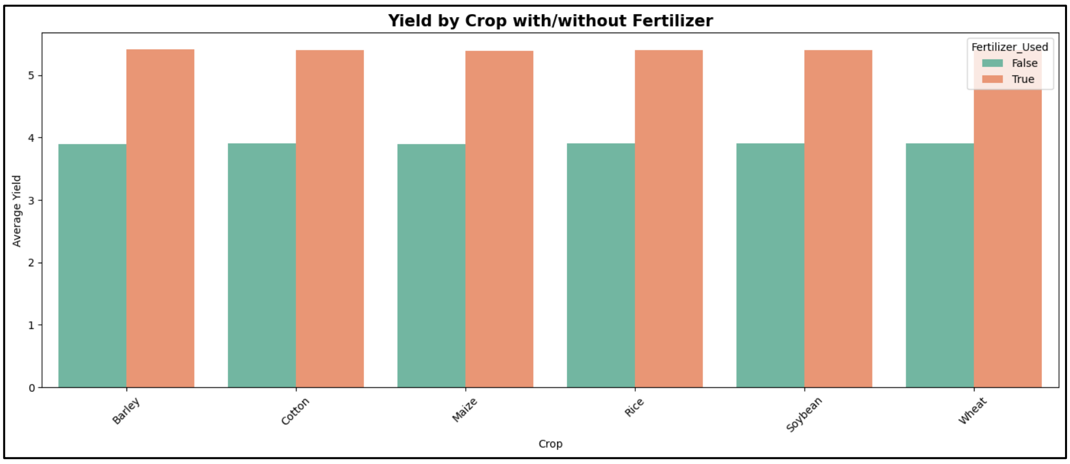

Figure 4: Distribution of Yield by Crop with/without Fertilizer

Correlation matrices were also formed to measure interdependencies between numerical variables, which aided in the determination of multicollinearity and possible feature redundancies.

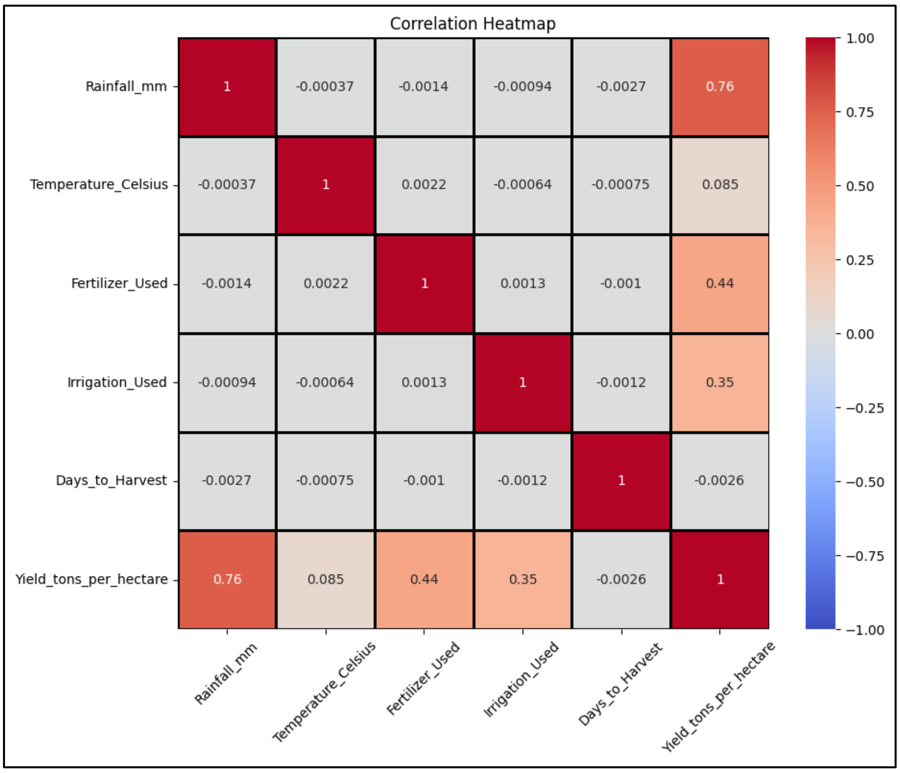


Figure 5: Correlation Heatmap for the features in the dataset

Histograms of temperature distribution and frequency charts of crops were also created to have a better insight into environmental conditions and cultivation trends.

