Paddy Yield prediction using Rainfall

Y.D Abeygunewardana.

IT20115784

B.Sc. (Hons) Degree in Information Technology

Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2023

Paddy Yield prediction using Rainfall

Y.D Abeygunewardana

IT20115784

B.Sc. (Hons) Degree in Information Technology

Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

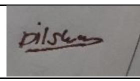
September 2023

Declaration

I declare that this is my own work and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

IT20115784	**Y.D Abeygunewardana**	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the Supervisor: ……………………………………

Date:…15/10/2023…

Abstract

In most of Asia, including Sri Lanka, rice is the main agricultural crop. It is well knowledge that climate change has a significant impact on the cultivation of paddy. Therefore, it is beneficial in many ways to understand the connections between climatic conditions and paddy yield. This study uses paddy harvesting data from the Sabaragamuwa Province, Sri Lanka, to demonstrate an artificial neural network (ANN) framework that may be utilized to assess the correlations between meteorological factors and the paddy yield. The district of Rathnapura produces the most paddy in Sri Lanka, thus the study has a lot of potential and interest. In this study, climate variables including rainfall were taken into account. In sum, our study, with a particular emphasis on the Rathnapura District, confirms the importance of rainfall data in predicting paddy yield. By developing resilience against the effects of climate change and supporting sustainable paddy farming within this particular geographic setting, it presents a personalized and region-specific approach to agricultural planning. Understanding the significance of rain in Rathnapura's agriculture is essential to assuring locals' access to food, economic success, and general wellbeing.

Keywords: Artificial neural network, Paddy yield, Mean squared error, rainfall;

Acknowledgement

I would like to thank my supervisor, Ms. Vindhya Kalapuge, for the guidance and motivation provided in order to make this research a success. I would also like to thank the Department of Information Technology of the Sri Lanka Institute of Information Technology as well as the CDAP lecturers and staff for the guidance and support provided.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
MASE	Mean Absolute Scaled Error
QoS	Quality of Service
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
ANN	Artificial Neural Network
LSTM	Long Short-Term Memory

List of Appendices

1.0 ITRODUCTION

The Rathnapura District is a region where nature's wealth converges with centuries of agrarian heritage. It is nestled in the lush embrace of Sri Lanka's verdant heartland. Rathnapura, known as the "City of Gems" due to its historical ties to precious stones, has treasures both below the earth's surface and in the rich soil that supports its population. None of these assets, however, shines more brilliantly than the age-old art of paddy farming. In addition to providing food for the people of Rathnapura, the sacred grain paddy also keeps the local economy humming.

The agriculture sector in the Rathnapura District has a difficult challenge: guaranteeing a reliable, resilient output of paddy while the global population soars and the threat of climate change casts a foreboding shadow. This mission goes beyond simple agriculture; it is a mandate of the utmost significance. The foundation of Rathnapura's food security is paddy rice, which is intertwined with locals' daily lives. Additionally, it supports the region's economic stability by giving jobs to numerous families and anchoring the local economy.

The analysis of a key aspect of contemporary agricultural methods in the Rathnapura District, specifically the accurate forecasting of paddy production using rainfall data, forms the core of this work. Beyond its direct impact on agriculture, this research has the potential to transform farming methods, improve food security, and promote sustainability in this particular geographic location.

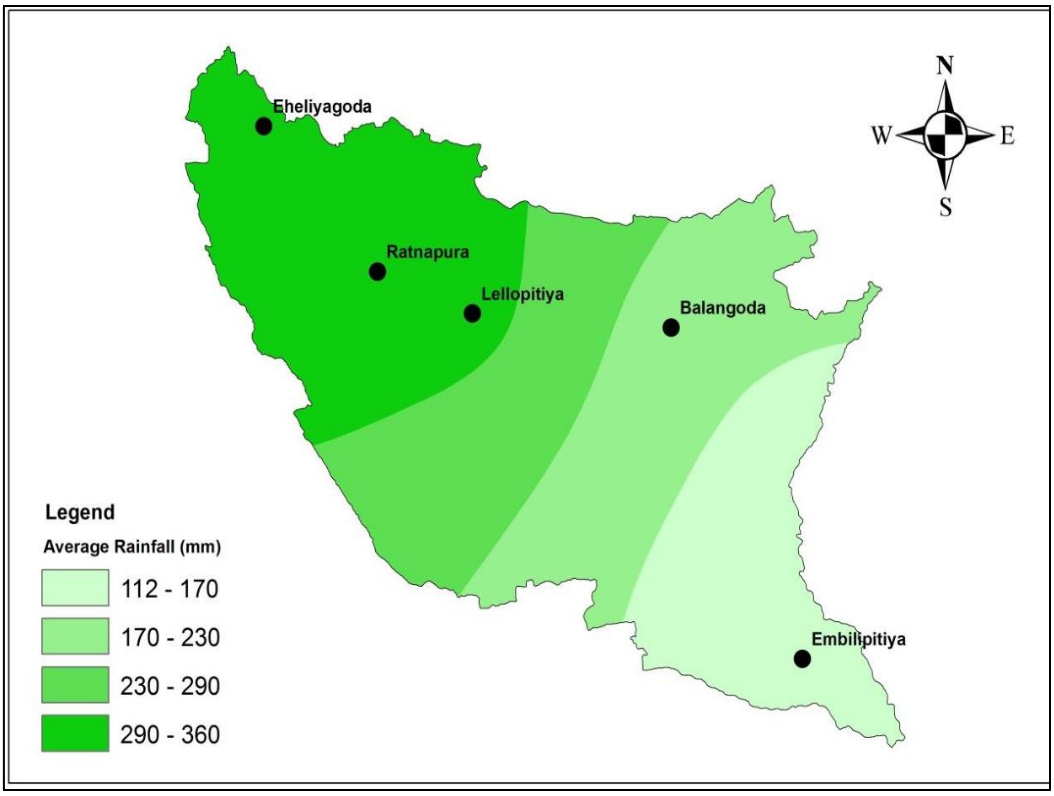


Figure 1:Rainfall Variability of South-West Monsoon: A Special Study Based on Ratnapura District, Sri Lanka

Rathnapura's geography unfolds like a tapestry, intertwined with heights, microclimates, and agro-ecological zones that characterize the area's varied farming circumstances. Local farmers have long understood that rainfall is the fundamental conductor composing the symphony of success or failure in paddy production in the face of such intricacy. The three components of rainfall—quantity, timing, and distribution—have a significant impact on paddy growth and production. As a result, the agrarian community of Rathnapura places a premium on comprehending and using rainfall statistics. figure 1

In order to reveal patterns and linkages particular to Rathnapura's agricultural landscape, our study carefully weaves together extensive paddy yield records with a plethora of historical and real-time rainfall data. We aim to demonstrate the fundamental importance of rainfall in determining the course of paddy agriculture in this area through thorough investigation.

Additionally, this research aims to present a cutting-edge predictive model that has been expertly calibrated to take into consideration Rathnapura's unique meteorological characteristics. This model provides local farmers with a practical and useful tool for decision-making by pinpointing the precise threshold rainfall values for various paddy growth phases. It signals the start of precision agriculture in Rathnapura, a paradigm change that gives farmers the power to maximize resource use, carefully plan irrigation schedules, and use efficient crop management techniques.

The Rathnapura District finds itself at a turning point in time characterized by growing climatic uncertainties. It is not immune to the effects of changing weather patterns, difficulties that echo through every field of rice and resonate around every dinner table. In light of this sobering reality, the significance of our study is increased even further. It gives the farming community in the Rathnapura District a proactive plan for resilience and adaptation to changing climatic circumstances.

Overall, this study charts the complex relationship between rainfall and paddy production in the Rathnapura District, setting off on an intriguing adventure. Its main goal is to emphasize the crucial and vital function that rainfall information plays in the field of agricultural planning. By doing this, it aims to spark a transformation that will lead to sustainable practices, strengthened food security, and stable livelihoods for the local population. We aim to significantly contribute to the prosperity and well-being of Rathnapura's dynamic farming communities by exploring this crucial intersection of climate, agriculture, and technology. Along the way, we learn about the art of sustainability, the fundamentals of resiliency, and the hope for a better future for Rathnapura in addition to the science of farming.

However, due to climate unpredictability, paddy is one of the crops that is most impacted [1].

1.1 Background and Literature

The lovely island nation of Sri Lanka's Rathnapura District is known for its fertile plains and verdant surroundings. The district's economy is based primarily on agriculture, with paddy cultivation being one of the main agricultural pursuits. Rice, commonly known as paddy, is a key crop in Sri Lanka, and a number of variables, including rainfall, affect its productivity.

Rathnapura District's agricultural environment is significantly shaped by rainfall. A distinct rainy season and dry season are present in the district's tropical rainforest climate. The success of paddy agriculture is directly impacted by the quantity and distribution of rainfall.

Heavy monsoonal rains during the wet season, which normally lasts from May to September, give paddy fields the moisture they require. When there is sufficient rainfall during this time, paddy plants receive the water they need for growth and development. On the other hand, a lack of rainfall during this crucial stage can result in lower yields, which will have an impact on the livelihoods of several farmers in the area.

Conversely, irrigation is necessary to maintain paddy crops during the dry season, which lasts from December to February. To effectively harness and manage water supplies, farmers in Rathnapura District have created sophisticated irrigation systems, including antiquated cascading tank systems. In the past, these systems have proven crucial for reducing the effects of unpredictable rainfall patterns.

Recent years have seen increased difficulties with paddy production in Rathnapura District due to climatic change. Crop yields are seriously threatened by unpredictable weather patterns, especially erratic and intense rainfall events. To lessen these risks, many farmers are implementing climate-resilient agricultural practices and diversifying their crop production.

In conclusion, rainfall patterns in the area play a significant role in predicting paddy yield in the Rathnapura District. The thorough monitoring and analysis of rainfall, which is the lifeblood of agriculture in this lush region, is crucial for ensuring food security and maintaining the livelihoods of the surrounding farming community. Innovative methods for predicting rainfall and adaptable farming techniques will be essential for the future of paddy production in Rathnapura District as climate change continues to have an impact on weather patterns.

ANN Structure is derived from biological neural processes in the human brain. The technique is developed using the relationships between the neurons. Several studies for crop prediction have discussed the potential of ANN use. A study forecasted rice yield using climatic predictions figures at most 40-60 kg per hectare. [2] One more by [3] has been cited as describing a straightforward and accurate tool for estimating rice production.

Rice production forecast using the application of the ANN model was also examined for South Asia. During training, ANN provides acceptable errors. That describes how accurate a model is. Next, in a comparison of the output, the ANN result reveals the same result that validates the accuracy of the forecast. In the Siraha district of Nepal, the anticipated effect might be applied to increase paddy yield [4].

