Deep Learning in Rice Yield Prediction

G. S. V. M. Ishan

2024

Deep Learning in Rice Yield Prediction

A dissertation submitted for the Degree of Master of Computer Science

G. S. V. M. Ishan

University of Colombo School of Computing

2024

DECLARATION

Name of the student: G. S. V. M. Ishan
Registration number: 2018/MCS/034
Name of the Degree Programme: Master of Computer Science
Project/Thesis Title: Deep Learning in Rice Yield Prediction

1. The project/thesis is my original work and has not been submitted previously for a degree at this or any other University/Institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.

2. I understand what plagiarism is, the various types of plagiarism, how to avoid it, what my resources are, who can help me if I am unsure about a research or plagiarism issue, as well as what the consequences are at University of Colombo School of Computing (UCSC) for plagiarism.

3. I understand that ignorance is not an excuse for plagiarism and that I am responsible for clarifying, asking questions, and utilizing all available resources in order to educate myself and prevent myself from plagiarizing.

4. I am also aware of the dangers of using online plagiarism checkers and sites that offer essays for sale. I understand that if I use these resources, I am solely responsible for the consequences of my actions.

5. I assure you that any work I submit with my name on it will reflect my own ideas and effort. I will properly cite all material that is not my own.

6. I understand that there is no acceptable excuse for committing plagiarism and that doing so is a violation of the Student Code of Conduct.

Signature of the Student	**Date (DD/MM/YYYY)**
Mehuka	29/09/2024

Certified by Supervisor(s)

This is to certify that this project/thesis is based on the work of the above-mentioned student under my/our supervision. The thesis has been prepared according to the format stipulated and is of an acceptable standard.

	Supervisor 1	**Supervisor 2**	**Supervisor 3**
Name	Dr. L.N.C. De Silva		
Signature			
Date	29 - 09 - 2024		

I would like to dedicate this thesis to my parents for always giving support in challenging times.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. L.N.C. De Silva for her invaluable guidance and support throughout the entire duration of my thesis. Her expertise, patience, and insightful feedback have been instrumental in shaping my project and enhancing its quality. I am truly fortunate to have had her as my supervisor. This research project would not have been possible without her guidance.

I owe my deepest gratitude to my family for their support and love in many ways and for encouraging me whenever I was down. Especially, I would like to convey my heartiest gratitude to my mother, who always supported me and provided me with a good education.

G. S. V. M. Ishan

ABSTRACT

Rice, being the staple food across much of Asia, holds paramount importance in countries like Sri Lanka. The yield of paddy rice is profoundly affected by climate variations, making accurate forecasting crucial for ensuring food security. However, predicting rice yield entails navigating through intricate non-linear relationships between climate factors and agricultural output. In this study, I leverage Convolutional Neural Networks (CNNs) to forecast rice yield using extensive historical weather and yield data from 25 key rice-producing cities in Sri Lanka. Our analysis encompasses a comprehensive array of climate variables, including temperature, radiation, precipitation, wind speed, and more, spanning the years 2004 to 2023. By employing CNNs, I demonstrate the efficacy of this advanced machine-learning technique in unraveling the complexities of rice yield prediction. This approach here not only provides valuable insights into the interplay between climate dynamics and rice cultivation but also offers a powerful tool for policymakers in formulating effective agricultural policies in Sri Lanka. This study underscores the significance of CNNs in enhancing rice yield prediction accuracy, thereby contributing to the sustainable management of food resources in Sri Lanka and beyond.

Keywords: Rice yield prediction, Convolutional Neural Networks, Weather data, Climate, Sri Lanka, Agricultural forecasting, Deep Learning, Paddy Yield

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1.1 INTRODUCTION

In Sri Lanka, Rice (Oryza sativa) yield has a significant role in the country's food supply. Therefore, rice yield prediction is vital for agriculture stakeholders to make optimal decisions regarding cultivation. On the other end, predicting rice yield is important to policymakers to make appropriate planning like storing, selling, fixing minimum price, and import and export decisions related to national food security. At present, the agriculture sector stakeholders experience many issues due to the lack of real-time (dynamic) information relevant to current production levels at the right time. As a result, they tend to make incorrect decisions at the crucial stages of the farming cycle, causing financial hardships, some situations even lead to tragic incidents such as suicide. Therefore, the availability of this type of information can also create proper coordination among all stakeholders in the agriculture sector.

Predicting the crop yield during the crop production life cycle is challenging as it depends on numerous factors such as plant health, fertilizer, pesticide effect, etc. (Nigam et al., 2019). In addition to that, factors that are unable to be controlled by the farmers, such as temperature, rainfall, soil type, etc., also play a significant role in predicting the crop yield. Hence, human intelligence alone may not be able to predict crop yield accurately in an economical sense due to complex nonlinear relationships among the factors (Amaratunga et al., 2020). In that regard, machine learning can be considered a promising decision-support tool for crop yield prediction.

According to the research publications, many researchers have used artificial neural networks to predict crop yield. Many of these models were based on factors such as temperature, rainfall, and soil. It was also observed that the research based on deep learning techniques had produced more promising results. However, the use of deep learning algorithms is limited to Convolutional Neural networks (CNN), Long-Short Term Memory (LSTM), and Deep Neural Networks (DNN). Hence investigating the use of deep learning approaches for crop yield prediction is essential for better decision-making.

The main objective of this research is to determine which deep learning approaches are better in predicting rice yield prediction for better decision-making. To achieve that we must answer 4 subproblems in this research.

- Identify data sources.
- Identify the factors that have a direct relationship with rice yield.
- Train Convolutional Neural Networks (CNN) model
- Evaluate the Convolutional Neural Networks model.

1.2 Motivation

Sri Lanka, a nation heavily dependent on rice as a staple food, confronts critical challenges in ensuring a consistent and adequate food supply for its population. The compounding effects of climate change, erratic weather patterns, and the imperative for sustainable agricultural practices present pressing concerns for the nation's food security. In this context, the traditional methods of predicting rice yield, reliant on historical data and statistical models, often fall short in adapting to Sri Lanka's unpredictable environmental conditions. The repercussions of inaccurate yield predictions are particularly severe in a country where rice holds immense cultural, dietary, and economic significance. The dire situation faced by farmers, marked by financial hardships and, tragically, instances of suicide due to the inability to cope with the consequences of poor rice yields, underscores the urgent need for more effective solutions.

Against this backdrop, the application of deep learning emerges as a beacon of hope. By harnessing the capabilities of neural networks to unravel complex relationships within diverse datasets, this research aims to bring about a transformative improvement in the precision of rice yield predictions in Sri Lanka. The outcomes of this work hold the potential not only to revolutionize local agricultural planning, empowering farmers with more accurate forecasts but also to address the human toll of financial hardships. By providing timely and reliable predictions, this research aspires to contribute meaningfully to alleviating the economic burden on farmers, mitigating the circumstances that lead to extreme measures, and fostering a more resilient and sustainable future for Sri Lanka's agricultural communities. Through the integration of advanced technology, this research aligns with the nation's commitment to safeguarding the well-being of its farmers and ensuring food security in the face of a changing climate.

.

1.3 Statement of the problem

For the purpose of agricultural planning, resource optimization, and food security, an accurate estimate of rice output is essential. Conventional techniques for estimating rice production frequently depend on statistical models and historical data or farmer's experience which may not adequately reflect the dynamic and complicated interactions between environmental conditions and paddy yield. Furthermore, more complex, and adaptable methods are required due to the growing variability in climatic patterns and the requirement for real-time decision-making in agriculture.

By utilizing deep learning, this study attempts to overcome the shortcomings of existing approaches by creating reliable and accurate models for predicting rice yield. In order to improve the accuracy and timeliness of rice production estimates, the issue is to efficiently integrate a variety of environmental variables, and historical data into a deep learning framework that can adjust to shifting agricultural conditions.

1.4 Research Aims and Objectives

1.4.1 Aim

The primary aim of this research project is to design and assess a 1D Convolutional Neural Network (CNN) model specifically tailored for predicting rice yield in Sri Lanka. Given that rice is a fundamental crop in Sri Lanka, accurate yield prediction holds immense significance for the country’s agricultural sector and overall food security. However, achieving precise and timely predictions of rice yield is a formidable task due to the intricate interplay of multiple factors, including climate variations, and diverse agricultural practices. The central research challenge lies in developing an efficient and dependable predictive model that can harness historical data related to climate patterns to forecast rice yields with precision.

To tackle this challenge, I employ deep learning techniques, particularly 1D CNNs. These neural networks are adept at capturing intricate patterns and dependencies within sequential data. By leveraging historical information and training the model on a rich dataset, I aim to create a robust predictive tool. This tool will not only provide accurate yield estimates but also offer valuable insights into rice yield variations.

Ultimately, my project endeavors to empower farmers and policymakers with informed decision-making tools, contributing to sustainable agricultural practices in Sri Lanka.

1.4.2 Objectives

The objective of this project proposal is to develop a robust predictive model that utilizes climate parameters to forecast rice yield in Sri Lanka. By analyzing historical climate data, including temperature, rainfall, humidity, and other relevant meteorological variables, our proposed model aims to provide accurate predictions for rice yield across different regions and seasons in the country. The focus will be on employing machine learning techniques, particularly regression models, to establish the relationship between climate parameters and rice yield. Additionally, the project seeks to enhance understanding of how climatic factors influence rice production and to provide valuable insights for farmers, agricultural policymakers, and stakeholders to optimize crop management strategies and enhance food security in Sri Lanka. Furthermore, the project aims to assess the performance of the predictive model and its potential for practical application in agricultural decision-making processes. The main objective can be divided into sub-objectives as mentioned below.

- **Algorithm Identification -** Identify and analyze deep learning algorithms commonly used for crop yield prediction and understand their strengths, limitations, and suitability for rice yield forecasting.
- **Feature Exploration** - Investigate the relevant features utilized in rice yield prediction models and determine which climate, soil, and agricultural factors significantly impact rice production.
- **Dataset Selection** - Locate and select a suitable dataset containing historical climate, soil, and yield data and ensure data quality, relevance, and representativeness.
- **Evaluation Criteria and Approaches** - Identify adequate evaluation criteria for assessing model performance and explore various approaches (e.g., accuracy, RMSE, MAE) to measure prediction quality.
- **Model Training** - Train machine learning models using selected deep learning algorithms and optimize model parameters and architecture for accurate predictions.

- **Comparison and Contrast** - Compare different evaluation criteria and approaches and then Contrast their effectiveness in capturing rice yield variations.

1.4.3 Scope

Food production in South Asian countries such as Sri Lanka and India is largely dependent on crops including rice, wheat, and various pulses (Gandhi et al., 2016) Among these crops grown in Sri Lanka, Rice is one of the most important cultivation in Sri Lanka. Hence, it is approximated that annual production of 2.7 million metric tons of rough rice (paddy) is necessary to meet the demand, which accounts for approximately 95% of the country's consumption, and Over 1.8 million farmers and their families are involved in the production process. (Amaratunga et al., 2020). Therefore, in this project, I chose Rice as a crop to yield prediction using data sets of Yala and Maha of the main 25 paddy-grown areas below mentioned.

- Colombo
- Gampaha
- Kalutara
- Kandy
- Matale
- Galle
- Matara
- Hambantota
- Jaffna
- Mannar
- Trincomalee
- Kurunegala
- Puttalam
- Badulla
- Ratnapura
- Nuwara Eliya
- Batticaloa
- Anuradhapura
- Monaragala
- Kegalle
- Polonnaruwa
- Ampara
- Kilinochchi
- Mullaitivu
- Vavuniya

The timeframe from the above scope will be 2004 (Maha) to 2023 (Maha)

CHAPTER 2 - LITERATURE REVIEW

2.1 Introduction

In Sri Lanka, agriculture and the related economic ecosystem are the primary sources of income for many Sri Lankans and Rice is the primary food source for many Sri Lankans. Therefore, changes in the annual national rice yield significantly impact Sri Lankan economic stability and food security. In the past, crop yield prediction was estimated by using farmers' experience. It was solely based on randomly selecting a plant, counting seeds in each bud, and then predicting the crop yield based on the farmer's experience. (Gandge and Sandhya, 2017). Advances in machine learning (ML) and artificial intelligence (AI) have led to the development of sophisticated models for predicting crop yields.

There are significant amounts of research papers on several crops in the Crop Yield prediction field due to their significance in any country's economy. The paddy yield prediction at the country level where rice is the staple food is reported in the literature, e.g. China, Bangladesh, Egypt, Korea, and South Korea. In some countries, studies on paddy yield prediction were confined only to a specific region, e.g. Ebro Delta in Spain. Moreover, the prediction of paddy yield in some states in India like Maharashtra, Tamil Nadu, and Andhra Pradesh was reported. (Ekanayake et al., 2022) Because no matter how developed the country is, it will be unstable if there is no food security for its people. This review explores the current state of research in paddy yield prediction using ML techniques and highlights the methodologies and findings of key studies.

2.2 Artificial Neural Networks (ANNs)

(Amaratunga et al., 2020) also emphasize paddy harvest extreme vulnerability to climate variance in their research paper. At the time of the research, the authors highlighted that there was no rice yield prediction research using climate data using ANNs in the Sri Lankan context. They used paddy yield and climatic parameters from Ampara, Batticaloa, Badulla, Bandarawela, Hambantota, Trincomalee, Kurunegala, and Puttalam to train 3 neural network algorithms which are Levenberg–Marquardt Algorithm (LM), Bayesian Regularization Algorithm (BR), and Scaled Conjugate Gradient Algorithm (SCG). They used the Correlation coefficient (R) and Mean Squared Error (MSE) as the performance indicators to evaluate the performance of the developed ANN models. They have used Rainfall (mm), morning and evening relative humidity (%), minimum and maximum temperature (°C), wind speed(km/hr), evaporation (mm), and sunshine hours (hr). In the research, authors witnessed that even though BR and SCG algorithms produced acceptable results, the LM training algorithm was better than them in accuracy.

In their paper (Baral et al., 2011), the authors employ Artificial Neural Networks with Particle Swarm Optimization as an optimization technique to predict paddy yield across three different districts situated in distinct climatic zones. They base their predictions on ten years of historical data, which includes information on paddy yields, daily temperature (mean and maximum), and precipitation (rainfall). The use of ANNs allows the model to approximate

and predict crop yield by learning from the dataset. They noticed that ANN with the PSO model produced consistently lower error rates with the variation of 8.24 ± 0.6 % for the 5-day dataset and 8.28 ± 0.08 % for the 3-day dataset. Testing of the prepared model gives an error rate of 8.3 ± 0.1 % on an overall basis. For a 5-day average dataset prepared from the test data, the error rate was 8.53 % and for the 3-day dataset, the error rate was 8.44%. The overall error rate for the model is 8 % which is in the allowable error range.

The paper of (Wickramasinghe et al., 2020) also discusses applications of ANNs in rice yield prediction using climate data in the context of Sri Lanka. They used 30 years of paddy yield data in the North-Western province especially Kurunegala and Puttalam districts from the Irrigation Department of Sri Lanka and Climate data such as rainfall (RF), minimum and maximum atmospheric temperatures (Tmin and Tmax), morning and evening wind speeds (WSmor and WSeve), sunshine hours (SH) and evaporation rates (ER) for same districts and same period from Department of Meteorology Sri Lanka. They used the ANN Levenberg Marquardt algorithm as the ANN algorithm. They noticed that the Results of 7 tests out of 10 show 100% accurate predictions of the rice yield using the models. It was evident that absolute percentage error varies between 0-12.1 % showing a good agreement between the predicted yield and the actual yield. The Mean Absolute Percentage Error (MAPE) is 2.99%. The output of the proposed ANN models can be considered as highly accurate prediction models. Therefore, the developed models can be used to forecast the paddy yield of Kurunegala and Puttalam Districts for the Yala and Maha seasons separately.

In this study of authors (Ruß et al., 2008), the authors explore the intersection of precision agriculture and information technology (IT). They emphasize the collection of agricultural data through sensors and GPS technology. Leveraging these raw data for meaningful insights requires effective information extraction. The paper specifically investigates the use of artificial neural networks to predict wheat yield using readily available in-season data. Once successful, this prediction can be directly applied to optimize fertilizer usage, benefiting both economic and environmental aspects. The data available in this work have been obtained in the years 2003 and 2004 on a field near Kothen, north of Halle, Germany. They collected seven input attributes from the fields, they are Nitrogen Fertilizer (N1, N2, N3), Vegetation (REIP32, REIP49), Electric Conductivity (EM38), and Yield (2003,2004) They got the mean error is 0.53, 0.49,0.48 for their 3 datasets while standard deviation on all data sets is 0.015. The study contributes valuable knowledge to the field of precision agriculture and underscores the potential of neural networks in crop yield estimation.

2.3 Support Vector Machines (SVMs)

Even though some parts of Sri Lanka use, man-made large tanks, or lakes to cultivate rice, most of the agriculture in Sri Lanka is heavily dependent on seasonal climate conditions such as monsoon rain seasons. Therefore, variableness in climate conditions can have a damaging impact on rice yield. (Gandhi et al., 2016) discuss the rice yield prediction in different climate conditions using publicly available Indian government records from 1998 to 2002. For research, they used climate factors such as precipitation, minimum temperature, average temperature, maximum temperature, reference crop evapotranspiration, and rice yield production for the Kharif season for the years 1998 to 2002. Data from 27 districts of Maharashtra state, India was considered for the experiment. They used the Mathews Correlation Coefficient (MCC) computed as a measure of the quality of classification. They noticed that the algorithm achieved an accuracy of 78.76%, a sensitivity of 68.17%, and a specificity of 83.97%. Mathews Correlation Coefficient was used to measure the quality of classification which resulted in 0.54. The error results of the classifier mean absolute error of 0.23, the root mean squared error of 0.39, the relative absolute error of 67.38%, and the root relative squared error of 82.51%. research has demonstrated the prediction of rice crop yield by applying one of the machine learning techniques, support vector machine (SVM). The experimental results showed that the other classifiers such as Naïve Bayes, Bayes Net, and Multilayer Perceptron performed better by achieving the highest accuracy, sensitivity, and specificity compared to the SMO classifier with the lowest accuracy, sensitivity, and specificity that has been reported earlier for the same data set.

2.4 Gaussian Process Regression

The paper by (Ekanayake et al., 2022) provides a comprehensive overview of the research landscape at the intersection of agriculture and meteorology. Given the critical role of paddy (rice) as a staple food in Sri Lanka, accurate yield prediction becomes essential for food security and livelihoods. Traditional methods often fall short due to their reliance on historical data and expert judgment. In response, the study introduces Gaussian Process Regression (GPR), a powerful machine-learning technique that captures complex relationships between weather variables and crop performance. In their research they used the climatic factors of rainfall, relative humidity, minimum temperature, maximum temperature, average wind speed, evaporation, and sunshine hours as input (independent) variables, while the paddy yield was the output (dependent) variable which sourced from Administrative Units of Ampara, Batticaloa, Trincomalee, Polonnaruwa, Kurunegala, Anuradhapura, Hambantota, Badulla and Monaragala. They sourced climate factors such as Rainfall, Relative Humidity, Minimum Temperature, Maximum Temperature, Average Wind Speed, Evaporation, and Sunshine hours. For each geographical region considered in this study, four GPR models were constructed based on four types of Kernel functions namely Rational Quadratic, Exponential, Squared Exponential, and Matern 5/2. The best-fitting model for a particular district was selected based on the correlation produced by each Kernel function. They observed that maximum temperature, evaporation, and morning wind are positively correlated with the paddy yield. Nevertheless, the RF, maximum RH, Tmin, and evening WS are negatively correlated with the paddy yield. In the research, GPR has outperformed ANN, SVMR, and statistical regression techniques as per the performance assessed in terms of statistical parameters such as R, R2, MSE, MAPE, and Nash number. Further, GPR was reportedly accurate when predicting the yield of other crops as well at the country level.

2.5 Multiple Algorithms

The paper of (Wickramasinghe et al., 2021) focuses on the critical intersection of climate science and agriculture. Rice, a staple food globally, plays a pivotal role in South and Southeast Asia. Accurate yield prediction is essential for food security, especially in major rice-producing regions like Sri Lanka. The study explores statistical and machine learning techniques, on rice harvest and yield data over the last three decades and monthly climatic data were used to develop the prediction model by applying artificial neural networks (ANNs), support vector machine regression (SVMR), multiple linear regression (MLR), Gaussian process regression (GPR), power regression (PR), and robust regression (RR). The performance of each model was assessed in terms of the mean squared error (MSE), correlation coefficient, mean absolute percentage error (MAPE), root mean squared error ratio (RSR), BIAS value, and the Nash number. They used the Kurunegala and Puttalam districts' paddy harvest. Rainfall, temperature (minimum and maximum), evaporation, average wind speed (morning and evening), and sunshine hours are the climatic factors considered for modeling. The climatic and rice yield data were categorized into four groups according to the district and agricultural season, and then utilized with each of the six techniques to determine the correlation between them. The effectiveness of each model was evaluated based on the Mean Squared Error (MSE) and correlation coefficient. The most influential climatic factors in the yield functions of the statistical models were determined to be rainfall, temperature, and average wind speed, based on the significance of their weights and exponents. The accuracy of the models was assessed using various metrics such as MSE, correlation coefficient, MAPE, RSR, BIAS value, and the Nash number. In terms of developing accurate relationship models, all machine learning techniques (ANN, SVMR, and GPR) outperformed statistical methods (MLR, PR, and RR). Among them, the GPR-based model exhibited the highest accuracy, with MSE, MAPE, RSR, and BIAS values approaching zero, and the Nash number and correlation coefficient approaching 1. To further validate the model, an independent dataset that was not used during model development was employed. The results demonstrated the GPR-based model's ability to predict rice yield using known or forecasted climatic data.

In their paper (Muthusinghe et al., 2018) emphasize how accurate yield prediction assumes paramount importance to Sri Lankan food security. Traditionally, yield estimation relied on historical data and expert judgment, often overlooking the dynamic interplay of climate variables such as temperature, precipitation, solar radiation, and wind patterns with crop performance. Authors have proposed a platform that targets the smart farming concept for paddy with the following modules: (1) a prediction module to predict paddy harvest and (2) a prediction module to predict rice demand. Smart Farming (SF) involves the incorporation of information and communication technologies into agricultural production systems as an enabler of more efficient, productive, and profitable farming enterprises. To realize this, they used the Recurrent Neural Network and Long Short-Term Memory (LSTM) model. The performances of algorithms were evaluated using real data sets for the Sri Lankan context. The results show that the prediction modules give accurate results in a short time. The performance of prediction modules was analyzed using the following matrices: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), training score, and test score of the models. The harvest prediction module gave 78% for the training score with a 0.04 MSE value and 75% test score with a 0.11 MSE value. The demand prediction module gave 79% for the training score with a 0.17 MSE value and 74% test score with a 0.34 MSE value.

Going further steps from only considering climate factors such as rainfall (RF), minimum and maximum temperature (TM, TX), solar radiation (SR), and relative humidity (RH) (Chandra et al., 2019) also consider non-weather data such as block level fertilizer consumption and soil type. The Authors attempted to develop kharif rice yield prediction models through Machine Learning approaches such as Artificial Neural Network and Random Forest for the 42 blocks covering 13,141 sq km upland rainfed area of Purulia and Bankura district, West Bengal. Models were developed integrating monthly NDVI with weather and non-weather variables at the block level for the period 2006 to 2015. The model correlation obtained was 0.702 with MSE 0.01. Though the weather variables vs NDVI models are quite satisfactory, NDVI vs kharif rice yield models, however, show relatively less correlation, about 0.6 revealing the requirement of varied additional farmer-controlled inputs. Development of NDVI vs crop yield models for different crop growth stages or fortnightly over a larger data set with selective adding of weather and non-weather variables to NDVI would be the most appropriate. The study reveals that for rainfed kharif rice yield prediction NDVI alone may not yield the desired model. Adding weather and non-weather predictor variables to NDVI values improves the model accuracy up to 0.7. Instead of monthly data, the models based on different time granularity such as crop growth stage-wise or fortnightly NDVI and weather variables are likely to have much better accuracy. The seed variety and soil texture type have shown little relevance. The present study is based on 9 years of data only. A large data size would bring in better accuracy and stability.

The authors (Gandge and Sandhya, 2017) explore data mining techniques for predicting crop yield. They consider factors such as climatic conditions (temperature, rainfall), soil nutrients, and agronomical parameters. This paper summarizes the results obtained by various algorithms that are being used by various authors for crop yield prediction, with their accuracy and recommendation. In the paper they investigate papers that used Machine learning algorithms such as Multiple Linear Regression, Decision tree analysis and ID3, Support Vector Regression model, Three models used APAR, SEBAL, Carnegie Institution Stanford model, Neural Networks, C4.5 algorithm and decision tree, Harmonic Analysis of NDVI Time Series algorithm, Gaussian Processes, Relational cluster Bee Hive algorithm, K-Means algorithm for clustering And for classification Linear Regression, k-NN, ANN model, J48, LADTree, K-Nearest Neighbor (KNN) and Naive Bayes (NB) which were used several crops including Rice, Wheat, Soybeans, etc. They noticed that there is still scope for improvement in the result. During the study that was carried out, it was observed that the algorithm which is used by most of the authors does not use a unified approach where all the factors affecting the crop yield can be utilized simultaneously for predicting the crop yield. There is still further scope for improvement as the dataset which is considered is small in some cases. Therefore, the result can be improved by using a large dataset.

As the world's most populated country, food security is one major concern in the Indian government. In the paper (Satpathi et al., 2023) authors also discuss weather variables and how they can provide reliable predictions of crop yields using Indian data. In their study, five alternative models, viz., stepwise multiple linear regression (SMLR), an artificial neural network (ANN), the least absolute shrinkage and selection operator (LASSO), an elastic net (ELNET), and ridge regression, were compared in order to discover the best model for rice yield prediction. The generalized linear model (GLM), random forest (RF), cubist, and ELNET approaches were utilized to construct ensemble models from the outputs of the individual models. They have used Weather data including maximum temperature (T_{max}), minimum temperature (T_{min}), rainfall, relative humidity at 7:20 am (RH I) and 2:20 pm (RH II), and sunshine hours of twenty-one years (1998–2018) for districts Surguja, Raipur and Bastar in Chhattisgarh state in India. In the study, they recognized that the mean weekly temperature for the study regions during the rice-growing season varied from 23.7 to 31.0 ◦C, 21.2 to 29.6 ◦C, and 21.2 to 28.9 ◦C for Raipur, Surguja, and Bastar, respectively, temperatures 30–33 ◦C given ideal temperature for optimum rice yield. The authors noticed that among SMLR, ANN, LASSO, ELNET, and ridge regression, the performance of LASSO was good at calibration for all three districts.

In their study, (Xu et al., 2019), the authors employed both Random Forest and Support Vector Machine techniques to develop an integrated climatic assessment indicator (ICAI) specifically tailored for evaluating the climatic suitability of wheat production in Jiangsu Province, China. The ICAI considers the comprehensive effects of multiple meteorological factors, providing a holistic view of climatic conditions. To determine a reasonable division in classification, distribution detection of climatic yield is carried out and Monte Carlo simulations are applied for the Kolmogorov–Smirnov (KS) test. The generated indicator includes three values: yield loss, normal, and yield increment, with the spatial and temporal prediction accuracy from 67.86% to 100% in the test set for the Northern, Central, and Southern Jiangsu. The ICAI is used to estimate the past climatic suitability of winter wheat and the future suitability under global warming conditions in Jiangsu Province. The results show that the climate in the 1990s had more adverse effects on wheat production than the other two sub-periods in Northern and Southern Jiangsu. The researchers found that the climate in the 1990s had more adverse effects on wheat production than subsequent sub-periods in both Northern and Southern Jiangsu. The ICAI serves as a valuable tool for assessing wheat production resilience in the face of changing climatic conditions. The authors highlighted that, with the increasingly accurate weather forecasts, the developing techniques of ML in the big data era, as well as the emergence of the driving datasets based on regional climate models, ICAI will be more accurate to serve in extensive applications for various crops. The design method applied in ICAI can also be used to assess the suitability of crop growth in other areas besides Jiangsu.

The authors (González-Sanchez et al., 2014) discuss how Crop yield prediction is a significant challenge in agriculture, especially during the planning phase. Accurate yield estimation is crucial for informed decision-making. In this study, the authors compare the predictive accuracy of several machine learning (ML) techniques and linear regression for crop yield prediction across ten different crop datasets. This paper compares the predictive accuracy of ML and linear regression techniques for crop yield prediction in ten crop datasets. Multiple linear regression, M5-Prime regression trees, perceptron multilayer neural networks, support vector regression, and k-nearest neighbor methods were ranked. Four accuracy metrics were used to validate the models: the root mean square error (RMS), root relative square error (RRSE), normalized mean absolute error (MAE), and correlation factor (R). Real data from an irrigation zone in Mexico were used for building the models. Models were tested with samples of two consecutive years. The results show that M5- Prime and k-nearest neighbor techniques obtain the lowest average RMSE errors (5.14 and 4.91), the lowest RRSE errors (79.46% and 79.78%), the lowest average MAE errors (18.12% and 19.42%), and the highest average correlation factors (0.41 and 0.42). Since M5-Prime achieves the largest number of crop yield models with the lowest errors, it is a very suitable tool for massive crop yield prediction in agricultural planning.

