Logo

Description automatically generated

**BIG DATA ANALYTICS AND APPLICATIONS(CS5542)**

**PROJECT REPORT**

**MONTHLY RAINFALL PREDICTION USING WAVELET NEURAL NETWORK ANALYSIS**

**TEAM MEMBERS:**

**JAGADEESH REDDY PANDHIKUNTA – 16326041**

**VENKATA SAIDEEP JAKKA- 16329410**

**AJAY PAVAN KUMAR REDDY GOPU – 16332139**

**YASWANTH GANDHAMUNENI -16327304**

**Table of Contents:**

Abstract

Introduction

Problem Statement

Literature Survey

Existing System

Proposed System

System Requirements

Uml Diagrams

Implementation

Screenshots

Conclusion

**ABSTRACT**

One of the most important elements in a hydrological model is rainfall. To study and predict the rainfall forecast, several models have been developed. Wavelet approaches have been frequently used in water resources research in recent years due to their temporal frequency representation. By merging the wavelet methodology with Artificial Neural Networks, this research attempts to establish an alternate way for rainfall prediction (ANN). Monthly rainfall data from the Darjeeling rain gauge station was used to test the wavelet and ANN models. The performance of the models in terms of calibration and validation is assessed using appropriate statistical approaches. The findings of monthly rainfall series modeling show that wavelet neural network models outperform ANN models in terms of effectiveness.

**INTRODUCTION**

In today's society, global warming is hurting people all over the world, having a tremendous impact on humanity and hastening climate change. As a result of this, the air and oceans are warming, sea levels are rising, and flooding and drought are becoming more common. Rainfall is one of the most devastating repercussions of climate change. Rainfall forecast is a difficult task these days, and most of the major world authorities are taking it into consideration. Rainfall is a climatic component that has an impact on a variety of human activities, including agricultural production, construction, power generation, and tourism. As a result, rainfall is a major concern, necessitating better rainfall forecasting.

In the past wavelet technology is used to determine the rainfall prediction using signal processing analysis. This technology is mostly used in water related applications. Then after updating continuous changes in technology the combination of both neural network and wavelet combination method is used to predict the rainfall more efficiently.

**Problem statement:**

Climate change is a significant factor in human life. As a result, the Prediction should be as accurate as feasible. In this study, we attempt to address the forecast of rainfall, which is also an important component of human life and provides the most important resource of human life, fresh water. Fresh water is an essential resource for human survival, not only for drinking but also for farming, washing, and a variety of other activities.

Because of climate change, making accurate climate predictions has become increasingly difficult.

Climate change is now the most pressing issue on the planet. People are trying to figure out how climate change affects the economy, from production to infrastructure. Predicting rainfall, like anything else, is a difficult undertaking that requires a high level of accuracy. Because traditional methods for predicting rainfall are ineffective, scientists are turning to machine learning and deep learning to uncover patterns for rainfall prediction.

A poor rainfall forecast can have an impact on agriculture, particularly for framers, whose entire crop is dependent on rainfall, and agriculture is always an important element of any economy. As a result, developing a somewhat accurate rainfall prediction is possible. Machine learning techniques are utilized in a variety of ways, but accuracy is always a challenge when making rainfall predictions. Rainfall has a variety of effects on the earth, including drought, flooding, and extreme summer heat, to name a few.

**LITERATURE SURVEY**

An investigation of rainfall-runoff forecasting using artificial neural networks and wavelet combination.

AUTHORS: François Anctil and Doha Guy Tape

ABSTRACT: The performance of multiple-layer artificial neurons and a neuro-wavelet combination system used in forecasting one-day-ahead stream flow at two sites is compared in this study. The period-scale structure of the available rainfall and runoff time series is found out using Morlet power spectra. The time series are wavelet divided into three sub-series: short, intermediate, and long wavelet periods, which describes the rainfall-runoff processes. Then, for each wavelet sub-series, multiple-layer artificial neurons are trained. The short wavelet periods are the main responsible for the most of the final neuro-wavelet hybrid forecasting inaccuracy, according to the observations. Any upcoming advancements in artificial neural network (ANN) rainfall-runoff forecasting models will depends on short-term variations. The neuro-wavelet combination forecasting system and the standard forecasting multiple-layer artificial neuron system have extremely same type of performance. The neuro-wavelet system's performance advantage can be lead to improved use of the evapotranspiration time series. Surface-water hydrology, rainfall-runoff, artificial neural networks, and wavelet decomposition are some of the main terms.

**EXISTING SYSTEM:**

Rainfall is a complicated atmospheric phenomenon that varies in space and time and is difficult to forecast. Rainfall series are frequently explained by a stochastic process due to their apparent random properties.

Wavelet approaches have been frequently used in water resources research in recent years due to their timefrequency representation.

**DISADVANTAGES OF EXISTING SYSTEM:**

* Because of their timefrequency, it has been widely utilized to various water resources studies.

**PROPOSED SYSTEM:**

In this paper, the author uses a wavelet signal dataset to train a Multilayer Perceptron Neural Network that, when compared to existing techniques, can reliably predict rainfall. All existing algorithms used a simple satellite dataset to estimate rainfall, which may or may not incorporate current weather information, resulting in inaccurate predictions. To solve this challenge, the author is employing a wavelet digital signature dataset that contains current situation data such that a neural network can accurately predict rainfall. Because signals contain current information, wavelet-based prediction has gained a lot of traction.

Because the author is combining Wavelet data with a neural network to predict rainfall, the approach is referred to as WNN.

**ADVANTAGES OF PROPOSED SYSTEM:**

neural network can predict accurate rainfall. Wavelet based prediction has gain lots of popularity due to signals contains present information.

**SYSTEM REQUIREMENTS:**

**SOFTWARE REQUIREMENT**

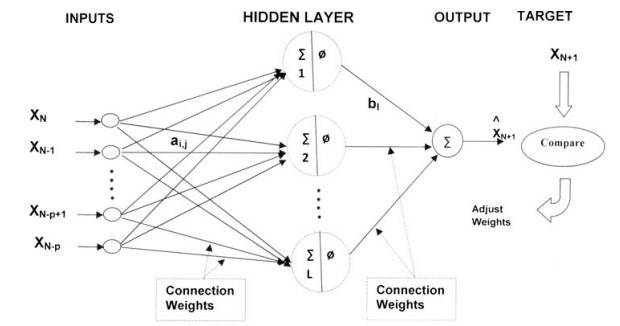
* Python idel 3.7 version (or)
* Anaconda 3.7 ( or)
* Jupiter (or)
* Google colab

**HARDWARE REQUIREMENTS**

* Operating system : windows, linux
* Processor : minimum intel i3
* Ram : minimum 4 gb
* Hard disk : minimum 250gb

**SYSTEM DESIGN**

**SYSTEM ARCHITECTURE:**

****

**DATA FLOW DIAGRAM:**

* A bubble chart is another name for a DFD. It is a basic graphical formalism that can be used to depict a system in terms of the data it receives, the processing it performs on that data, and the data it generates as output.
* If the user wants to check the rainfall chances, checks with data set. After this Preprocessing is done on dataset like finding out the missing values and everything.
* In the next process the preprocessed data is sent into neural network model which gives the prediction of rainfall for 30 days after doing all the operations in the algorithm.