Other recent studies have shown the effectiveness of a DSS in Thailand's Phimai district that employed ANN to forecast rice production [5]. Details from another learning that proved the use of ANNs with feed forward back propagation for farming produce forecasting were related results

1.1.1 Centrality Measures

Using rainfall data in the Rathnapura District, it will be essential to identify critical variables and their impact on paddy yield forecast using centrality metrics including degree, betweenness, and eigenvector centrality. These steps will aid in the prioritization of the meteorological aspects that have the greatest influence, allowing for more efficient resource allocation and model improvement for farmers and decision-makers to make better-informed choices.

1.1.1.1 Degree Centrality

Degree centrality will identify the most interconnected meteorological parameters while projecting paddy output with rainfall in Rathnapura District. We can determine which variables have the strongest associations with paddy yield by counting the number of direct connections each variable has. The forecast model will be greatly influenced by these intricately linked variables, which will help us narrow our attention to the crucial climatic factors that have the most impact on crop outcomes. For stakeholders in agriculture and government, degree centrality will be a useful tool for prioritizing data gathering, model improvement, and resource allocation, resulting in a more precise and usable prediction framework.

1.1.1.2 Closeness Centrality

We will use closeness centrality to discover the meteorological dataset variables that are most closely related to one another in order to estimate paddy yield using rainfall data in Rathnapura District. This metric measures the speed at which knowledge or power can move from a given variable to the rest of the network. In this study, closeness centrality will identify the meteorological variables that are crucial in connecting rainfall effects to final paddy yield results. By concentrating on these crucial elements, we can learn important lessons about the immediate and direct effects of climatic conditions on crop yield, assisting farmers and policymakers in making better decisions.

1.1.1.3 Betweenness Centrality

Due to the region's reliance on rain-fed agriculture, the prediction of paddy production in Sri Lanka's Rathnapura District heavily depends on rainfall patterns. An important factor in estimating crop productivity is rainfall. A network analysis metric called the Betweenness Centrality of rainfall data can provide important insights into how various meteorological variables interact. We can pinpoint important rain-dependent times that have a significant impact on paddy farming by evaluating the Betweenness Centrality of the rainfall in this situation. Understanding the significance of specific rainy seasons enables farmers to organize their planting and harvesting operations wisely, thus increasing agricultural productivity and ensuring food security in Rathnapura District.

1.1.1.4 Eigenvector Centrality

Predicting paddy production is an important agricultural task in Sri Lanka's Rathnapura District, where crop results are greatly influenced by rainfall. A network analysis metric called eigenvector centrality can be used to evaluate how different environmental parameters affect yield. The significance of the rainfall in the complex interaction of meteorological factors affecting crop development is indicated in this context by the eigenvector centrality of the rainfall. We can create forecast models that take into account the unique characteristics of the area by examining historical rainfall data and its importance within the larger meteorological network. These models offer more precise projections of paddy output, assisting local farmers in making decisions, allocating resources efficiently, and improving food security.

1.2 Research Gap

A growing body of research has focused on predicting paddy yield using artificial neural networks (ANN) and information about rainfall. The significance of paddy production for global food security and the livelihoods of millions of people has led to this research focus. More than half of the world's population depends on paddy, or rice, as a staple crop, and the output of this crop is quite sensitive to changes in the weather, especially rainfall. For the purpose of guaranteeing food security and effectively managing agricultural resources, accurate paddy yield prediction is crucial. Several research gaps need to be filled despite the advancements made in this area.

The requirement for reliable and accurate predictive models that can capture the intricate relationship between rainfall patterns and paddy crop yields is a significant research gap in the field of paddy yield prediction utilizing rainfall data and ANN. Even while ANNs have showed promise in simulating this relationship, current models frequently lack the depth and complexity needed to take into consideration the wide range of factors that affect paddy farming. In addition, it is necessary to investigate several ANN types, such as deep neural networks, convolutional neural networks, and recurrent neural networks, to ascertain which architecture is most appropriate for this particular task.

Lack of comprehensive, high-quality datasets that include several facets of paddy farming, such as soil type, crop management techniques, and regional environmental factors, is another significant research gap. It is necessary to incorporate these variables into the dataset in order to create precise ANN models for paddy yield prediction. Long-term historical data are also necessary to take climate variability and trends into consideration because they have a long-term impact on rice yields.

Additionally, there is a need for study to examine the integration of remote sensing data into the ANN-based paddy yield prediction models, such as satellite photos and aerial photography. The health, stage of growth, and environmental conditions of the crop can all be learned via remote sensing data. Incorporating such data into the modeling process could improve the precision and timeliness of paddy output projections, enabling farmers and policymakers to make more informed decisions.

Another significant study need is the paddy yield prediction's temporal component. Understanding the temporal dynamics of rainfall patterns and their effects on paddy crops is crucial for accurate forecast. To enable more accurate predictions and early interventions, research should concentrate on creating ANN models that can capture the seasonality and patterns in paddy yields.

A research vacuum exists in addition to the technical issues regarding the adoption and application of ANN-based paddy yield prediction systems in actual agricultural practices. Although there are viable models in research settings, it is important to consider how farmers, particularly those in low-resource areas, may use and access them. Focusing on user-friendly interfaces, integrating data with local knowledge, and enhancing the capacity of farmers and extension agents are necessary to close this gap.

The potential effects of climate change on the forecasting of paddy yields must also be taken into account. Understanding how these changes affect rainfall patterns and, in turn, paddy agriculture, is of utmost relevance as global climate patterns continue to change. It should be investigated how ANN models might be modified to take climate change scenarios into account and support the development of resilient agricultural practices.

In conclusion, the research gap in paddy yield prediction using rainfall data and ANN is a complicated issue that involves model complexity, data quality, temporal dynamics, the integration of remote sensing, real-world application, and climate change adaptation. To improve the precision and applicability of paddy yield prediction models and eventually contribute to food security and sustainable agriculture, it is crucial to fill these research gaps. Researchers and stakeholders should work together to investigate creative solutions to these problems and guarantee the advancement of this important field of study.

2.0 RESEARCH PROBLEM

A crucial and complex issue, the research subject of estimating rice production using rainfall data in the Rathnapura District has substantial consequences for Sri Lankan agriculture, food security, and rural lives. Accurate yield forecasts are crucial because this area, well known for its significant contribution to the nation's rice production, mainly depends on the monsoon rains. With better resource allocation, risk reduction, and overall agricultural sustainability possible, this research challenge aims to address the urgent need for a trustworthy and locally customized prediction model that can inform farmers, regulators, and stakeholders about prospective crop outcomes.

The Rathnapura District experiences different wet and dry seasons due to its location in a tropical monsoon climate zone. The volume, distribution, and timing of rainfall directly affect the soil moisture levels necessary for rice growth in this region, which are directly correlated with the success of paddy agriculture. Therefore, the main focus of this research is on creating a reliable predictive model that can use historical rainfall data to reliably estimate paddy yield results.

The great variation in rainfall patterns within the Rathnapura District is one of the most noticeable difficulties posed by this study subject. The area has microclimates that, even within very small geographic areas, can result in significant differences in the distribution of rainfall. Therefore, to develop a trustworthy predictive model, it is essential to gather and examine high-resolution rainfall data at the sub-district or even village level. Additionally, to improve the model's precision and thoroughness, additional meteorological elements like temperature, humidity, and wind patterns may need to be taken into account.

To build the prediction model, the proposed research would make use of cutting-edge machine learning and data science methodologies. This model will be built using painstakingly matched historical rainfall data over multiple decades and paddy yield records. Regression analysis, neural networks, and ensemble approaches are just a few examples of the machine learning algorithms that will be used to create a solid, data-driven relationship between rainfall and paddy output. To guarantee the model's dependability and generalizability, it will undergo rigorous training and validation using cross-validation approaches.

Additionally, this research issue acknowledges how the Rathnapura District's rainfall patterns may soon be impacted by climate change. Understanding how global climate change affects paddy production is becoming more and more important as precipitation patterns continue to change. In order to make the predictive model more adaptable and produce more precise long-term predictions, the research will investigate creative ways to incorporate climate change projections into it.

The findings of this project are expected to have important practical ramifications. The Rathnapura District's farmers stand to gain a great deal from accurate and timely production estimates. With this knowledge, they are better equipped to choose crops, arrange planting times, and allocate resources, thus improving their chances of farming success. The model's insights can also be used by politicians and agricultural authorities to establish well-informed plans for disaster preparedness, climate adaptation, and agricultural development—all of which are essential for the region's sustainable growth and food security.

Using rainfall data to estimate paddy output in the Rathnapura District is a complex research subject that necessitates a multidisciplinary approach. This study intends to provide important insights into the complex interaction between rainfall patterns and agricultural outcomes by leveraging the capabilities of data science, machine learning, and climate modeling. In the end, the creation of a precise predictive model will considerably boost the resilience and sustainability of the rice industry in this crucial agricultural region, ensuring Sri Lanka's economic stability and food security.

3.0 OBJECTIVES

3.1 Main Objective

This study's main goal is to build a reliable predictive model for paddy output using rainfall data from the Rathnapura District. With the help of this model, farmers and policymakers will be able to make informed choices about crop planning, resource allocation, and disaster preparedness. The project aims to uncover the intricate interaction between climatic conditions and agricultural outcomes, thereby boosting the region's resilience, sustainability, and food security. It does this by utilizing cutting-edge techniques in data science, machine learning, and centrality metrics. This key goal emphasizes how crucial it is to use data-driven insights to improve rice agriculture in this crucial agricultural region.

3.2 Specific Objectives

The specific objectives of this research are as follows

- To collect and analyze historical rainfall data at a high spatial resolution in the Rathnapura District.
- To compile and correlate this rainfall data with historical paddy yield records spanning several decades.

- To develop and validate a predictive model using machine learning techniques to forecast paddy yield based on rainfall patterns.
- To integrate centrality measures, including degree, betweenness, and eigenvector centrality, to identify key meteorological factors influencing paddy yield.
- To assess the potential impact of climate change on rainfall patterns and incorporate climate projections into the predictive model.
- To provide actionable insights to farmers, policymakers, and stakeholders, enabling them to optimize agricultural practices and ensure food security in the region.

4.0 METHODOLOGY

4.1 Requirement Gathering

The research on paddy yield prediction utilizing rainfall in Rathnapura District must be carried out successfully, and this requires effective requirement gathering. In order to complete this process, detailed information must be gathered regarding historical rainfall patterns, paddy cultivation techniques, and yield records. For model development, it also needs access to sophisticated data analytics tools and computer resources. To gain domain-specific insights and guarantee data accuracy, collaboration with local agricultural authorities, meteorological agencies, and farmers is essential. For model validation and integration of climate change projections, specialists in machine learning and climate science are also required. A strong and useful research framework is built on these requirements.