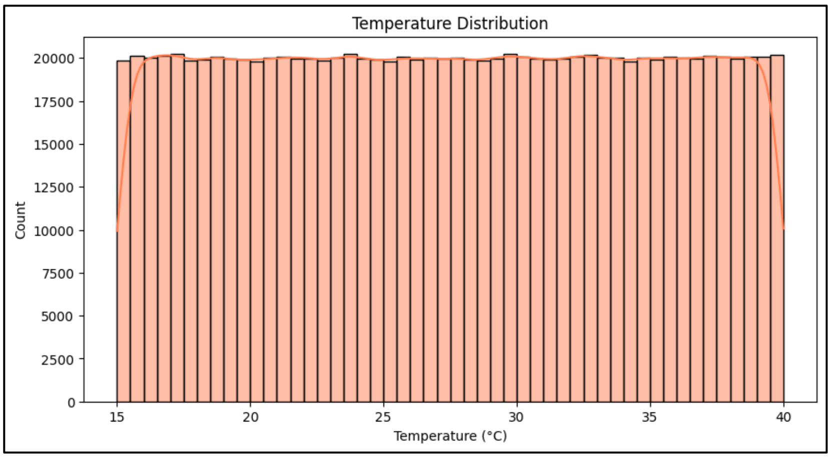


Figure 6: Distribution of Temperature

The presented visual and statistical analyses provided the foundation of the feature selection and guided the structure of the machine learning and deep learning models that were created in the next steps.

## 3.4 Feature Engineering and Transformation

After obtaining the insights according to EDA, the second step was to process the data to be used in machine learning algorithms. All the numerical values such as rainfall, temperature, and number of days to harvest were standardized in order to make each feature equally contributing in the training of the model via StandardScaler. Normalization was especially important in deep learning models, where input features have a tendency to be sensitive to scale. The LabelEncoder stage had been used to encode categorical variables. This was informed by the objective of ensuring interpretability of every feature of input given the agricultural setting where transparency in decision making is vital. The preprocessed dataset was composed of nine features, which taken together, were used as predictors of the target variable, crop yield. The data was further split into training and testing dataset in ratio of 80:20 so as to be able to evaluate the model sufficiently without overfitting.

## 3.5 Model Development

In order to determine the predictive capability of various methodologies, a diverse set of models were applied, including standard regression models, and modern deep learning models.

Lasso Regression, K-Nearest Neighbors (KNN) and Decision Tree Regressor were considered as traditional models. Lasso Regression was selected as it can both engage in regularization and feature selection which is useful in decreasing the overfitting by penalizing the less important variables. KNN is an easy to use but powerful algorithm that offered a non-parametric solution to learn the local patterns within the data. The Decision Tree Regressor was most suitable to model non-linear relationships and work with mixed data types, thus, it was a good candidate to be tested first.

In addition to these baseline models, deep learning architectures were created in order to capture more complicated and non-linear interactions between features. The initial deep learning model built was a fully connected Deep Neural Network (DNN) consisting of three hidden layers with 128, 64 and 32 neurons respectively. The ReLU activation was used in each layer and dropout layers were added to minimize overfitting. The model was trained with Adam optimizer and 10 epochs and a validation split to track the performance.

An LSTM (Long Short-Term Memory) network was also applied. Despite the LSTM models normally being used with sequential data, the architecture was modified to apply to this problem in order to test whether any underlying sequential structure might be identified in the crop data. The data was converted into 3-D form to match the LSTM specification and the model was trained as it was in the case of the DNN.