2.6 Deep Learning

In this study of (Nevavuori et al., 2019), the authors apply Convolutional Neural Networks (CNNs) a deep learning methodology known for outstanding performance in image classification tasks to build a model for crop yield prediction. They utilize Normalized Difference Vegetation Index (NDVI) and RGB data acquired from Unmanned Aerial Vehicles (UAVs). By leveraging these data sources, the model aims to estimate crop yield accurately. The Authors used the Adadelta training algorithm, L2 regularization with early stopping, and a CNN with 6 convolutional layers, The mean absolute error (MAE) in yield prediction of 484.3 kg/ha and the mean absolute percentage error (MAPE) of 8.8% was achieved for data acquired during the early period of the growth season (i.e., in June of 2017, growth phase < 25%) with RGB data. When using data acquired later in July and August of 2017 (growth phase > 25%), an MAE of 624.3 kg/ha (MAPE: 12.6%) was obtained. Significantly, the CNN architecture performed better with RGB data than the NDVI data. The integration of deep learning techniques with remote sensing data holds promises for optimizing agricultural practices and enhancing food security.

(Han et al., 2022) investigates the critical relationship between rice yield and climate variables. Rice, as a staple crop globally, holds immense significance for food security. China’s prominence as the largest rice producer and consumer underscores the need for accurate yield predictions. Traditional methods, relying on destructive field measurements, fall short of capturing the dynamic interplay of climate factors. Enter deep learning, a powerful tool for image recognition and processing. By combining remote sensing images with deep learning, the study aims to simulate the four key factors affecting rice yield. The research explores various regression models, both linear and nonlinear, to predict actual rice yield. In this experiment, a total of 207 paddy plots in the field were measured. Each plot planted 20 rice plants of the same variety, but the rice varieties in different plots were

different. Each rice plot separately extracts the image features of the rice ear plot, the detailed image features of a single rice ear, and the seed test features of a single grain. Then, the regression equation between image features and plot total output is constructed, to realize the purpose of rice plot yield prediction. The study demonstrates the potential of deep learning and remote sensing in improving rice yield prediction. By integrating advanced image processing techniques with comprehensive regression models, the research provides a promising approach to enhancing agricultural productivity and food security. This method not only reduces the need for physical labor but also offers higher precision and scalability in yield estimation.

Saeed and Wang discuss the impact of genotype, environment, and their interactions on crop yield using the dataset released in the 2018 Syngenta Crop Challenge. They discovered that when using the deep neural network approach their model shows a higher prediction success rate with root-mean-square-error (RMSE) 12% of the average yield and 50% standard deviation for the validation dataset using predicted weather data. With perfect weather data, the RMSE would be reduced to 11% of the average yield and 46% of the standard deviation. Their research indicated that deep neural networks outperform popular machine learning methods such as regression trees, Lasso, and shallow neural networks.

Khaki and Wang (Khaki and Wang, 2019) discuss the impact of genotype, environment, and their interactions on crop yield using the dataset released in the 2018 Syngenta Crop Challenge. They investigated the use of deep neural networks (DNNs) for crop yield prediction. The approach used deep neural networks to make yield predictions (including yield, check yield, and yield difference) based on genotype and environment data. The carefully designed deep neural networks were able to learn nonlinear and complex relationships between genes, environmental conditions, as well as their interactions from historical data and make reasonably accurate predictions of yields for new hybrids planted in new locations with known weather conditions. The performance of the model was found to be relatively sensitive to the quality of weather prediction, which suggested the importance of weather prediction techniques. Their study found that DNNs outperform other popular methods such as Lasso, shallow neural networks (SNNs), and regression trees (RTs). This suggests that DNNs can be a powerful tool for capturing the complex relationships between multiple factors that influence crop yield.

2.7 Comparison between several research papers and their used algorithms

Author, Year	Research Paper	Techniques used
(Amaratunga et al., 2020)	Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data	Artificial Neural Networks (Levenberg–Marquardt, Bayesian Regularization, and Scaled Conjugated Gradient)
(Wickramasinghe et al., 2020)	Artificial Neural Network Approach for Paddy Yield Prediction	Artificial Neural Networks (Levenberg–Marquardt)
(Wickramasinghe et al., 2021)	Modeling the Relationship between Rice Yield and Climate Variables Using Statistical and Machine Learning Techniques	Artificial Neural Networks, Support Vector Machine Regression, Multiple Linear Regression, Gaussian Process Regression, Power Regression
(Ekanayake et al., 2022)	Development of Crop-Weather Models Using Gaussian Process Regression for the Prediction of Paddy Yield in Sri Lanka	Gaussian Process Regression
(Muthusinghe et al., 2018)	Towards Smart Farming: Accurate Prediction of Paddy Harvest and Rice Demand	Recurrent Neural Network, Long Short-Term Memory
(Gandge and Sandhya, 2017)	A Study on Various Data Mining Techniques for Crop Yield Prediction	Multiple Linear Regression, Decision tree analysis, ID3, Support Vector Regression model, Neural Networks, K-Means algorithm
(Ruß et al., 2008)	Data Mining with Neural Networks for Wheat Yield Prediction	Neural Networks
(Han et al., 2022)	Research on Rice Yield Prediction Model Based on Deep Learning	Convolutional neural network
(Baral et al., 2011)	Rice crop yield prediction using artificial neural networks	Neural Networks
(González-Sanchez et al., 2014)	Predictive ability of machine learning methods for massive crop yield prediction	M5-prime regression tree, k-nearest neighbor, support vector machine
(Khaki and Wang, 2019)	Crop Yield Prediction Using Deep Neural Networks	Neural Networks
(Nevavuori et al., 2019)	Crop yield prediction with deep convolutional neural networks	Convolutional neural network
(Gandhi et al., 2016)	Rice Crop Yield Prediction in India Using Support Vector Machines	Support Vector Machines

(Ramesh, n.d.)	ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES	Multiple Linear Regression
(Xu et al., 2019)	Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China	Random Forest, Support Vector Machine

Table 1: Research Paper technique comparison table

2.8 Most used machine learning algorithms

- Neural Networks
- Support Vector Machine
- Linear Regression
- K-Means algorithm

2.9 Conclusion

The literature on rice yield prediction in Sri Lanka concludes that machine learning (ML) and artificial intelligence (AI) techniques have significant potential to improve agricultural productivity and food security. Traditional methods, which relied on farmers' experience and manual data collection, are quickly being replaced by advanced predictive models that utilize large datasets and sophisticated algorithms. Several studies have demonstrated the efficacy of various ML models, including Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Gaussian Process Regression (GPR), and deep learning techniques, in predicting rice yields under varying climatic conditions. These models have shown considerable accuracy improvements, with GPR and ANNs particularly standing out due to their high predictive performance. For instance, ANNs optimized with Particle Swarm Optimization have yielded low error rates in predicting paddy yields, while GPR models have effectively captured the complex relationships between weather variables and crop performance, outperforming other methods in statistical assessments. Furthermore, the integration of non-climatic factors such as soil type, fertilizer consumption, and remote sensing data with climatic variables has also been explored, enhancing the precision of yield predictions. Studies like those by Chandra et al. (2019) illustrate the benefits of incorporating diverse datasets, although they also highlight the need for large and varied data samples to improve model stability and accuracy. The application of deep learning methods, such as Convolutional Neural Networks (CNNs), has shown promise in utilizing UAV-acquired imagery for yield prediction, further advancing the potential for precision agriculture. These techniques have not only improved yield estimation accuracy but have also offered scalable and less labor-intensive solutions.

Overall, the research underscores the transformative impact of ML and AI on rice yield prediction in Sri Lanka, pointing towards a future where these technologies could significantly bolster the country's agricultural resilience and food security. The continuous development and refinement of these predictive models, supported by comprehensive data collection and integration, will be crucial in addressing the challenges posed by climate variability and ensuring sustainable agricultural practices.

CHAPTER 3 - METHODOLOGY

3.1 Introduction

This chapter describes the design approach for the aforementioned problem. The research objectives are centered around evaluating and comparing the performance of the CNN model's ability to predict paddy yield and to improve decision-making processes related to national food security. The project design will follow an empirical research methodology.

3.2 Problem identification and motivation

The primary aim of this project is to build a CNN model that can predict rice yield when given necessary weather parameters. The motivation for this research stems from the need to enhance the decision-making process regarding national food security, specifically concerning rice yield.

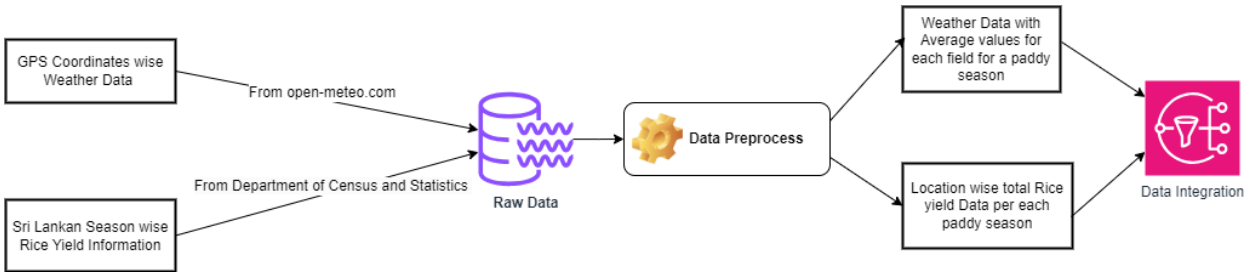
3.3 Formulating Research Objectives

The primary objective is to evaluate and compare the performance of different deep-learning algorithms for rice yield prediction. This objective can be further broken down into specific sub-objectives.

3.3 Data collection and preprocessing

The data set which will be used for training the CNN model will consist of two data sets.

- Paddy Yield data of relevant areas
- Historical Meteorological data of the relevant areas



3.4 Paddy Yield data

The paddy data will be sourced from the Department of Census and Statistics, Sri Lanka's official website (http://www.statistics.gov.lk/Agriculture/StaticalInformation/Paddy_Statistics#gsc.tab=0)

From the Paddy yield statics below two stats were sourced for model training.

- Annual Paddy yield per area (Metric Tons)
- Total area used for paddy cultivation (Hectares)

From year 2004 to 2023 historical paddy yield data will be collected for the area below.

- Colombo
- Gampaha
- Kalutara
- Kandy
- Matale
- Galle
- Matara
- Hambantota
- Jaffna
- Mannar
- Trincomalee
- Kurunegala
- Puttalam
- Badulla
- Ratnapura
- Nuwara Eliya
- Batticaloa
- Anuradhapura
- Monaragala
- Kegalle
- Polonnaruwa
- Ampara
- Kilinochchi
- Mullaitivu
- Vavuniya

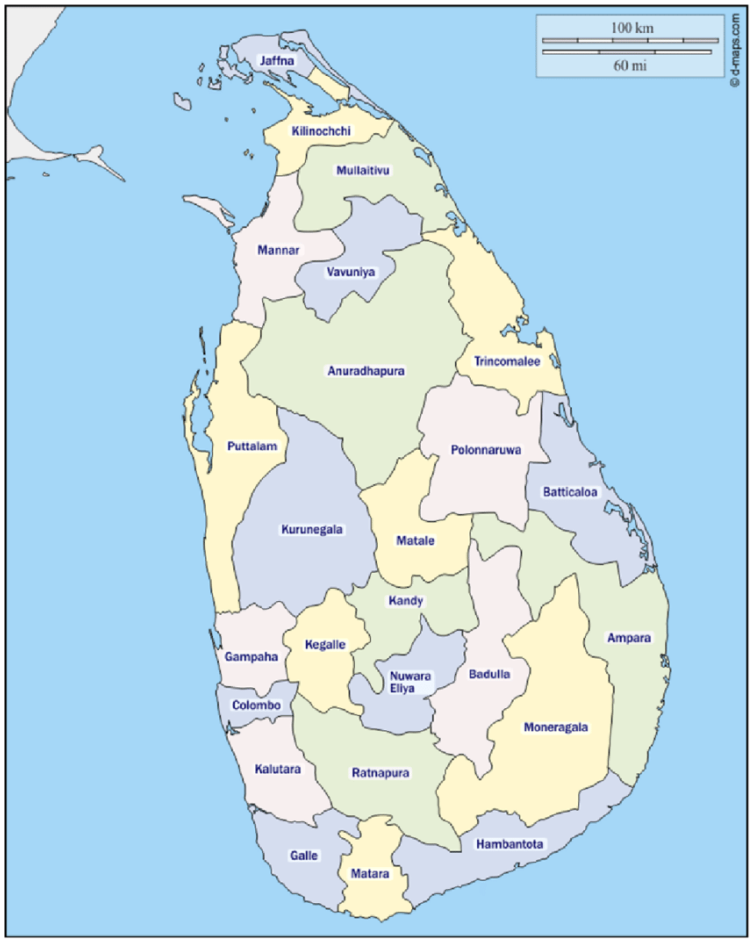


Figure 1- Areas which involved in Paddy Harvest

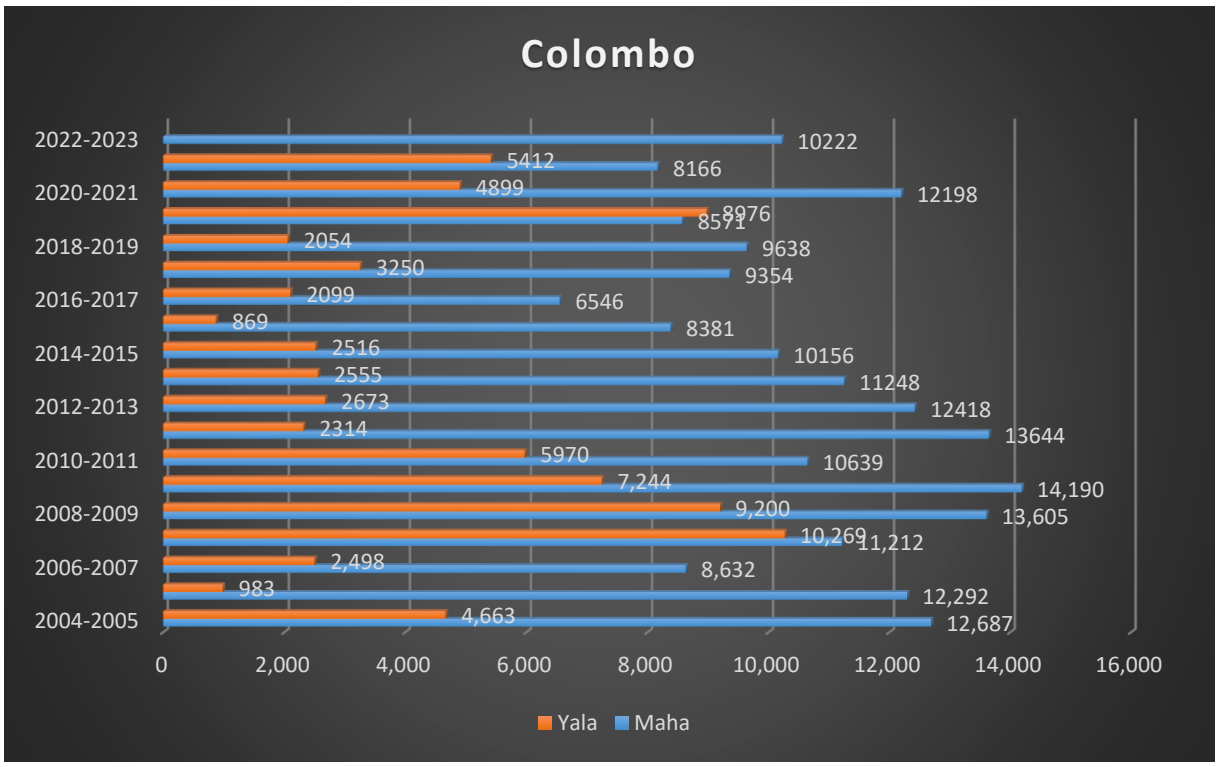


Figure 2- Colombo Season-wise total paddy yield.

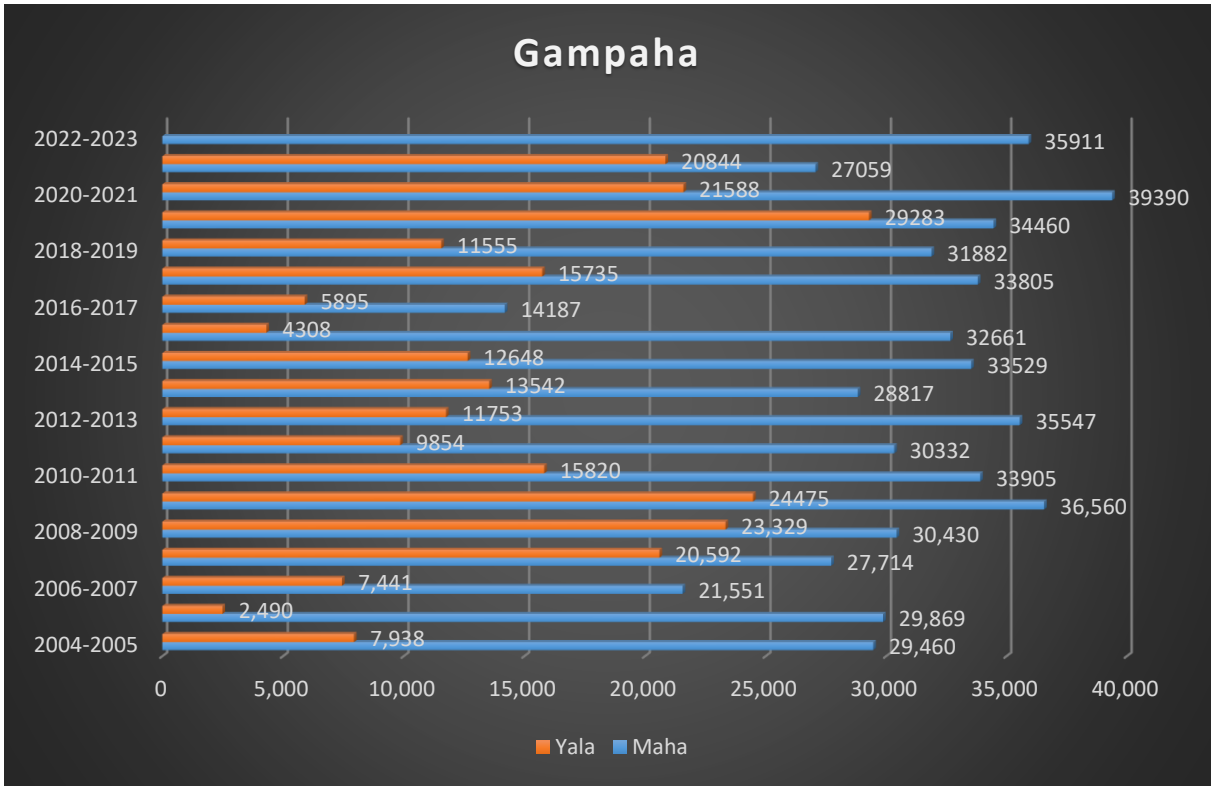


Figure 3- Gampaha Season-wise total paddy yield.

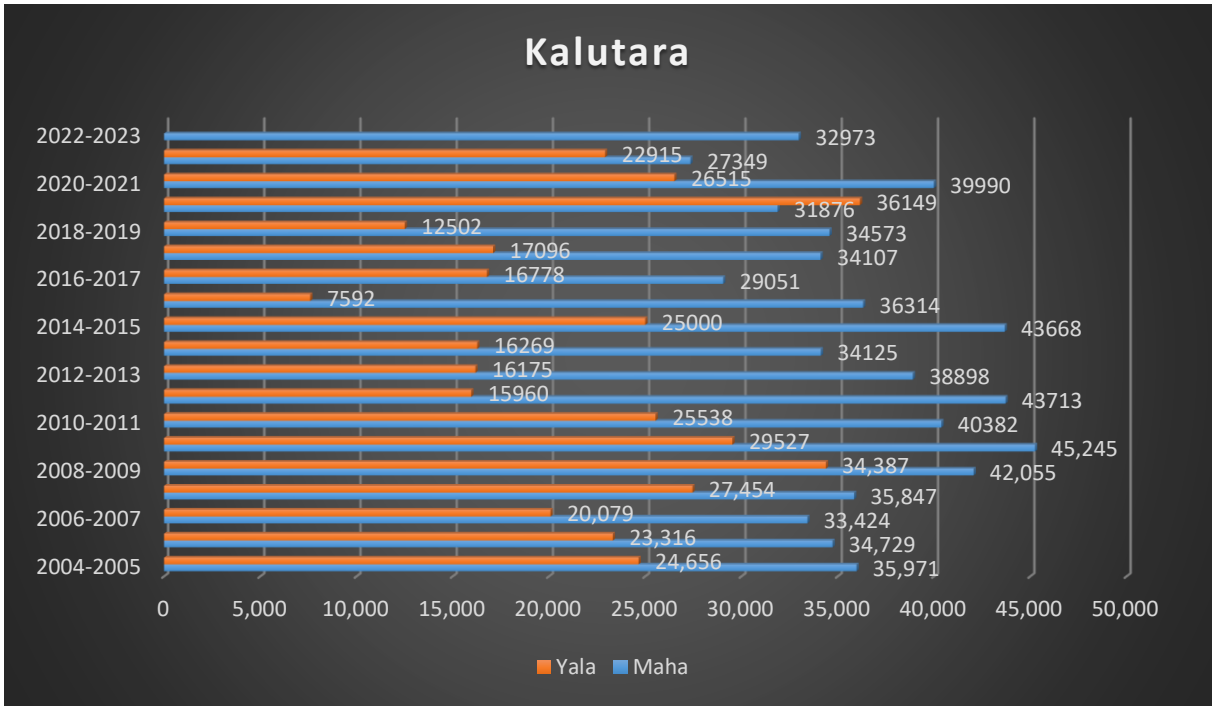


Figure 4 - Kalutara Season-wise total paddy yield.

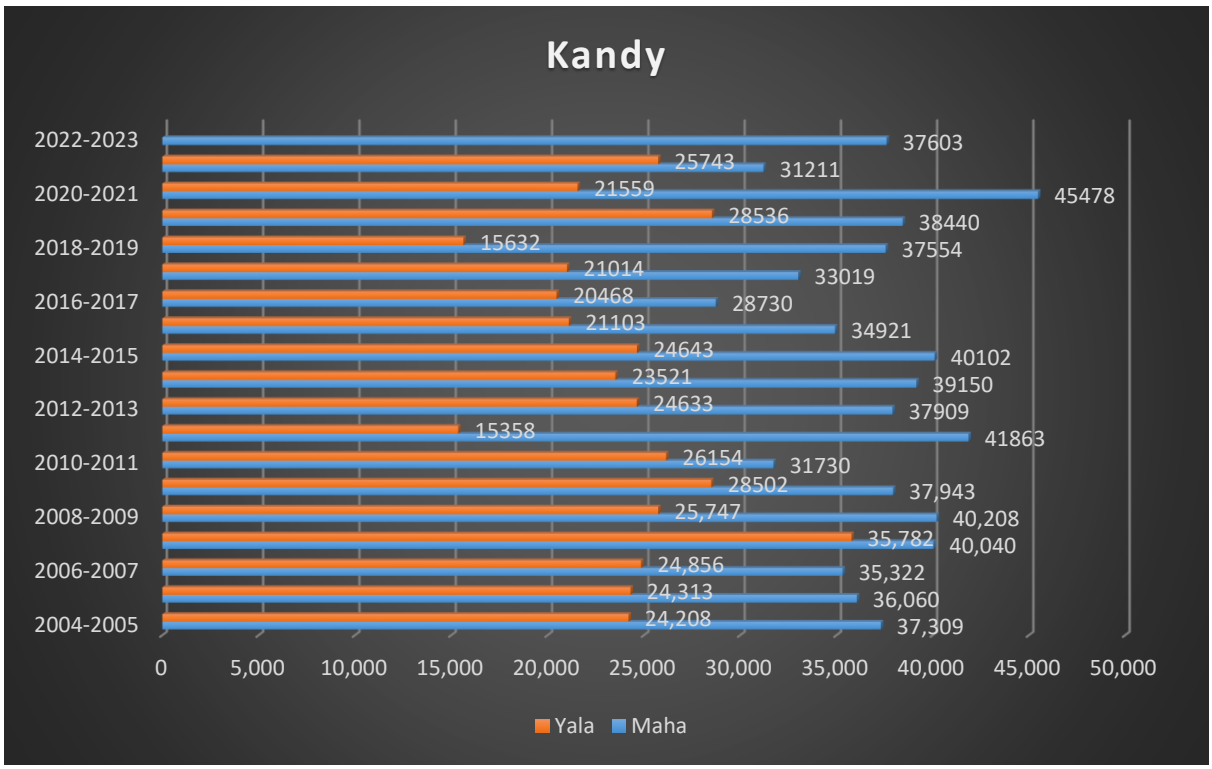


Figure 5 - Kandy season-wise total paddy yield.

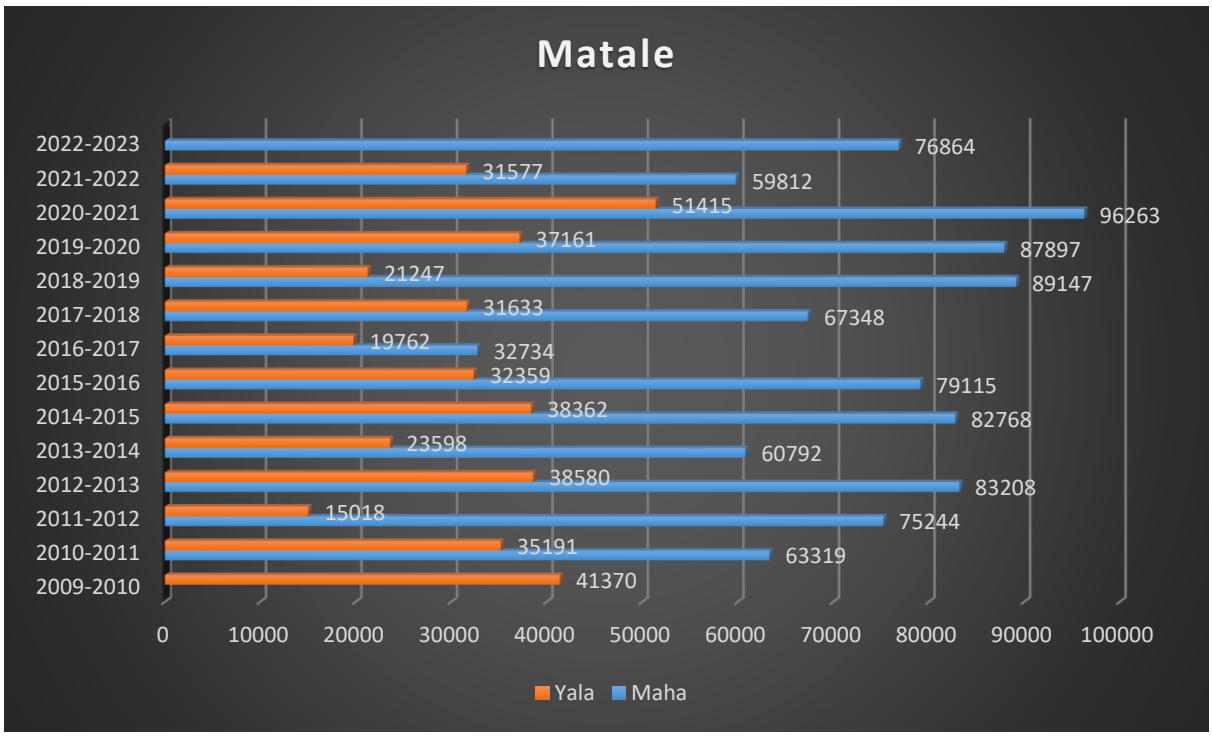


Figure 6 - Matale Season-wise total paddy yield.

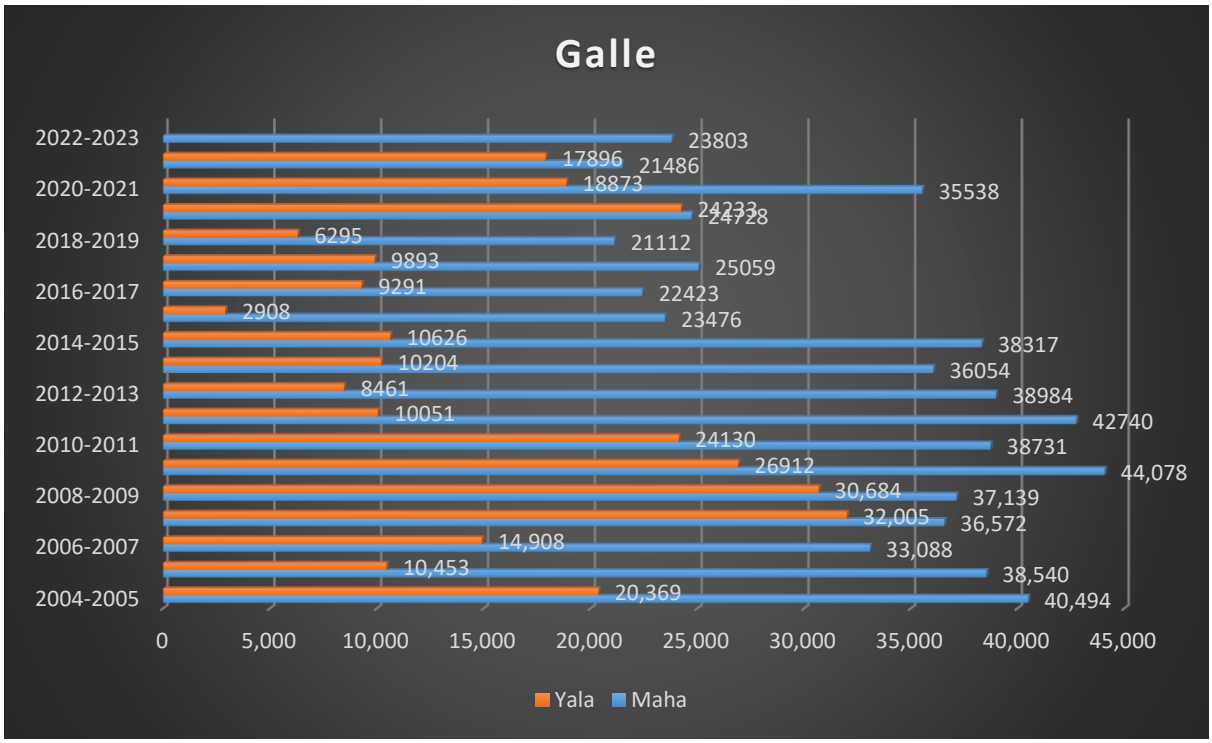


Figure 7 - Galle season-wise total paddy yield.