4.1.1 Past Research Analysis

Our current study was made possible by earlier investigations on the forecasting of paddy yields in the Rathnapura District. The relationship between rainfall and crop results has been the primary focus of earlier studies, underscoring the crucial role that climatic conditions play in the cultivation of rice. Our research aims to expand the field by adding machine learning and centrality measurements for more precise predictions.

Previous studies have used statistical models and historical data to find correlations. Furthermore, earlier studies frequently failed to consider the effects of climate change, which is something we intend to remedy by incorporating climate projections. Our research intends to develop a more thorough and practical predictive model for increased agricultural sustainability in the area by building on these earlier studies.

4.1.2 Identifying Existing Systems

It became clear that conventional statistical models had been applied when methods for paddy yield prediction utilizing rainfall in the Rathnapura District were identified. These models frequently rely on basic statistical studies and historical rainfall data. But more advanced data-driven systems and machine learning strategies are starting to emerge as viable substitutes. Existing systems frequently do not use cutting-edge methodologies or centrality measurements for finding important variables. Additionally, the majority of these systems do not take the effects of climate change into account. Our study intends to close these gaps and provide a more thorough and precise forecasting system for better-informed decision-making in the Rathnapura District's agriculture sector.

4.2 Feasibility Study

4.2.1 Technical Feasibility

It is technically possible to forecast paddy yield in Rathnapura District depending on rainfall. Farmers and agricultural specialists can create efficient plans for maximizing rice production by examining historical rainfall data and its impact on crop yields. In order to find links between rainfall patterns and the productivity of paddy crops, this involves gathering and examining rainfall records over a number of years. By revealing crucial growth stages of the crop where rainfall has the most impact, this data enables farmers to make well-informed choices. Farmers can plan for anticipated weather-related difficulties, such as drought or excessive rainfall, by incorporating current and forecasted rainfall data into predictive models. In general, using rainfall data to anticipate paddy production in Rathnapura District has the potential to enhance agricultural practices, boost crop yields, and guarantee food security in the area.

4.2.2 Schedule Feasibility

For agricultural schedule viability, predicting paddy production in Rathnapura District using rainfall data is essential. We can generate precise projections by looking at past rainfall patterns and taking into account the sorts of crops, the soil, and cutting-edge technology. With the help of these forecasts, farmers may optimize their planting and harvesting timetables, resulting in more productive crops and environmentally friendly farming methods.

4.2.3 Economic Feasibility

To determine whether paddy farming is economically feasible, Rathnapura District rainfall data must be used to predict paddy production. Farmers can reduce crop losses in dry years and ensure efficient water usage in rainy years by using historical rainfall patterns and their association with yield to make informed decisions, optimize resource allocation, and increase economic viability. The region's agriculture is more profitable, requires less inputs, and is more economically sustainable thanks to this predictive strategy.

4.3 Requirement Analysis

The complex task of predicting paddy (rice) production from rainfall data necessitates a methodical and in-depth evaluation of project requirements. The accomplishment of this task depends on a thorough comprehension of the different factors that influence the project's scope, objectives, and restrictions. The project's requirement analysis acts as the cornerstone, directing the choice of data sources, modeling strategies, and decision support technologies.

This requirement analysis tries to identify the key elements that must be taken into account while developing and implementing a project to predict paddy output using rainfall data. It acts as a road map for project stakeholders, researchers, and developers, identifying the crucial issues that must be resolved to successfully complete the project's goals.

4.4 System Analysis

The performance of paddy production, which is crucial to maintaining global food security, is closely related to climate variables, particularly rainfall. For secure food supply, accurate paddy yield forecast models are crucial. In this study, system analytic approaches are used to improve paddy production forecast, with a particular emphasis on the role played by rainfall patterns.

A comprehensive strategy called system analysis looks at the dynamic interactions that occur inside complex systems. By using this model to anticipate paddy yields, we want to reveal the complex connections between rainfall, farming methods, and crop results. Understanding these processes helps improve forecasts and guide policies for allocating resources, reducing risks, and adapting to climate change.

The purpose of this introduction is to set the tone for the following parts, which will discuss our research methodologies, data sources, modeling strategies, and findings. It also provides an overview of the significance of paddy agriculture and the function of rainfall. The urgency of accurate paddy yield estimates increases as climate change accelerates. Our research aids in the creation of adaptable models, helping farmers and decision-makers choose sustainable paddy production practices in the face of changing environmental conditions.

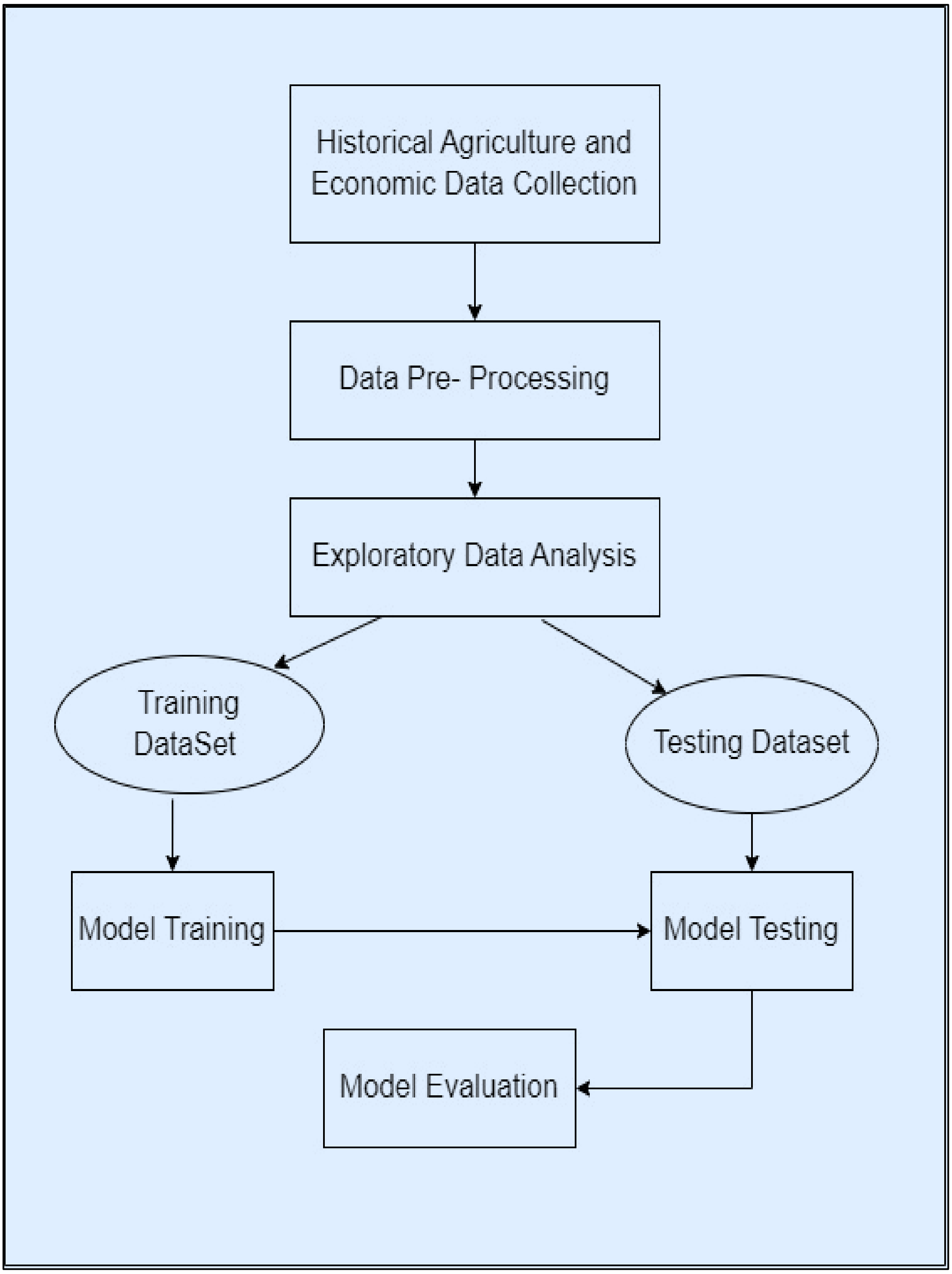


Figure 2 Proposed architecture for paddy yield prediction using rainfall

In Sri Lanaka, there are two periods of paddy cultivation harvesting namely "Yala" and "Maha." This creates a machine learning model to predict the rice yield in the above seasons in the Ratnapura district according to the precipitation over the period.

a. Data Collection

This study seeks to give a general overview of the procedure for gathering data in order to forecast paddy yields in Sri Lanka. For both farmers and governments, accurate forecasts are essential because they can help to improve crop management techniques, allocate resources wisely, and prevent possible food shortages.

Ratnapura			
Sown Extent (ha)	Gross Harvested Extent (ha)	Net Harvested Extent (ha)	Production (MT)
30,781	30,587	26,000	78,089
31,015	30,819	26,197	74,234
27,649	26,612	22,619	68,746
28,956	28,701	24,397	74,525
25,933	25,567	21,731	66,122
29,203	29,045	24,688	77,708
28,878	28,563	24,279	65,000
25,624	24,757	21,044	58,400
28,326	28,082	23,869	69,050
28,312	28,165	23,940	69,350
27,527	26,910	22,873	71,040
26,233	25,474	21,653	64,850
26,530	26,215	22,283	71,492
27,759	26,886	22,853	63,138
28,314	27,600	23,460	68,302
27,806	27,128	23,057	73,884
25,842	25,519	21,690	71,555

Figure 3:sample Data set for paddy yield prediction

Required data is collected from the agricultural department of Sri Lanka in Peradeniya. The amount of paddy cultivated in the Yala and Maha seasons and the rice yield was obtained from 2010 to 2021 as “Yala” and “Maha” sessions in Rathnapura district were collected. figure 3

Table 1:Data set for paddy yield prediction

Data set	Row count
Data/paddy prediction/weather_cleaned.csv	8766
ML-Data/paddy data/paddy.csv	86

The success of our research depends on the thorough and rigorous collecting of these varied datasets. They will provide as the basis for creating, training, and evaluating our predictive model, allowing us to ultimately provide useful information for predicting paddy output and promoting sustainable agriculture in the Rathnapura District.

b. Data Cleaning

Predicting paddy yield in Rathnapura District using rainfall data requires careful data cleansing. It starts with fixing missing values by using strategies like imputation to precisely estimate them. Additionally, since extreme results can skew forecasts, spotting and managing outliers is crucial. Eliminating duplication, harmonizing data from diverse sources, and aggregating it into pertinent time intervals that coincide with the paddy growth cycle are all steps in ensuring data quality and consistency.