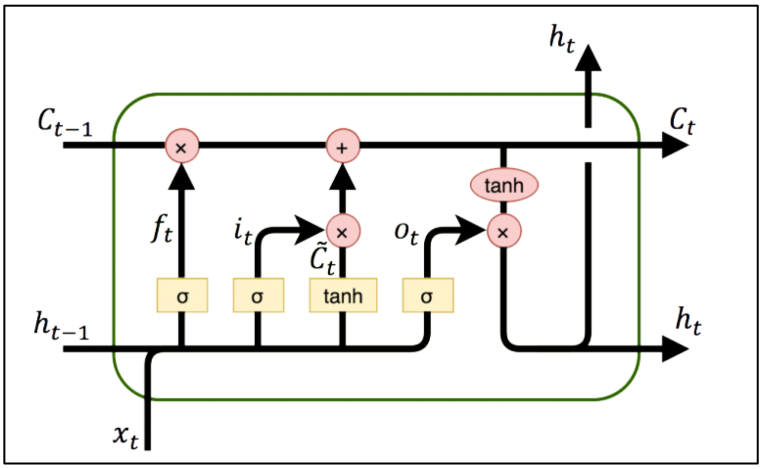


Figure 7: Architecture of the LSTM Model

As well, a prosed CNN-GRU (Convolutional Neural Network -Gated Recurrent Unit) model was designed to integrate the advantages of convolutional layers in feature extraction with the temporal learning ability of GRUs. The data was transformed and remodeled to fit 1D convolutional layers and the layers were stacked in sequence with batch normalization and dropout to have stable learning and generalization.

One of the distinctive features was to combine a statistical time series model and deep learning using this proposed ARIMA-GRU hybrid architecture. First, a smooth and sorted time series of crop yields was fitted to an ARIMA model. The remaining errors of this model were taken and fed into a GRU network, so the hybrid model would capture the linear (through ARIMA) and non-linear (through GRU) parts. The last forecast was obtained by adding the ARIMA forecast to the residuals predicted by GRU and provides a new ensemble approach to yield forecasting.

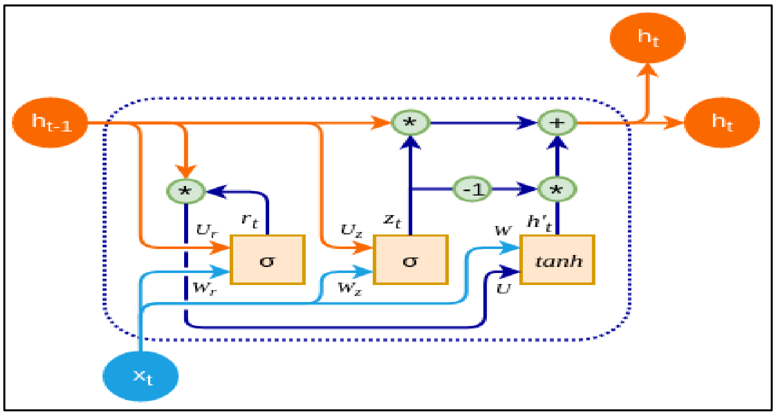

Figure 8: GRU Model Architecture

While the Long Short-Term Memory (LSTM) networks are widely recognized for their effectiveness in the modeling sequential data, the Gated Recurrent Unit (GRU) was selected in this study due to its computational efficiency and comparable predictive power. GRUs have a simpler architecture than LSTMs-using fewer gates and parameters-yet they are capable of capturing long-term dependencies in time-series data effectively. This transforms it to the particularly suitable for modeling complex agricultural sequences where the training time and computational resources are constrained. Empirical evidence from the prior research proposes that GRUs often perform on par with, or even better than, the LSTMs in tasks which involving nonlinear temporal dependencies with the relatively small or moderately sized type of datasets. By opting for the GRUs, the models in this research study benefited from faster convergence and reduced any risk of overfitting, where on the another side maintaining the high forecasting accuracy.

## 3.6 Model Evaluation

All the models were measured with similar parameters in the test set. The main evaluation measures included Mean Absolute Error (MAE), Mean Squared Error (MSE) and coefficient of determination (R 2 score). The metrics were selected to give an overview of the accuracy, robustness, and generalization of the models.