**User**

**Check**

**Unauthorized user**

**Yes NO**

**Upload Wavelet Dataset**

**Preprocess Dataset**

**Build MLP Neural Network Model**

**Rainfall Prediction for 30 Days**

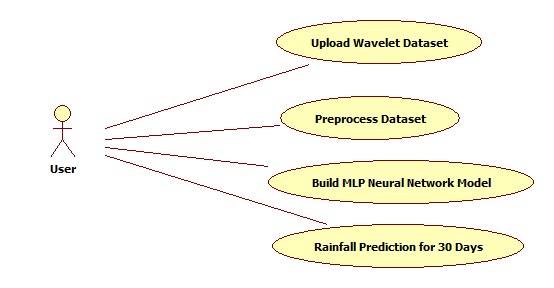
**End process**

**UML DIAGRAMS**

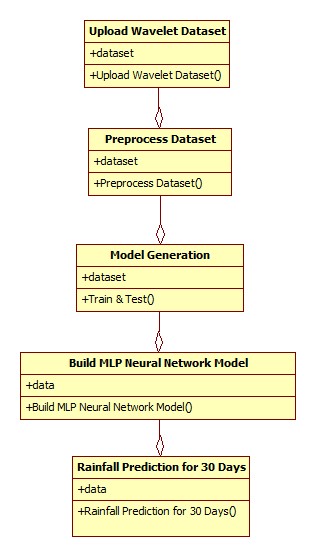
UML is known for "Unified Modeling Language." In the realm of object-oriented software, UML is a standardized general-purpose modeling language. The main purpose of uml diagrams implementation is better understanding of the model and representing it in visual form. This helps in development of model more efficiently and fast. So this uml diagrams helps us to better implement the model.

.

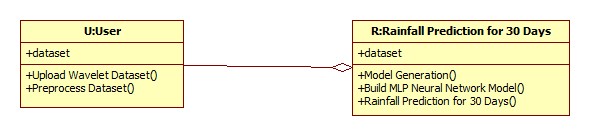
**Use case diagram:**



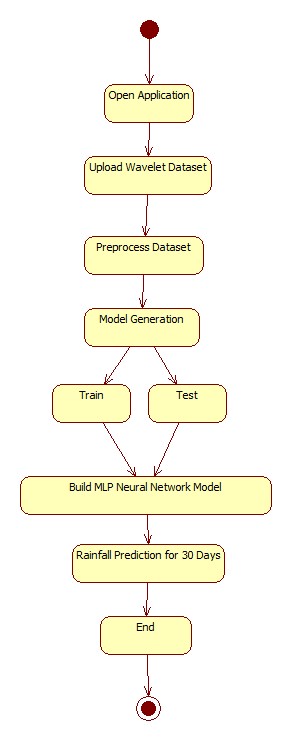
**Class diagram:**



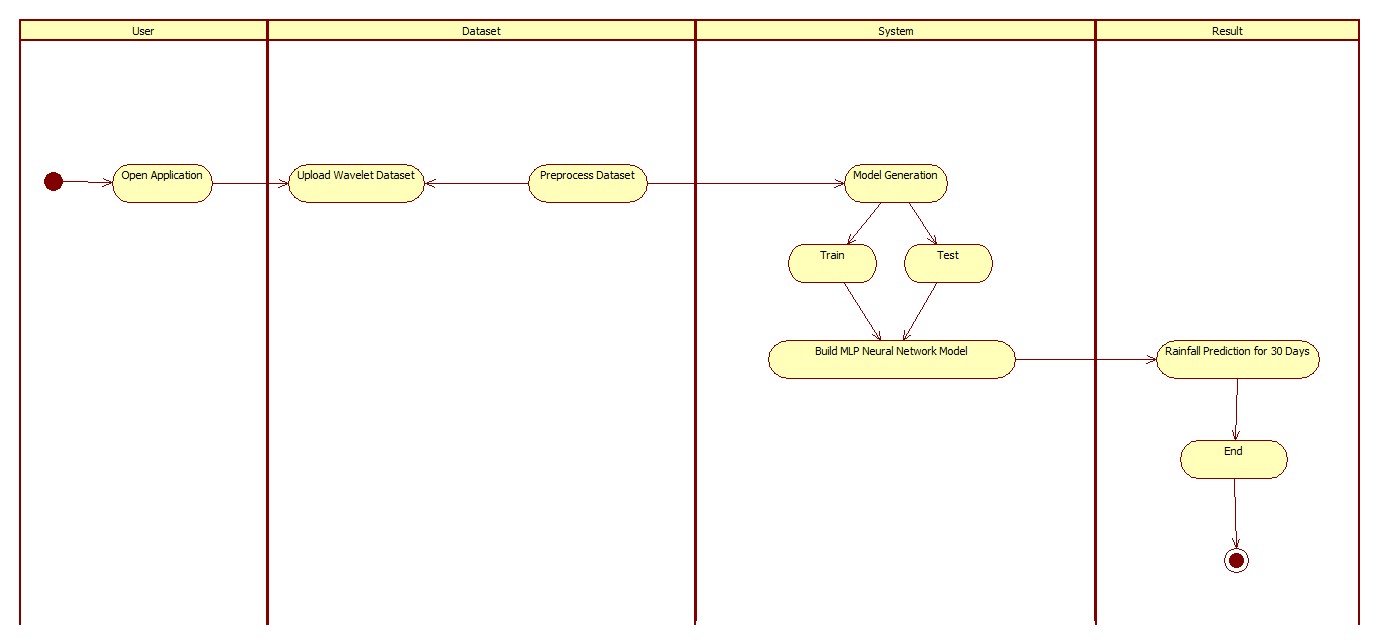
**Object diagram:**



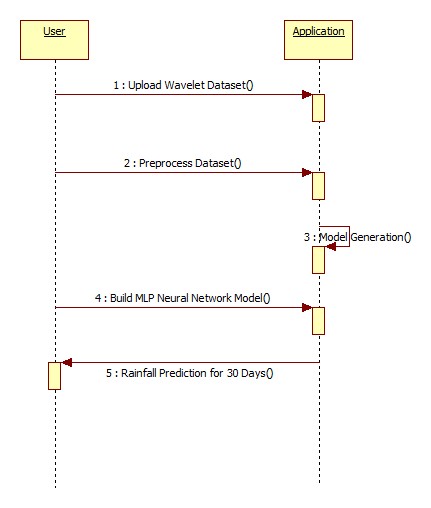
**State diagram:**



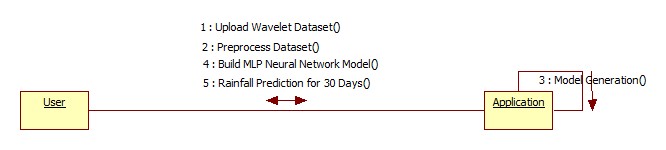
**Activity diagram:**



**Sequence diagram:**



**Collaboration diagram:**

****

**IMPLEMENTATION:**

* Upload Wavelet Dataset
* Preprocess Dataset
* Build MLP Neural Network Model
* Rainfall Prediction for 30 Days