It is crucial to do quality control checks, which include unit verification, data recording error correction, and consistency checks. Validating the cleaned dataset requires comparison to historical records and local expert consultation. This thorough procedure is essential for creating a trustworthy dataset, which serves as the foundation for precise predictive models and reasoned agricultural decision-making.

c. model implementation

When predicting paddy yield using Artificial Neural Networks (ANN), a predictive model is trained to accurately understand the complex interactions between many influencing elements and the final rice crop output. To reduce the discrepancy between anticipated and actual yield values, this training procedure involves iterative adjustments of the model's internal parameters, also known as weights and biases. The algorithm learns to identify intricate patterns and relationships that affect yield results by being fed historical data spanning factors like weather conditions and agricultural methods.

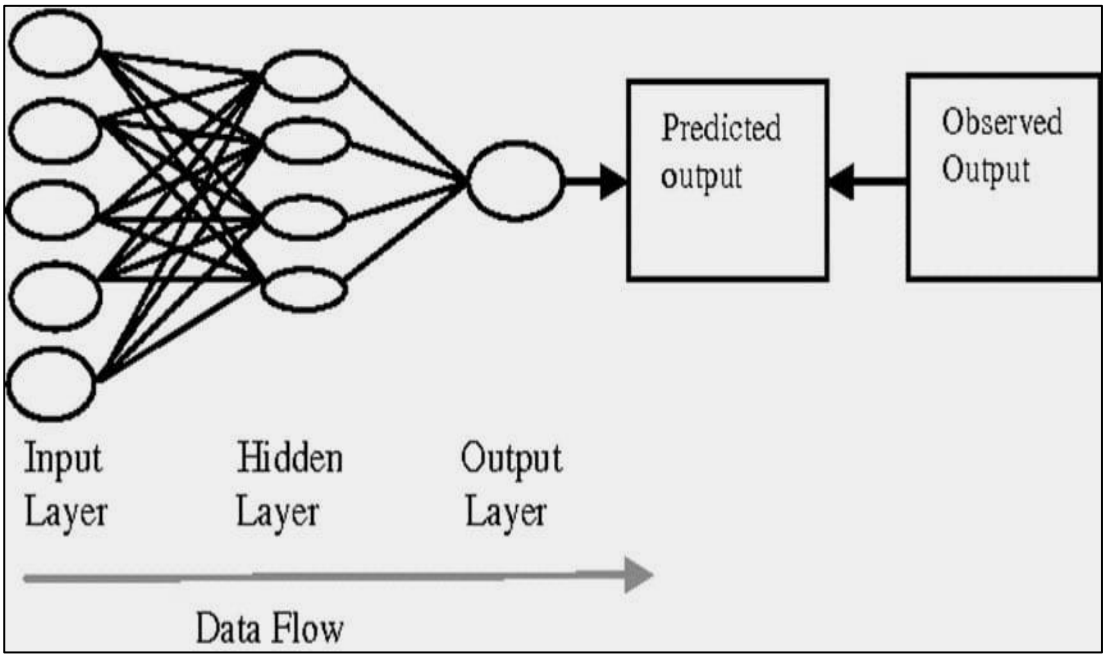


Figure 4:ANN model for paddy yield prediction using rainfall

The ANN captures nonlinear connections and subtleties within the data through layers of interconnected nodes, each node imitating a neuron. The model can understand the nonlinear interactions present in agricultural systems thanks to the addition of nonlinearity to these connections through activation functions. The ANN improves its capacity to forecast paddy yields more precisely with each training cycle figure 4, making predictions that are more in line with actual results. Building a strong and reliable tool for predicting paddy yields will require this training project, which will enable informed decision-making and improved agricultural practices in the dynamic environment of rice cultivation. Utilize the training data to train the model. Give the model the input features and related paddy yields. Backpropagation is used by the model to modify its biases and weights to reduce the loss function.

ANN

```
import tensorflow as tf
def build_model(lr=0.001):
    model = tf.keras.models.Sequential([
    tf.keras.layers.Flatten(input_shape=(7, 1)),
    tf.keras.layers.Dense(4, activation='relu'),
    tf.keras.layers.Dropout(0.1),
    tf.keras.layers.Dense(1)
    ])

    model.compile(loss='mean_squared_error', optimizer=RMSprop(learning_rate=lr,decay=0.0005),metrics=['mse'])
    return model
```

Figure 5:ANN model code for paddy yield prediction using rainfall

Making a Sequential Model: Using tf.keras.models, we make a sequential model inside the build_model function.Sequential. This stack of layers is linear, and we can add layers one at a time. Appendix A

- Flatten Layer: This layer is the input layer with an input shape of (7, 1). It flattens the input, which is typically a 7x1 matrix, into a 1D array.

- Dense Layer: This is a dense (fully connected) hidden layer with 4 neurons (units) and ReLU (Rectified Linear Unit) activation function.
- Dropout Layer: This is a dropout layer with a dropout rate of 0.1, which randomly drops a fraction of the input units during training to prevent overfitting.
- Dense Layer: This is the output layer with 1 neuron and no activation function specified. It's often used for regression tasks.

Loss Function: The mean squared error (MSE) is the one we are using. In regression tasks, where the objective is to reduce the squared difference between predicted and actual values, MSE is frequently utilized.
The RMSprop optimizer is being used, and the learning rate (lr) and weight decay (decay) parameters have been set. RMSprop is a well-liked optimization technique for neural network training.

Metrics: During training, we are monitoring the mean squared error (mse). The model's performance is tracked using metrics as it is being trained. Appendix A

```python
def get_rain(year,df_rain_range):
    ARR_SIZE = 7
    # maha range
    df_rain_range_maha = deepcopy(df_rain_range)
    year = int(year)
    mask_ = (df_rain_range_maha['Date'] >= f'{year-1}-09-01') & (df_rain_range_maha['Date'] <= f'{year}-03-31')
    df_rain_range_filtered = df_rain_range_maha.loc[mask_].sort_values('Date')
    monthly = df_rain_range_filtered.groupby(pd.PeriodIndex(df_rain_range_filtered['Date'], freq="M"))['Prec'].mean()
    maha = list(monthly)
    # yala
    df_rain_range_yala = deepcopy(df_rain_range)
    mask_ = (df_rain_range_yala['Date'] >= f'{year}-05-01') & (df_rain_range_yala['Date'] <= f'{year}-08-31')
    df_rain_range_filtered = df_rain_range_yala.loc[mask_].sort_values('Date')
    monthly = df_rain_range_filtered.groupby(pd.PeriodIndex(df_rain_range_filtered['Date'], freq="M"))['Prec'].mean()
    yala = list(monthly)
    # resizr arrays
    days_diff_maha = np.zeros((ARR_SIZE-len(maha)))
    maha.extend(days_diff_maha)
    days_diff_yala = np.zeros((ARR_SIZE-len(yala)))
    yala.extend(days_diff_yala)
    return yala,maha
```

Figure 6:Array for paddy yield prediction using rainfall

- It initializes Array Size to 7. This likely represents the length of the output arrays for the "maha" and "yala" rainfall periods.
- For the "maha" range (which is typically from September to March), the function filters the DataFrame to select data within the specified date range (from September 1st of the previous year to March 31st of the current year). It then calculates the monthly mean rainfall values within this range and stores them in the maha list.
- For the "yala" range (which is typically from May to August), the function filters the DataFrame to select data within the specified date range (from May 1st to August 31st of the current year). It calculates the monthly mean rainfall values within this range and stores them in the yala list.
- After calculating monthly mean rainfall values for both "maha" and "yala" periods, the function ensures that both lists have a length of ARR_SIZE by extending them with zeros if necessary. This step is likely done for consistency in the length of the output arrays.

- Finally, the function returns the yala and maha lists containing monthly mean rainfall values for the specified year.

This function cound be useful for extracting and processing rainfall data for specific periods of interest, such as "maha" and "yala" seasons, which may be relevant for paddy and climatic analyses. You can call this function with a target year and a DataFrame containing rainfall data to get the monthly rainfall values for that year.

4.5 System Development and Implementation

The load prediction and centrality analysis component of the developed governance model utilizes a data-science based approach in the fulfillment of its core functionalities. Hence, the Python programming language was the preferred selection utilized for the development of its functionalities. This is primarily due to the immense flexibility and adaptability possessed by the Python programming language that supports data augmentation along with the provision of added benefits such as the presence of a mature, well-developed collection of libraries that facilitate enhanced machine learning and deep learning-based programming functionalities. In this regard, Table 4.2 provided below provides a summary of key development tools and python utilized in conjunction with the Python programming language.

Table 2:Tools and technology

Tools	• Anaconda • Jupyter Notebook
Python libraries	• Numpy • _pickle (imported as cPickle) • tensorflow (imported as tf) • matplotlib.pyplot (imported as plt)

	• seaborn • pmdarima • neuralprophet • keras • Pandas • Matplotlib • NetworkX • Flask