Visual inspection of scatter plots of actual and predicted values of the yield was done. Plots of residuals also showed the amount and pattern of error and assisted in detecting possible biases or systematic departures in predictive accuracy of the model. The ARIMA-GRU hybrid architecture was the best performing model among all tested models in terms of reducing prediction errors and detecting the complex yield patterns. The multi-layered models such as CNN-GRU achieved good results, confirming that complex data is best represented using a multi-layered model.

### 3.7 Summary

To conclude, the research approach used in this paper entailed a comprehensive and multidimensional approach to the prediction of crop yields. It has been shown that a powerful and flexible modeling framework was created by combining classical statistical methods with the newest deep learning models. The procedure placed an emphasis on data cleanliness, significant feature extraction, and prudent selection and assessment of models. The methodology offers a strong basis of high-accuracy agricultural forecasting through wide experimentation and analysis and can be of benefit to all stakeholders including the farmer and policymakers.

## 4. Results and Discussion

### 4.1 Introduction

The chapter offers the findings of the predictive models created to estimate crop yield and explains its performance in detail. The models were compared using standard regression statistics, that is Mean Absolute Error (MAE), Mean Squared Error (MSE) and Coefficient of Determination (R2 Score). These measures provide an understanding of the accuracy and reliability of each model so that a balanced comparison can be made. The purpose of the evaluation was not only to find out which model achieved the best results in an absolute sense but also the extent to which each method was able to characterize the complex relationships that exist in agricultural data.

### 4.2 Overview of Model Performance

The table below summarizes the performance measures of all the models that were developed in this study:

Table 3: Performance of Various Models for Crop Yield Forecasting, evaluated by MAE, MSE and R-Squared

| Model | MAE | MSE | $R^2$ Score |
|---|---|---|---|
| Lasso | 0.889085 | 1.187171 | 0.586770 |
| K-Nearest Neighbors (KNN) | 0.870164 | 1.150233 | 0.599628 |
| Decision Tree | 0.582484 | 0.531282 | 0.815072 |
| Deep Neural Network (DNN) | 0.508287 | 0.397923 | 0.861491 |
| Long Short-Term Memory (LSTM) | 0.591495 | 0.535705 | 0.813532 |
| Convolutional Neural Network - GRU (CNN-GRU) | 0.458649 | 0.329227 | 0.885403 |
| **ARIMA-GRU Hybrid** | **0.159848** | **0.040122** | **0.986223** |

The table shows clearly how the accuracy of the models has improved with the traditional regression models to more complex deep learning and hybrid models. The hybrid model (ARIMA-GRU) performed the best of all and is much superior to the other models in all three metrics.

## 4.3 Analysis of Traditional Models

The simplest methods Lasso Regression, K-Nearest Neighbors (KNN) and Decision Tree worked as a good benchmark against more complicated methods as the results demonstrated on Table 3.

The Lasso Regression delivered rather high values of the error, an MAE of 0.889, and an $R^2$ score of 0.586. This implies that although Lasso was capable of generalizing the data to a certain degree, it failed to represent the nonlinear dependencies that are present in the data. Its regularization ability, which is useful to avoid overfitting, might have caused underfitting in this instance, particularly in a dataset with many interactions between the variables.

KNN was a little bit better, having less MAE and higher R 2 value. Its approach of instance-based learning enabled it to capture local structures in the data, but the absence of an underlying model led it to be vulnerable to noise and feature scaling problems, particularly in high-dimensional data or non-homogeneous data.

On the other hand, Decision Tree Regressor performed significantly better with $R^2$ score higher than 0.81. The model can split the feature space and manage both categorical and numerical variables, which explains this performance. The decision trees are especially well adapted to problems whose decision boundaries are hierarchical or rule based, which is possible in agricultural systems.