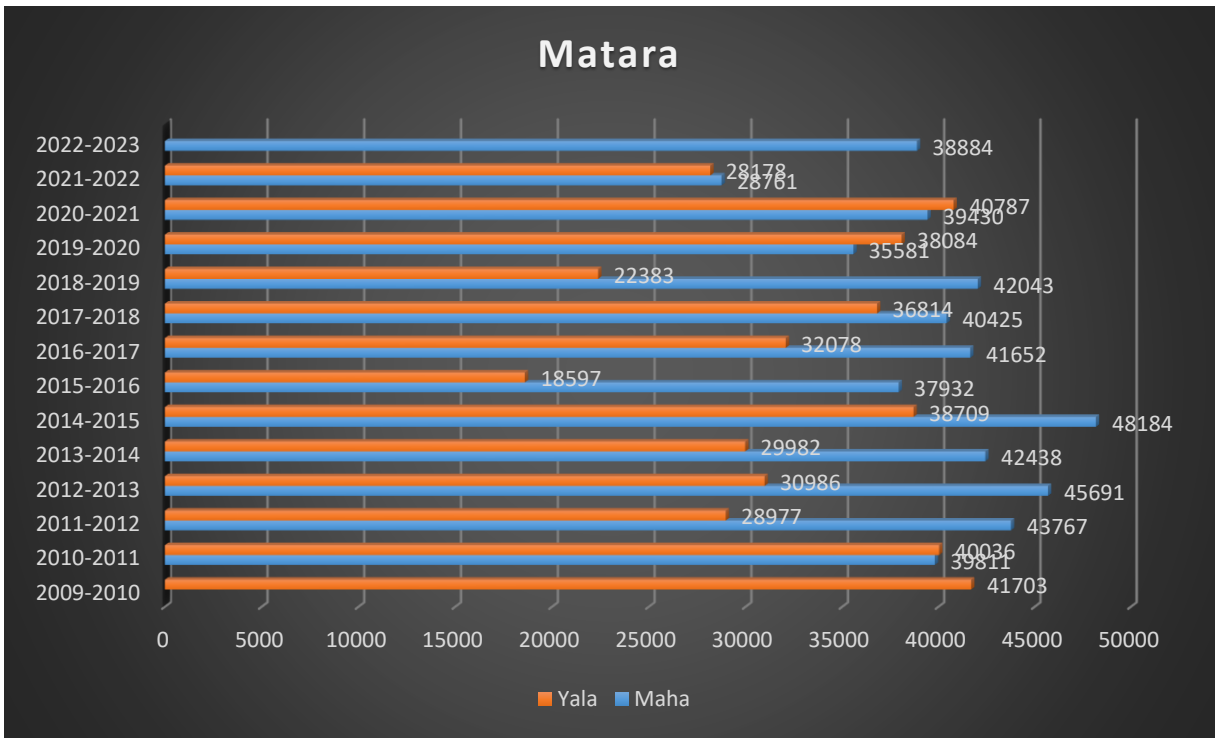


Figure 8 – Matara Season-wise total paddy yield.

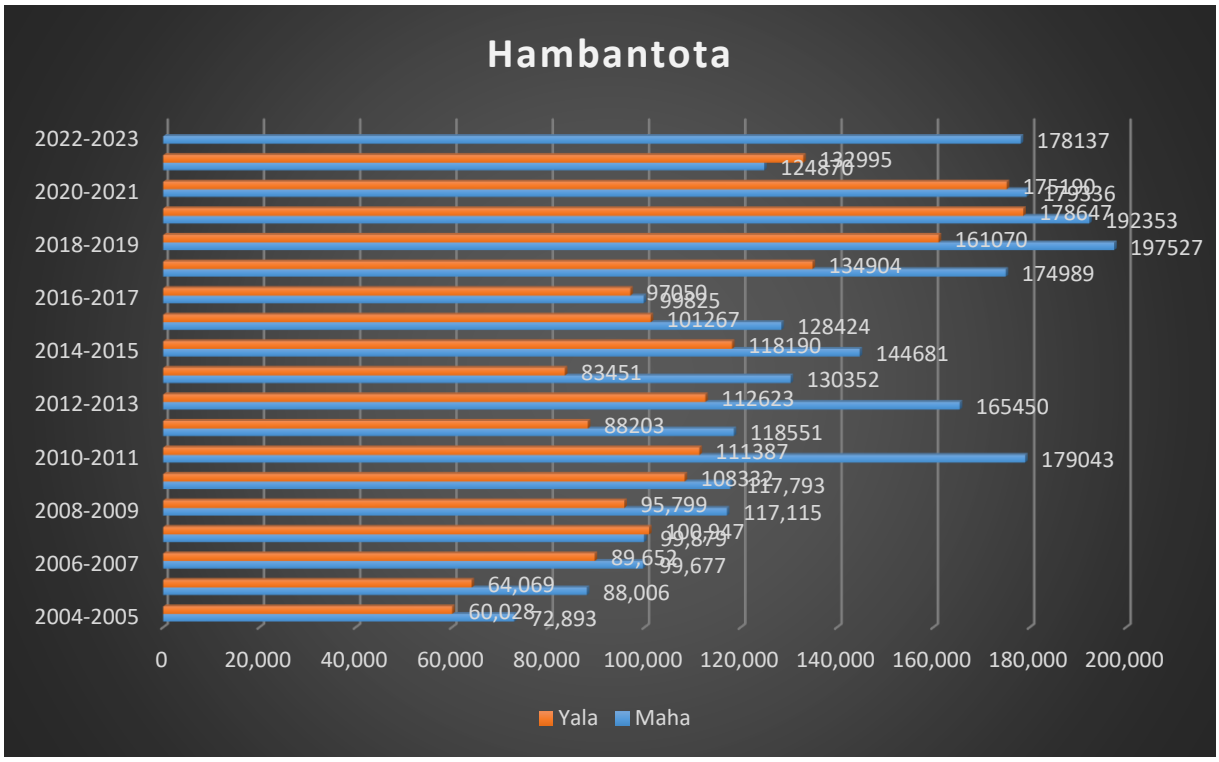


Figure 9 - Hambantota season-wise total paddy yield.

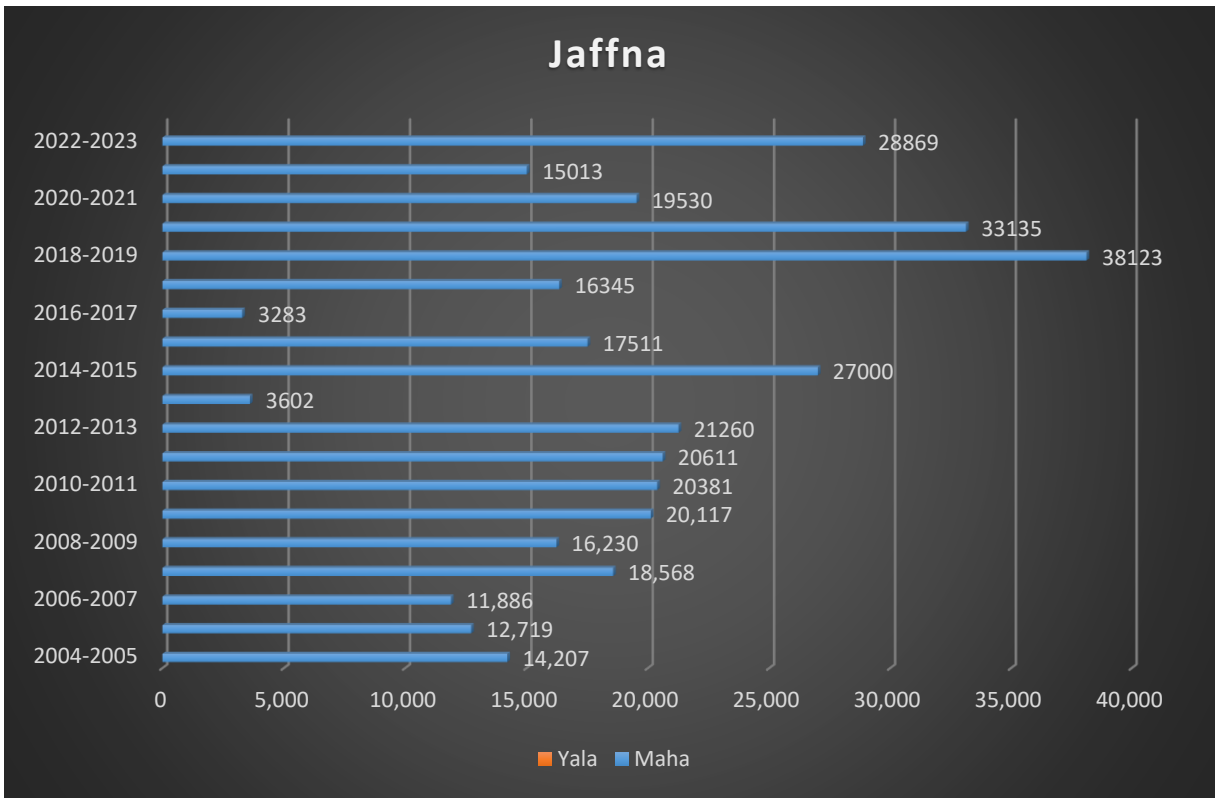


Figure 10 - Jaffna Season-wise total paddy yield.

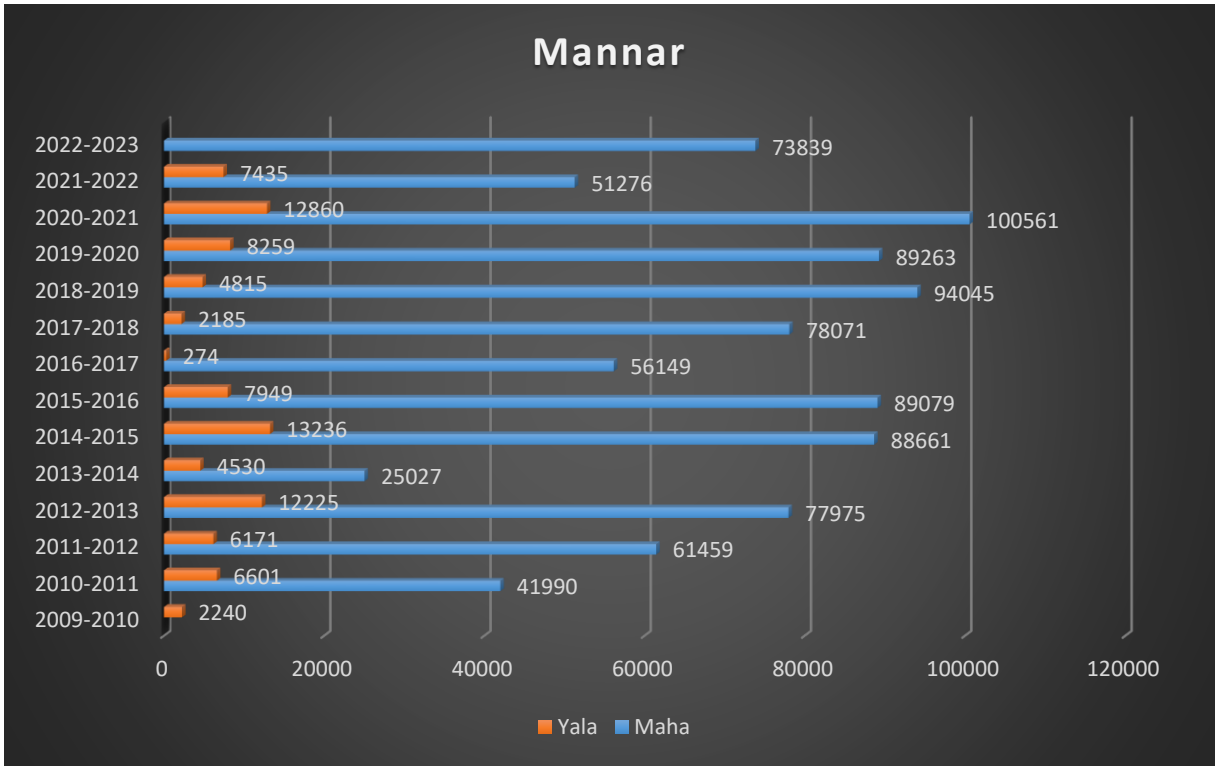


Figure 11 - Mannar Season-wise total paddy yield.

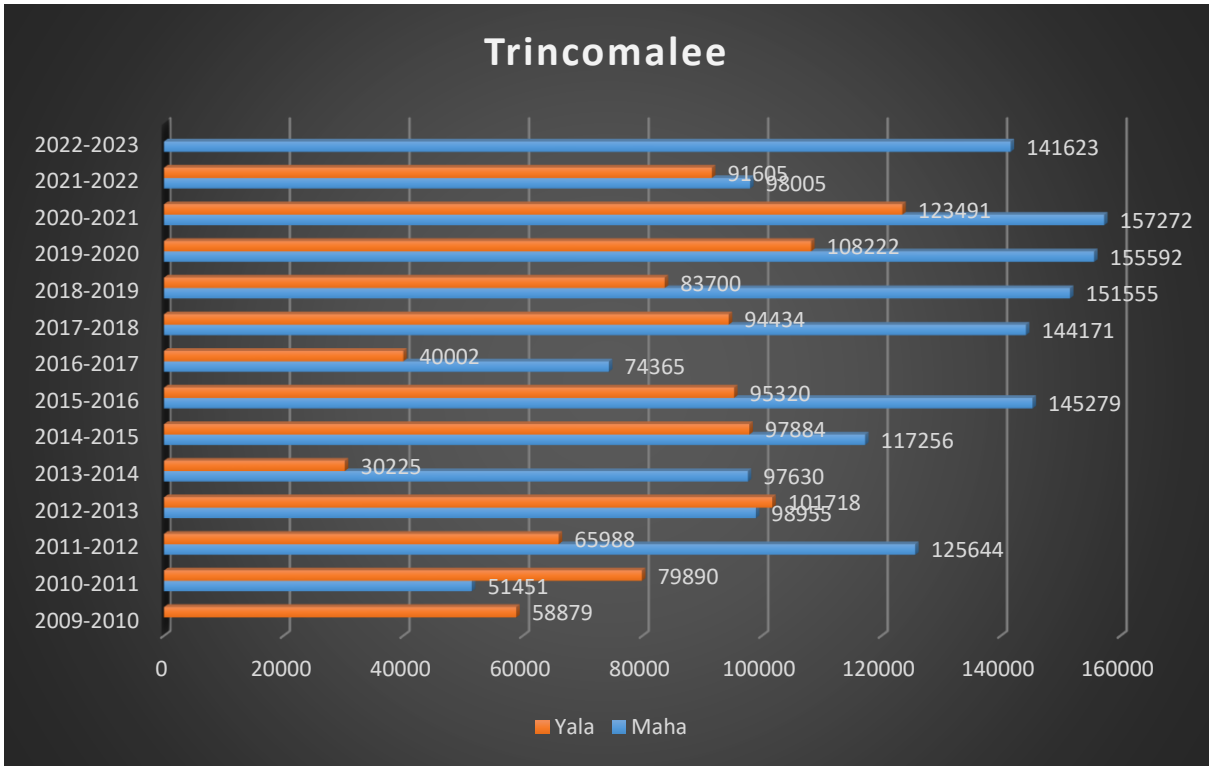


Figure 12 - Trincomalee Season-wise total paddy yield.

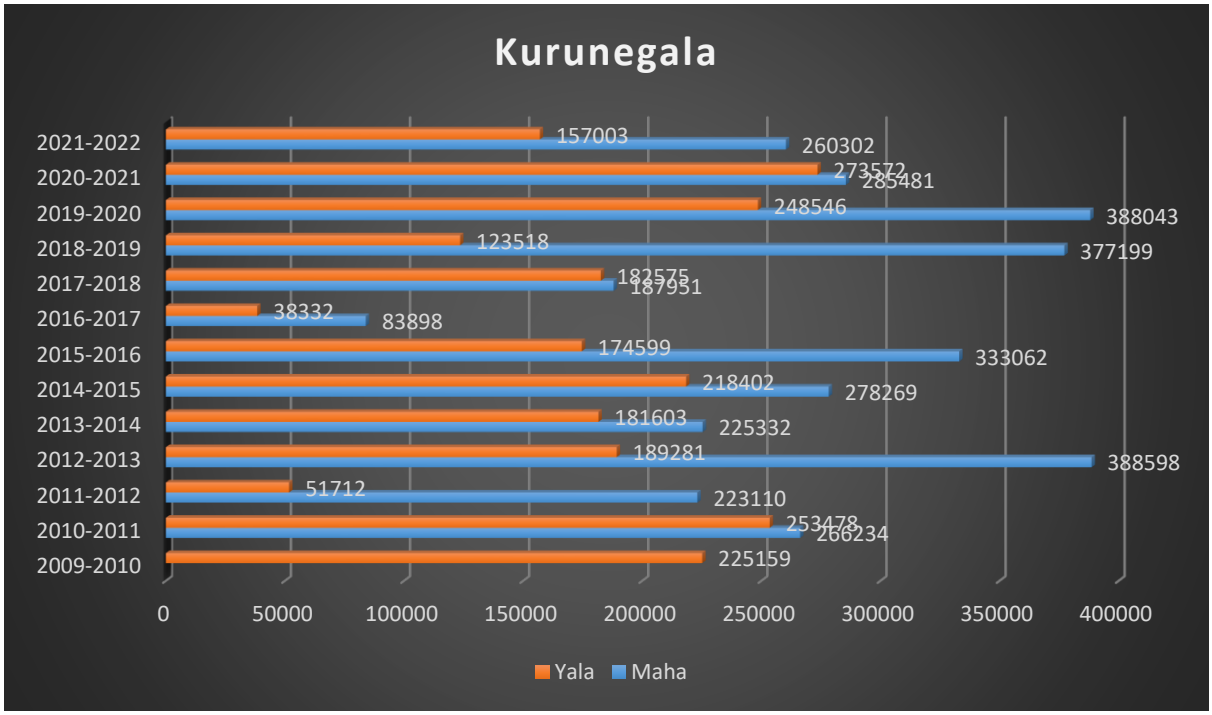


Figure 13 - Kurunegala Season-wise total paddy yield.

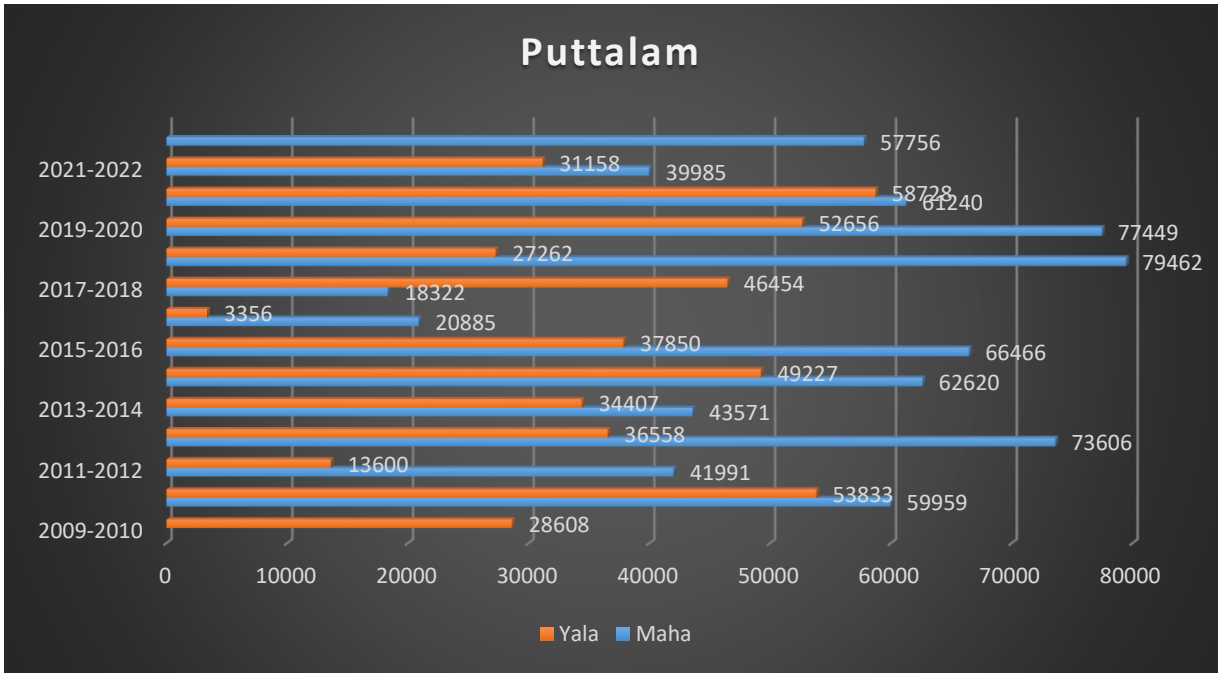


Figure 14 - Puttalam Season-wise total paddy yield.

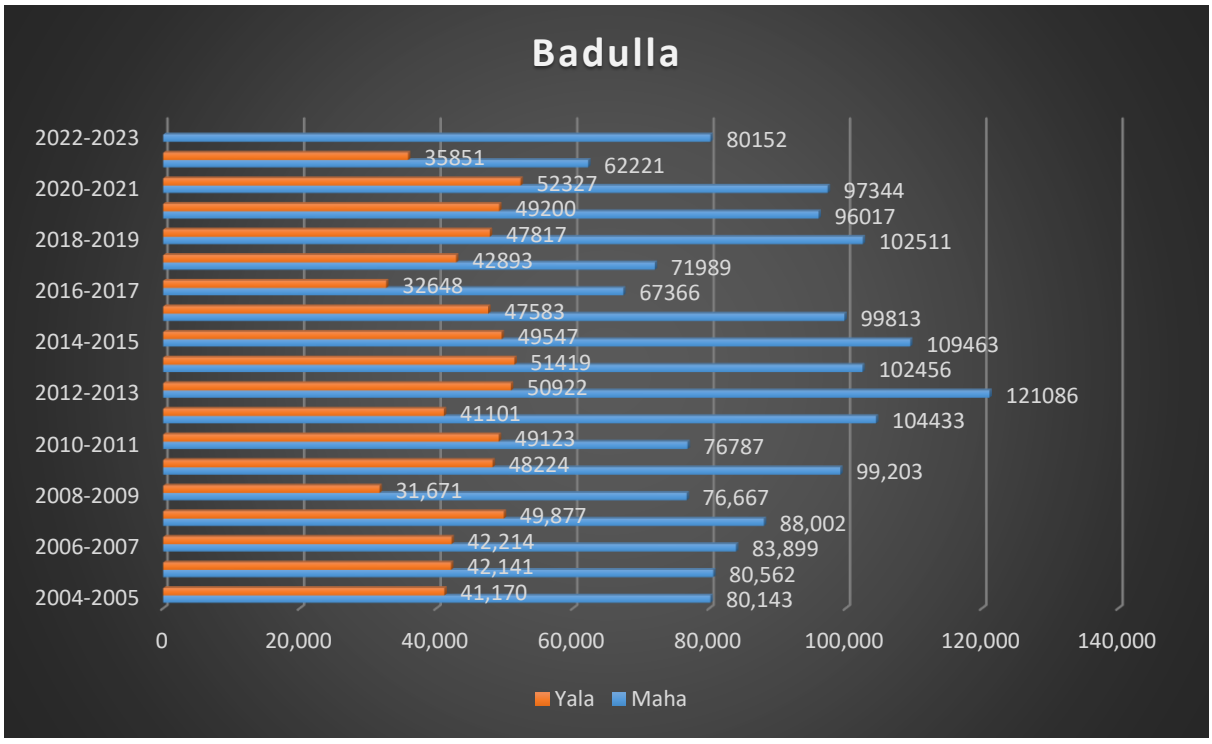


Figure 15 - Badulla Season-wise total paddy yield.

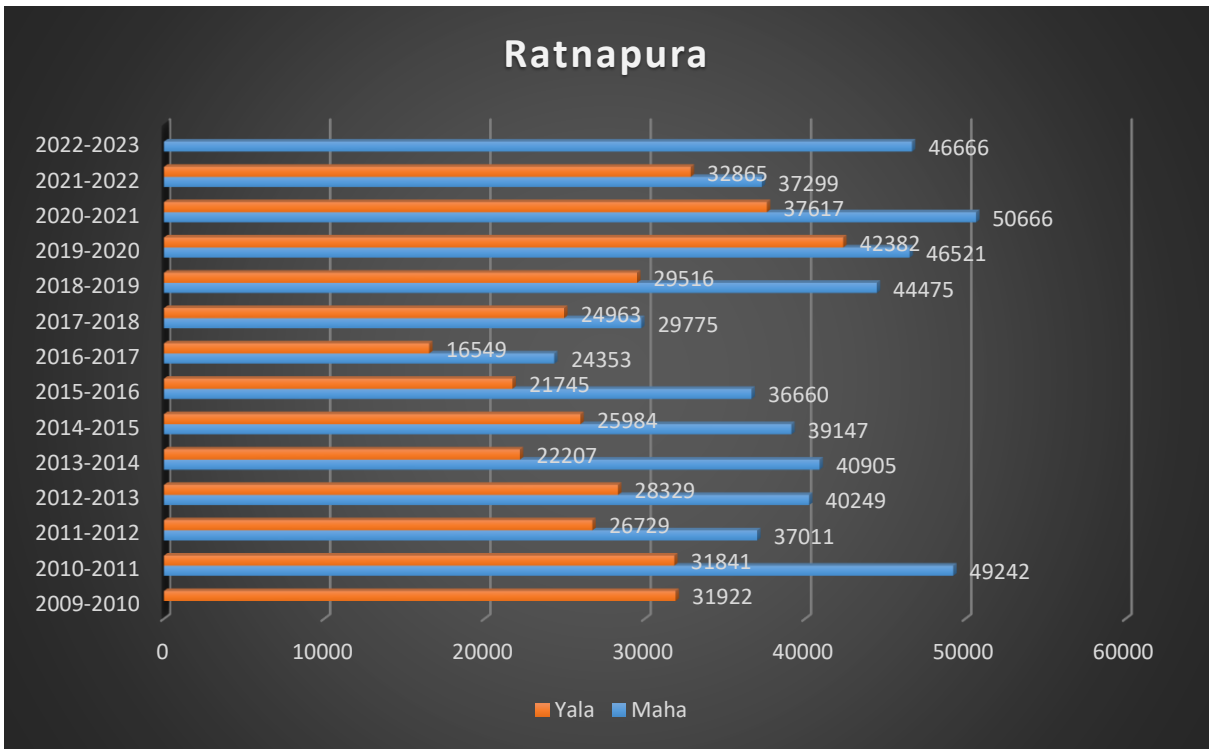


Figure 16 - Ratnapura Season-wise total paddy yield.

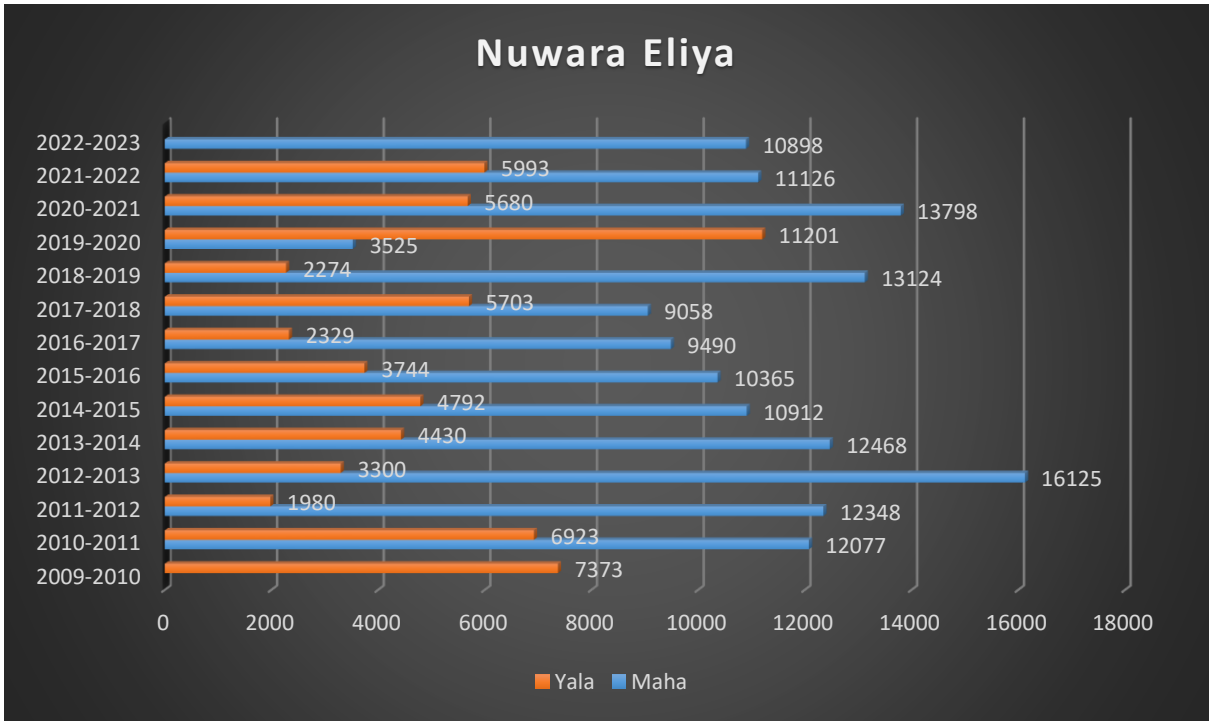


Figure 17 - Nuwara Eliya Season-wise total paddy yield.