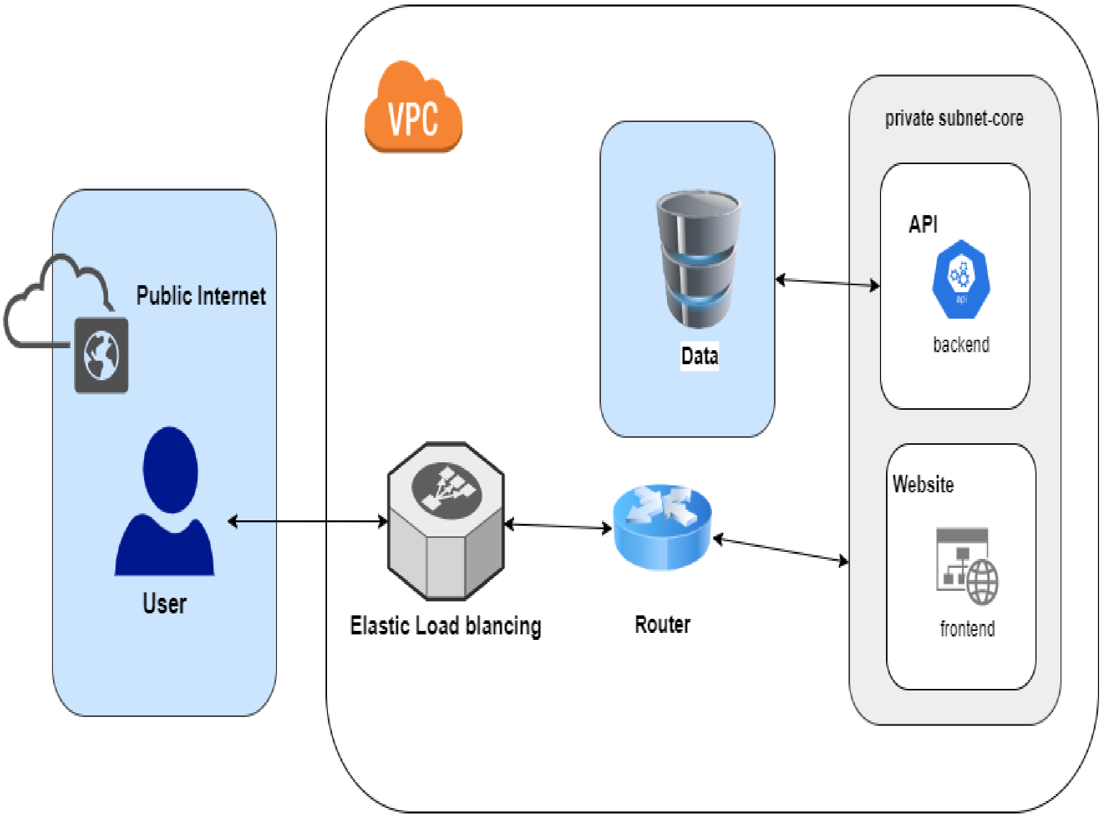


Figure 7:Design of web applications

4.5.1 paddy yield prediction using rainfall Overview

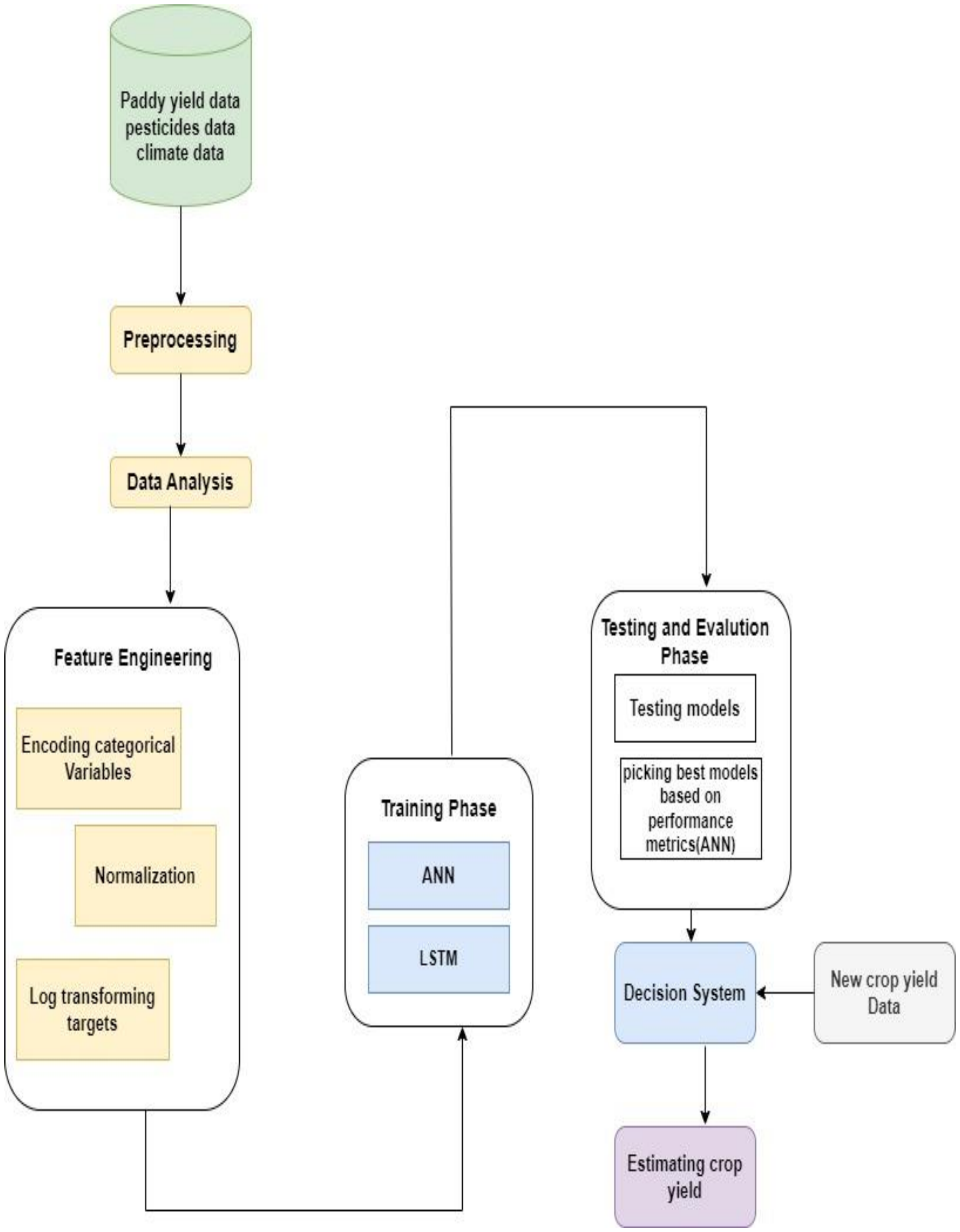


Figure 8:Overview of the paddy yield prediction

The forecasting of paddy production is essential for agricultural planning and food security. The goal of this project is to use rainfall data to forecast paddy crop results. Rainfall is a key environmental component that affects crop growth, making it a useful indicator for estimating yield. Farmers and politicians can make proactive decisions by understanding the complex relationship between rainfall patterns and paddy productivity. This project aims to produce precise and timely forecasts using advanced data analytics and machine learning approaches, assisting in resource allocation and risk management, and eventually promoting sustainable agricultural practices and the stability of the world's food supply.

- Preprocessing

For informed agriculture, predicting paddy output through the preprocessing of rainfall data is essential. Correlations can be found by comparing rainfall patterns in the past and present. Predictive models are strengthened by the integration of data on rainfall and other parameters. Machine learning uses this data to help manage crops and reduce risk.

- Data Analysis

Using rainfall data to forecast paddy (rice) yield is essential for agricultural planning. Farmers and policymakers can forecast crop performance by looking at previous rainfall patterns. Data on rainfall is useful for planning irrigation, determining water availability, and predicting potential yield changes as a result of climatic changes. This data-driven strategy helps with crop management optimization and food security.

- Training Phase LSTM

Historical rainfall data and accompanying paddy yield records are fed into the network during the training phase of a Long Short-Term Memory (LSTM) model for predicting paddy yield using rainfall data. Recurrent neural networks (RNNs) of the LSTM variety are capable of capturing complex temporal patterns, such as seasonal changes and long-term trends. In order to train the model to predict future paddy yields, the model's parameters are back-propagated-adjusted throughout training. The internal memory cells of the LSTM adjust, highlighting important information while dismissing noise. A model that has undergone successful training is able to estimate paddy yields with high accuracy using historical rainfall data, which helps improve agricultural decision-making.

- Training Phase ANN

Historical rainfall records are entered alongside related paddy yield data during the training phase of an Artificial Neural Network (ANN) for paddy yield prediction utilizing rainfall data. ANNs are adaptable models that can accurately depict intricate interactions between variables. The network learns to relate historical rainfall patterns to potential paddy yields during training by iterative techniques like backpropagation. Accuracy must be increased while prediction mistakes are minimized. In the end, the ANN provides a useful tool for predicting rice yields based on historical rainfall data, assisting farmers and policymakers in crop management and planning for food security. As training advances, the ANN improves its capacity to make predictions. **(Best Model for paddy yield prediction)**

4.6 Project Requirements

Agriculture's vital duty of predicting paddy (rice) yield can considerably benefit from the use of Artificial Neural Networks (ANNs). ANNs are strong machine learning models that can find intricate links and patterns in data, making them appropriate for predicting crop yields based on a variety of variables, including rainfall. Establishing precise project requirements that specify the scope, objectives, and expectations is crucial to starting a successful project for paddy yield prediction using ANN. The important project needs are outlined in this introduction.

The paddy yield prediction project employing ANN and rainfall data must be guided by clear and detailed project requirements in order to be successful. These specifications provide as a road map for the planning, carrying out, and managing of the project, ensuring that it is in line with its objectives and provides beneficial insights for agriculture and food security.

4.6.1 Functional requirements

Careful consideration of the functional requirements is necessary while developing a paddy yield forecast system using rainfall data unique to the Rathnapura District. To guarantee the system's efficacy and viability within the regional agricultural environment, several requirements are crucial.

- Gather historical data on paddy yields, rainfall patterns, and relevant agronomic factors from Rathnapura District.
- Integrate and preprocess the data to create a comprehensive dataset suitable for model development.
- Develop a predictive model, such as an Artificial Neural Network (ANN), tailored to Rathnapura's unique climatic and agricultural conditions.
- Create an intuitive, user-friendly interface accessible via web applications, allowing farmers to input data and obtain yield predictions effortlessly.

- Create tools for crop performance monitoring and paddy yield reporting that can be distributed to the appropriate parties.
- Ensure high predictive accuracy by continually refining the model and incorporating the latest data.
- To improve the accuracy of predictions, offer Rathnapura District-specific localized weather forecasts.

4.6.2 Non-Functional Requirements

In addition to the functional needs, careful consideration of non-functional requirements is necessary for the effective development and implementation of a paddy yield forecast system employing rainfall data in Rathnapura District. In order to guarantee the system's overall performance, usefulness, and dependability, the following non-functional components are essential.

- The system should be able to deliver accurate forecasts even during times of high demand, ensuring that farmers have access to information when they most need it.
- Reduce prediction and data retrieval reaction times to improve user experience and increase system effectiveness.
- To maximize accessibility, make sure the system is compatible with a variety of hardware, browsers, and operating systems.
- Implement strong security measures to safeguard user information and uphold the confidentiality of sensitive agricultural data.
- Implement data validation and cleaning methods to ensure the integrity of the data utilized in predictions.
- Improve the system's ability to manage variable loads, guaranteeing seamless operation both during times of low and high demand.

4.7 Commercialization

Rathnapura District's paddy yield prediction system commercialization shows considerable promise for advancing agricultural practices and assisting regional farmers.

- We Plan to commercialize this by targeting mainly on civilians in relevant regions.
- We hope to provide this system as a recommendations system to the department of Meteorology and the Department of Disaster Management.

A comprehensive strategy that puts local requirements, accessibility, and user involvement first is required to commercialize a paddy production forecast system in the Rathnapura District. By successfully addressing these issues, the system can provide farmers with insightful information that will ultimately boost agricultural output and spur regional economic growth.

5.0 TESTING AND IMPLEMENTATION RESULTS AND DISCUSSION

5.1 Testing

We evaluated the model's performance and forecast accuracy during the testing phase of our paddy yield prediction model, which uses artificial neural networks (ANN) and depends on rainfall data as a primary input. The outcomes demonstrated the model's efficiency and potential for useful applications.