## 4.4 Performance of Deep Learning Models

When the modeling was changed into deep learning methodologies, its performance drastically increased. The Deep Neural Network (DNN) obtained a significant decrease in MAE and MSE with a resultant R2 score of 0.861. Such a boost highlights the potential of the DNN to model complex, non-linear relationships that other regression models cannot model. The layers of neurons enabled the model to learn high-level interactions between input features e.g. how rainfall, soil type and temperature interact to affect the yield outcomes.

LSTM model was a bit inferior to the performance of DNN but gave good results. It had an MAE of 0.591 and $R^2$ of 0.813. Though LSTMs are typically used with temporal or sequential data, the model was re-purposed in this work in order to determine whether any sequential organization in the feature interactions might be exploited. The relatively larger error than that of the DNN means that although the LSTM worked, the time aspect might not have been emphasized greatly in this data set to warrant its complexity.

The CNN-GRU architecture was the best of the deep learning models. Having an MAE of 0.458 and an $R^2$ score of close to 0.89, this hybrid model was able to pair the feature extraction ability of Convolutional Neural Networks with the memory efficient nature of GRUs. The convolutional layers extracted the spatial patterns in the data and the GRU layers worked over such patterns sequentially resulting in more robust and accurate prediction.

## 4.5 Superiority of the ARIMA-GRU Hybrid Model

The best performance was recorded by ARIMA-GRU hybrid model. This model had the remarkably small MAE of 0.159, the insignificant MSE of 0.040, and the $R^2$ of 0.986, which means nearly ideal fitting. The effectiveness of this hybrid architecture is that it is capable of combining the advantages of both statistical and deep learning paradigms.

ARIMA was applied to model the linear trends and autocorrelations in the yield data and represent the conventional time-series behavior. Nevertheless, it is not very useful in modeling non-linear relationships. To make up this, the ARIMA model residuals were fed to GRU network which took the non-linear features that were not captured by ARIMA. The last prediction of the yield was calculated as the combination of the ARIMA forecast and the residual prediction of the GRU. This mix produced an effective model, which could respond to the organized and random variations of crop yields data. The substantial decrease in the metric of errors and extremely high R2 index prove the superiority of the hybrid model and its viability in complex tasks of agricultural prediction. In addition, its performance indicates that there are both linear tendencies and nonlinear interactions in the crop yield data set-an observation that would have been lost had it been applied by any single model (statistical or neural).

## 4.6 Discussion and Implications

The steady increase in the performance of the models along the spectrum of the models, Lasso Regression, to the ARIMA-GRU hybrid, shows that model selection is of paramount importance in agricultural prediction tasks. Although conventional approaches can be simpler and easier to interpret, in many cases they fail to work with multidimensional, mutually correlated data. Conversely, the more complicated and computationally intensive models, i.e. deep learning, are very strong in learning hidden patterns and interactions.

The effectiveness of the hybrid model of the ARIMA-GRU model has various practical implications. To the agricultural stakeholders, such a model can give very precise yield predictions and thus they can plan their resources, supply chain logistics, and market strategies better. To researchers and data scientists, the results also demonstrate the usefulness of hybrid architectures in the trade-off between explainability and performance, in particular, in areas where both past trends and more recent inputs are present. Also, the findings support the general idea of using several modeling techniques in the framework of a single predictive model. The approach of statistical rigor and neural flexibility sets the methodology that leads to more robust and generalizable models in agricultural analytics.

## 4.7 Summary

Finally, this chapter has given a detailed overview of the results of the model performances and has brought out the relative merits and shortcomings of each method. The conventional regressions gave baseline comparisons which were outdone by the neural networks. The ARIMA-GRU hybrid was the best and most precise crop yield prediction model, proving that the combination of statistical and deep learning models is very effective. These results constitute a step in the right direction towards developing smart farming systems that can make informed decisions and predictive planning.

# 5. Conclusion and Future Work

## 5.1 Conclusion

This study explored the problem of the accurate prediction of the crop yield and used both traditional regression models and deep learning architectures to make this prediction. This study was motivated by the fact that there is a dire need of reliable agricultural forecasting tools that can aid in the optimization of planning and allocation of resources and design of policies in the agricultural sector. In building an end-to-end modeling pipeline, including data collection and preprocessing, model evaluation, this work was intended to determine the most effective predictive approach in estimating yield.