**MODULES:**

**Modules Description:**

**1.Data Collection**

The dataset for this project was collected from the survey conducted in Darjeeling. A graph depicting the primary features as well as the whole dataset. The dataset is then divided into two parts: one for training and the other for testing the algorithms. Furthermore, each class in the whole dataset is represented in about the right proportion in both the training and testing datasets in order to generate a representative sample. The fractions of the training and testing datasets that were used in the study.

**2.Data Preprocessing**

There is a chance that the data obtained contains missing values, which could lead to inconsistencies. To get better results, data must be preprocessed to boost the algorithm's efficiency. Outliers must be deleted, and variable conversion must be performed. We utilize the map function to solve these problems.

**3.Model Selection**

Machine learning is the process of predicting and detecting patterns in order to generate appropriate results after fully comprehending them. Algorithms that use machine learning look for patterns in data and learn from them. With each attempt, an ML model will learn and improve. To determine a model's effectiveness, the data must first be divided into training and test sets. So, before training our models, we divided the data into two sets: the Training set, which comprised 70% of the total dataset, and the Test set, which comprised the remaining 30%. It was therefore necessary to apply a variety of performance indicators to our model's predictions.

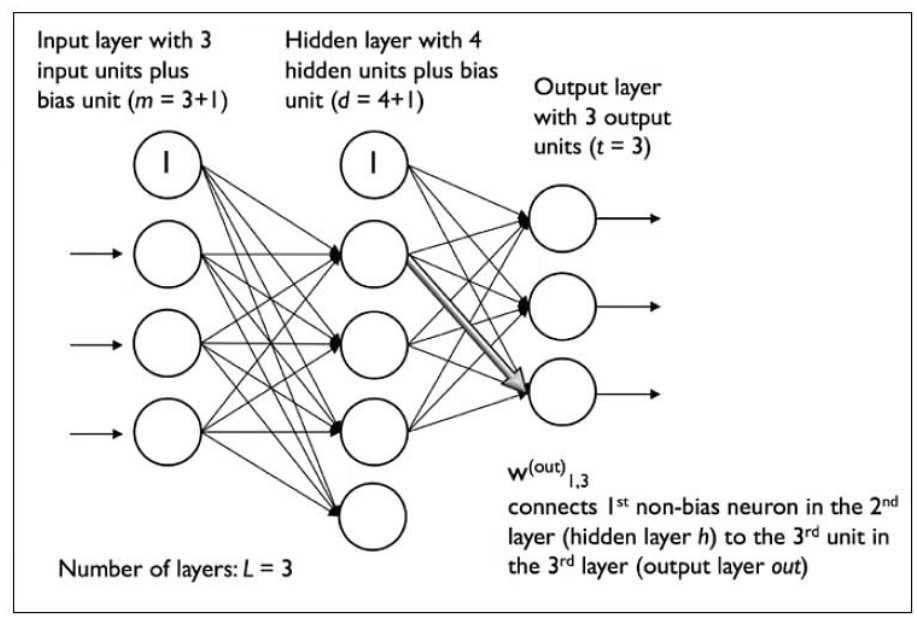
**4.Predict the results**

The designed system is put through its paces with a test set, and its performance is guaranteed. The description and modeling of regularities or trends for objects whose behavior evolves over time is referred to as evolution analysis. Precision and accuracy are two common measures derived from the confusion matrix.

**Algorithm:**

**Multi-layer ANN**

Deep Learning, also known as Deep Neural Networks, is the process of training multi-layer artificial neural networks. The backpropagation algorithm was created by Dr. Hinton and his colleagues to train a multilayer neural network. Today, it's a hot topic, with companies like Google, Facebook, and Microsoft investing extensively in deep neural network applications. A Multilayer Perceptron is a fully connected multi-layer neural network (MLP).



It contains three levels, one of which is deeper. A deep ANN is one that has more than one hidden layer. A feedforward artificial neural network, such as an MLP, is a common example.

The amount of layers and neurons in a neural network are referred to as hyper parameters, and they must be tuned. To determine appropriate values for these, cross-validation techniques must be applied.

Back propagation is used for weight adjustment training. Data processing is improved by using deeper neural networks. Deeper layers, on the other hand, can cause vanishing gradient issues. This problem necessitates the use of special algorithms.