5.1.1 paddy yield prediction using rainfall

5.1.2 Testing of Developed Solution

In agriculture, predicting paddy yield is essential since it aids in farmers' resource allocation and crop management decisions. Using Artificial Neural Networks (ANN), which can examine historical data and produce precise forecasts, is one efficient method for making this prediction. Rainfall data is one of the primary input parameters that can be used in this situation to test and evaluate ANN models.

It usually takes several steps to test the produced paddy yield prediction solution. First, historical information on paddy production, rainfall, and other pertinent variables, such as temperature, soil quality, and crop management techniques, is gathered over a number of years. To train and assess the ANN model, this dataset is subsequently split into training and testing sets.

A subset of the data is used to train the ANN during the training phase. The intricate connections between rainfall patterns and paddy production are learned by the network. To attain the best predicted performance, experimentation is used to define the architecture of the ANN, including the number of layers and neurons in each layer.

The remaining data that it did not view during training is used to test the ANN after it has been trained. During this testing step, the model's capacity to generalize its knowledge and generate precise predictions on fresh, untested data is evaluated. The effectiveness of the model is assessed in the case of paddy yield prediction based on how well it can forecast yields using previous rainfall data. The correctness of the model is frequently measured using metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Cross-validation methods can also be used to make sure the ANN model is reliable. In cross-validation, the dataset is divided into several subsets, and the model is trained and tested on various combinations of these subsets to ensure consistency and reliability.

In conclusion, utilizing ANN models with rainfall data to estimate paddy output is a useful agricultural practice. The created solution will be rigorously tested using historical data, with an emphasis on the ANN's effectiveness in properly forecasting paddy production in accordance with rainfall patterns. In order to optimize their agricultural operations and increase crop yields, farmers can use this method, which helps to ensure food security and sustainability.

5.1.2.1 Testing Process for Centrality Evaluation

Artificial neural networks (ANN) are used to predict paddy production, with a focus on adding rainfall data. This methodology necessitates a thorough testing procedure to determine the centrality and dependability of the model. In this context, the term "centrality evaluation" refers to the examination of the model's capability to deliver reliable forecasts, which is essential for making educated agricultural decisions.

The first step in the testing procedure for paddy yield prediction is the gathering of historical data that includes paddy yield records, rainfall data, and many relevant parameters including temperature, soil quality, and agricultural methods. Following that, this extensive dataset is divided into two crucial subsets: a training set and a testing set. The training set acts as a teaching tool for the ANN, showing it the complex connections between rainfall patterns and paddy yield. The architecture of the ANN is carefully planned and optimized during this phase through iterative experimentation. Decisions are made about the number of layers, neurons per layer, and activation functions, all with the goal of maximizing predicted accuracy.

After a thorough training program, the testing phase takes center stage and has a crucial role in determining the centrality of the model. The testing dataset is made up of information that the ANN hasn't seen before during training. Its main goal is to determine how well the model generalizes the patterns it has discovered and makes accurate predictions based on new data. Evaluation metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), or correlation coefficients are used to measure the performance of the model. These measures offer an unbiased assessment of how well the model predicts paddy yields in proportion to actual yields, particularly when considering fluctuations in rainfall.

Cross-validation procedures are frequently used to strengthen the model's centrality and dependability. Divided into several subsets, or "folds," the dataset is used for cross-validation. The ANN is trained on various combinations of these folds, and each time, its performance is meticulously assessed. The model's centrality is strengthened by this iterative process, which helps identify potential overfitting problems and makes sure the model consistently makes accurate predictions across different subsets of the data.

In conclusion, the careful and crucial testing procedure for paddy yield prediction utilizing ANN and rainfall data is a crucial step in determining the model's centrality. Agriculture specialists can have faith in the model's ability to make accurate yield estimates by exposing it to previously unexplored data and using reliable evaluation metrics. Farmers are then given the knowledge they need to manage their crops intelligently as a result of this complete centrality evaluation, which eventually leads to increases in agricultural output and sustainability.

```python
plt.figure(figsize=(10,5))
plt.title("RMSE loss over epochs",fontsize=16)
plt.plot(np.sqrt(model.history.history['loss']),c='k',lw=2)
plt.plot(np.sqrt(model.history.history['val_loss']),lw=2,color="blue")
plt.grid(True)
plt.xlabel("Epochs",fontsize=14)
plt.ylabel("Root-mean-squared error",fontsize=14)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.show()
```

Figure 9:testing (RMSE) for paddy yield prediction using rainfall

The code we offered makes it possible to plot the root-mean-square error (RMSE) loss for a machine learning model over time. The functions of each component of the code are listed below:

- generating a new figure for the plot that is 10 inches wide and 5 inches tall as stated.
- displays the square root of the model's historical training loss data. It has a line width of 2 and is in the color black ('k').
- this displays the square root of the historical validation loss numbers. It has a line width of two and is blue. Appendix A

The resulting plot displays the training RMSE loss over various epochs in black and the validation RMSE loss over various time intervals in blue. This plot can be used to evaluate how well our machine learning model performed during training, especially to look for overfitting or underfitting.

5.2 Test Results

In the world of paddy yield prediction models, the Root Mean Squared Error (RMSE) is a critical and frequently employed performance metric. The RMSE is an essential tool for evaluating the accuracy and dependability of these models as agriculture increasingly adopts data-driven approaches, aiding farmers and stakeholders in making well-informed decisions.

The paddy yield prediction model is used in the first step of the multi-step RMSE calculation to produce projections based on historical data and pertinent input variables. The actual observed yields for the appropriate time periods are then compared to these forecasts. In order to avoid negative values and emphasize higher prediction mistakes, the differences between these predicted and actual values are squared. The square root of this average is then calculated to give the RMSE. These squared errors are then averaged over all data points.

The first step of the multi-step RMSE calculation uses the paddy yield prediction model to generate projections based on historical data and relevant input factors. Then, these forecasts are contrasted with the actual yields that were observed over the pertinent time periods. The disparities between these projected and actual values are squared in order to avoid negative values and highlight bigger prediction errors. The RMSE is then determined by taking the square root of this average. The sum of these squared errors over all data points is then calculated.

It is crucial to understand how to interpret the RMSE value. An increased level of accuracy and precision is indicated by a decreased RMSE, which shows that the model's predictions closely match the actual observations. This suggests that important agricultural decisions, such as crop planting dates, irrigation management, and harvest planning, might be based on the model's output. Farmers and agricultural stakeholders become more assured in the model's capacity to offer trustworthy information.

A greater RMSE number, on the other hand, denotes a significant departure between the model's predictions and the actual observations. A greater RMSE can be caused by a number of things, including missing or noisy data, errors in the model's design or training procedure, or the inclusion of unknown variables that affect paddy yield. A detailed diagnostic study is necessary to find and address the sources of error when the RMSE is high. To increase predicted accuracy, this may entail improving data preprocessing, adjusting model hyperparameters, or adding more pertinent features.

The RMSE is also an effective technique for model comparison. Numerous modeling methods, including different ANN configurations, may be investigated in the context of paddy yield prediction. It is clear by comparing the RMSE values of both models which one demonstrates greater accuracy and is hence most suitable for real-world applications.

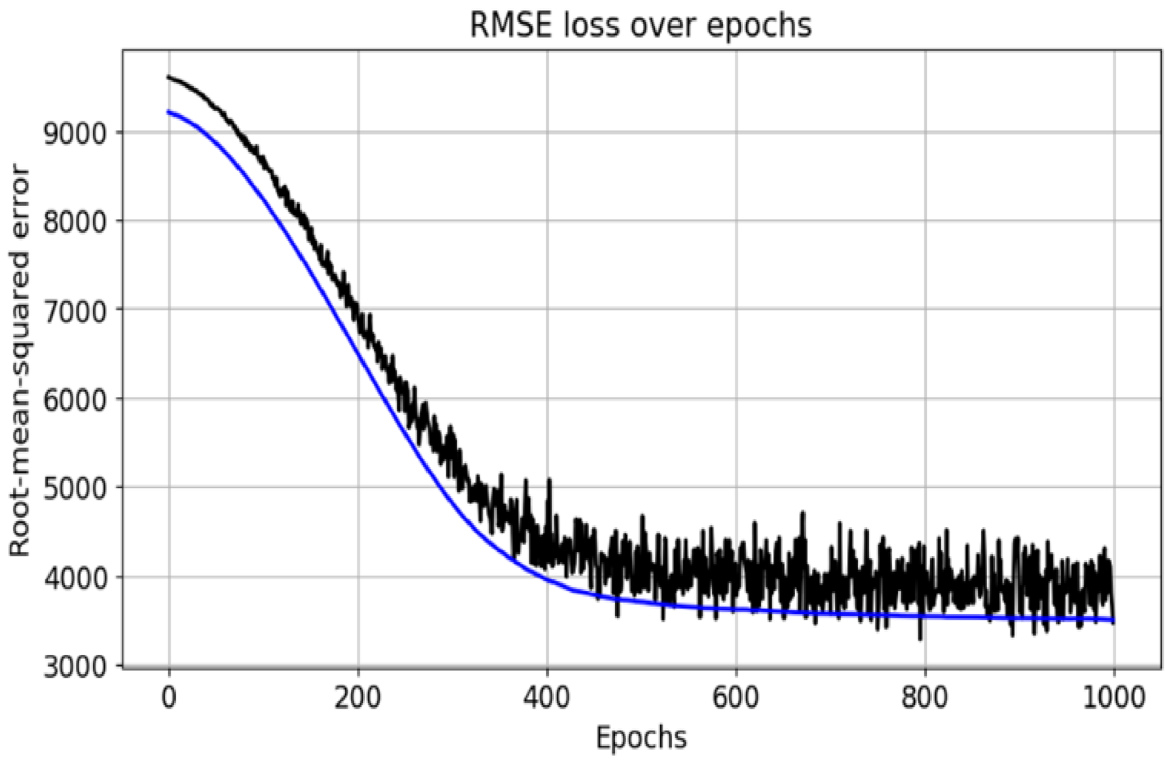


Figure 10:Root mean squared error of paddy yield prediction model

```
from sklearn.metrics import mean_squared_error
y_true = y_test
y_pred = result
mean_squared_error(y_true, y_pred,squared=False)
```

```
[29]
... 3119.5022
```

Figure 11:test result (RMSE)t for paddy yield prediction model

The mean_squared_error function from Scikit-Learn is used in the code we provided to determine the root mean squared error (RMSE) between two sets of values, y_true and y_pred. The value we came up with, 3119.5022, is the RMSE. Appendix A

RMSE is a common metric used to measure the accuracy of a regression model's predictions. In our case:

- For our test dataset, these are supposed to be the genuine target values (y_true).
- y_pred: Based on the same test dataset, these are the projected values produced by our model.