The results indicated that traditional models including Lasso Regression and K-Nearest Neighbors provide a basic insight and facilitate interpretations but do not perform as well, in terms of accuracy, in the presence of complex, non-linear and interdependent variables that are typical of agricultural data. On the contrary, the deep learning models, especially the CNN-GRU and ARIMA-GRU hybrid models, could learn the complex patterns and dependencies in the data and resulted in a massive boost in predictive accuracy.

The ARIMA-GRU hybrid model has been the most resilient one among the models tested, with error measures that are exceptionally low, and the R 2 value greater than 0.98. The model is a unique combination of the strengths of deep learning and statistical time-series modeling. The hybrid model that included ARIMA to represent linear trends and seasonality and GRU to represent the non-linear residuals provided an overall comprehension of the variability in yield. The success of this model does not only support the worth of ensemble methods but also leads to the conclusion that the behavior of crop yields is characterized both by systematic effects and irregular ones and each needs a different approach to modeling.

In sum, the study points to the fact that hybrid models that combine interpretability, adaptability, and learning capacity are the most promising way of predicting crop yield. The approach to methodology formulated in the paper can have a wider application in precision agriculture, food security planning, and climate impact assessment.

## 5.2 Limitations of the Study

Although the results of the research were encouraging, there are a number of limitations that should be mentioned. The data set was time limited in its granularity and geographically limited in spite of being rich in environmental and agronomic variables. Sub-seasonal or daily data, and higher spatial resolution, e.g. remote sensing or satellite imagery, may result in more information about microclimatic and location-specific influences. In addition, although holdout datasets were used to validate the models, the external validation was not done in different regions or crop types which limits the applicability of the results to the existing dataset.

The other constraint is the computational requirements of deep learning model. The processing power needs of training hybrid models like CNN-GRU or ARIMA-GRU is high, and increases with either larger datasets, or more complex inputs. To be adopted in the real world, especially in poorer environments, optimization and model simplification might be required.

Lastly, although the models have high accuracy, they are mostly black boxes. Interpretability of deep learning models has been a limitation, especially in areas of application where transparency and trust is a requirement towards end user adoption, such as in agriculture.

## 5.3 Future Work

The knowledge obtained in the course of this study presents numerous opportunities in terms of future research and practice. The inclusion of a wider variety of sources of data and more granular data is one of the directions. The models have the potential to be greatly improved by high-resolution weather data, satellite-derived vegetation indices, and real-time sensor data. Another solution would be to extend the dataset to several seasons and regions in order to create more robust models that are not only accurate, but also robust in different situations.

The other potential extension would be to enhance the interpretability of deep learning models. Methods like SHAP (Shapley Additive Explanations) or attention mechanisms may be included to emphasize which of the factors have the biggest impact on the predictions of the yield and thus to enhance transparency and trust that a user will have towards the results of the model.

Technically, it may be possible to find other hybrid architectures that are even more accurate and generalizable, e.g. the combination of convolutional transformers with statistical models. Also, the models used in a decision support system of farmers and agricultural planners could significantly increase their practical value. This would entail creating friendly interfaces or mobile applications through which real time data and yield prediction can be done in the field.

Finally, future research might examine how the overhead of computation can be minimized in order to facilitate greater adoption. These advanced models can be an accessible option to users with low processing capability or bandwidth through lightweight neural networks, model pruning, or transfer learning.

## 5.4 Final Remarks

Finally, this research paper makes a valuable contribution to the agricultural analytics field by showing that by combining statistical and deep learning models, it is possible to improve crop yield prediction considerably. The findings evidently prove the excellence of hybrid architectures in describing the complexity of the dynamics of yield-related variables. Despite the still-existing issues of generalization, computational burden, and interpretability, the work done in this study is a step in the direction where predictive modeling will become a key tool in the optimization of agricultural processes. Such models can be further refined and expanded, and with additional development, they could revolutionize how agricultural decisions are made, leading to more robust and more efficient food production systems in many parts of the world.