**Libraries Used in Project:**

Tensorflow

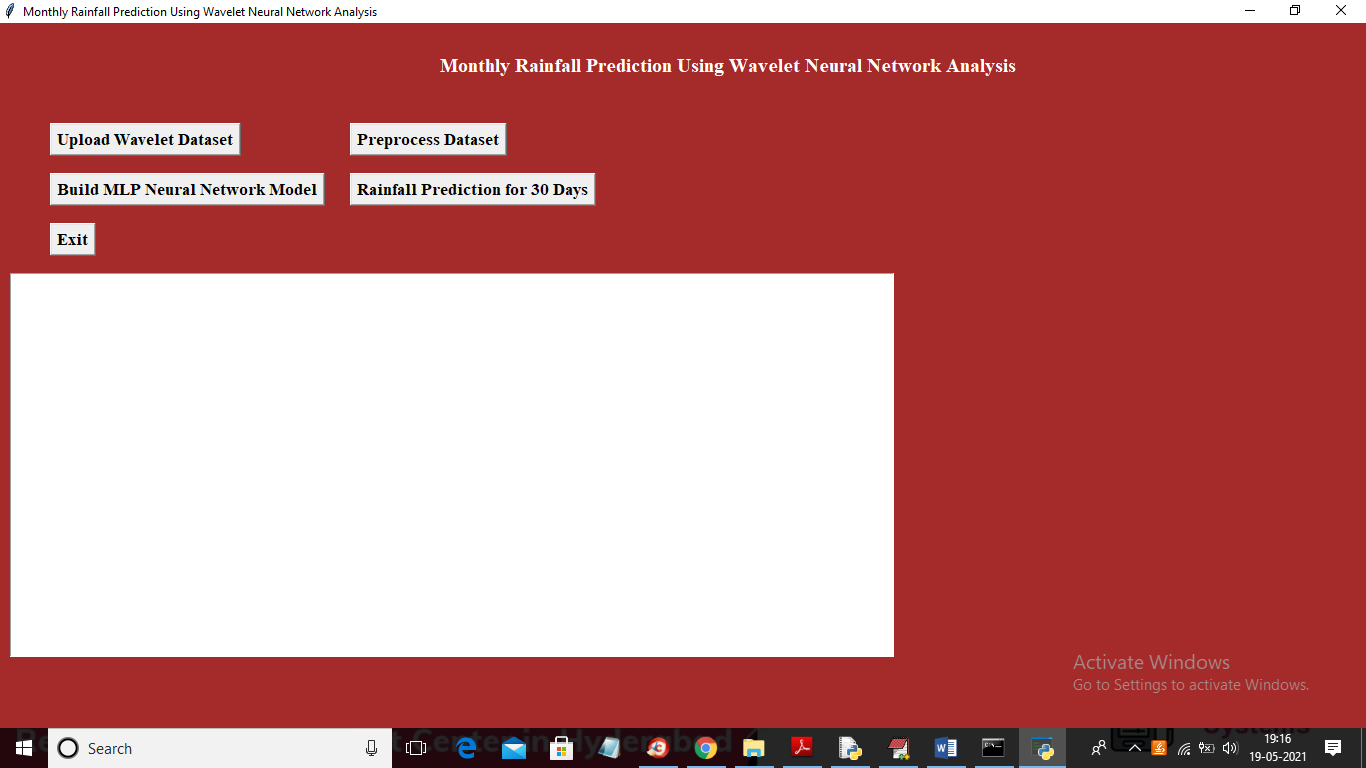
Numpy

Pandas

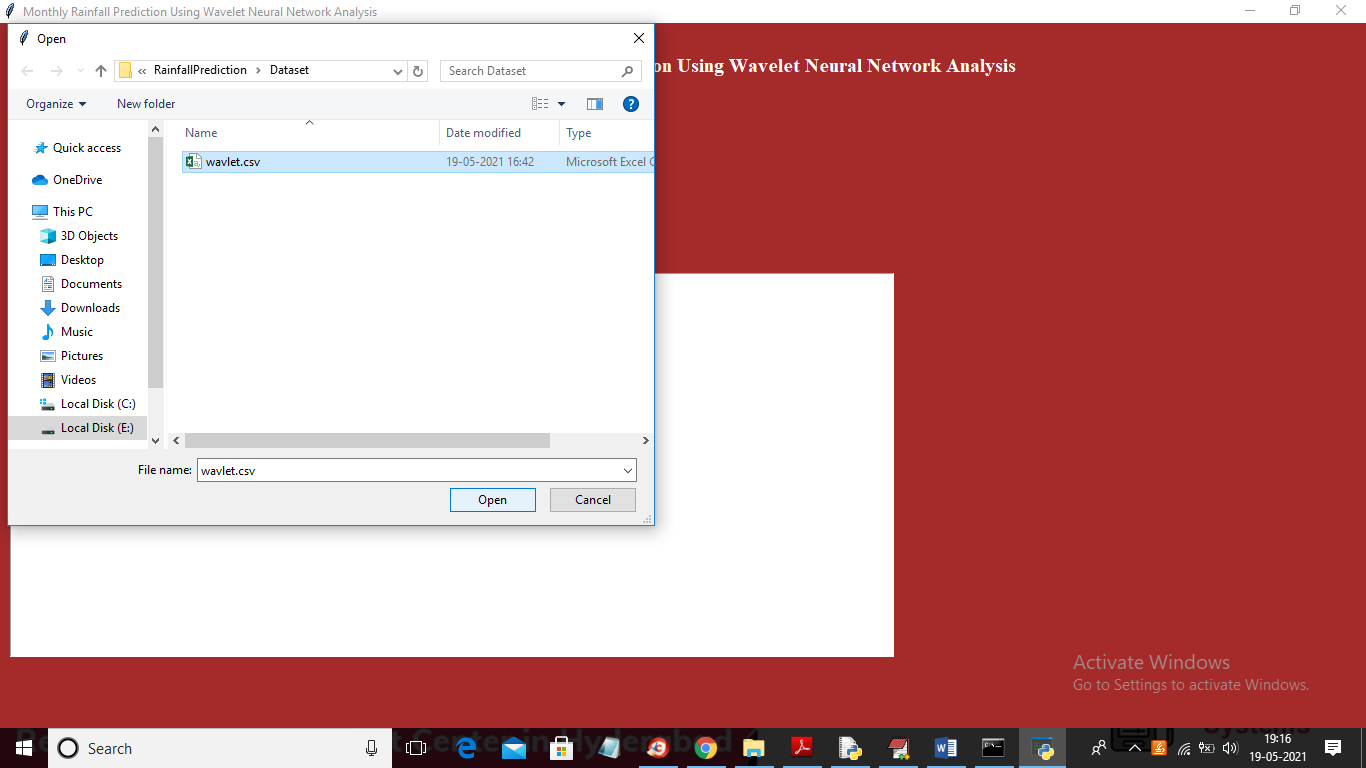
Matplotlib

Scikit – learn