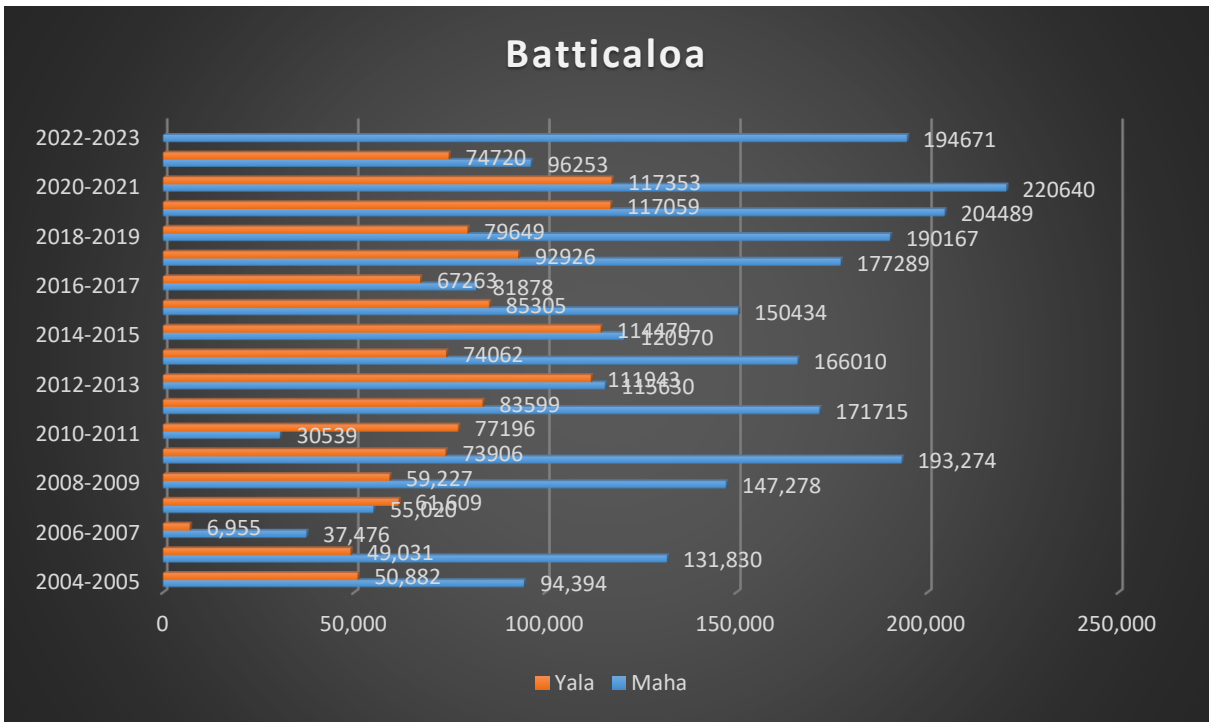


Figure 18 - Batticaloa Season-wise total paddy yield.

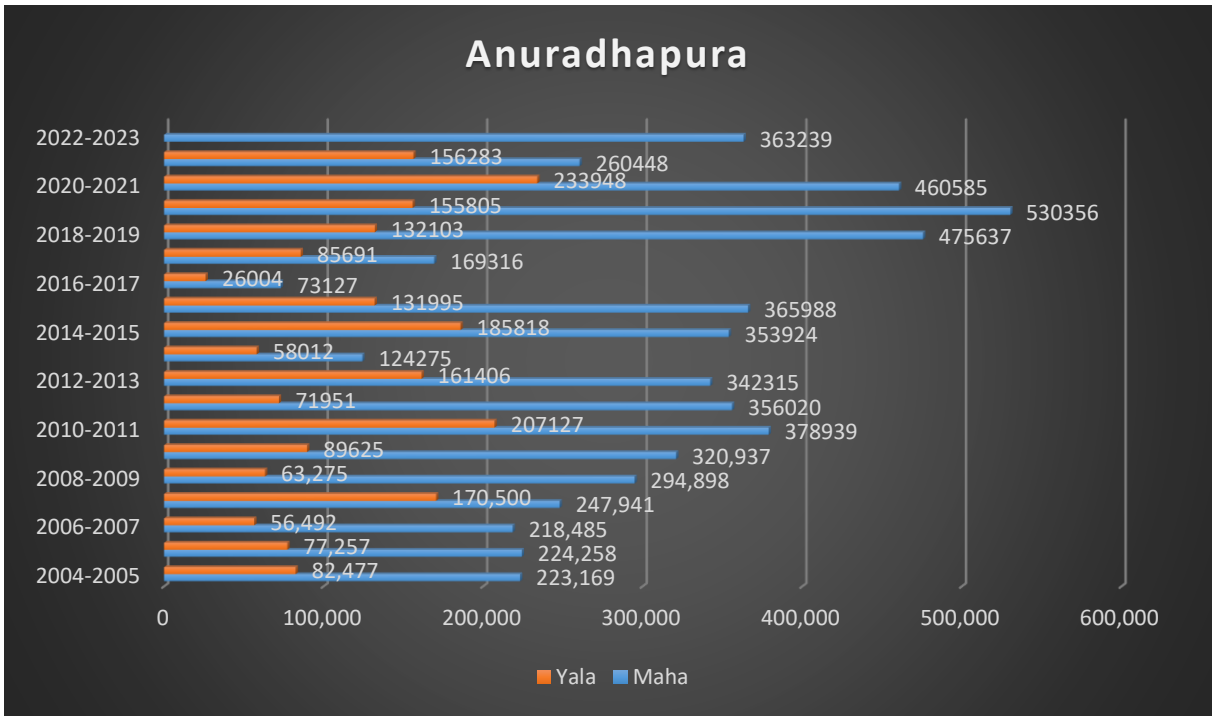


Figure 19 - Anuradhapura Season-wise total paddy yield.

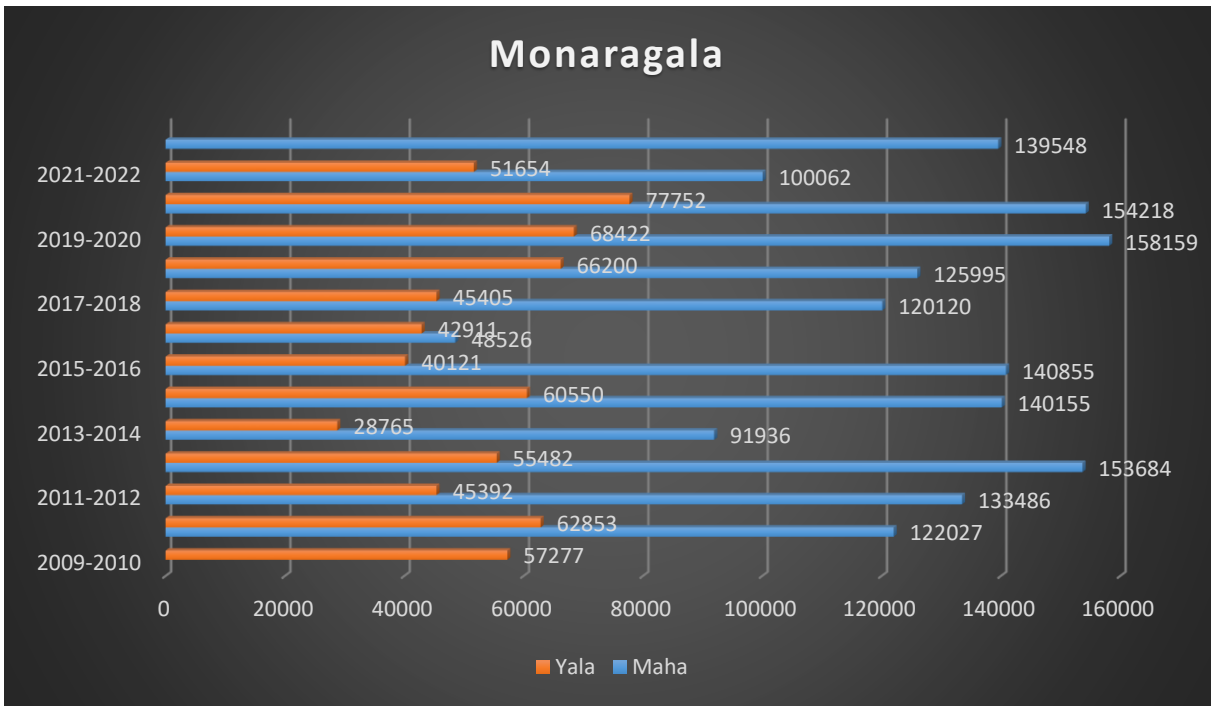


Figure 20 - Monaragala Season-wise total paddy yield.

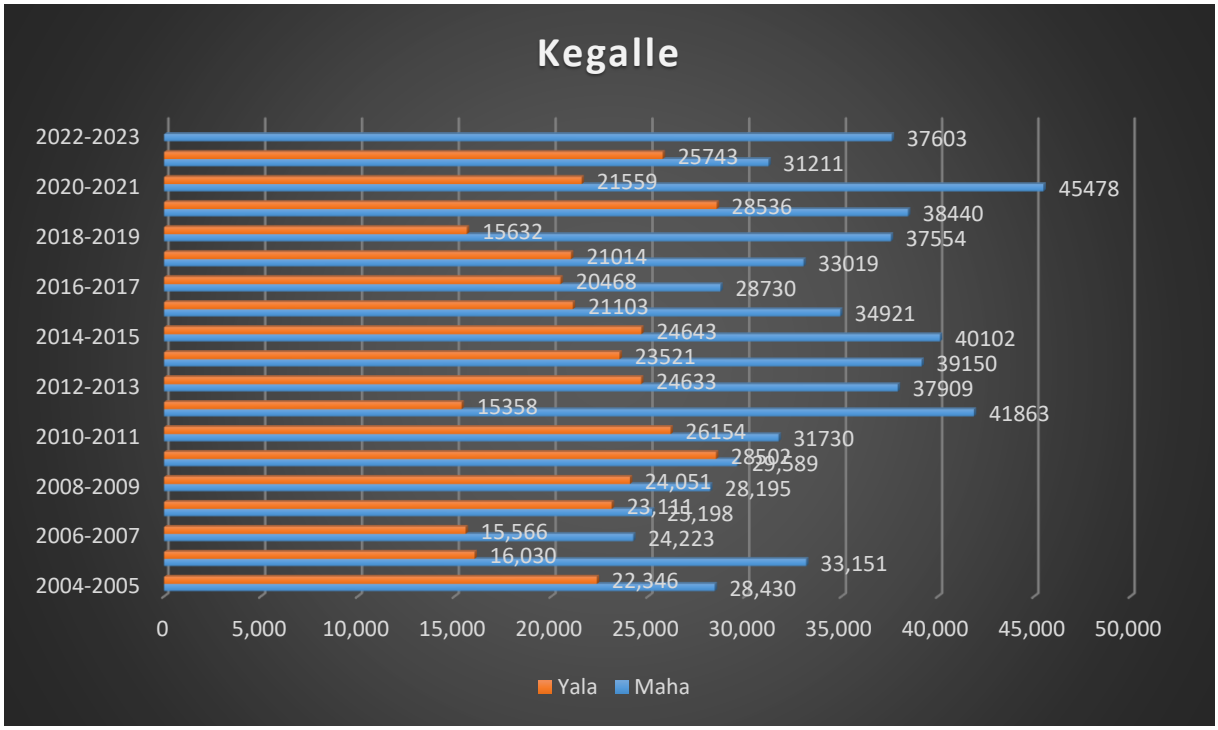


Figure 21 - Kegalle Season-wise total paddy yield.

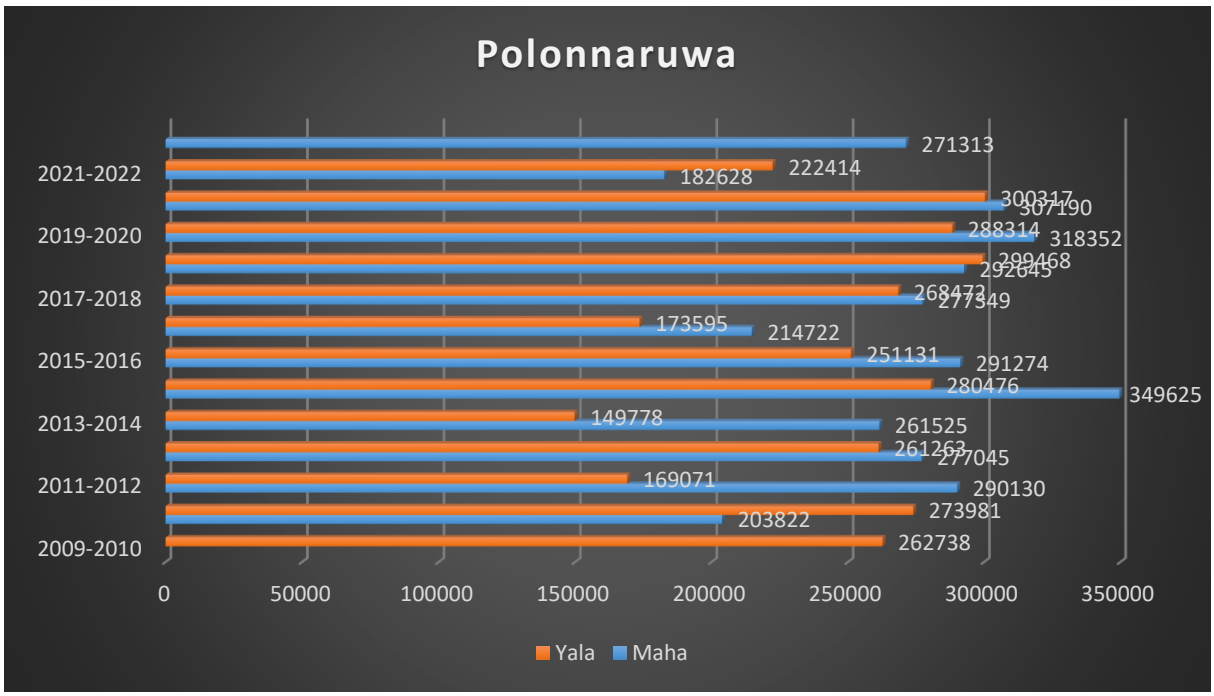


Figure 22 - Polonnaruwa Season-wise total paddy yield.

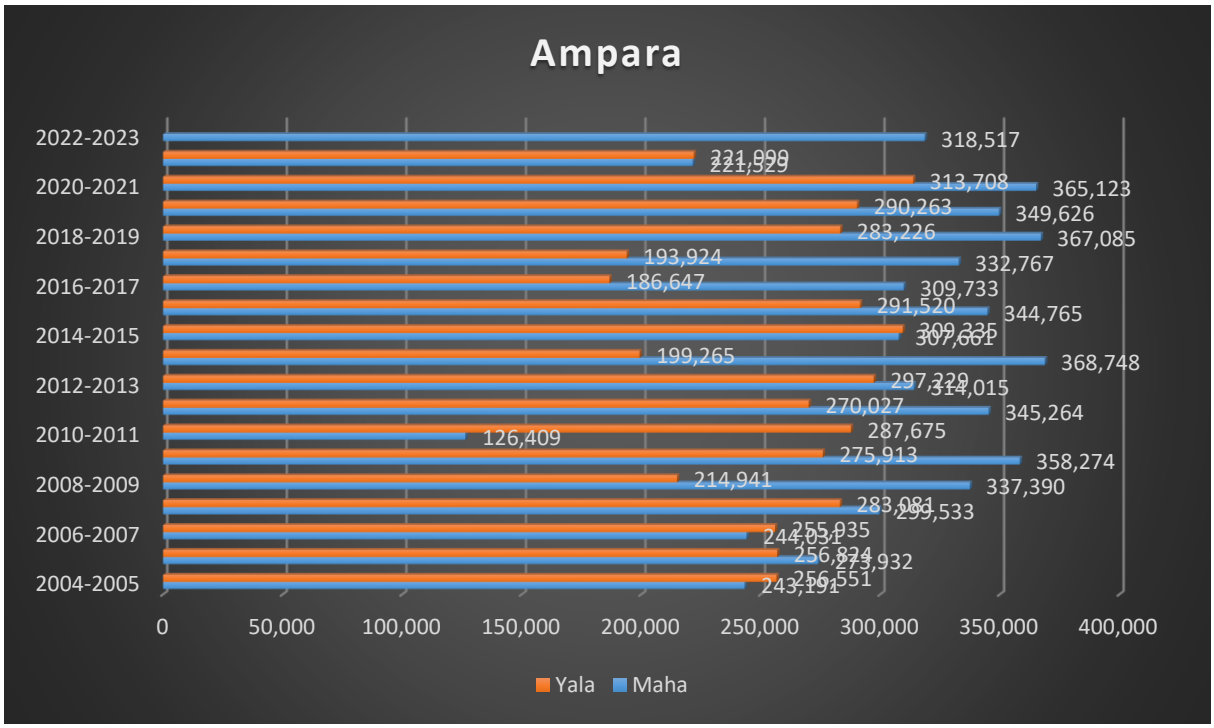


Figure 23 - Ampara season-wise total paddy yield.

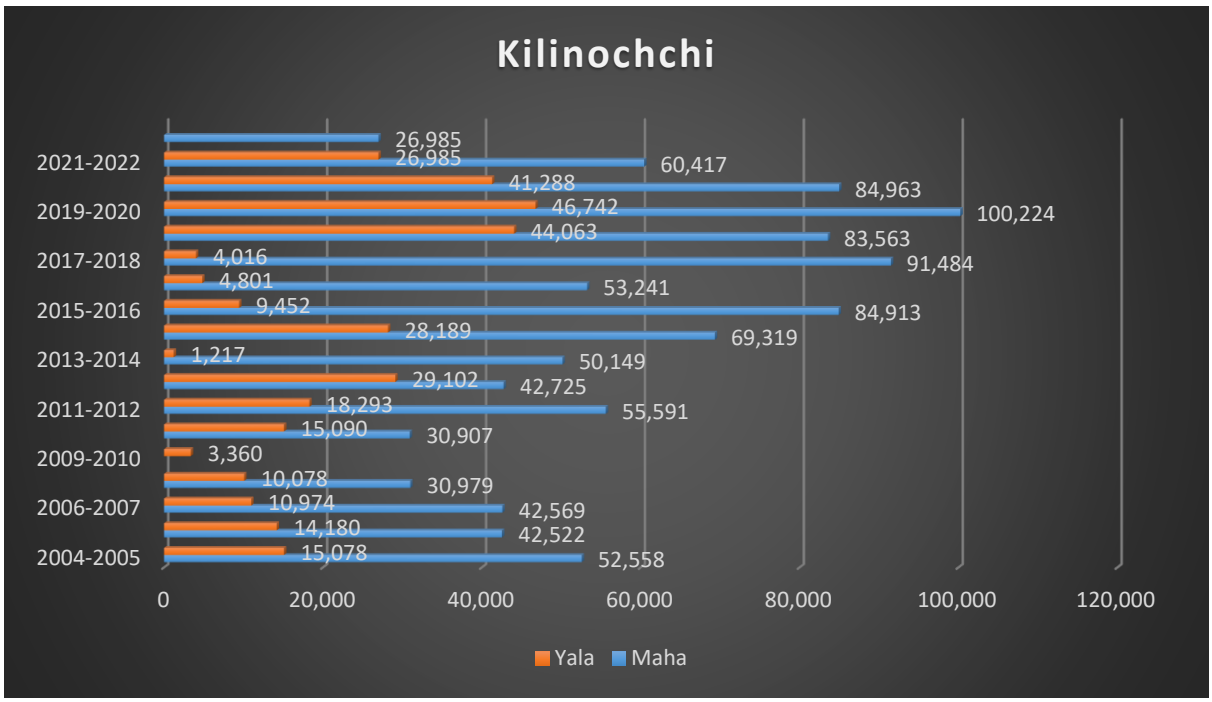


Figure 24 - Kilinochchi Season-wise total paddy yield.

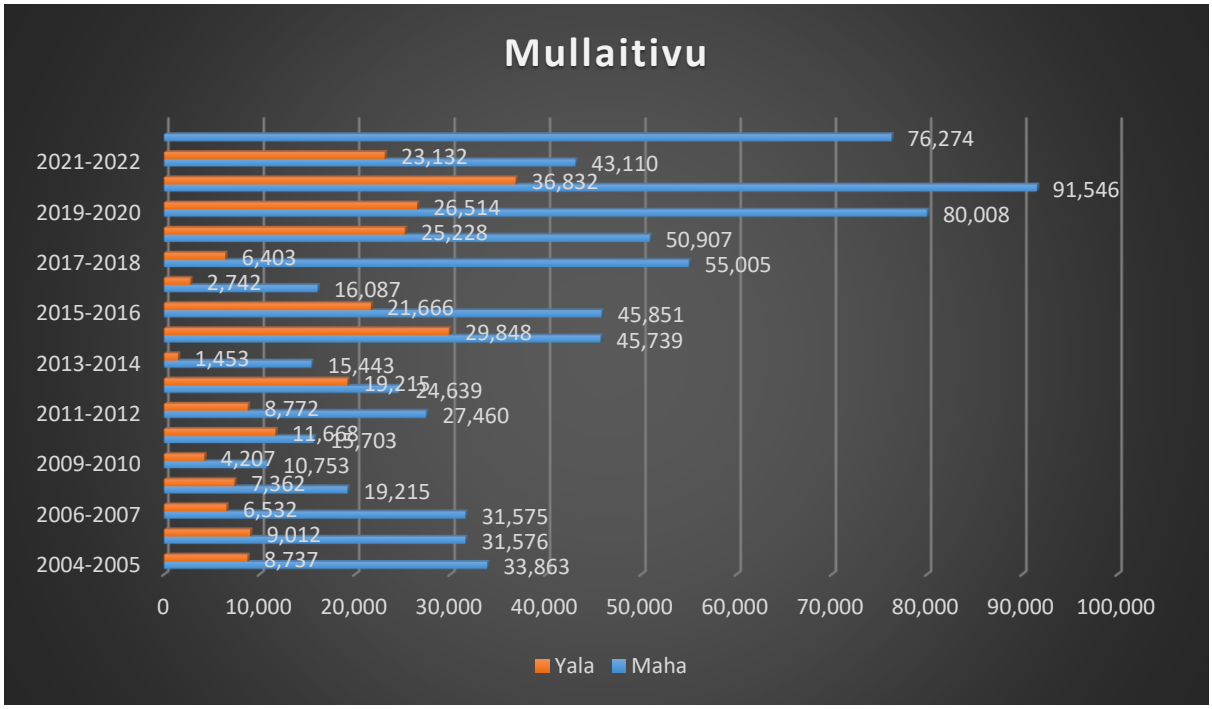


Figure 25 - Mullaitivu Season-wise total paddy yield.

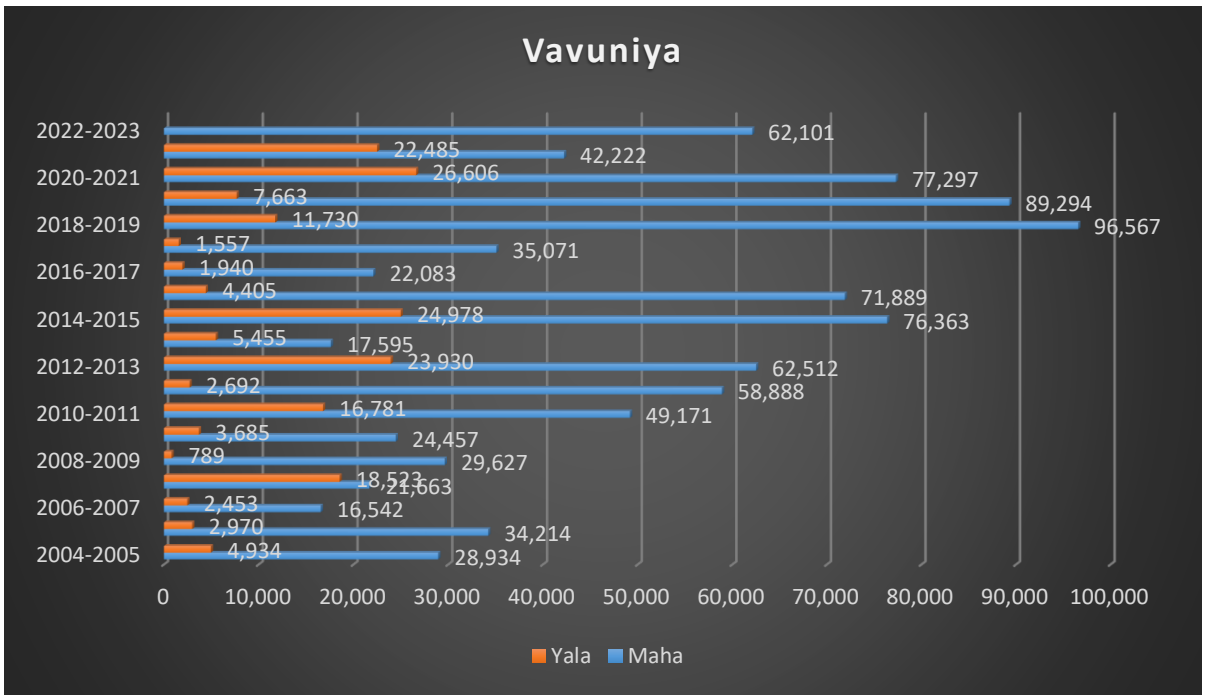


Figure 26 - Vavuniya Season-wise total paddy yield.

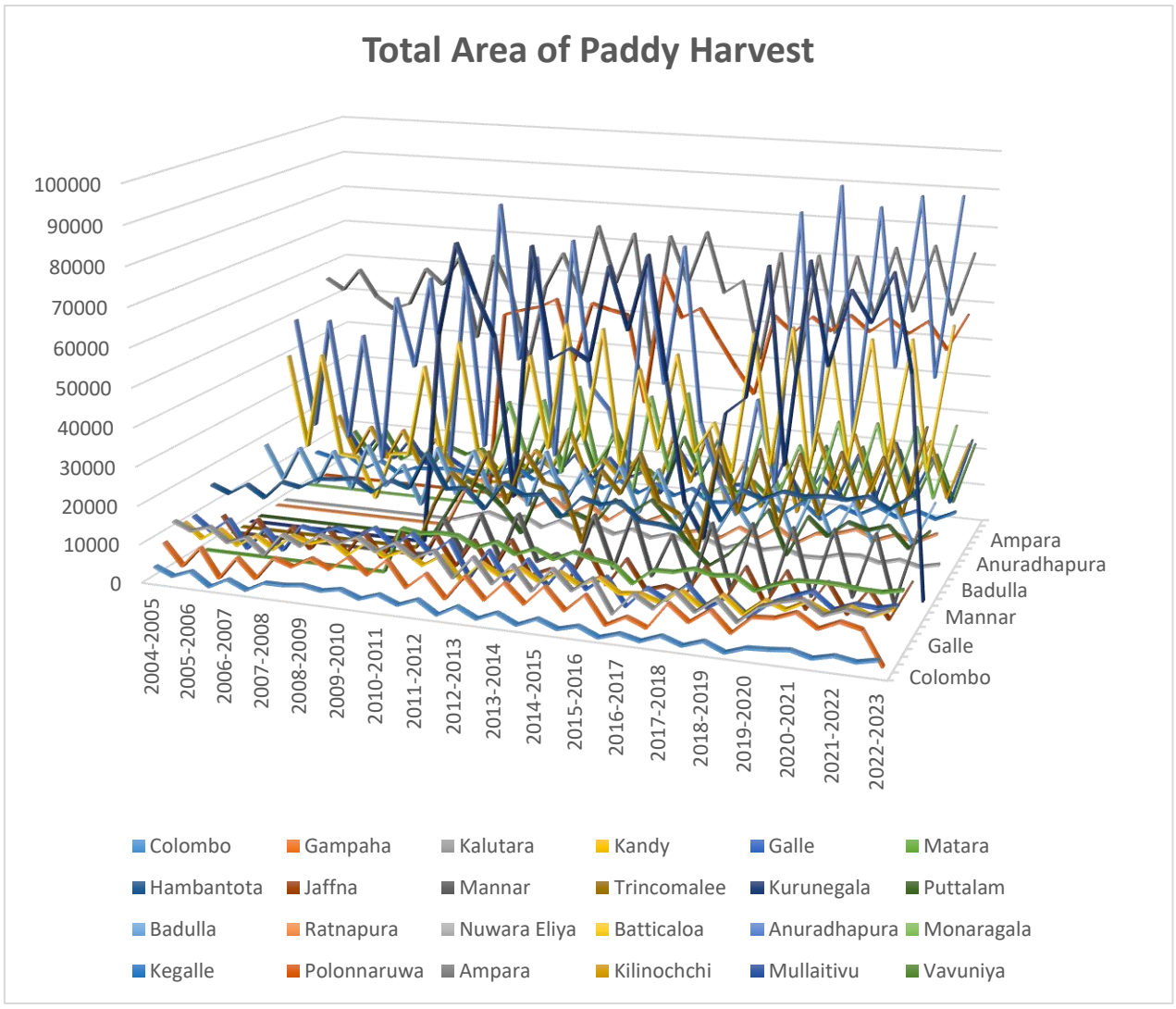


Figure 27- Location-wise Total paddy harvest is per season.