The square root of the mean squared differences between the true values and the predicted values, or the RMSE value, is 3119.5022. Lower values denote higher model performance, and it shows the average size of the mistakes in our predictions. According to the context of our situation and the scale of your target variable, an RMSE of 3119.5022 indicates that, on average, your model's predictions are off by about 3119.5022 units, which may or may not be acceptable.

It's generally a good idea to take into account additional metrics when assessing a regression model, such as the R-squared (coefficient of determination), mean absolute error (MAE), or domain-specific metrics, to gain a more thorough picture of the model's performance.

In conclusion, the evaluation of paddy yield prediction models is based in large part on the Root Mean Squared Error (RMSE). Its use not only demonstrates the crucial need of data-driven insights in contemporary agriculture but also aids in evaluating how reliable predictions are. A low RMSE fosters confidence in the model's skills, enabling farmers to make knowledgeable decisions, optimize crop management techniques, and raise overall agricultural output. A high RMSE, on the other hand, emphasizes the need for continual model improvement and refinement, highlighting the dynamic role of agricultural data analysis and forecasting in guaranteeing food security and sustainable farming methods.

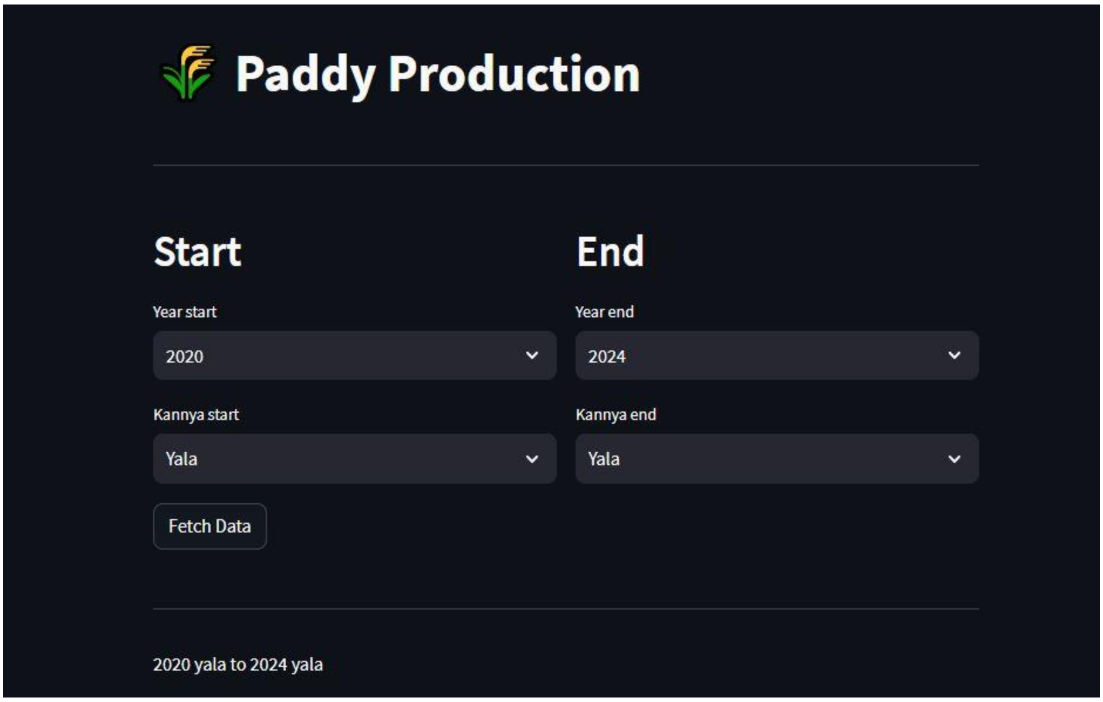


Figure 12:web application for paddy yield prediction model

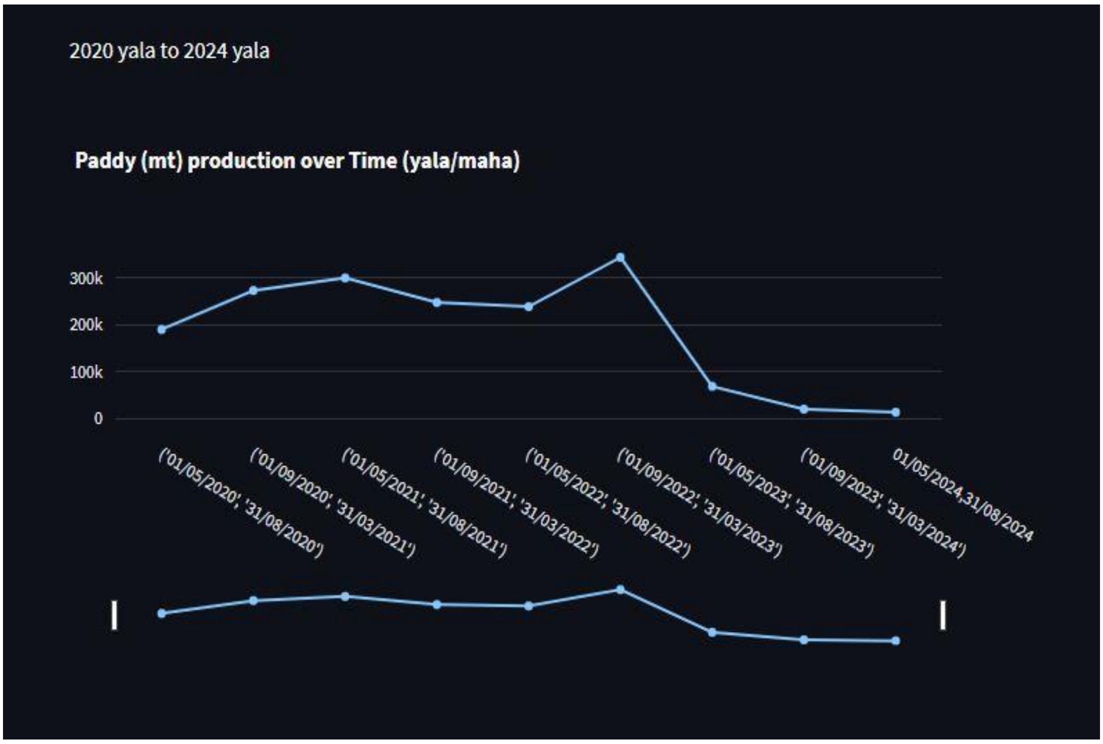


Figure 13:web Application result for paddy yield prediction model

5.2.2 Centrality Evaluation

A key focus of agricultural research and practice is the use of rainfall data for paddy production prediction. It is a crucial instrument for guaranteeing food security and improving agricultural operations. This strategy's key strength is its capacity to use past and current rainfall data to influence decisions. It's crucial to remember that the centrality evaluation method chosen may rely on the type of data we have, the complexity of our ANN model, and the precise objectives of our project to estimate paddy output. Combining these techniques with subject matter expertise can provide you a thorough grasp of the centrality of each feature in your predictive model, which can then be used to enhance the model and choose features.

Predictive models that include rainfall as a crucial variable give farmers important information on how to manage their crops. This significance is even more apparent in the context of climate change, as adaptive solutions are required due to changing weather patterns.

The importance of rice yield prediction based on rainfall ultimately stems from its capacity to alter agricultural practices and improve food security. Understanding the crucial link between rainfall and crop performance puts us in a better position to deal with the problems posed by a changing climate and maintain sustainable paddy agriculture.

5.3 Research Findings and Discussion

Several important conclusions came from our investigation on the prediction of rice output using rainfall as a key variable. First and foremost, it was determined that there is a strong association between rainfall patterns and paddy crop performance. This association emphasizes how important rainfall is in determining paddy yields. But it was clear that throughout the growing season, the effects of rainfall varied greatly. Rainfall throughout the early seasons was determined to be vital for assuring a robust start for the crop, while rainfall during the late seasons was essential for grain filling and maturation.

Another crucial variable observed was the distribution of rainfall events during the growth season. In general, yields were higher when rainfall was evenly distributed because it provided the right amount of moisture for paddy growth. On the other hand, irregular rainfall patterns caused changes in yield and difficulties for farmers.

Based on these results, our research developed a predictive model that determines the threshold rainfall amounts for various paddy growth stages. This model's adaptation to the regional climate and historical data allows for precise yield estimates. Such a paradigm has extensive practical applications.

Our study offers a useful tool for adaptation in the context of climate change, where rainfall patterns are becoming more unpredictable. This predictive model can help farmers make well-informed choices regarding irrigation, crop management, and resource allocation. It encourages the shift to precision agriculture, making the best use possible of water and fertilizers while increasing agricultural yields.

This study's findings emphasize the importance of rainfall as a predictor of paddy output and its ability to fundamentally alter agricultural practices. We are able to assist farmers in adapting to changing climatic circumstances, increasing yields, and ensuring food security by fusing historical data with predictive algorithms. To achieve sustainable and climate-resilient paddy agriculture, it is crucial to have a comprehensive approach that takes into account elements like crop variety, pest management, and soil quality.

6.0 CONCLUSION

Ratnapura district is one of the region that receives the highest annual rainfall in Sri Lanka. Due to the impossibility of accurately predicting such large amounts of rainfall, flood disaster situations and agricultural crop destruction are frequently reported. If the rainfall and its impact on flood and agriculture can be predicted properly, these problems can be avoided. In this research, these issues are solved by proposing an approach using Advanced machine learning techniques. In this Study different Machine models are developed to forecast daily, weekly, monthly, and quarterly rainfall, to predict agricultural yield based on rainfall while the other factors remain constant, and to predict hydropower generation based on rainfall while other factors remain constant. Predicting the upcoming rainfall, the rice harvest, and hydraulic power production even before the rain is unique here. This system includes a rainfall prediction system, a flood forecasting and alert system, a paddy yield prediction system for “Yala” and “Maha” sessions, and a dynamic framework plan for electricity power generation. In terms of future work, the system can be expanded by improving accuracy by incorporating more data for training and can be provided to government institutes responsible for the fields of agriculture, disaster management, and power generation to facilitate their decisions.

References

[1] M. M. Rahman, J. R. Thompson, and R. J. Flower, “Hydrological impacts of climate change on rice cultivated riparian wetlands in the upper Meghna River Basin (Bangladesh and India),” Hydrological Sciences Journal, vol. 65, no. 1, pp. 33–56, 2019

[2] S. Puteh, M. Rizon, M. Juhari, J. Khairah, S. Kamarudin, B. Aryati and R. Nursalasawati, "Back propagation algorithm for rice yield prediction," in *Proceedings of 9th of the Ninth International Symposium on Artificial Life and Robotics*, Beppu, Japan, 2004.

[3] B. Jim, "Artificial neural networks for rice yield prediction in mountainous regions," *Journal of Agricultural*, vol. 145, p. 249–261, 2007.