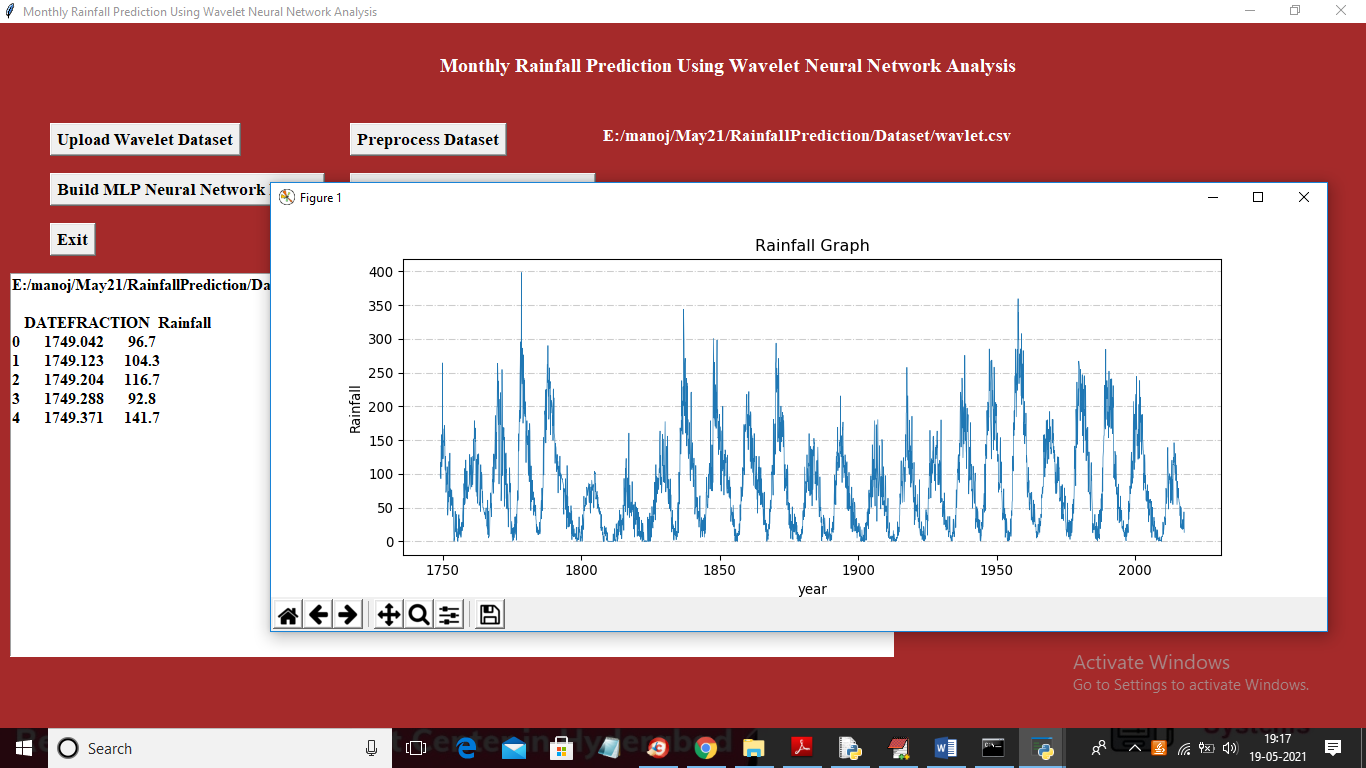
**SCREENSHOTS**



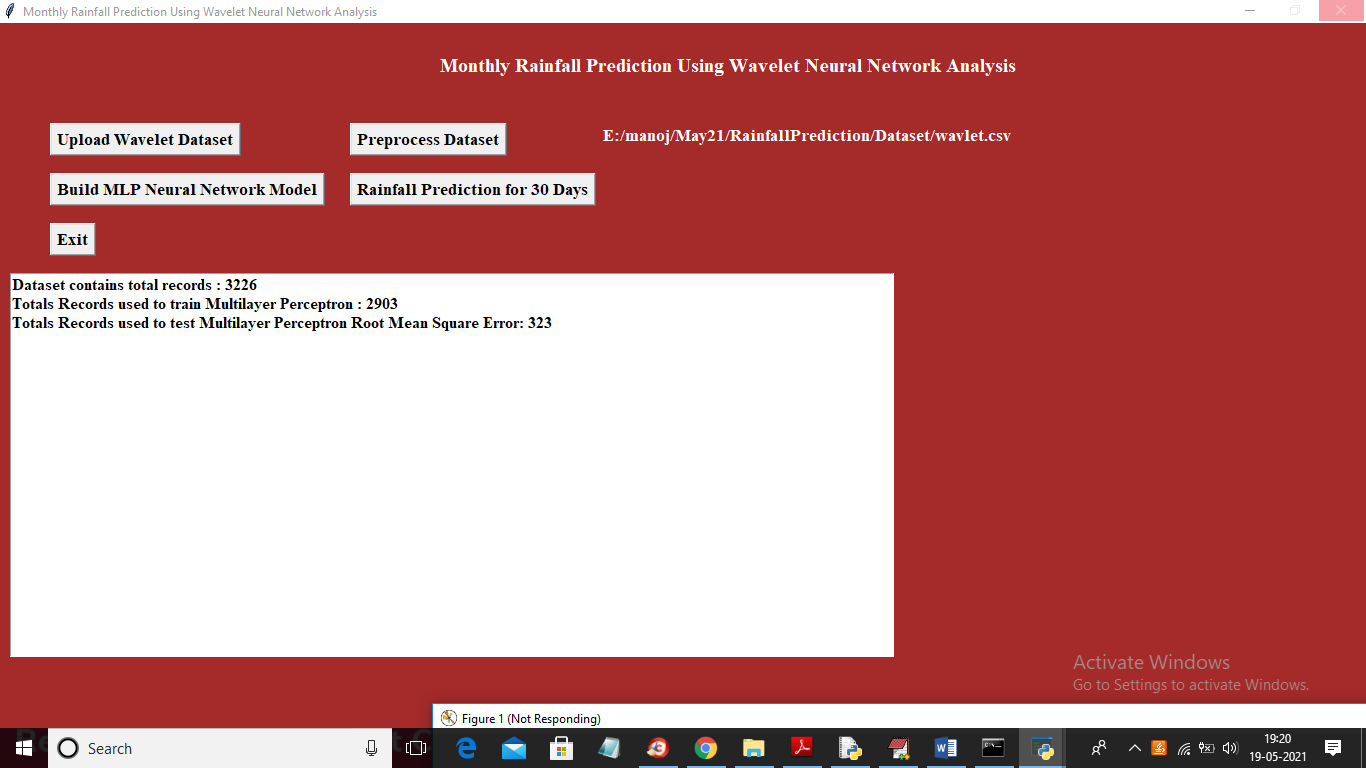
In above screen click on ‘Upload Wavelet Dataset’ button to load dataset