The above 25 areas are involved in paddy harvesting in Sri Lanka and they represent almost all the areas that cultivate paddy. When analyzing *Figure 2* to *Figure 26* we can notice that the highest contribution comes from the Ampara district, which amounts to 15% of the overall production of the country, while the Polonnaruwa, Kurunegala, and Anuradhapura districts produced the next highest harvest. Further, the contribution from each of the five districts viz. Hambantota, Batticaloa, Trincomalee, Badulla, and Monaragala and from each special agricultural zone accounted for more than 3% of the overall paddy harvest of the island. Over 80% of the total paddy production is covered by the above 11 regions.

3.4.1 Paddy Yield Collection Process

- Access the Department of Census and Statistics of Sri Lanka (http://www.statistics.gov.lk/Agriculture/StaticalInformation/Paddy_Statistics#gsc.tab=0)
- Download pdf files of Rice yield data in Metrics Units from 2004 to 2023.
- Create a CSV file using the above PDFs which has the location name, paddy season, total area of harvesting, and total paddy season yields in MT.
- In some years there was no data for the Total seasonal paddy yield and in Jaffna, only the Maha season was favorable for paddy no paddy cultivation in the Yala season. Such data was ignored in the data collection process.
- All Paddy yield data was stored in separate dedicated CSV files.

The rice yield data is publicly available and falls under the jurisdiction of the government. No specific permissions are required for its use, as it is considered open data provided by the Department of Census and Statistics, Sri Lanka. The Department of Census and Statistics employs robust validation and quality control procedures for rice yield data. These procedures include cross-verification of data through multiple sources, on-site inspections, and periodic audits to maintain data accuracy.

3.5 Historical Meteorological data

Historical Meteorological data were sourced from https://open-meteo.com archival data. Historical daily data from 2004-2023 were sourced for the below-mentioned areas and taken average for the Yala and Maha seasons for each year to match the Seasons of paddy yield.

- Colombo
- Gampaha
- Kalutara
- Kandy
- Matale
- Galle
- Matara
- Hambantota
- Jaffna
- Mannar
- Trincomalee
- Kurunegala
- Puttalam
- Badulla
- Ratnapura
- Nuwara Eliya
- Batticaloa
- Anuradhapura
- Monaragala
- Kegalle
- Polonnaruwa
- Ampara
- Kilinochchi
- Mullaitivu
- Vavuniya

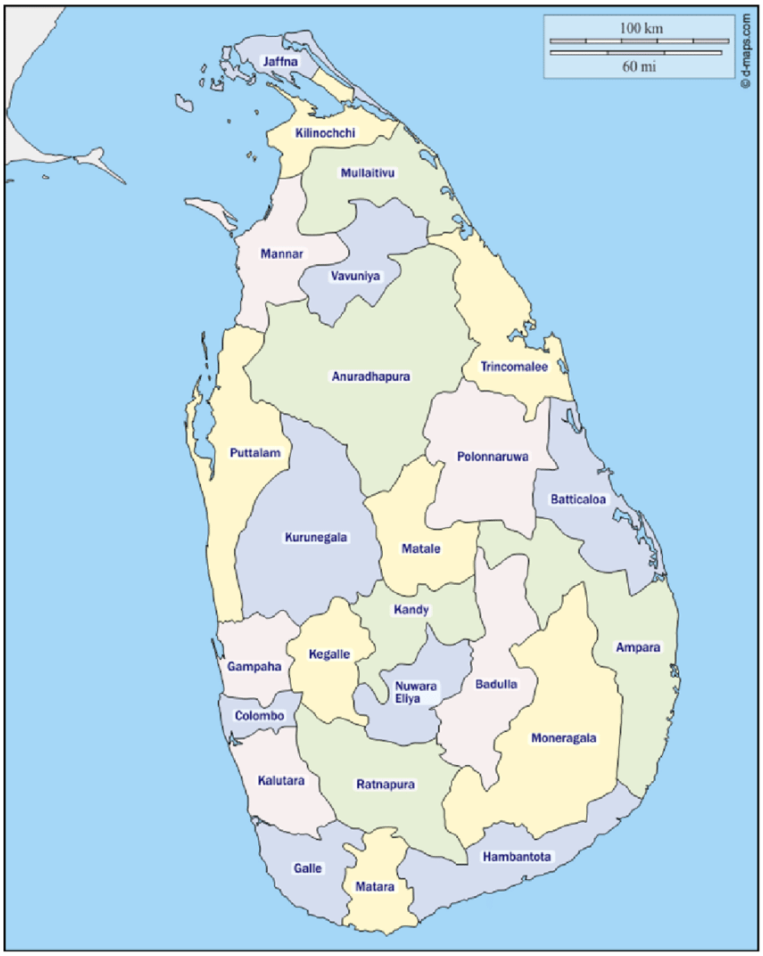


Figure 28- Weather Data collected areas.

Period for each season.

- Yala Season – From May to the end of August of the year.
- Maha Season – From September to March in the following year

This project required the climate parameters below, which were sourced from open-meteo.com.

- temperature_2m_max (°C)
- temperature_2m_min (°C)
- temperature_2m_mean (°C)
- shortwave_radiation_sum (MJ/m²)
- precipitation_sum (mm)
- precipitation_hours (h)
- windspeed_10m_max (km/h)
- windgusts_10m_max (km/h)
- et0_fao_evapotranspiration (mm)