# References

El-Kenawy, E. S. M., Alhussan, A. A., Khodadadi, N., Mirjalili, S., & Eid, M. M. (2024). Predicting potato crop yield with machine learning and deep learning for sustainable agriculture. *Potato Research*.

Filippi, P., Han, S. Y., & Bishop, T. F. A. (2025). On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. *Precision Agriculture, 26*(1).

Guo, W. W., & Xue, H. (2014). Crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models. *Mathematical Problems in Engineering, 2014*, Article 857865.

Hasanat, S. M., Ullah, K., Yousaf, H., Munir, K., Abid, S., Bokhari, S. A. S., Aziz, M. M., Naqvi, S. F. M., & Ullah, Z. (2024). Enhancing short-term load forecasting with a CNN-GRU hybrid model: A comparative analysis. *IEEE Access*.

Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems, 187*, 103016.

Paudel, D., Boogaard, H., de Wit, A., van der Velde, M., Claverie, M., Nisini, L., Janssen, S., Osinga, S., & Athanasiadis, I. N. (2022). Machine learning for regional crop yield forecasting in Europe. *Field Crops Research, 276*, 108377.

Paudel, D., de Wit, A., Boogaard, H., Marcos, D., Osinga, S., & Athanasiadis, I. N. (2023). Interpretability of deep learning models for crop yield forecasting. *Computers and Electronics in Agriculture, 206*, 107663.

Gong, Y., Zhang, Y., Wang, F., & Lee, C. H. (2024). Deep learning for weather forecasting: A cnn-lstm hybrid model for predicting historical temperature data. arXiv preprint arXiv:2410.14963.

Vallileka, N., Rajkumar, G. V., Krishnan, R. S., Shankar, S. V., Raj, J. R. F., & Karthikeyan, M. S. (2025, February). Hybrid CNN-LSTM Model for Enhanced Weather Forecasting: Leveraging Spatial and Temporal Dependencies. In 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL) (pp. 1188-1195). IEEE.

Waqas, M., & Humphries, U. W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. MethodsX, 102946.

Zhou, Q., Guo, Y., Li, G., Xu, K., & Wang, K. (2025). Multimodal Hybrid Aero-Engine Mechanical Wear Fault Diagnosis Algorithm Based on Two-Channel Data Input Types. International Journal for Numerical Methods in Engineering, 126(10), e70046.

Elsamie, M. A., Ali, T., & Zhou, D. Y. (2021). Using a dynamic time series model (ARIMA) for forecasting of Egyptian cotton crop variables.Elsamie, M. A., Ali, T., & Zhou, D. Y. (2021). Using a dynamic time series model (ARIMA) for forecasting of Egyptian cotton crop variables.

Patrick, S. (2024). Time series and ensemble models for forecasting Tanzanian banana crop yield under various effects of Climate change (Doctoral dissertation, NM-AIST).

Venkatanaresh, M., & Kullayamma, I. (2024). Deep learning based concurrent excited gated recurrent unit for crop recommendation based on soil and climatic conditions. Multimedia Tools and Applications, 83(24), 64109-64138.

Slater, L., Arnal, L., Boucher, M. A., Chang, A. Y. Y., Moulds, S., Murphy, C., ... & Zappa, M. (2022). Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models. Hydrology and Earth System Sciences Discussions, 2022, 1-35.

Abhishek, A. (2023). Dynamics of Seasonal Crop Yield Prediction Under Weather and Climate Extremes. Michigan State University.

Badshah, A., Alkazemi, B. Y., Din, F., Zamli, K. Z., & Haris, M. (2024). Crop classification and yield prediction using robust machine learning models for agricultural sustainability. *IEEE Access*.

Pant, J., Pant, R. P., Singh, M. K., Singh, D. P., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings, 46, 10922-10926.

Kowalska, A., & Ashraf, H. (2023). Advances in deep learning algorithms for agricultural monitoring and management. *Applied Research in Artificial Intelligence and Cloud Computing*, *6*(1), 68-88.