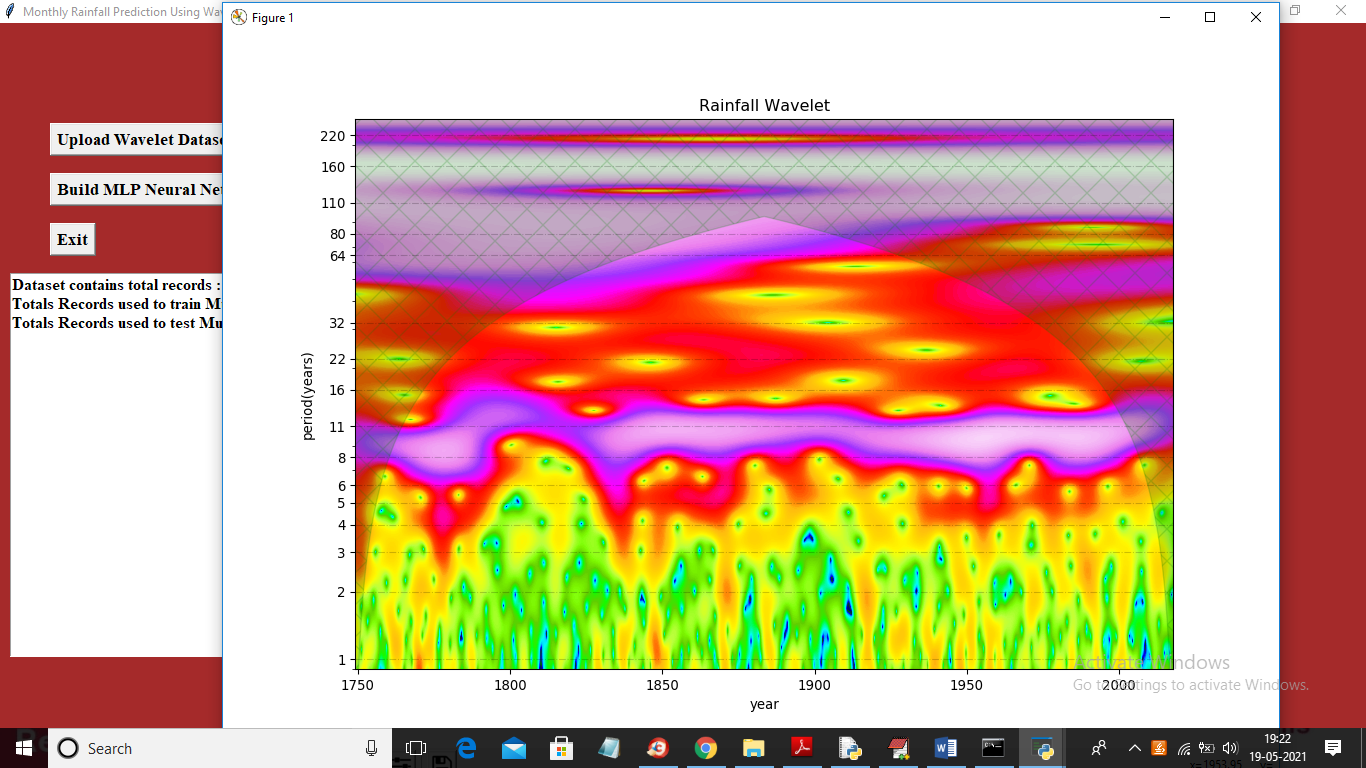
In above screen selecting and uploading ‘wavelet.csv’ file and then click on ‘Open’ button to load dataset and to get below screen



In above screen text area we can see year and rainfall from dataset and same thing we are plotting in graph where x-axis represents year and y-axis represents rainfall occurred in that year and now close above graph and then click on ‘Preprocess Dataset’ button to removing missing values and then split dataset into train and test part and then display wavelet rainfall graph from dataset like below screen

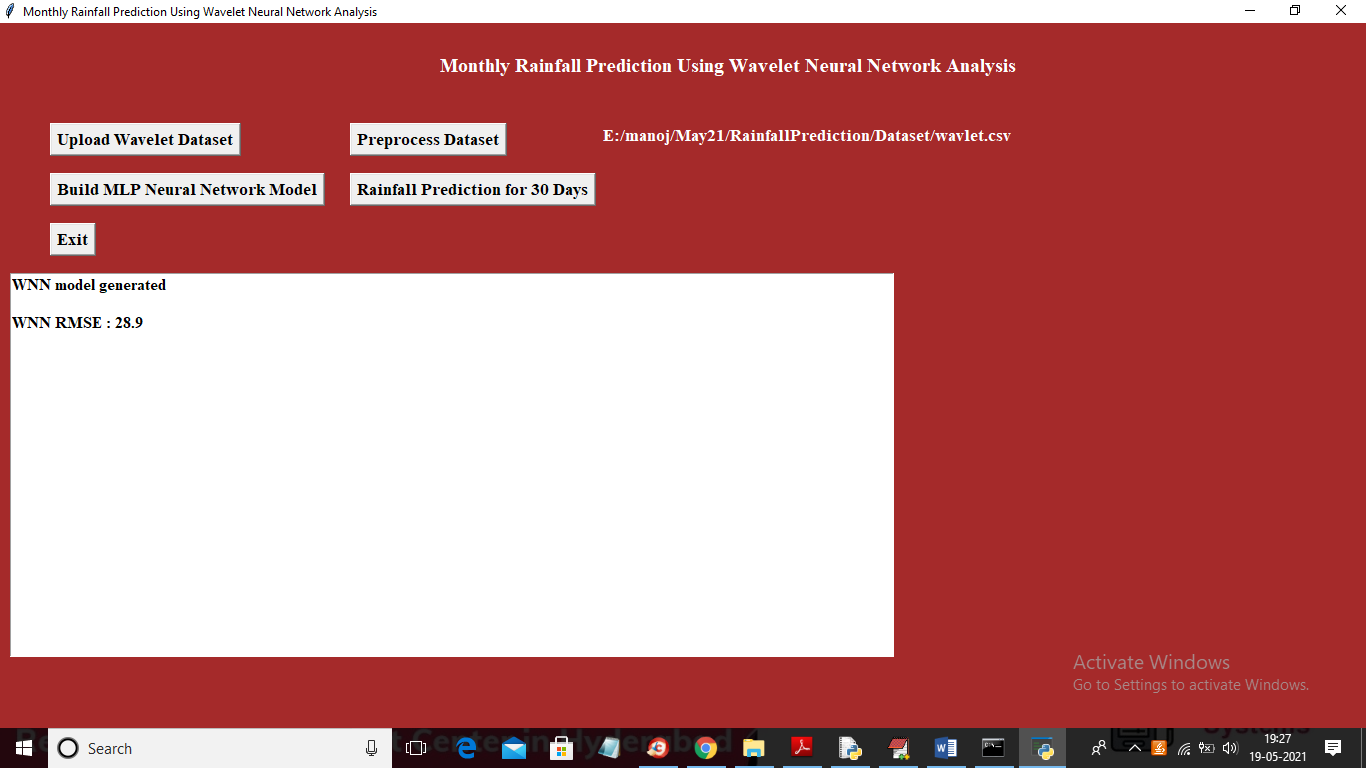


In above screen we can see dataset contains total 3226 records and application using 2903 (80%) records for training MLP and 323 records (20%) to test MLP prediction performance and then calculate RMSE (root mean square error). RMSE represents difference between actual and predicted values. If algorithm predict exact value then different will low and if predict wrong value then difference will be high