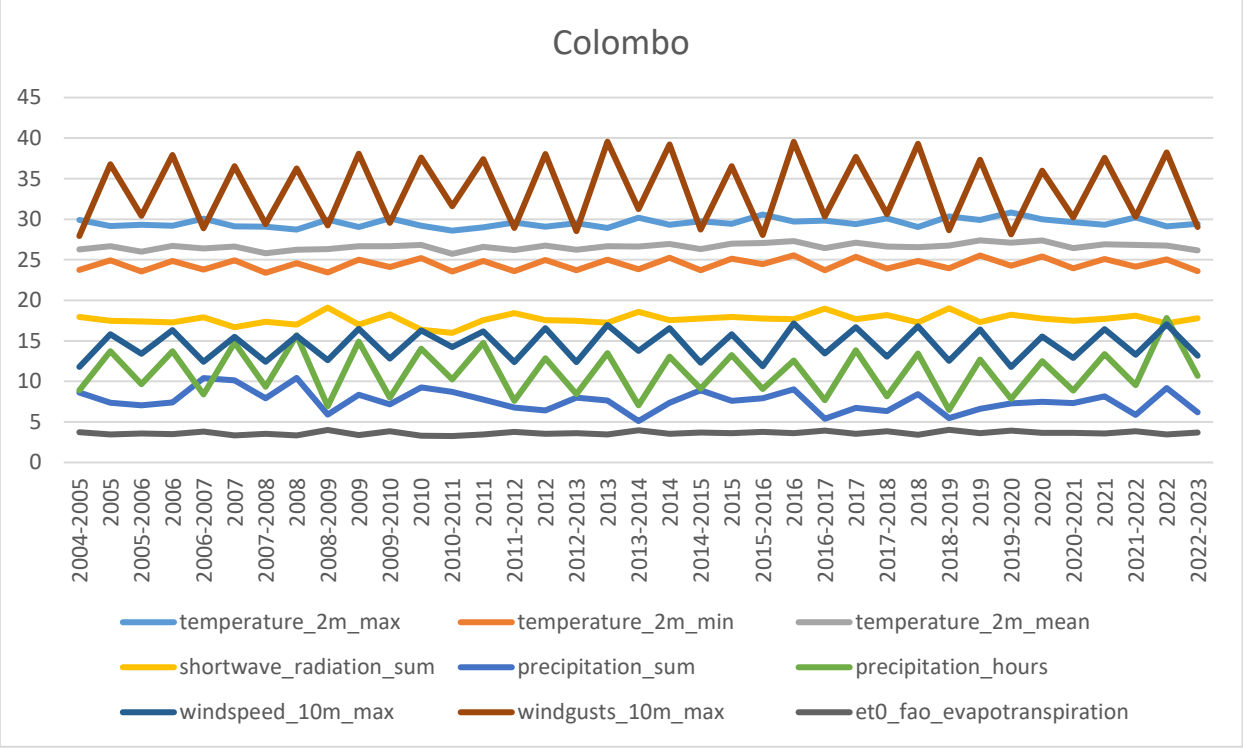


Figure 29 - Colombo Weather Data

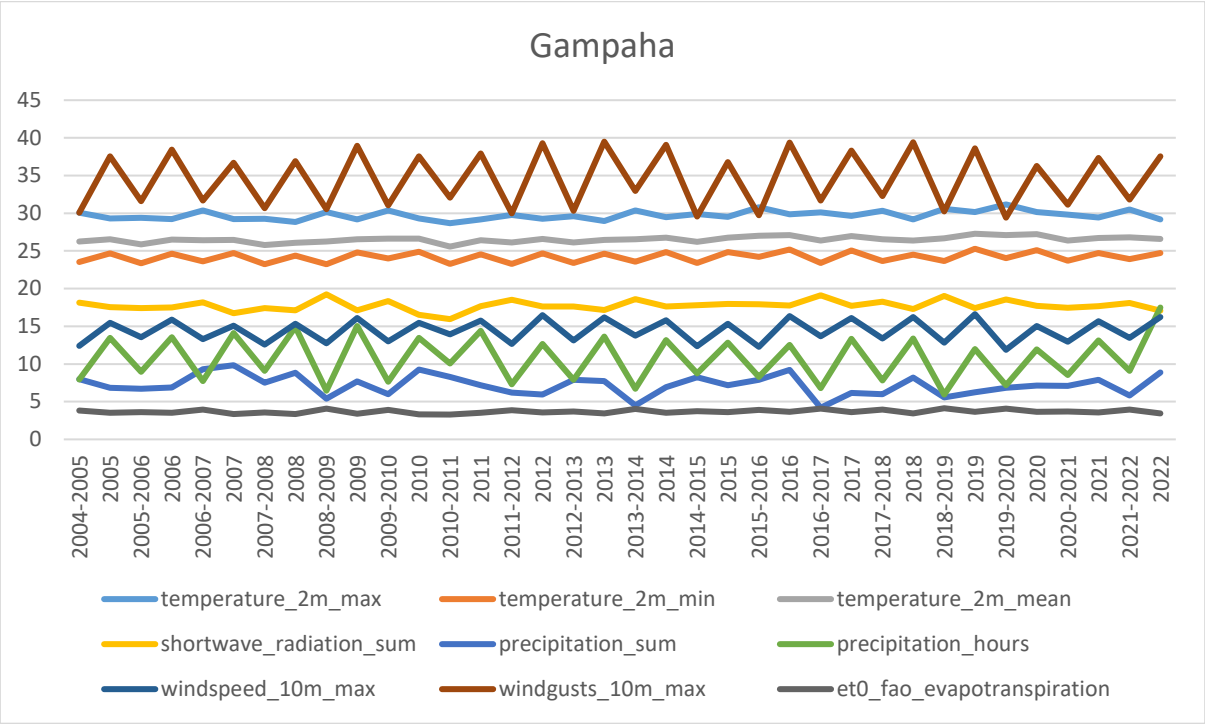


Figure 30 - Gampaha Weather Data

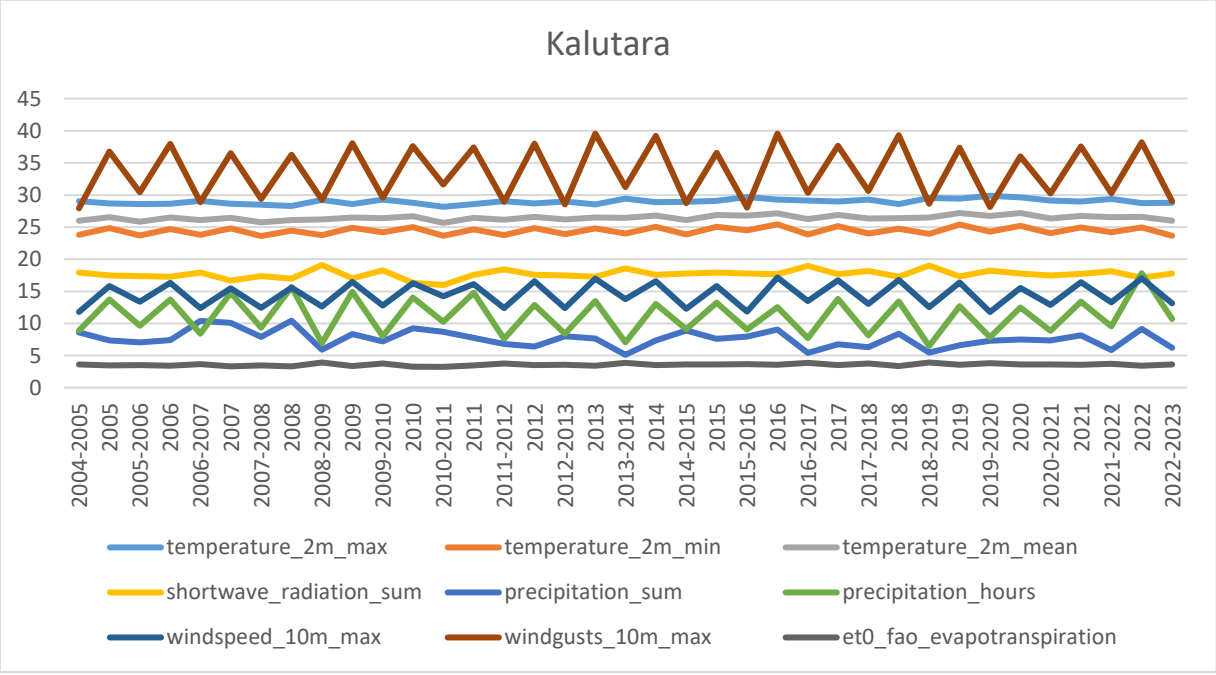


Figure 31 - Kalutara Weather Data

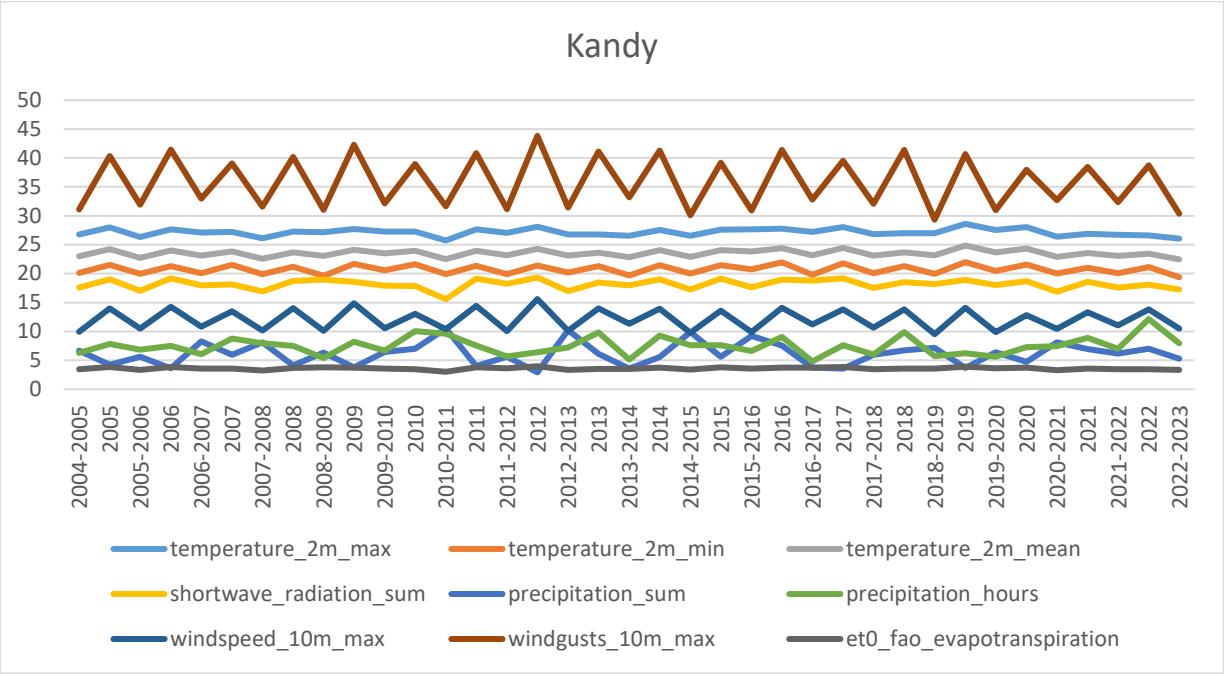


Figure 32 - Kandy Weather Data

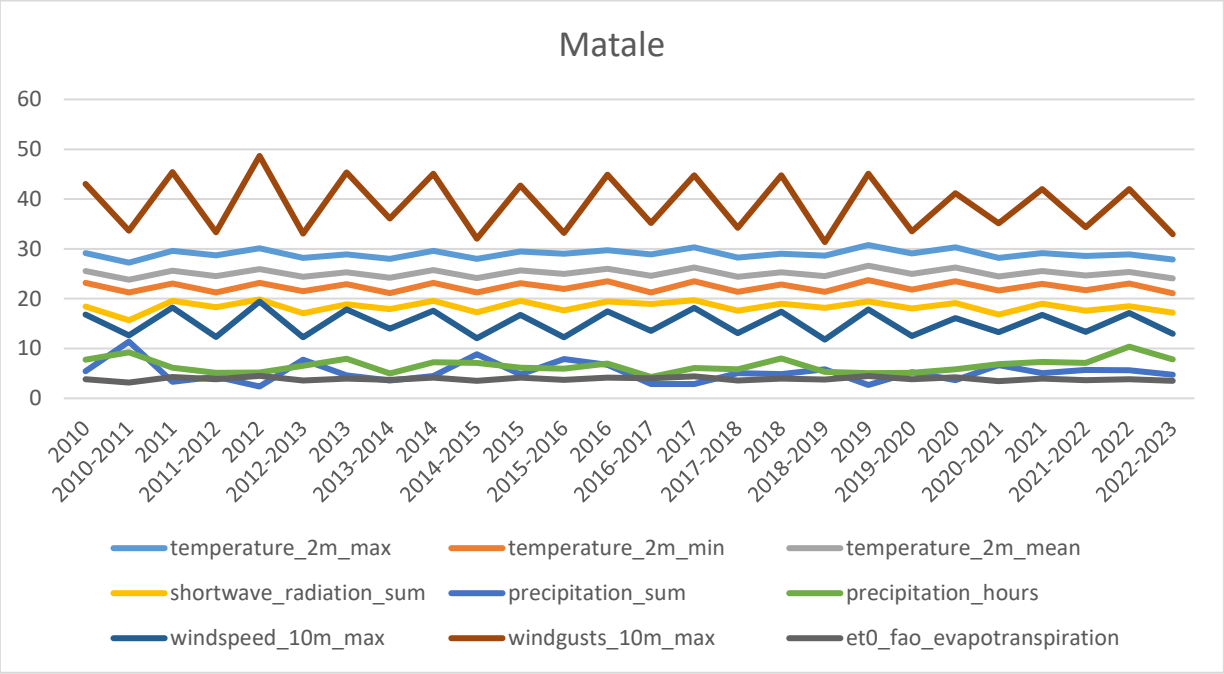


Figure 33 - Matale Weather Data

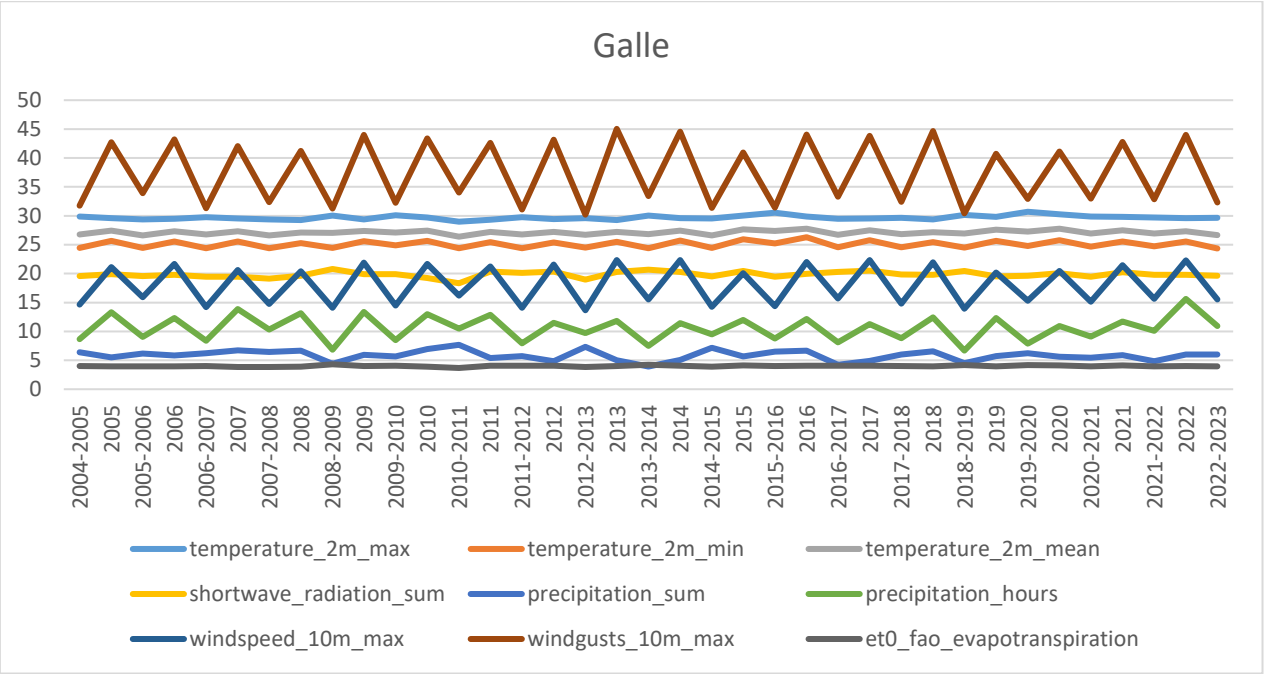


Figure 34 - Galle Weather Data

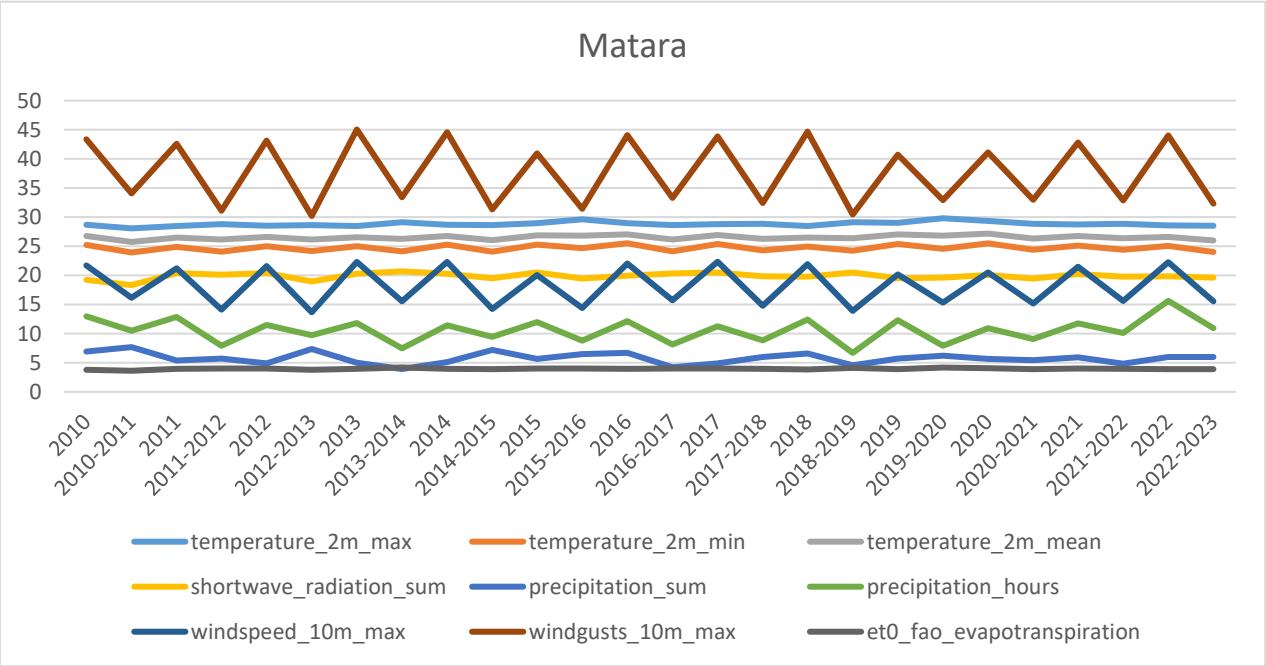


Figure 35 - Matara Weather Data

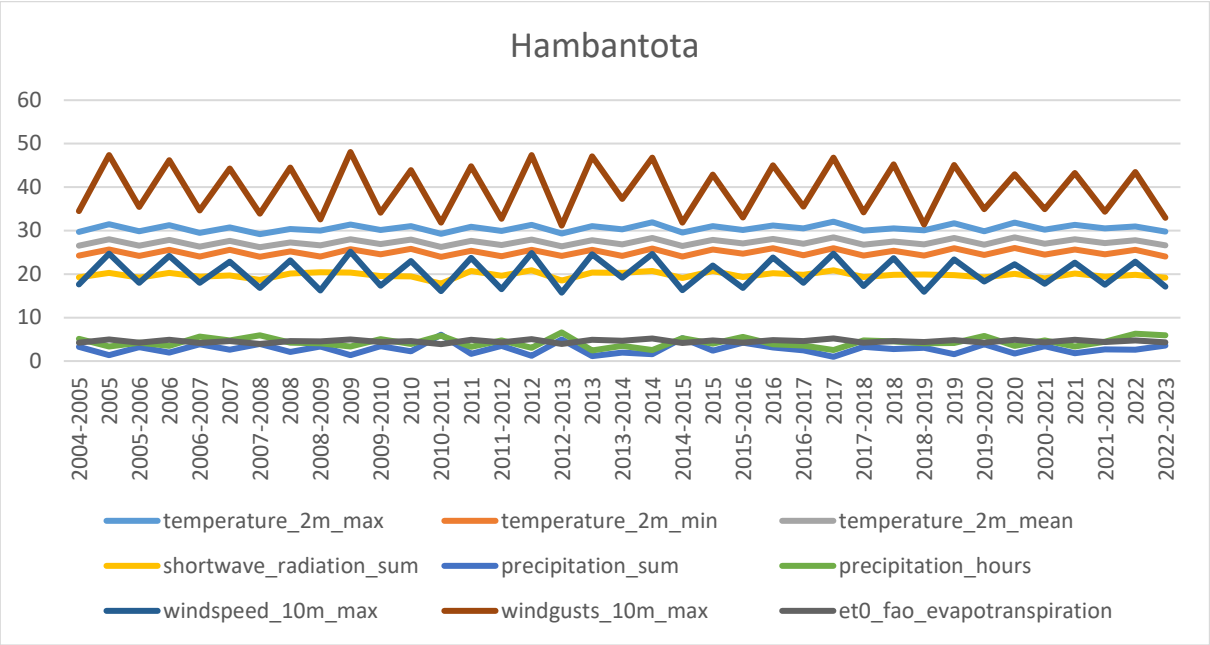


Figure 36 - Hambantota Weather Data

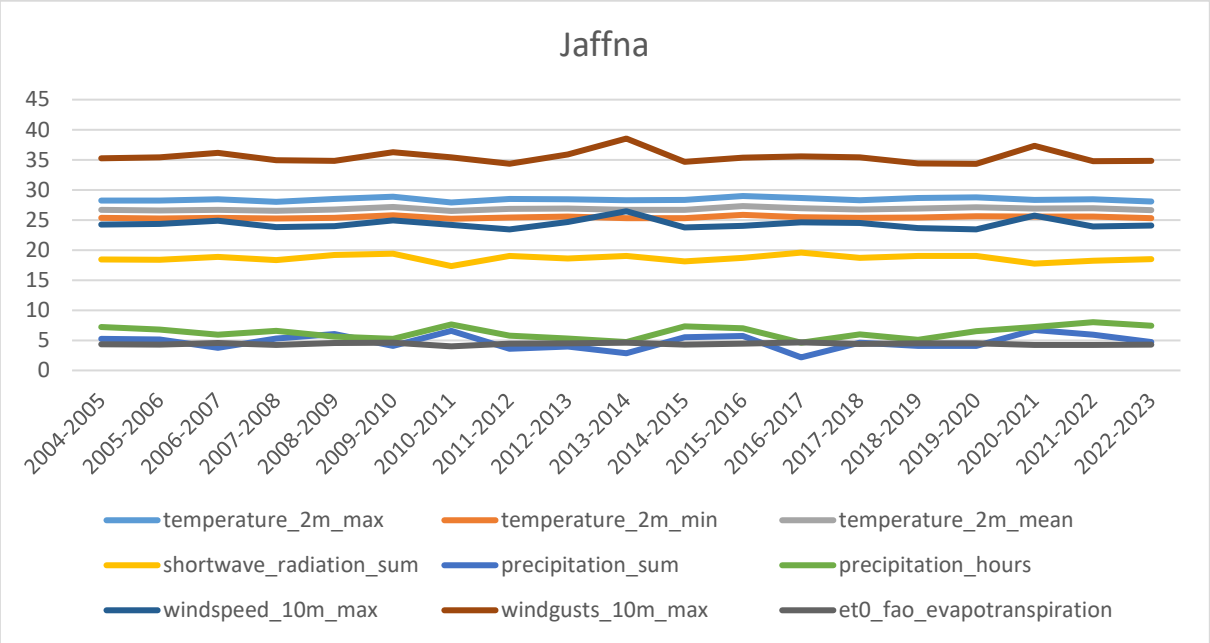


Figure 37 - Jaffna Weather Data

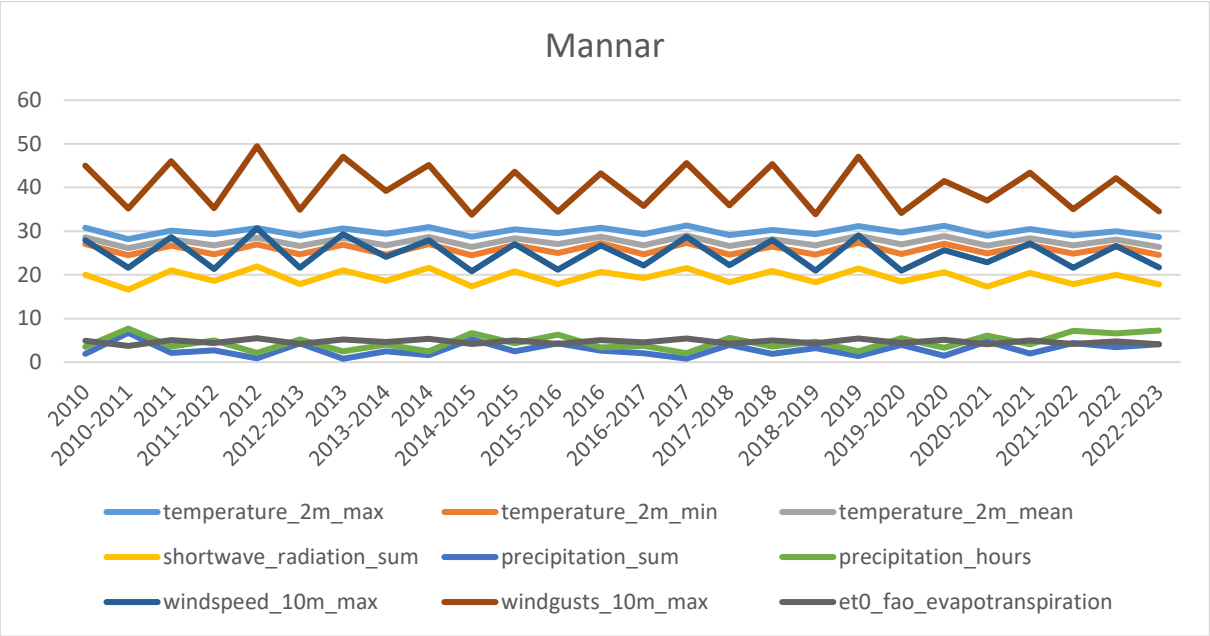


Figure 38 - Mannar Weather Data

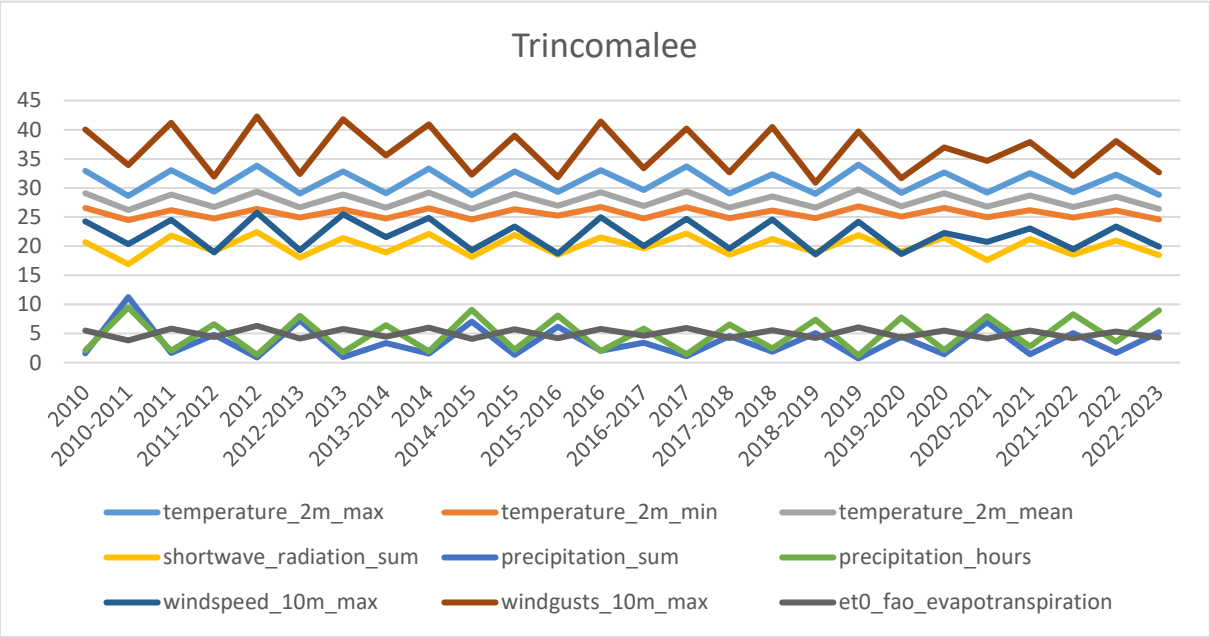


Figure 39 - Trincomalee Weather Data

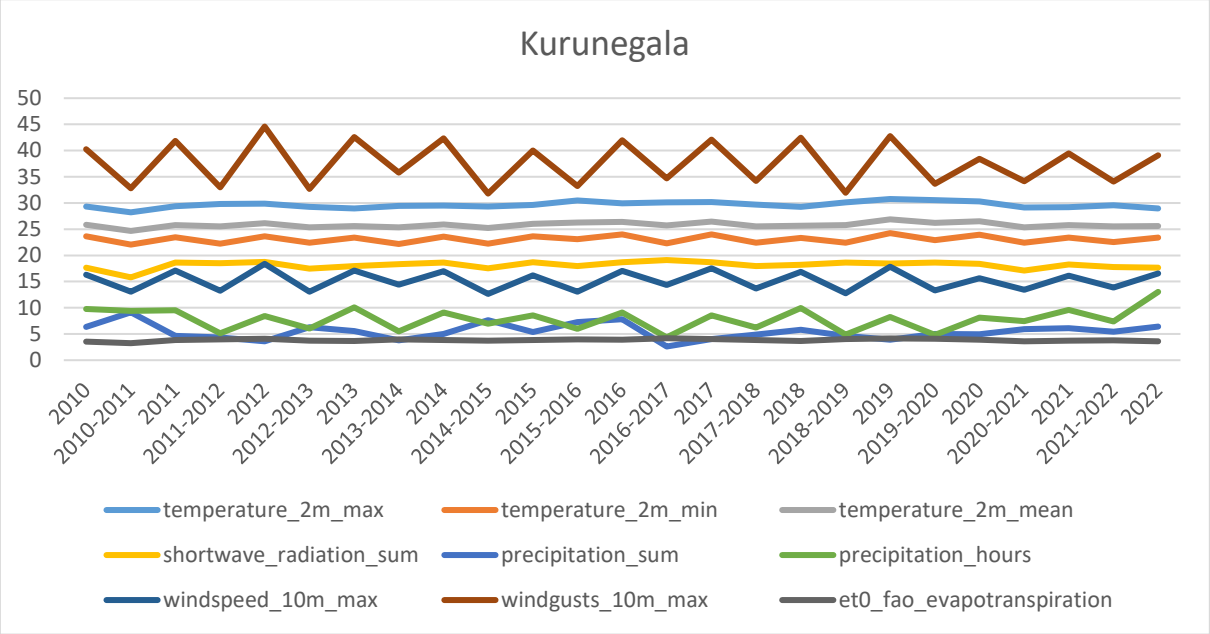


Figure 40 - Kurunegala Weather Data

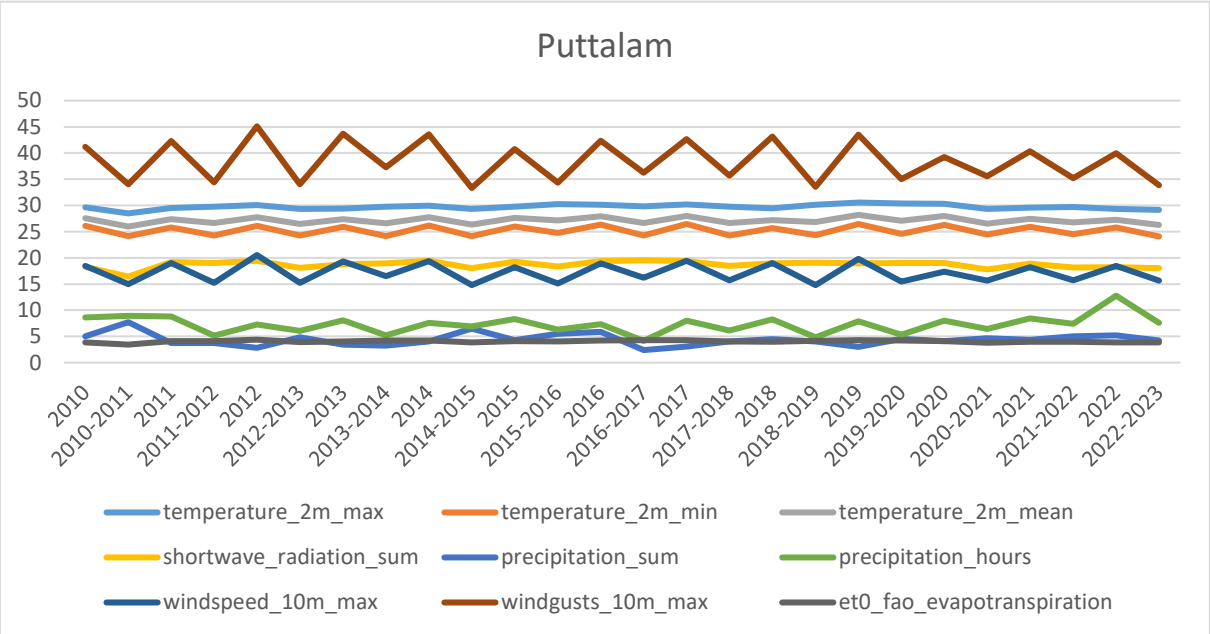


Figure 41 - Puttalam Weather Data

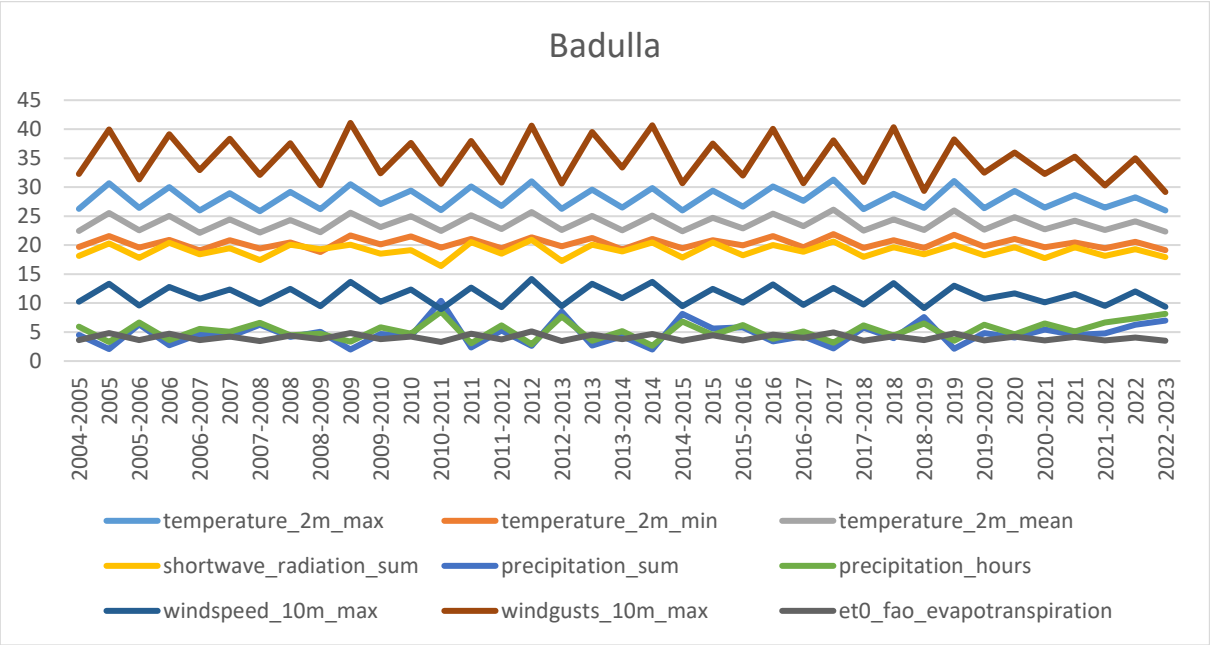


Figure 42 - Badulla Weather Data

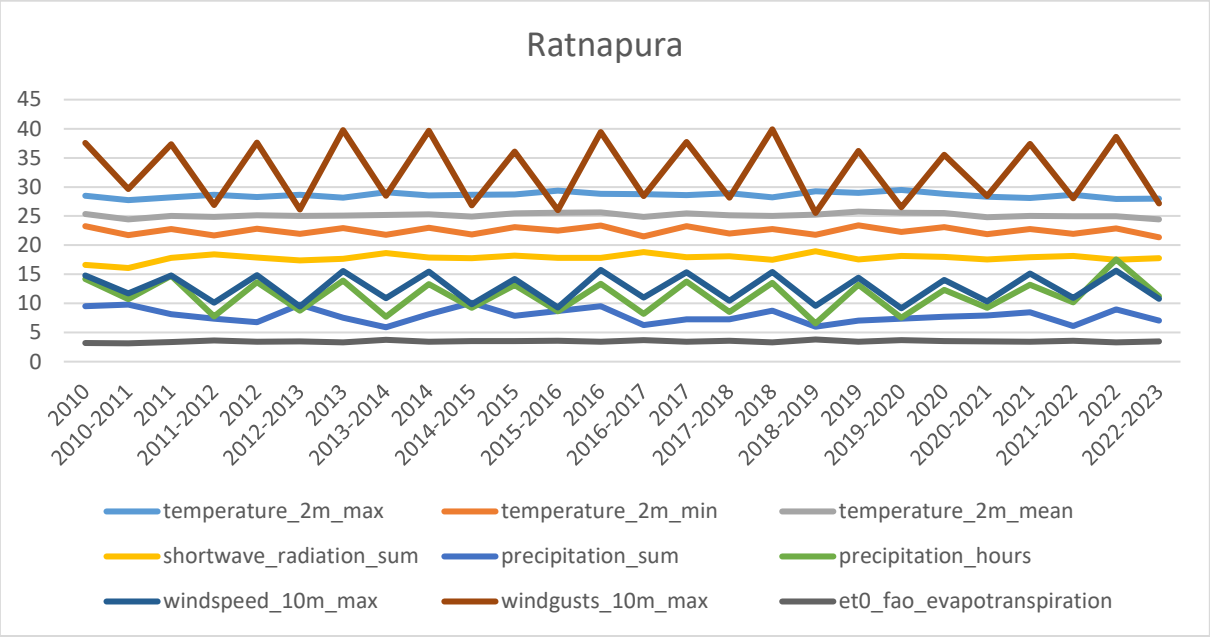


Figure 43 - Ratnapura Weather Data

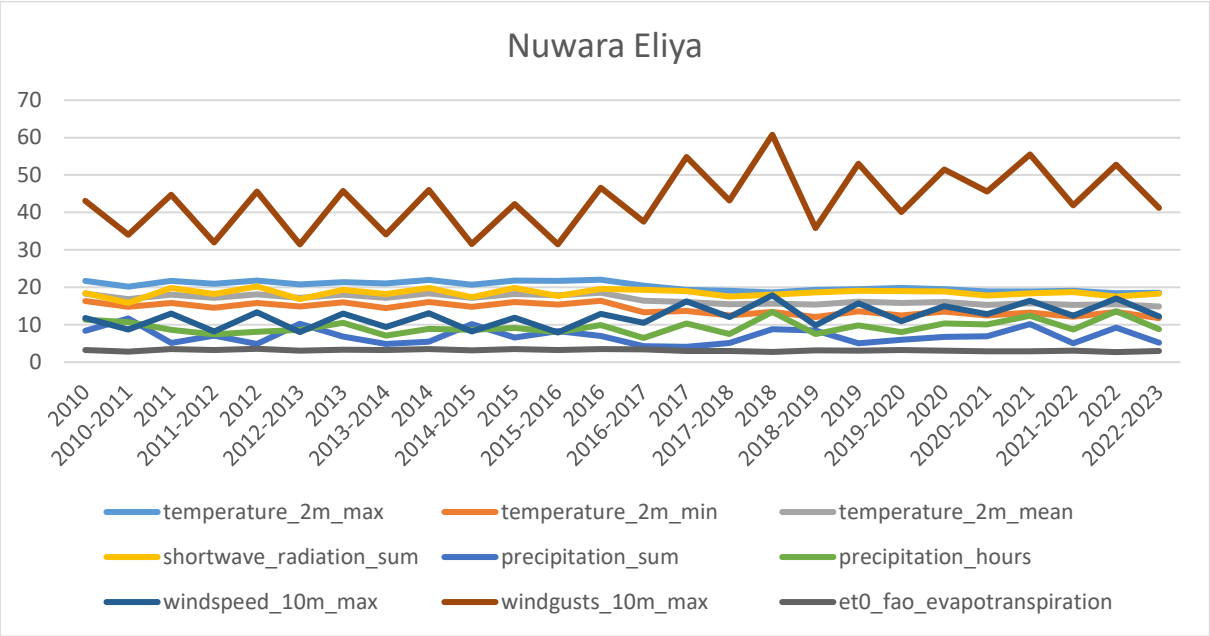


Figure 44 - Nuwara Eliya Weather Data

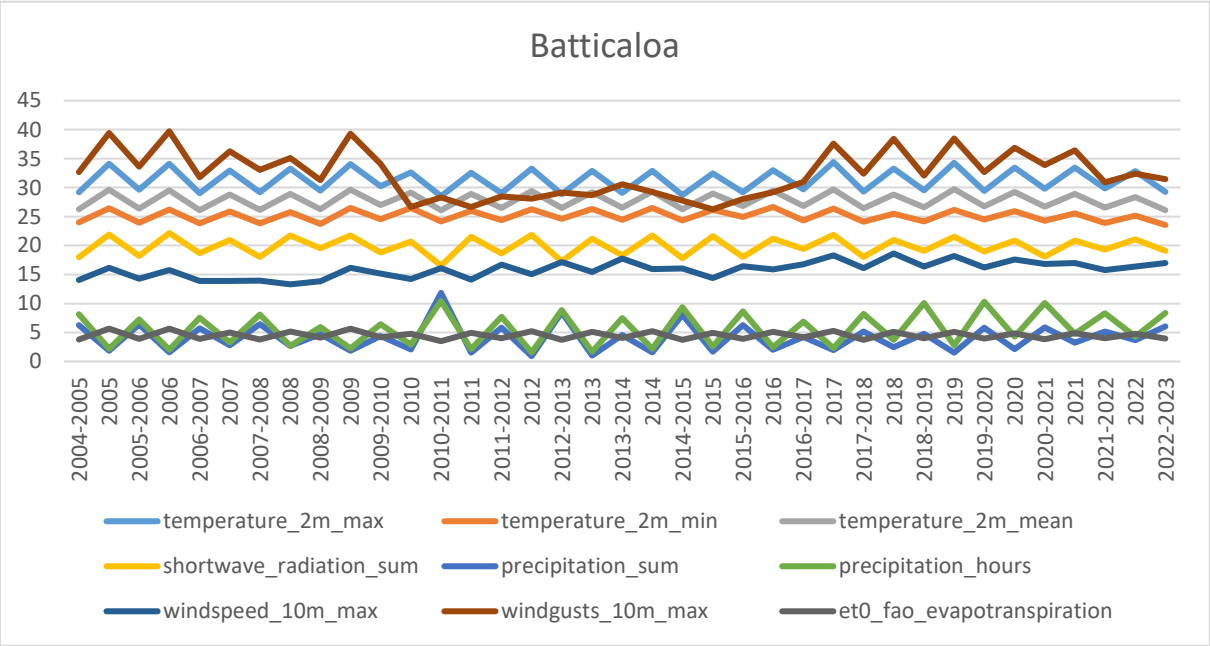


Figure 45 - Batticaloa Weather Data

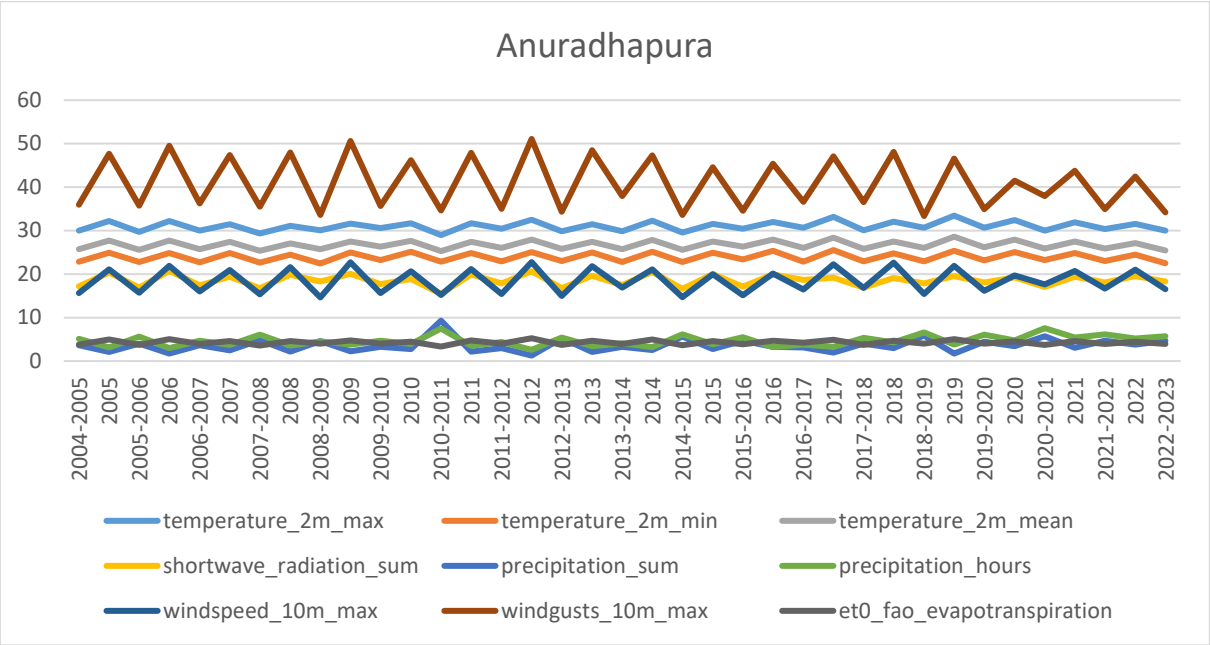


Figure 46 - Anuradhapura Weather Data

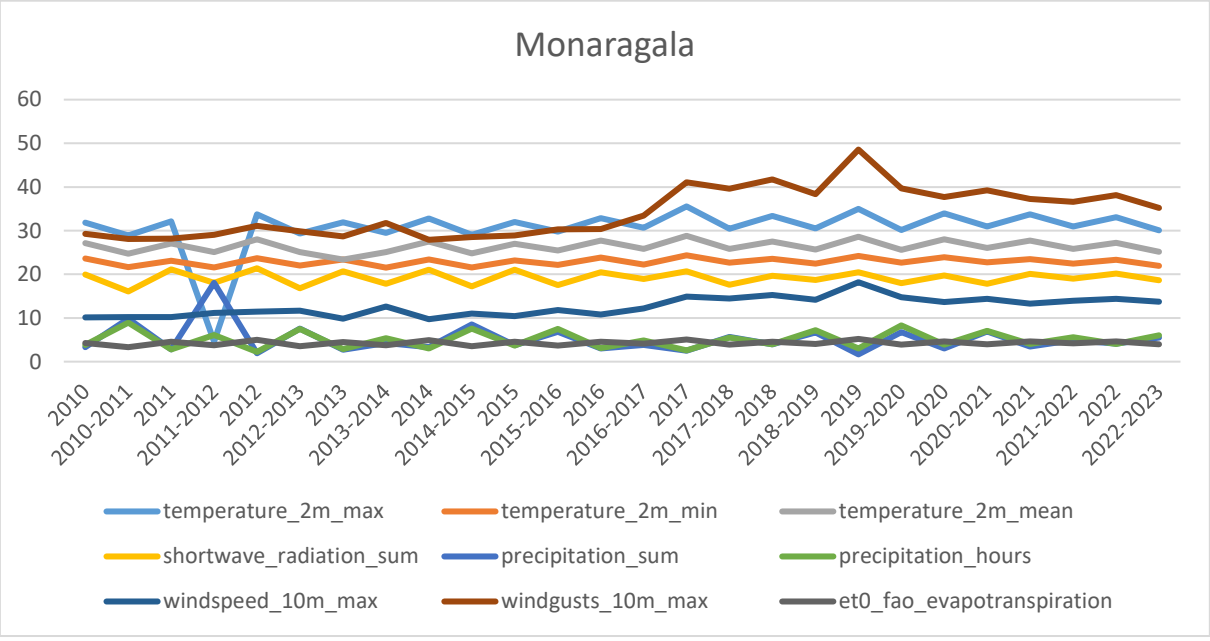


Figure 47 - Monaragala Weather Data

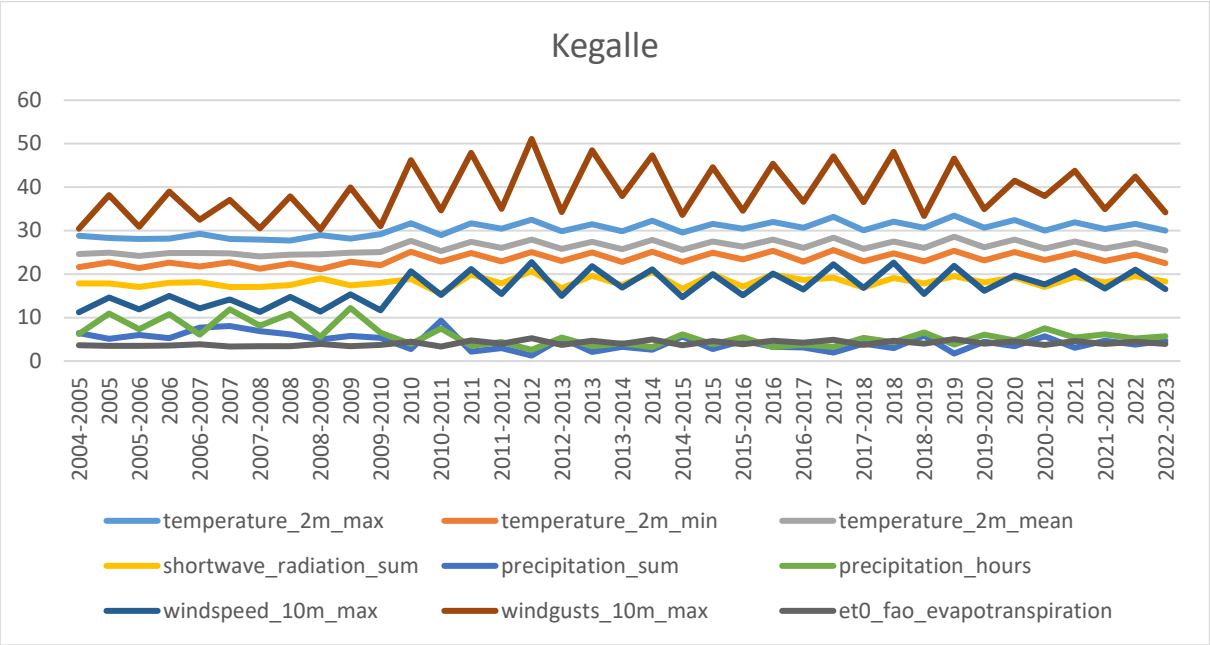


Figure 48 - Kegalle Weather Data

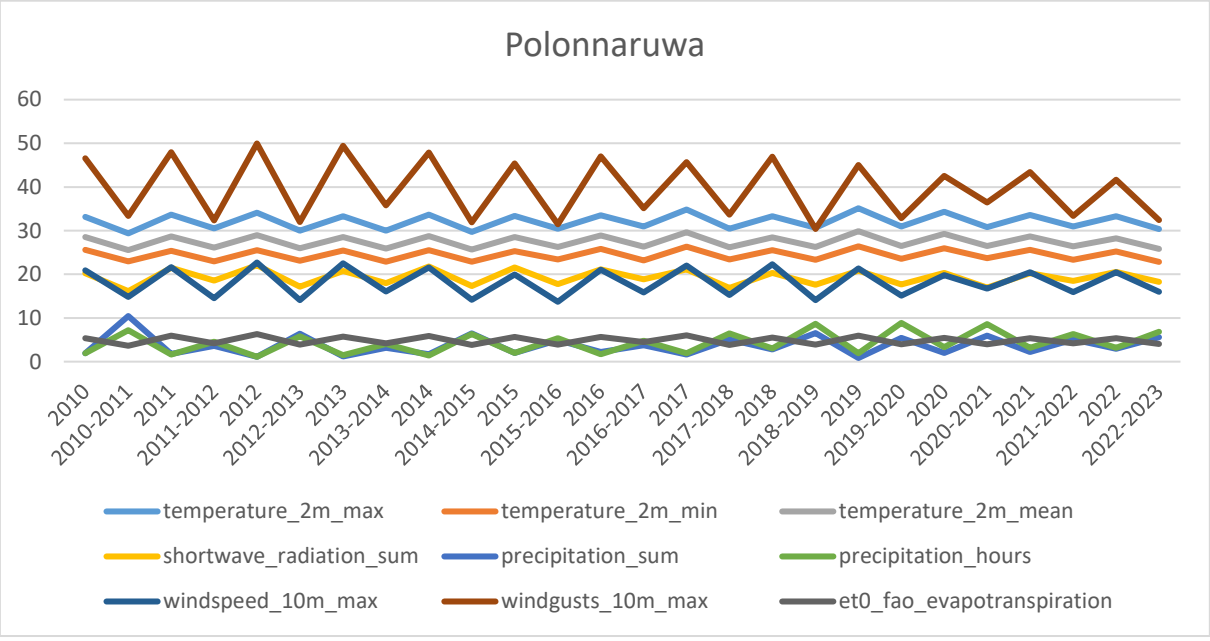


Figure 49 - Polonnaruwa Weather Data

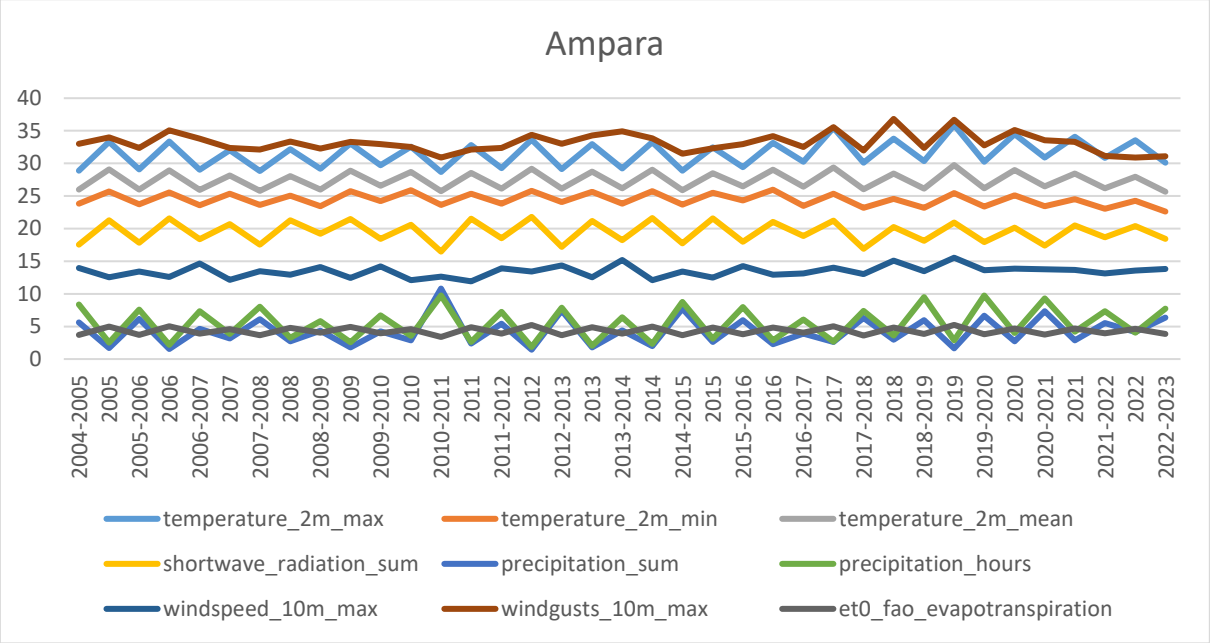


Figure 50 - Ampara Weather Data

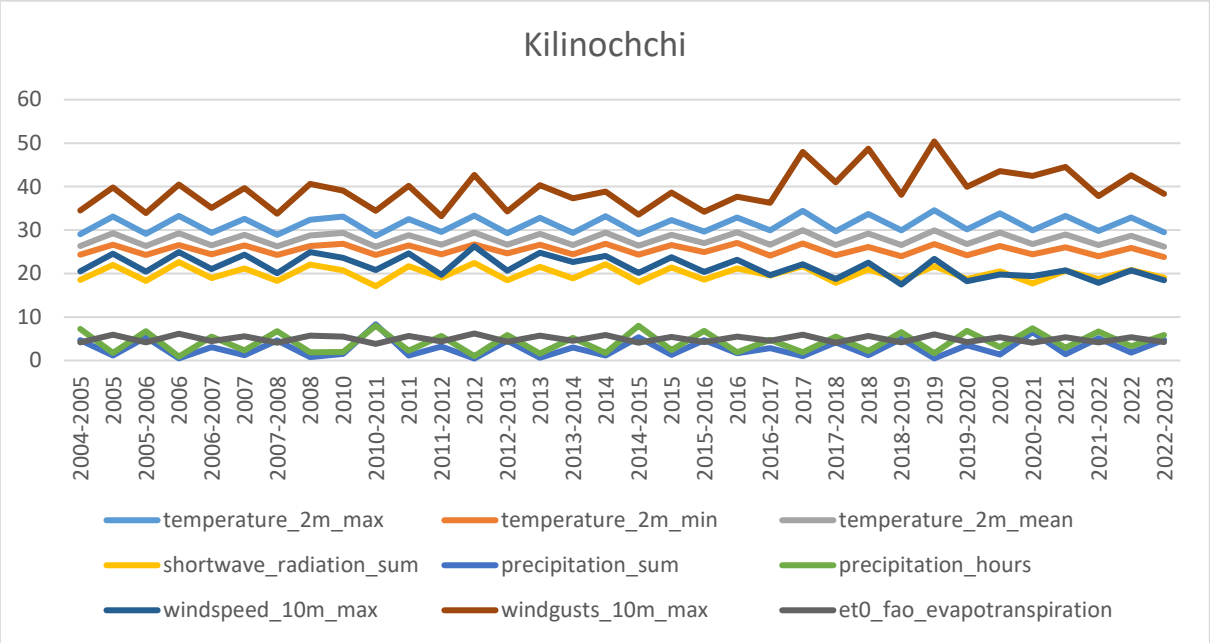


Figure 51 - Kilinochchi Weather Data

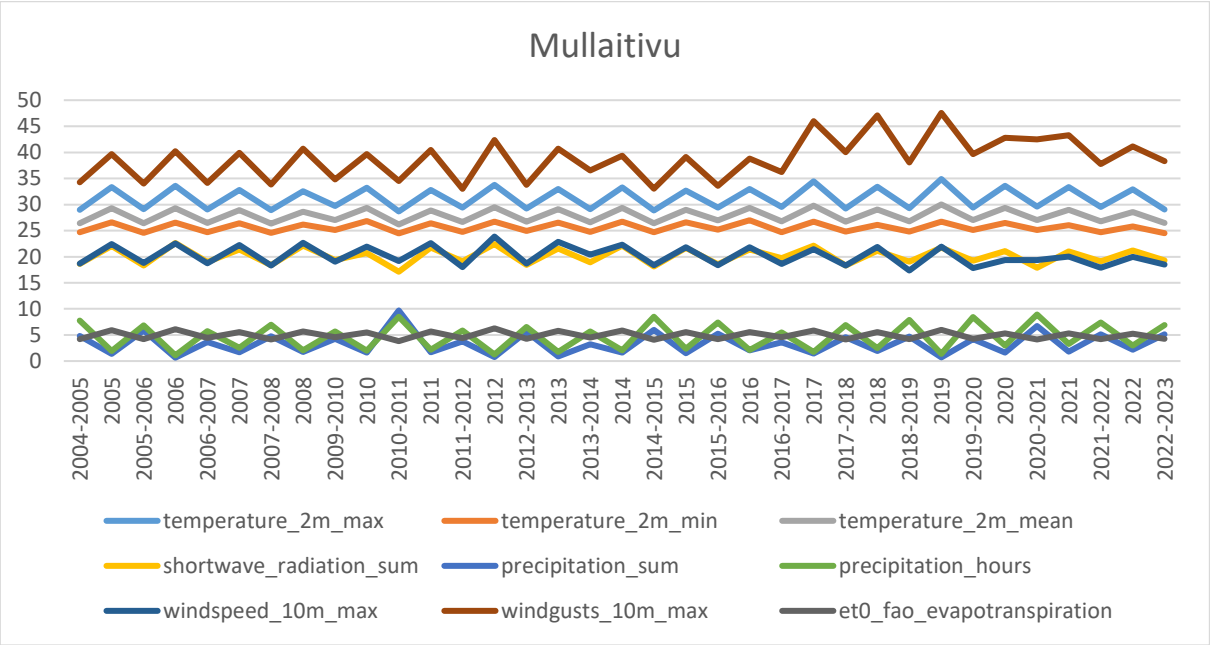


Figure 52 - Mullaitivu Weather Data

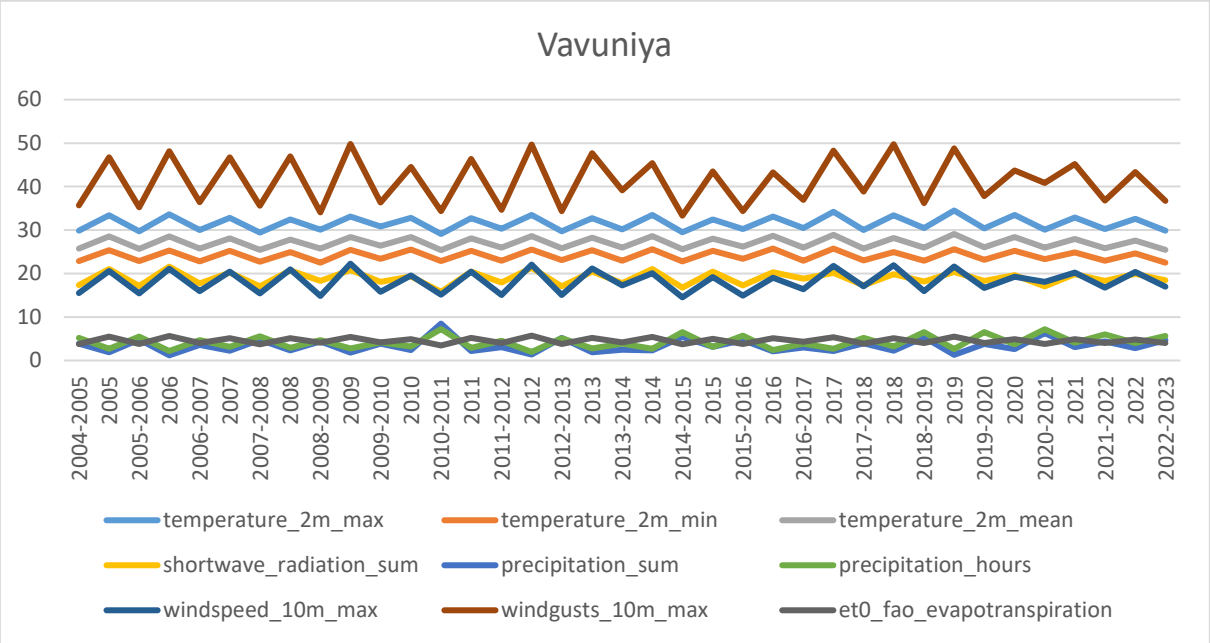


Figure 53 - Vavuniya Weather Data

3.5.1 Weather Data Collection Process

- Access the Historical Weather API page of the open-meteo site (https://open-meteo.com/en/docs/historical-weather-api)
- Give the Latitude and Longitude relevant to the location which you can retrieve from https://www.countrycoordinate.com
- Select all the necessary weather parameters from the *Daily Weather Variables* section for the respective paddy season date range.
    - temperature_2m_max
    - temperature_2m_min
    - temperature_2m_mean
    - shortwave_radiation_sum
    - precipitation_sum
    - precipitation_hours
    - windspeed_10m_max
    - windgusts_10m_max
    - et0_fao_evapotranspiration
- Then Download the CSV file.
- The current CSV file has daily weather parameters, but we need season-wise weather, which is close to the average weather of the paddy season's period.
- Take the Average from each weather parameter for each paddy season's period. (Yala Season is May to the end of August and Maha is September to March in the following year)
- Even though Weather data was available for all the locations for the selected length of time, some locations' weather data was not collected, if they didn't have corresponding paddy yield for that season in paddy yield data
- All the weather data was stored in a separate dedicated CSV file.

open-meteo.com is one of the reputable weather sites that provide climate parameters and forecasting in the world. The site has a long-standing history and has weather data as far back as 1940. Open-Meteo is climate data that is provided free of charge under freemium. Data in the sites were sourced from each region's climate authorities which are responsible for gathering them.

3.6 Combined Data Set

The next step was to combine data from the above two CSV files, in respect to their time and relevant paddy season. After that, I got combined row data set with the below fields, from 2004 to 2023.

- District
- Season
- temperature_2m_max (°C)
- temperature_2m_min (°C)
- temperature_2m_mean (°C)
- shortwave_radiation_sum (MJ/m²)
- precipitation_sum (mm)
- precipitation_hours (h)
- windspeed_10m_max (km/h)
- windgusts_10m_max (km/h)
- et0_fao_evapotranspiration (mm)
- total_space (Metric Tons)
- annual_rice_yield (Hectares)

For model training can only use numerical values, but District and Season are string values. However, it is important to add those values to model training as the climate factors a significantly different from district to district and season to season.

Therefore, the Seasons were changed as

- Yala = 1
- Maha = 2

Districts were changed to numeric alphabetic order as below.

- Ampara - 1
- Anuradhapura - 2
- Badulla - 3
- Batticaloa - 4
- Colombo - 5
- Galle - 6
- Gampaha - 7
- Hambantota - 8
- Jaffna - 9
- Kalutara - 10
- Kandy - 11
- Kegalle - 12
- Kilinochchi - 13
- Kurunegala - 14
- Mannar - 15

- Matale - 16
- Matara - 17
- Monaragala - 18
- Mullaitivu - 19
- Nuwara Eliya - 20
- Polonnaruwa - 21
- Puttalam - 22
- Ratnapura - 23
- Trincomalee - 24
- Vavuniya – 25

Even though the season of Paddy yield or the location of the area where they were grown has no direct impact on paddy growth like other variables like climate conditions, I decided to include them in the deep learning process, because this trained model is supposed to predict for whole Sri Lankan context, where the district-to-district weather conditions are varied disregard to relatively small size of the country. So, when using location and paddy season, the model will be able to identify the differences in the conditions.

Now the combined data set consists of only numerical values, which is easy to train in the 1D CNN machine learning algorithm.

3.6.1 Combined Data Set Analysis

After combining both weather and climate data by the respective paddy fields, now I need to clarify whether each attribute has a significant impact on rice yield. For this process Weka Machine learning is suitable.

In Weka, I used the *Select Attributes* feature. Among the Several Evaluator algorithms, *CorrelationAttributeEval* is a good fit, which supports numerical, nominal, and hybrid data sets to figure out potentially relevant features. CorrelationAttributeEval evaluates the worth of an attribute by measuring the correlation (Pearson's correlation coefficient) between it and the class (in this instance, Seasonal rice yield). It's particularly useful for identifying relevant features when building predictive models.

According to Weka, all the selected attributes have significant importance with the below ranking.

- 0.9702 - 12 total_space
- 0.1899 - 2 Season
- 0.1893 - 3 temperature_2m_max
- 0.1117 - 5 temperature_2m_mean
- 0.0575 - 11 et0_fao_evapotranspiration
- 0.0384 - 4 temperature_2m_min
- -0.0563 - 6 shortwave_radiation_sum
- -0.1031 - 7 precipitation_sum
- -0.1179 - 9 windspeed_10m_max
- -0.1626 - 10 windgusts_10m_max
- -0.2174 - 8 precipitation_hours
- -0.2344 - 1 city_no

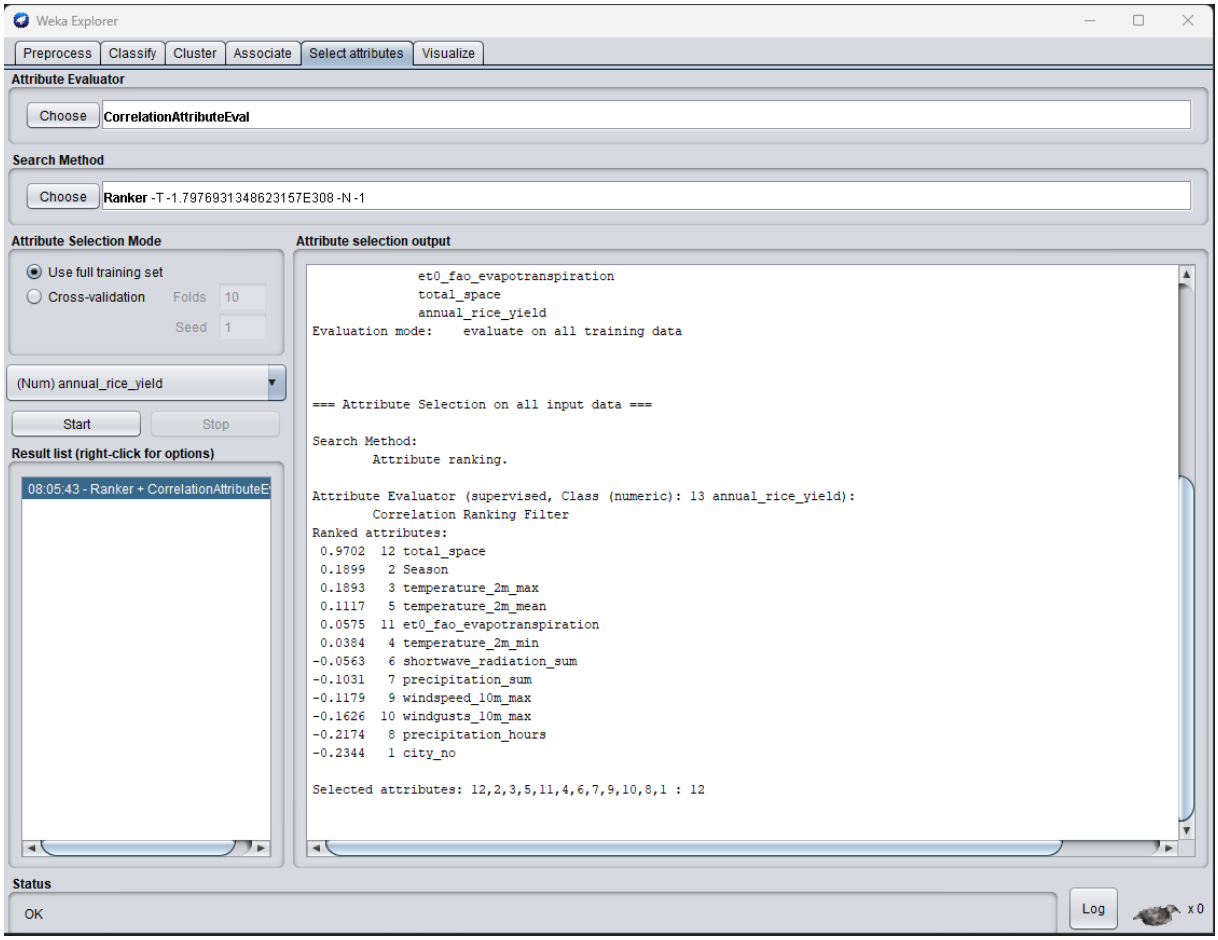


Figure 54 - Weka Outcome for Attribute Selection

3.7 Deep learning model development