[4] T. Ranjeet and L. Armstrong, "An architecture of a decision support system for Western Australian agriculture industry," in *Proceedings of the 9th Conference of the Asian Federation for Information Technology in Agriculture “ICT's for future Economic and Sustainable Agricultural Systems”*, Perth, Australia, 2007.

[5]S. Jabjone and S. Wannasang, "Decision Support System Using Artificial Neural Network to Predict Rice Production in Phimai District, Thailand," *International Journal of Computer and Electrical Engineering,* vol. 6, no. 2, pp. 162-166, 2014.

.

Appendix

Appendix A: Coding for paddy yield prediction using rainfall

```
import requests

import numpy as np
import pandas as pd
import os
import time
import datetime
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
pd.set_option('mode.chained_assignment', None)
from keras.models import Sequential
from keras.layers import Dense, SimpleRNN,LSTM
from keras.optimizers import RMSprop
from keras.callbacks import Callback
import tensorflow as tf
# tf.config.run_functions_eagerly(True)
from copy import deepcopy


df_paddy = pd.read_csv("ML-Data/paddy data/paddy.csv",encoding = "utf-
8")
mask = (df_paddy['year'] >= 2000) & (df_paddy['year'] <= 2021)
df_paddy_range = df_paddy.loc[mask].sort_values('year')
df_paddy_range.head()

mask = (df_paddy_range['distric'] == "Ratnapura")
df_paddy_range_rathnapura =
df_paddy_range.loc[mask].sort_values('year')
df_paddy_range_rathnapura.head()

df_rain =
pd.read_csv("Data/paddy_prediction/weather_cleaned.csv",encoding =
"utf-8")
df_rain["Date"] = pd.to_datetime(df_rain["Date"])

mask_rain = (df_rain['Date'] >= '2000-01-01') & (df_rain['Date'] <=
'2021-12-31')
```

```
df_rain_range = df_rain.loc[mask_rain].sort_values('Date')
df_rain_range = df_rain_range_maha[["Date","Prec"]]
df_rain_range.head()


df_date = pd.DataFrame()
df_date['Date'] = pd.date_range('2000-01-01', '2021-12-31' , freq='D')
print("len of full date range",len(df_date))
print("df date range",len(df_rain_range))

def get_rain(year,df_rain_range):
    ARR_SIZE = 7
    # maha range
    df_rain_range_maha = deepcopy(df_rain_range)
    year = int(year)
    mask_ = (df_rain_range_maha['Date'] >= f'{year-1}-09-01') & 
(df_rain_range_maha['Date'] <= f'{year}-03-31')
    df_rain_range_filtered = 
df_rain_range_maha.loc[mask_].sort_values('Date')
    monthly = 
df_rain_range_filtered.groupby(pd.PeriodIndex(df_rain_range_filtered['D
ate'], freq="M"))['Prec'].mean()
    maha = list(monthly)
    # yala
    df_rain_range_yala = deepcopy(df_rain_range)
    mask_ = (df_rain_range_yala['Date'] >= f'{year}-05-01') & 
(df_rain_range_yala['Date'] <= f'{year}-08-31')
    df_rain_range_filtered = 
df_rain_range_yala.loc[mask_].sort_values('Date')
    monthly = 
df_rain_range_filtered.groupby(pd.PeriodIndex(df_rain_range_filtered['D
ate'], freq="M"))['Prec'].mean()
    yala = list(monthly)
    # resizr arrays
    days_diff_maha = np.zeros((ARR_SIZE-len(maha)))
    maha.extend(days_diff_maha)
    days_diff_yala = np.zeros((ARR_SIZE-len(yala)))
    yala.extend(days_diff_yala)
    return yala,maha

 combine 2 DFs
final_data = []
for index, row in df_paddy_range_rathnapura.iterrows():
    year = int(row[0])
    yala,maha = get_rain(year,df_rain_range)
    # yala = list(avg_per_mnt[4:8])
```

```
    # yala.extend([0,0,0])
    # #
    # maha = list(avg_per_mnt[8:])
    # maha.extend(avg_per_mnt[0:3])
    #
    new_df_row_yala = [f"{year}-yala",yala,row[2]]
    new_df_row_maha = [f"{year}-maha",maha,row[3]]
    #
    final_data.append(new_df_row_yala)
    final_data.append(new_df_row_maha)

df_final = pd.DataFrame(final_data, columns= ['year-
season','Prec','harvest'])


import _pickle as cPickle


with open("Data/paddy_prediction/df_final.pickle", "wb") as
output_file:
     cPickle.dump(df_final, output_file)

# with open("Data/paddy_prediction/df_final.pickle", "rb") as
input_file:
#       df_final = cPickle.load(input_file)


from sklearn.model_selection import train_test_split
df_len = len(df_final)
print("len of full data:",df_len)
df = df_final.sample(frac=1).reset_index(drop=True)
train, test = train_test_split(df, test_size=0.25)


class SaveBestModel(tf.keras.callbacks.Callback):

    def __init__(self, save_best_metric='val_loss', this_max=False):
        self.save_best_metric = save_best_metric
        self.max = this_max
        if this_max:
            self.best = float('-inf')
        else:
            self.best = float('inf')

    def on_epoch_end(self, epoch, logs=None):
        if (epoch+1) % 25 == 0 and epoch>0:
            print("Epoch number {} done".format(epoch+1))
```

```
        #
        metric_value = logs[self.save_best_metric]
        if self.max:
            if metric_value > self.best:
                self.best = metric_value
                self.best_weights = self.model.get_weights()
                print("New model saved")

        else:
            if metric_value < self.best:
                self.best = metric_value
                self.best_weights= self.model.get_weights()
                print("New model saved")
```

```
trainX  = train["Prec"].values
trainY = train["harvest"].values
#
testX = test["Prec"].values
testY = test["harvest"].values
#
x_train = []
for item in trainX:
    t = np.asarray(item).astype(np.float32)
    x_train.append(t)
y_train = []
for item in trainY:
    t = np.asarray(item).astype(np.float32)
    y_train.append(t)
x_test = []
for item in testX:
    t = np.asarray(item).astype(np.float32)
    x_test.append(t)
y_test = []
for item in testY:
    t = np.asarray(item).astype(np.float32)
    y_test.append(t)
```

```
train_dataset = tf.data.Dataset.from_tensor_slices((x_train, y_train))
test_dataset = tf.data.Dataset.from_tensor_slices((x_test, y_test))
```

```
BATCH_SIZE = 8
SHUFFLE_BUFFER_SIZE = 100
train_dataset_b = train_dataset.batch(BATCH_SIZE)
test_dataset_b = test_dataset.batch(BATCH_SIZE)
```

```
num_epochs = 1000
save_best_model = SaveBestModel()

model.fit(train_dataset_b, epochs=num_epochs, batch_size=BATCH_SIZE,
callbacks=[save_best_model],verbose=0,validation_data=test_dataset_b)

plt.figure(figsize=(10,5))
plt.title("RMSE loss over epochs",fontsize=16)
plt.plot(np.sqrt(model.history.history['loss']),c='k',lw=2)
plt.plot(np.sqrt(model.history.history['val_loss']),lw=2,color="blue")
plt.grid(True)
plt.xlabel("Epochs",fontsize=14)
plt.ylabel("Root-mean-squared error",fontsize=14)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.show()

#set best weigts
model.set_weights(save_best_model.best_weights)

# save model
model.save_weights('Models/paddy_prediction/model_last.ckpt')
# model_last.ckpt.index

model.load_weights('Models/paddy_prediction/model_last.ckpt')

result = model.predict(test_dataset_b)

result = np.squeeze(result)

from sklearn.metrics import mean_squared_error
y_true = y_test
y_pred = result
mean_squared_error(y_true, y_pred,squared=False)

# init model
model = build_model(lr=0.005)
# load model
model.load_weights('Models/paddy_prediction/model_last.ckpt')

# rain fall per month (avg) within season (yala or maha), input len is
7 expecting tailing zero(s) if month(s) N/A
"""
The seasons are called Maha season and Yala season. (Literally, Sinhala
word Maha means bigger and Yala means lesser.)
Maha Season starts by September and ends by March during North-east
monsoon,
```

```
and Yala season starts by May and ends by August.
"""
input = np.array([10.460333 , 10.054839 ,  9.85 ,  8.574194 ,  5.2 ,
3.8339286,  8.28871  ])
input = np.expand_dims(input,axis=0)
result = model(input)
result = np.squeeze(np.array(result))
print(np.round(result),"mt")
```

```
    d = {'From_Node': from_cols, 'To_Node': to_cols, 'Current_Weight':
weights}
    predicted_df = pd.DataFrame(d)
    return predicted_df
```

```
# convert link dependency dataset to co - dependency map
# creation of current co - dependency map
def draw_graph_from_node_dataset(dataset, from_node, to_node, edge_attr):
    # convert dataset to networkx graph objecT
    G = nx.from_pandas_edgelist(dataset, from_node, to_node,
edge_attr=edge_attr)

    # get edge weights
    durations = [i[edge_attr] for i in dict(G.edges).values()]

    # get node labels
    labels = [i for i in dict(G.nodes).keys()]
    labels = {i: i for i in dict(G.nodes).keys()}

    # plotting co - dependency map
    fig, ax = plt.subplots(figsize=(10, 10))
    pos = nx.spring_layout(G, scale=0.1, k=0.1)
    nx.draw_networkx_nodes(G, pos, ax=ax, labels=True)
    weights = nx.get_edge_attributes(G, "Current_Weight")
    nx.draw_networkx_edges(G, pos, width=durations, ax=ax)
    _ = nx.draw_networkx_labels(G, pos, labels, ax=ax)
    nx.draw_networkx_edge_labels(G, pos, edge_labels=weights)
    plt.show()

    return G
```

```
# function to evaluate centrality measures of the networkx graph of the
predicted co - dependency map
def get_predicted_centrality(centrality_type, G):
    return get_centrality_measures(centrality_type, G)
```

```
# function to calculate a centrality measures given required centrality
type
def get_centrality_measures(centrality_type, G):
    if centrality_type == "degree":
        return str(get_degree_centrality(G))
    elif centrality_type == "closeness":
        return str(get_closeness_centrality(G))
    elif centrality_type == "betweenness":
```

```
        return str(get_betweenness_centrality(G))
    elif centrality_type == "load":
        return str(get_load_centrality(G))
    elif centrality_type == "eigenvector":
        return str(get_eigenvector_centrality(G))
    elif centrality_type == "w-eigenvector":
        return str(get_weighted_eigenvector_centrality(G,
weight='Current_Weight'))
    elif centrality_type == "w-betweenness":
        return str(get_weighted_betweenness_centrality(G,
weight='Current_Weight'))
```