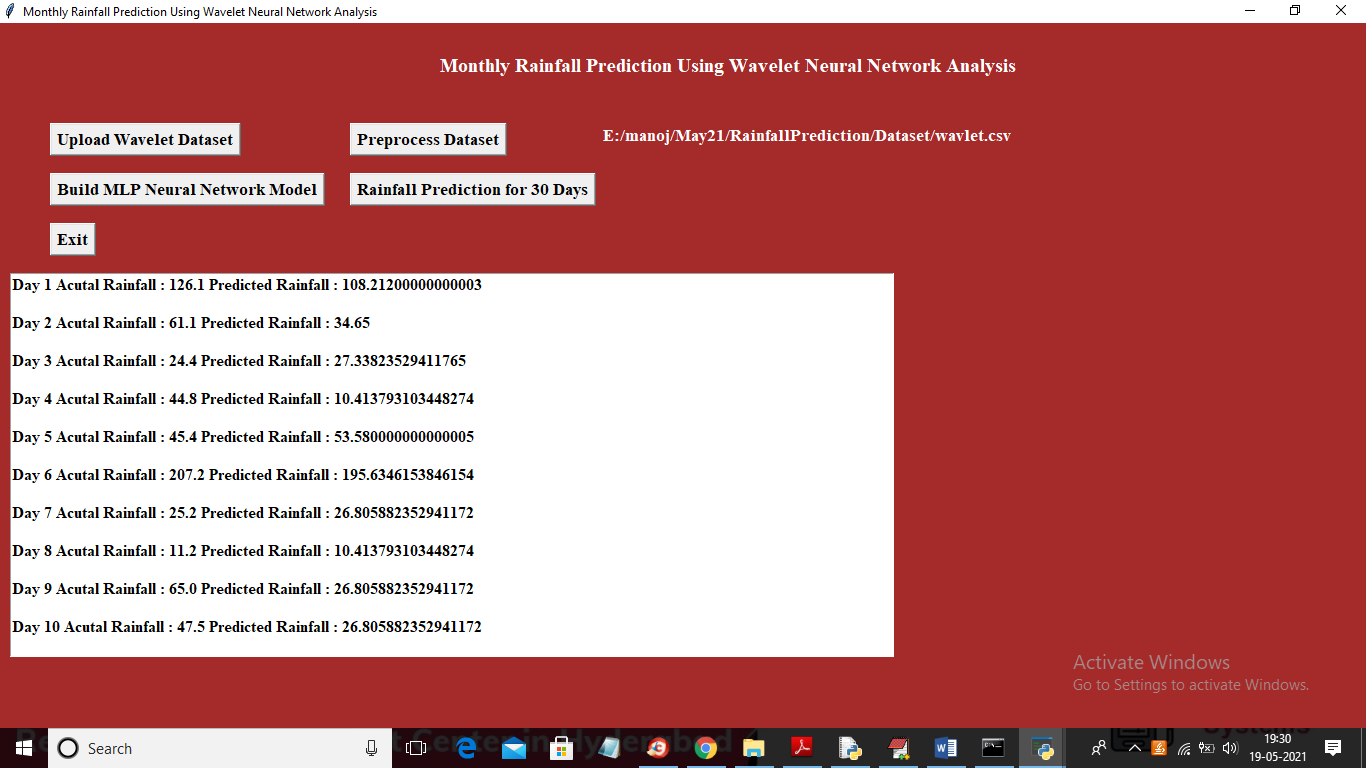


After preprocessing we got Wavelet rainfall graph where green dots represents rainfall and in above graph x-axis represents year and y-axis represents rate of receive rainfall.

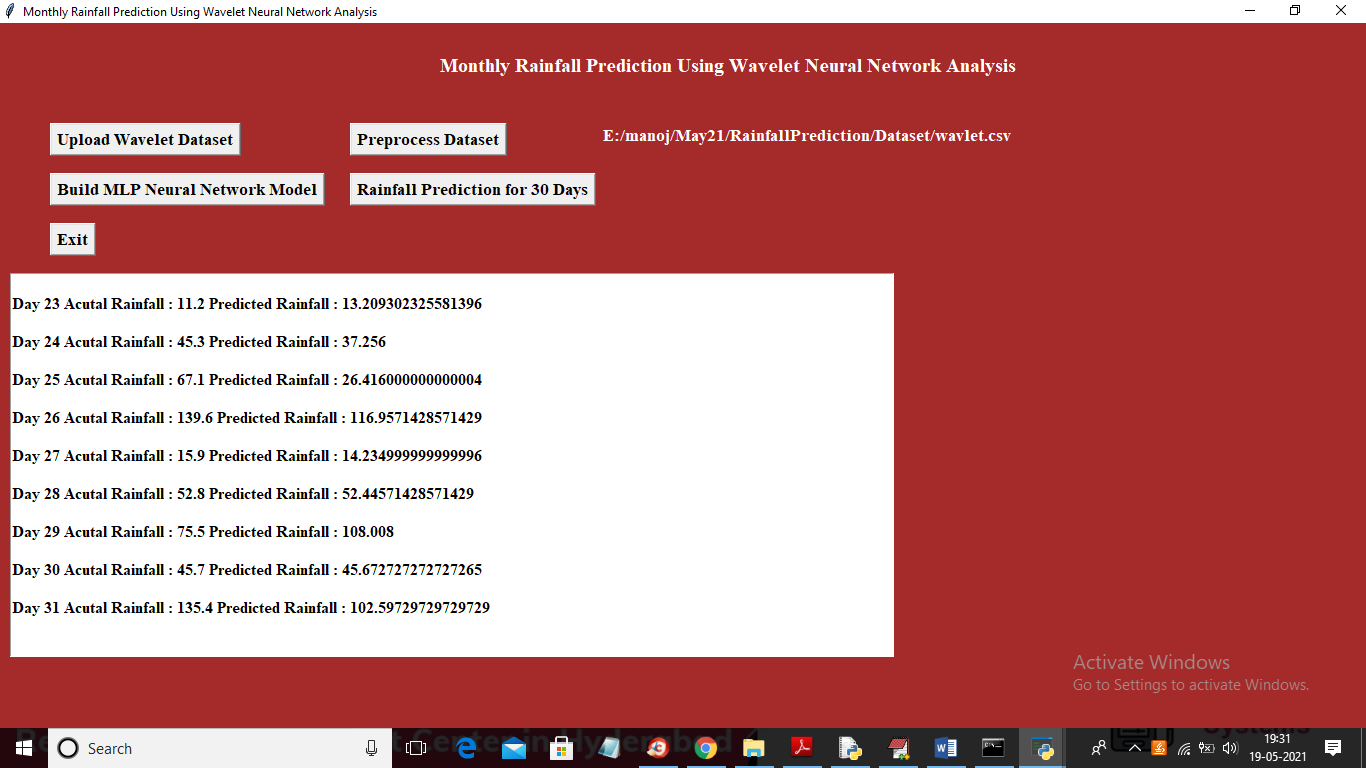
Note: To get above graph output application may take 2 to 4 minutes time so please wait till graph appear correctly. Now close above graph and dataset train and test is also ready and now click on ‘Build MLP Neural Network Model’ button to generate WNN model on above dataset