Deep learning offers numerous advantages for rice yield prediction, including its ability to handle complex non-linear relationships, integrate multiple data sources, automatically extract features, and process large datasets. Moreover, deep learning models are well-suited for analyzing temporal and spatial data, adapting to new information, and providing improved prediction accuracy. These strengths make deep learning a powerful and effective approach for enhancing the precision and reliability of rice yield predictions, ultimately contributing to better agricultural planning and decision-making.

3.7.1 Deep Learning Algorithm Selection Clarification

In this project, I will be using One Dimension Convolutional Neural Networks (CNN) as the deep learning algorithm for model development. Using CNNs for rice yield prediction using weather data offers several advantages due to their inherent architecture and capabilities. Here are some reasons why CNNs can be particularly effective for this task

- **Handling Spatio-Temporal Data**
    - Spatial Patterns: CNNs are well-suited to recognize spatial patterns in data. In the context of weather data, spatial relationships between different weather variables across various regions can be crucial for accurate yield prediction.
    - Temporal Patterns: While CNNs are typically associated with spatial data, they can also be adapted to handle temporal data by treating time as another dimension. Weather data often involves time series, and CNNs can capture temporal dependencies when used in conjunction with other layers or architectures.
- **Feature Extraction** - CNNs automatically learn hierarchical feature representations from the data. For weather data, this means the network can learn complex features that might be important for predicting rice yield without the need for manual feature engineering.
- **Data Dimensionality** - Weather data can be multi-dimensional (e.g., temperature, precipitation, humidity, etc., across different locations and times). CNNs can process this multi-dimensional data efficiently, capturing the interactions between these dimensions.
- **Scalability and Efficiency**
    - Efficient Computation: Due to their weight-sharing mechanism and the use of filters, CNNs are computationally efficient and can be scaled to handle large datasets, which is often the case with weather data.
    - Parallel Processing: Modern deep learning frameworks optimize CNN operations for parallel processing on GPUs, making them efficient for training on large datasets.

Therefore, using CNNs for rice yield prediction leveraging weather data can lead to more accurate, efficient, and scalable models capable of capturing complex patterns and relationships inherent in the data. Their ability to handle multi-dimensional inputs, learn hierarchical features, and integrate seamlessly with other architectures makes them a powerful tool in agricultural forecasting and decision-making.

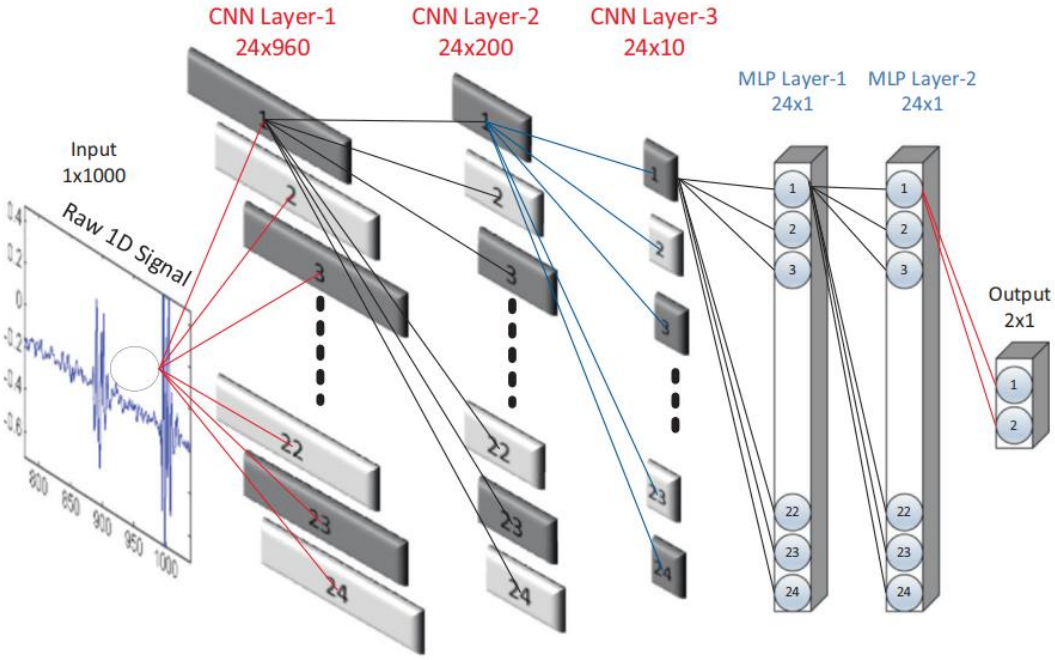


Figure 55 - 1D CNN configuration with 3 CNN and 2 MLP layers.

The current project will be using Kera's Sequential API for the implementation of the Deep learning model.

3.8 Model evaluation and performance analysis.

Rice yield prediction is a regression task. There for the project, I will be using the mentioned measures below to evaluate the model after training.

3.8.1 Mean Absolute Error (MAE)

This metric calculates the average absolute differences between the predicted and actual values. It gives a good indication of the average error magnitude.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$$

Where:

$\hat{y}_i$ = Predicted value for the i^{th} data point

y_i = Actual value for the i^{th} data point

n = number of observations

Figure 56 - Mean Absolute Error

3.8.2 Mean Squared Error (MSE)

MSE calculates the average of the squared differences between the predicted and actual values. It penalizes larger errors more significantly than MAE.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

Where:

$\hat{y}_i$ = Predicted value for the i^{th} data point

y_i = Actual value for the i^{th} data point

n = number of observations

Figure 57 - Mean Squared Error

3.8.3 Root Mean Squared Error (RMSE)

RMSE is the square root of the MSE and provides a measure of the standard deviation of the errors. It is in the same unit as the target variable.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

$$SS_{\text{res}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Figure 58 - Root Mean Squared Error

Where:

SS_{res}= The sum of squares of the residual errors, representing the variation that the model fails to explain

SS_{tot} = The total sum of squares, representing the total variation in the data

$\hat{y}_i$ = Predicted value for the i^{th} data point

y_i= Actual value for the i^{th} data point

$\bar{y}$ = Mean value of the dependent variable

n =Number of observations

Figure 59- Coefficient of Determination

3.8.4 Coefficient of Determination (R^2 or R-squared)

R-squared measures the proportion of the variance in the dependent variable (rice yield) that is predictable from the independent variables (features). It indicates how well the model captures the variance in the data.

Within the project, CNN will be trained in different parameter configurations and compare the above-mentioned measures to compare the efficiency of the model.

3.9 Result communication

Given the critical importance of rice yield to Sri Lanka's national food security, accurate prediction is essential. The findings of this research, including the performance evaluation of different deep learning algorithms, will be communicated to stakeholders and decision-makers, aiding them in making informed choices concerning national food security and leveraging the predictive capabilities of deep learning models.

By integrating deep learning algorithms into the empirical research methodology, this project aims to provide insights into the most effective approaches for predicting rice yield using advanced machine learning techniques. The research outcomes will contribute to enhancing decision-making processes and supporting national food security efforts in Sri Lanka.

3.10 Real-world Implementation

The goal of the project is to ultimately predict seasonal rice yield using climate data. However, the Goal of Real-world implementation should describe how it should be implemented so that the research has an outcome.

3.10.1 Implementation Goal

Provide actionable insights to farmers and other stakeholders can use in the real world.

3.10.2 Implementation Process

The implementation process should have the following points from start to end.

- Data Collection
- Data Preprocessing
- Select Attributes from the processed data collection.
- Train Models using selected 1D CNN algorithm in Keras.
- Evaluate Models and select the best-performing model.
- Build and Expose REST API which consumes the selected model and provides Rice Yield predictions.

3.10.3 Technologies Involved in Implementation.

- Data Preprocessing – MS Excel
- Attribute Select – Weka Machine Learning Software
- Model Training – Keras Deep Learning API, TensorFlow
- REST API – Python, Flask Framework

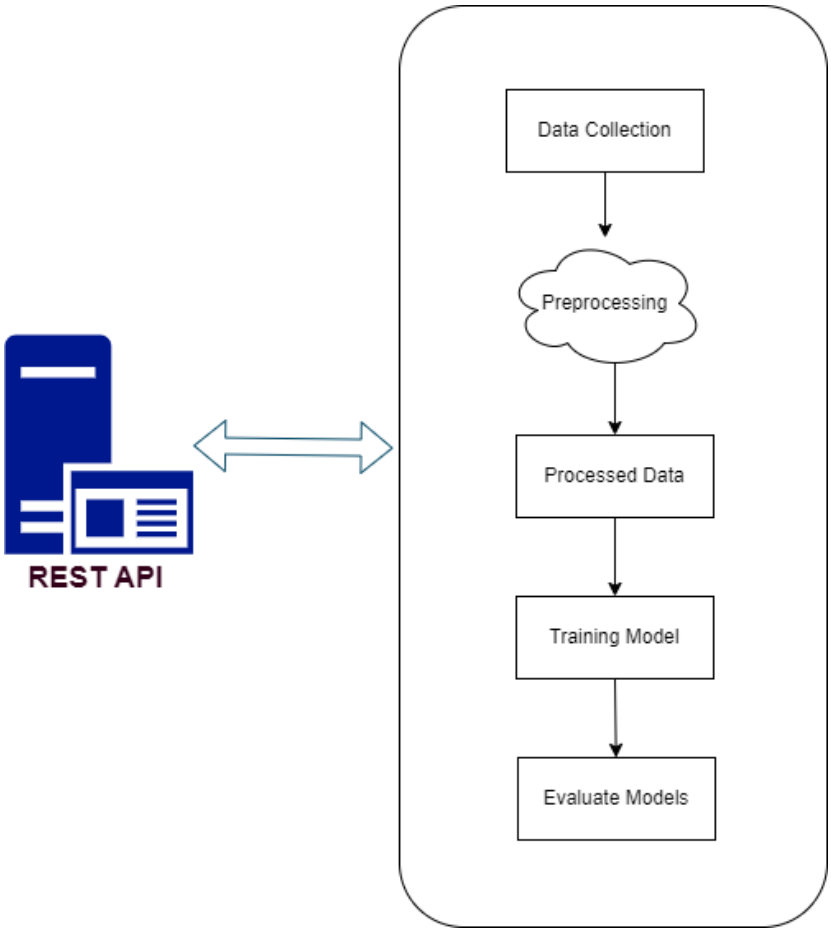


Figure 60 - Rice Yield Prediction Implementation.

CHAPTER 4 - EVALUATION AND RESULTS

4.1 Introduction

This chapter is focused on how to evaluate the trained models' outcomes. In the implementation, I used the 1D CNN architecture for the project with the architecture below.

1. **Sequential Model Initialization**: The architecture begins by creating a sequential neural network model using Keras. In this case, we're designing a model specifically for predicting rice yield. The Sequential class allows us to stack layers one after the other, forming a linear architecture.

2. **Convolutional Layers**: The first two layers are 1D convolutional layers (Conv1D). These layers are useful for extracting local patterns from sequential data, which aligns well with the temporal nature of rice yield prediction. Let's delve into the details of each layer:

First Conv1D Layer:

- filters=64: This parameter specifies the number of filters (kernels). Each filter learns different features from the input data.
- kernel_size=3: The size of the convolutional window (filter). It slides over the input data to capture relevant patterns.
- activation='relu': The Rectified Linear Unit (ReLU) activation function is applied after the convolution operation.
- input_shape=(X_train.shape[1], 1): Describes the shape of the input data. For rice yield prediction, we assume a single-channel input (monochannel data).

Second Conv1D Layer:

- Similar to the first layer but with filters=128.

3. **Flattening Layer**: After the convolutional layers, we add a Flatten layer. Its purpose is to reshape the output from the convolutional layers into a 1D vector. This transformation prepares the data for subsequent dense layers.

4. **Dense Layers**: incorporate three fully connected (dense) layers:

- First Dense Layer (128 Units):

This layer contains 128 neurons, each with a ReLU activation function. It learns higher-level features from the flattened data.

- Second Dense Layer (64 Units):

Like the first dense layer, but with 64 neurons.

- Output Layer:

The final dense layer has a single unit (output layer). Since there's no specified activation function, we assume this model is designed for a regression task. It predicts a continuous numeric value, which in this context represents rice yield.

4.2 Model Performance under different parameters.

Here I have trained in the 1D CNN model for different Learning rates, Epochs, and Batch sizes, using the below Python code which utilizes Keras for 1D CNN model training.

```
X_train, X_test, y_train, y_test = train_test_split( *arrays: X_scaled, y, test_size=0.2, random_state=42)
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], 1)
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], 1)
model = Sequential([
    Conv1D(filters=64, kernel_size=3, activation='relu', input_shape=(X_train.shape[1], 1)),
    Conv1D(filters=128, kernel_size=3, activation='relu'),
    Flatten(),
    Dense( units: 128, activation='relu'),
    Dense( units: 64, activation='relu'),
    Dense(1)  # Output layer, assuming a regression task
])

learning_rate = 0.001  # Define your learning rate
optimizer = Adam(learning_rate=learning_rate)
model.compile(optimizer=optimizer, loss='mean_squared_error')
epochs = 600
batchSize = 10
model.fit(X_train, y_train, epochs=epochs, batch_size=batchSize, validation_split=0.1)

model.save_weights(
    "trained_model_weights" + "-" + str(epochs) + "-" + str(learning_rate) + ".h5")
y_pred = model.predict(X_test)
print("Actual Values vs Predicted Values:")
for actual, predicted in zip(y_test.values, y_pred.flatten()):
    print(f"Actual: {actual}\tPredicted: {predicted}")

mape = np.mean(np.abs((y_test - y_pred.flatten()) / y_test)) * 100
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error (MSE):", mse)
print("Mean Absolute Error (MAE):", mae)
print("Root Mean Squared Error (RMSE):", rmse)
print("R-squared (R^2):", r2)
print("Mean Absolute Percentage Error (MAPE):", mape)
```

Figure 61 - Code used for training.

The performance measurement of each model is listed in the table below.

Learning rate	Epochs	Batch Size	MSE	MAE	RMSE	R-squared
0.0001	200	10	5704336020	55925.2313	75527.05489	0.181591888
0.0001	200	15	6153417239	59503.83715	78443.7202	0.117161653
0.0001	200	20	6125561575	58969.68055	78265.9669	0.121158139
0.0001	200	25	6355163324	60475.62334	79719.27824	0.088216894
0.0001	200	30	6605260007	61681.21232	81272.7507	0.052335215
0.00001	200	10	7689975844	57817.16352	87692.50734	-0.1032903
0.00001	200	15	9618716304	58767.38204	98075.05444	-0.38000907
0.00001	200	20	11094227813	65757.63904	105329.1404	-0.59170252
0.00001	200	25	11156791128	66038.61925	105625.7124	-0.60067856
0.00001	200	30	11999349547	70885.44287	109541.5426	-0.72156145
0.001	200	10	4085352726	43462.57962	63916.76405	0.413869415
0.001	200	15	4757425511	48245.82681	68974.09304	0.317446305
0.001	200	20	4781028995	49372.32361	69144.98532	0.314059884
0.001	200	25	4809052812	48286.68202	69347.33457	0.310039272
0.001	200	30	5165659247	50040.27446	71872.52081	0.258876508
0.001	300	10	3177476174	39078.6392	56369.10655	0.544123581
0.001	300	15	3411684538	40631.40945	58409.6271	0.510521419
0.001	300	20	3564320340	41275.56322	59701.92912	0.488622572
0.001	300	25	3856417155	42244.43688	62100.0576	0.446715083
0.001	300	30	4074178419	43435.27468	63829.29123	0.415472606
0.001	400	10	2415420380	33922.10212	49146.92645	0.653456664
0.001	500	10	2574676180	34116.25153	50741.26704	0.630608039
0.001	600	10	262477264.86319897	9216.519811316382	16201.150109273074	0.9671274487609389

0.001	700	10	2497199439	32732.19622	49971.98654	0.641723722
0.001	800	10	2803558836	35115.89763	52948.64339	0.597769962
0.001	900	10	2299955576	28855.13921	47957.85207	0.670022542
0.001	1000	10	2304592298	29547.49194	48006.16937	0.669357306
0.001	1100	10	2215660805	29614.57317	47070.8063	0.682116416
0.001	1200	10	2049342286	27658.11041	45269.66187	0.743341164
0.001	1300	10	2529195758.11326	29764.633904952036	50291.110130054396	0.683244500450252 5
0.001	2000	10	2336590232	29904.2885	48338.2895	0.664766523

Table 2: Performance Outcome of 1D CNN Model

For ease of visualization of each evaluation metric, I have separated them against the learning rate, but they have similar variance with Epoch and Batch size. For clarity, I have also included the Table of Learning rate, Epoch, and Batch size at the end of the below diagrams.

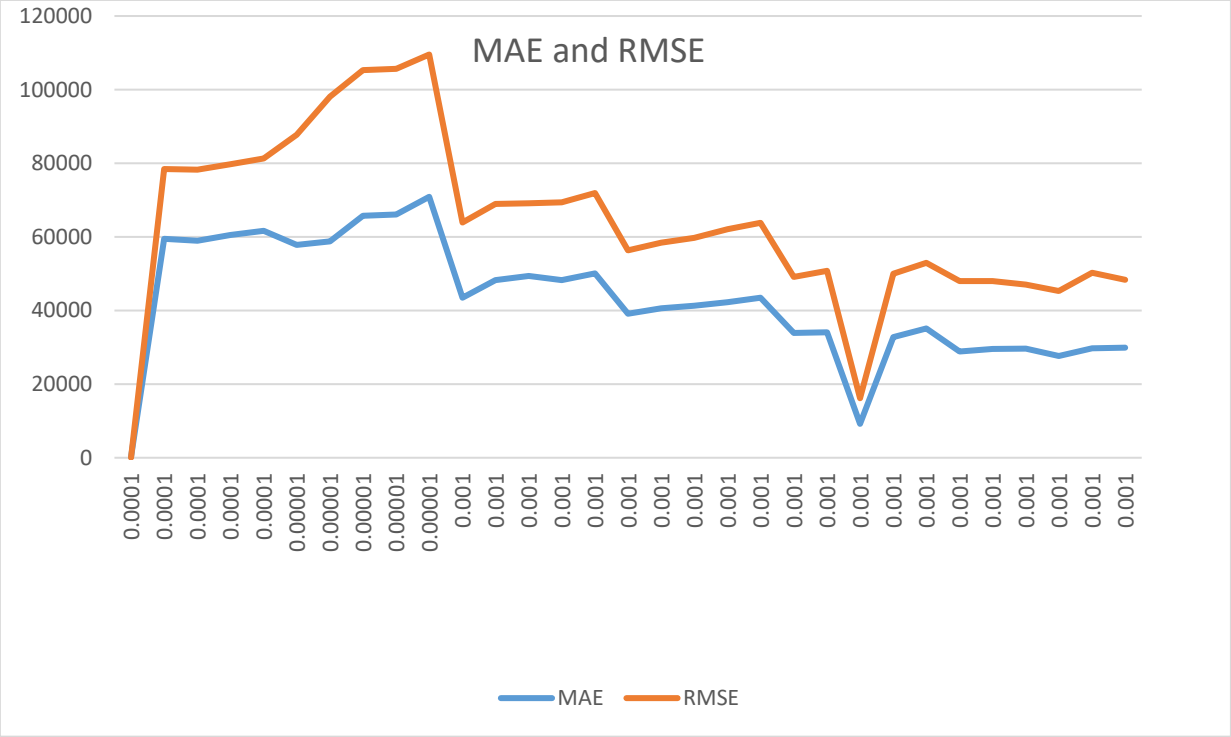


Figure 62 - MAE and RMSE against Learning rate.

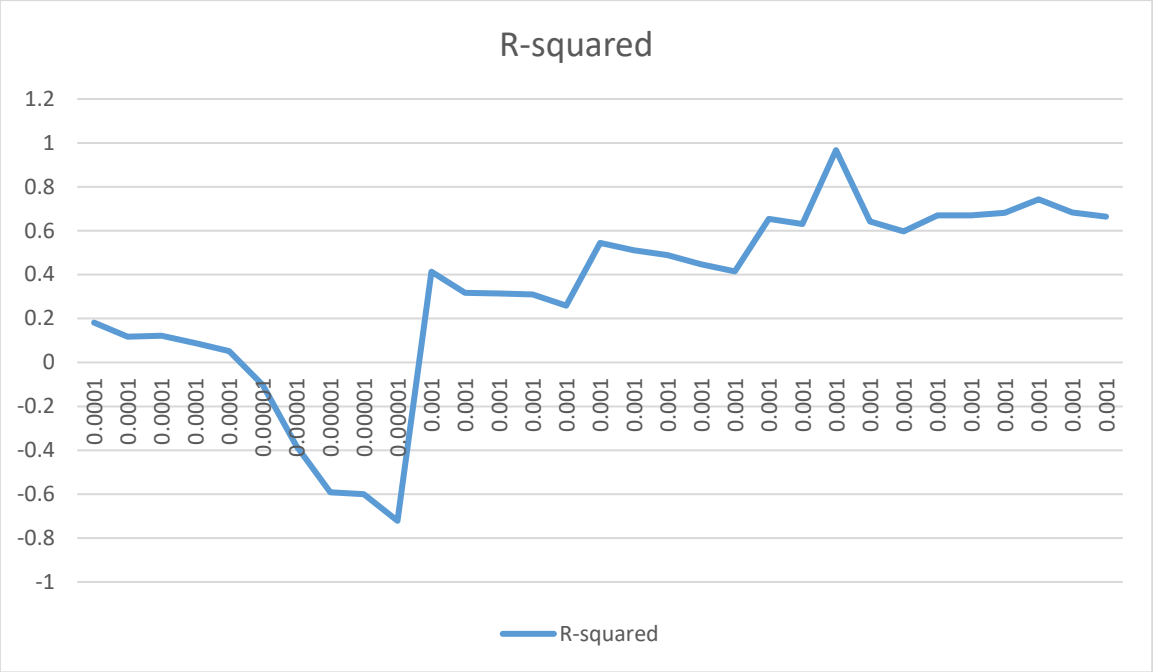


Figure 63 - R-squared against Learning rate.

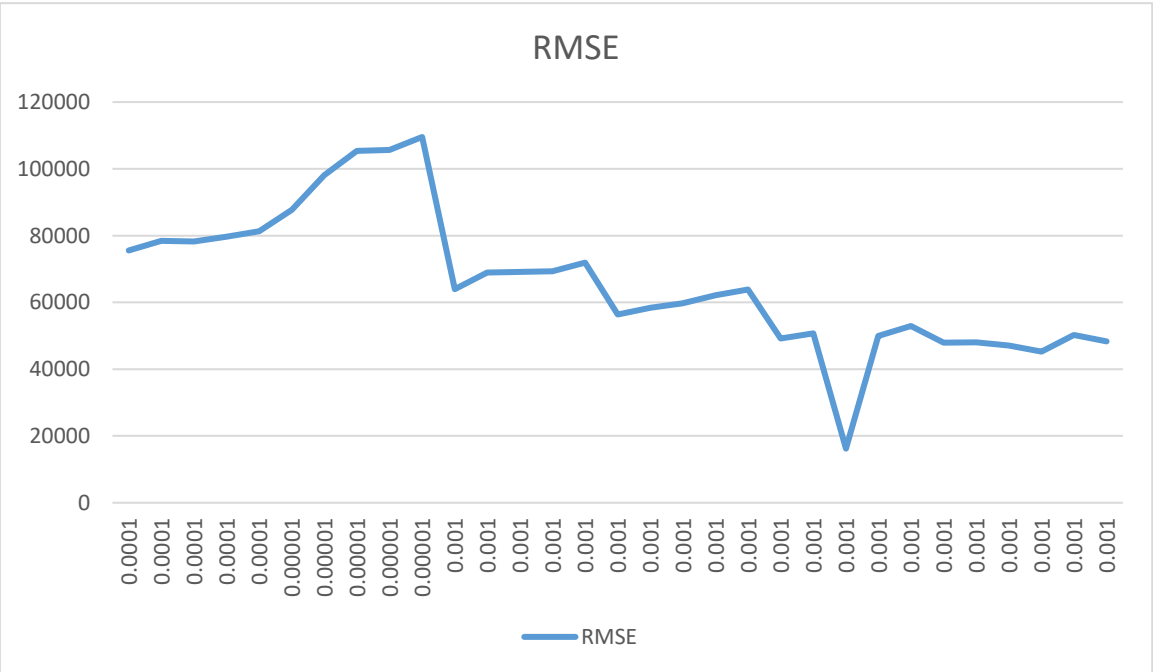


Figure 64 - RMSE against Learning rate.

Learning rate	Epochs	Batch Size
0.0001	200	10
0.0001	200	15
0.0001	200	20
0.0001	200	25
0.0001	200	30
0.00001	200	10
0.00001	200	15
0.00001	200	20
0.00001	200	25
0.00001	200	30
0.001	200	10
0.001	200	15
0.001	200	20
0.001	200	25
0.001	200	30
0.001	300	10
0.001	300	15
0.001	300	20
0.001	300	25
0.001	300	30
0.001	400	10
0.001	500	10
0.001	600	10
0.001	700	10
0.001	800	10
0.001	900	10
0.001	1000	10

0.001	1100	10
0.001	1200	10
0.001	1300	10
0.001	2000	10

Table 3 – Learning Rate, Epoch, and Batch Size Behavior

4.3 How to analyze the model using the above measurements.

When the Predicted Rice yield against the actual Rice Yield graph was drawn, it looks like below.

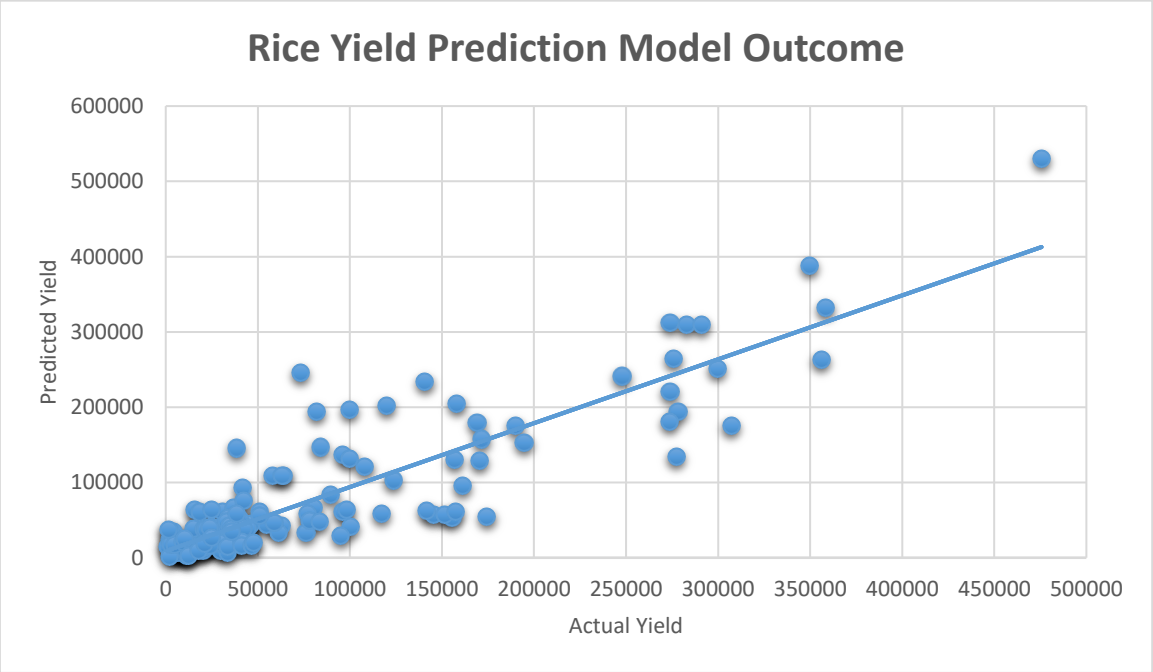


Figure 65 - Rice Yield Prediction Model Outcome

After analyzing evaluation metrics graphs in *Figure 62 - MAE and RMSE against Learning rate.*, *Figure 63 - R-squared against Learning rate.*, and *Figure 64 - RMSE against Learning rate.* of each iteration, I observed that the optimal learning rate is 0.001 and batch size is 10 for the 1D CNN model from all the models I have trained.

Next conclusion I have observed that epoch count 600 is a good choice for the model training, even though some epoch counts like 1200 have better values for measurements, there is a risk of overfitting to data.

The above X-Y scatter plot is drawn with Predicted against actual values for the highest performing model I could train out of this dataset, which are learning rate = 0.001, epoch=600, and batch size = 10.

As the graph shows, the model should be further improved in order to get accurate predictions for the intended goal, which is to predict the total seasonal rice yield of any location in Sri Lanka.

CHAPTER 5 - CONCLUSION AND FUTURE WORK

5.1 Conclusion

When analyzing *Table 2* and evaluation metrics behaviors, which are *Figure 62 - MAE and RMSE against Learning rate.*, *Figure 63 - R-squared against Learning rate.*, and *Figure 64 - RMSE against Learning rate.* I can speculate that my best performing model was when the Algorithm Learning rate was 0.001, Batch Size was 10 and Epochs was 600 for CNN with below evaluation metrics.

- MSE - 262477264.86
- MAE - 9216.52
- RMSE - 16201.15
- R-Squared - 0.9671274487609389

According to *Figure 65* When the actual yield is less than 50000 MT, the model is more prone to produce near correct results. Especially when the actual Seasonal yield is under 50000 MT. But when the actual total yield surpasses 100000 MT model seems to predict inaccurate rice yields. But if I analyze historical rice yield data from 2004 to 2023 from *Figure 2* to *Figure 26*, I noticed that only Ampara, Anuradhapura, Hambantota, Mannar, Trincomalee, Kurunegala, Badulla, Batticaloa, Monaragala, and Polonnaruwa had ever produced the total seasonal yield over 100000 MT, making my current model suitable for all other locations' rice yield prediction.

Even though the project intended to develop a CNN model which able to predict the total rice yield of areas that have total rice yields of less than 100000 MT, such as Colombo, Gampaha, Kalutara, Kandy, Matale, Galle, Matara, Jaffna, Puttalam, Ratnapura, Nuwara Eliya, Kegalle, Kilinochchi, Mullaitivu, and Vavuniya.

5.2 Limitations of the project

In the current project, I assumed that except for weather conditions other conditions are the same. Even though Sri Lanka is a relatively small country, it has diverse climate changes among districts. The southwest region, known as the wet zone, receives an average annual rainfall exceeding 2,500 mm, largely influenced by the southwest monsoon. On the other hand, the dry zones located in the south and northwest receive less than 1,750 mm of rainfall. The intermediate zones, found in the eastern and central regions, receive rainfall ranging from 1,750 mm to 2,500 mm, primarily due to the northeast monsoon. In certain areas of the southwest slopes of the central hills, the annual rainfall can reach up to 5,000 mm, and the amount of rainfall can vary by more than 1,000 – 2,000 mm within a distance of less than 100 km. Some areas have access to water for cultivation throughout the year, even though no rain for days. Some have seasonal drying rivers like the Mahawali River, some have rarely dried-out rivers like the Nilwala River. Some districts like Anuradhapura, and Polonnaruwa have man-made lakes for irrigation. Some districts depend solely on rain. Therefore, assuming weather conditions are equal for everyone using climate data could impact on ML model which trained for data from Sri Lanka

In the current research, I didn't discard the data that had an external impact. Such as below.

- Drought times when selecting data sets – Drought has a severe impact on the paddy because of its semi-aquatic nature. Even though Sri Lanka is a small island nation, it would be a rare occasion for a drought could impact the island due to its high geographical variability in the island nation. Therefore, using drought times data without considering the effective area could work as outliers while training the data.
- Politically impact scenarios – The fertilizer bans in the time of former president Gotabaya also had a significant impact on rice yield production as per farmers' comments. This is natural because paddy yield has a significant impact on soil conditions such as pH, soil nutrients, etc.
- Changes in agricultural practices – Change in agricultural practices also has a significant impact on Rice yield. For example, moving away from traditional fertilizers to tea fertilizers. As per the government's suggestion farmers recently moved to tea fertilizers due to their significant improvement in yield. Therefore, using such data with traditional fertilizer rice yield data would appear data anomaly. On the other hand, I can't ignore the data, because if farmers continue to use tea fertilizer, then the selected model will produce higher incorrect predictions.
- Data Outliers – From Analyzing Rice yield data from *Figure 2* to *Figure 26*, I can see I didn't remove significant differences in rice yield outlier data such as Gampaha Yala in time 2005-2006 (*Figure 3*) or Jaffna-Yala 2013-2014 (*Figure 10*) etc.

5.3 Future Work

In the current study, I used historical climate factors and rice yield to develop a rice yield prediction CNN model. By considering my experience and literature around rice yield I can identify several future works that can be undertaken to improve Rice yield prediction models or precision agriculture.

- Rice yield prediction Classifier for Sri Lanka - As of now there is no model of any kind of Machine Learning algorithm, that predicts any crop yield and covers all the Sri Lanka. Most of the literature is scoped around part of Sri Lanka.
- Total Annual Sri Lankan Rice Prediction Model – even in my project, I considered only predicting rice yield for certain areas, even though I used data that covered whole Sri Lankan rice-producing areas. There is no literature on how to predict the total Sri Lankan Annual or Seasonal Rice yield.
- Train district-wise model – If we trained district-wise models, we could guarantee that most common conditions are homogeneous for the dataset, like extra water conditions, soil conditions, or impactful weather conditions.
- Use soil conditions for the model – Even though I couldn't find any source that could give ph. Value of each district, there are hourly data for Soil Temperature and soil Moisture available on open-meteo.com. Which can be used for the training model.
- Use other soil conditions for models – In the Sri Lankan context, there was no literature found that uses pH values or Soil nutrients such as Nitrogen (N), Phosphorus (P), or Potassium (K).

REFERENCES

Amaratunga, V., Wickramasinghe, L., Perera, A., Jayasinghe, J.M.J.W., Rathnayake, U., 2020. Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data. Mathematical Problems in Engineering 2020. https://doi.org/10.1155/2020/8627824

Baral, S., Kumar Tripathy, A., Bijayasingh, P., 2011. Yield Prediction Using Artificial Neural Networks, in: Das, V.V., Stephen, J., Chaba, Y. (Eds.), Computer Networks and Information Technologies, Communications in Computer and Information Science. Springer, Berlin, Heidelberg, pp. 315–317. https://doi.org/10.1007/978-3-642-19542-6_57

Chandra, A., Mitra, P., Dubey, S.K., Ray, S.S., 2019. Machine Learning Approach for Kharif Rice Yield Prediction Integrating Multi-Temporal Vegetation Indices and Weather and Non-Weather Variables. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 423, 187–194. https://doi.org/10.5194/isprs-archives-XLII-3-W6-187-2019

Ekanayake, P., Wickramasinghe, L., Jayasinghe, J.M.J.W., 2022. Development of Crop-Weather Models Using Gaussian Process Regression for the Prediction of Paddy Yield in Sri Lanka. International Journal of Intelligent Systems and Applications 14, 52–665. https://doi.org/10.5815/ijisa.2022.04.05

Gandge, Y., Sandhya, 2017. A study on various data mining techniques for crop yield prediction, in: 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT). Presented at the 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 420–423. https://doi.org/10.1109/ICEECCOT.2017.8284541

Gandhi, N., Petkar, O., Armstrong, L., Tripathy, A., 2016. Rice crop yield prediction in India using support vector machines. pp. 1–5. https://doi.org/10.1109/JCSSE.2016.7748856

González-Sanchez, A., Frausto-Solis, J., Ojeda, W., 2014. Predictive ability of machine learning methods for massive crop yield prediction. SPANISH JOURNAL OF AGRICULTURAL RESEARCH. https://doi.org/10.5424/sjar/2014122-4439

Han, X., Liu, F., He, X., Ling, F., 2022. Research on Rice Yield Prediction Model Based on Deep Learning. Computational Intelligence and Neuroscience 2022, 1922561. https://doi.org/10.1155/2022/1922561

Khaki, S., Wang, L., 2019. Crop Yield Prediction Using Deep Neural Networks. Frontiers in Plant Science 10, 621. https://doi.org/10.3389/fpls.2019.00621

Muthusinghe, M.R.S., S.T., P., Weerakkody, W.A.N.D., Saranga, A.M.H., Rankothge, W.H., 2018. Towards Smart Farming: Accurate Prediction of Paddy Harvest and Rice Demand, in: 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). pp. 1–6. https://doi.org/10.1109/R10-HTC.2018.8629843

Nevavuori, P., Narra, N., Lipping, T., 2019. Crop yield prediction with deep convolutional neural networks. Computers and Electronics in Agriculture 163, 104859. https://doi.org/10.1016/j.compag.2019.104859

Ramesh, D., n.d. ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES. International Journal of Research in Engineering and Technology 04, 470.

Ruß, G., Kruse, R., Schneider, M., Wagner, P., 2008. Data Mining with Neural Networks for Wheat Yield Prediction. pp. 47–56. https://doi.org/10.1007/978-3-540-70720-2_4

Satpathi, A., Setiya, P., Das, B., Nain, A., Jha, P.K., Singh, Surendra, Singh, Shikha, 2023. Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield

Forecasting for Chhattisgarh, India. Sustainability 15, 2786. https://doi.org/10.3390/su15032786
Wickramasinghe, L., Jayasinghe, J.M.J.W., Rathnayake, U., 2020. Artificial Neural Network Approach for Paddy Yield Prediction.
Wickramasinghe, L., Weliwatta, R., Ekanayake, P., Jayasinghe, J., 2021. Modeling the Relationship between Rice Yield and Climate Variables Using Statistical and Machine Learning Techniques. Journal of Mathematics 2021, 6646126. https://doi.org/10.1155/2021/6646126
Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., Zhu, M., Wu, X., 2019. Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China. Ecological Indicators 101, 943–953. https://doi.org/10.1016/j.ecolind.2019.01.059