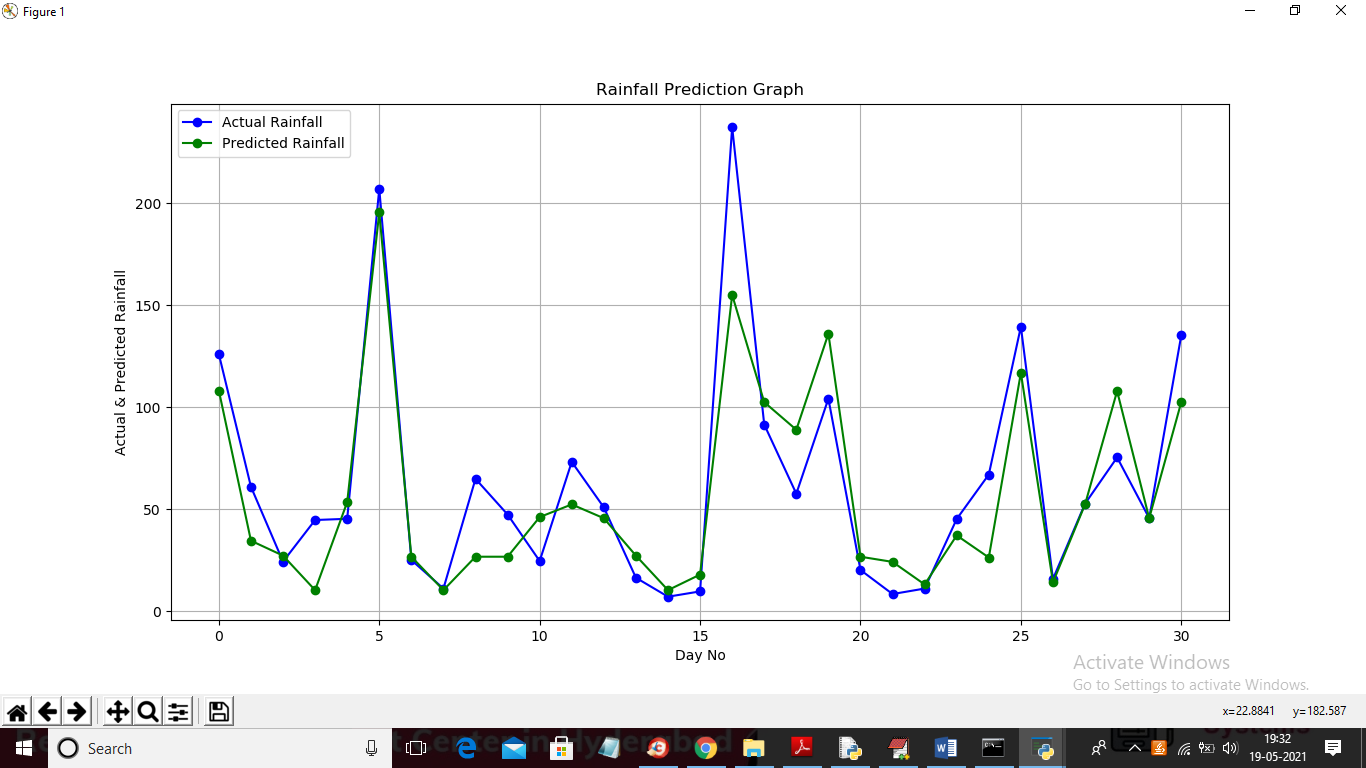
In above screen WNN model generated and we got RMSE error as 28% which is less than RMSE given in paper so model is accurate and now click on ‘Rainfall Prediction for 30 Days’ button to predict rainfall using WNN model



In above screen for each day we are printing actual rainfall from test data and predicted rainfall from WNN model and we can see difference between them is closed and you scroll down above text area to view all records



In above screen you can see for day 26, 27 and 28 prediction is closed to actual value and then will get below graph between actual and predicted rainfall



In above graph x-axis represents day and y-axis represents rainfall and in above graph blue line represents actual rainfall from test data and green line represents predicted rainfall and in above graph we can see difference between two lines are closed so we can conclude that prediction is accurate.

**CONCLUSION:**

This research presents a hybrid model for monthly rainfall time series modeling dubbed the wavelet neural network model (WNN). Wavelet analysis and an artificial neural network are combined in the suggested model. Wavelet decomposes time series into many levels of detail, allowing it to do multiresolution analysis and effectively diagnose the signal's main frequency component and abstract local information. Monthly rainfall at the Darjeeling rain gauge station was used to test the proposed WNN model. DWT was used to break rainfall time series data into subseries. Appropriate sub-series of the variable were used as inputs to the ANN model, while the variable's original time series was used as output. The model parameters are calibrated using 44 years of data, and the remaining data is used to validate the model. The efficiency index for WNN models is over 94 percent, whereas the efficiency index for ANN models is 64 percent, according to this analysis. Overall, the performance of ANN models is inferior to that of WNN models; this could be explained by variations in nonlinear temperature dynamics, which play a significant role in the rainfall process in hilly locations and are successfully mapped by Wavelet based models. It's worth noting that the WNN model's hydrological data has been divided into details and approximations, which could lead to a better representation of rainfall processes. The study only used data from one rain gauge station, thus more research using data from many rain gauge stations in different places may be needed to back up these findings.

**REFERENCES**

1. Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J Hydrol 390(1–2):85–91
2. Anctil F, Tape DG (2004) An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. J Environ Eng Sci 3:121–128
3. Antonios A, Constantine EV (2003) Wavelet Exploratory Analysis of the FTSE ALL SHARE Index. In proceedings of the 2nd WSEAS international conference on non-linear analysis. Non-linear systems and Chaos, Athens
4. ASCE Task Committee (2000a) Artificial neural networks in hydrology-I: preliminary concepts. J Hydrol Eng 5(2):115–123 ASCE Task Committee (2000b) Artificial neural networks in hydrology-II: hydrologic applications. J Hydrol Eng 5(2):124–137
5. Aussem A, Murtagh F (1997) Combining neural network forecasts on wavelet transformed series. Connect Sci 9(1):113–121
6. Benaouda D, Murtagh F, Starck JL, Renaud O (2006) Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting. Neurocomputing 70(1–3):139–154
7. Bhakar SR, Singh RV, Neeraj C, Bansal AK (2006) Stochastic modeling of monthly rainfall at kota region. ARPN J Eng Appl Sci 1(3):36–44