Predictive Analytics on Rainfall using Long Short Term Memory for Identification of Drought

Vignesh Karthikeyan, S.Poornima, M.Pushpalatha

Abstract: ***A drought is duration of below-average precipitation in a certain region, resulting in prolonged shortages in the supply of water. It occurs naturally and has perilous impacts on the society. Observation of patterns of droughts in the past and using it to predict the ones likely to occur in the future can be very helpful. Preparations can be made to try and limit their effects on the society. Drought is however random and dependent on drought variables that possess a non-linear nature. With development in neural networks in the past years, it has shown good scope for time-series prediction with non-linear models. This research approaches the drought prediction problem with the use of Recurrent Neural Networks. The proposed model makes use of past years rainfall values to predict the risk of shortage of rainfall in the given region. The model is expected to show better performance over the existing traditional methods.***

Keywords: ***Drought, Neural Networks, Recurrent Neural networks, LSTM, Standardized Precipitation Index***

I. INTRODUCTION

Drought is a time frame when the precipitation level in a particular region is below average, resulting in a dearth of water supply in any form. It has an impact on the society in several ways, produces negative effects on the environmental, economic and social levels [1].

It has a slow development process, it takes a lot of time for the effects on the economic sector to be observed. Out of all the extreme climate events, drought is reckoned to be the most convoluted and affects a wide range of the population [2].

The Standardized Precipitation Index (SPI) was introduced [3] to help understand and keep track of the relative wetness and dryness of a region along different time periods. Short-term observations of SPI are connected with soil moisture while long-term observations are can help in analyzing ground water and reservoir storage. SPI can be used across regions irrespective of their climate conditions. This is a result of the quantification of observed precipitation as a selected probability distribution function that is modeled over raw rainfall data. SPI can be used around a 1-36 month time scale and is explained as the count of standard deviations by which the observed variation deviates from long term mean. As SPI doesn't take into account the climatic changes associated with evapotranspiration, the SPEI (Standardized Precipitation-Evapotranspiration Index) has been additionally proposed.

Table 1: SPI classes

SPI Value	Classification
$\leq$ -2	Extreme dryness
-1.99 to -1.5	Severe dryness
-1.49 to -1	Moderate dryness
-0.99 to 0	Near average
0 to 0.99	Normal
1 to 1.49	Moderate wet
1.5 to 1.99	Severe wet
$\geq$ 2	Extreme wet

II. LITERATURE SURVEY

Several studies in the past few years have worked on methods to improve drought forecasting.

A.K. Mishra and V.R. Desai used linear stochastic models such as ARIMA and multiplicative SARIMA (Seasonal Autoregressive Integrated Moving Average) for predicting droughts, they used standardized precipitation index (SPI) series in the Kansabati river basin, West Bengal, India [6]. They concluded that the stochastic models gave good results upto a lead-time of 2 months and that the results were better at higher SPI series. They also stated the advantages of SPI over other indices.

In [5], the model was a seasonal drought prediction using a Bayesian framework and was used to study the hydrological drought in the region of the Gunnison River Basin. Standardized Streamflow Index (SSI) was used by the authors to characterize drought events.

Manuscript published on 30 September 2019
* Correspondence Author
Vignesh Karthikeyan*, Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology,Chennai, India
S.Poornima, Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.
M.Pushpalatha, Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India
Corresponding Author; Email: 1vignesh2108@gmail.com

The probabilistic approach was expected to prove useful to develop appropriate drought mitigation strategies by having an insight about future drought status. The proposed method showed promising results in the presence of adequately correlated and dependent variables. Therefore, the drawback observed in the model was that, the essential components of the conditional forecast methodologies are correlation and dependency and without dependent variables, the conditional forecast would lose its meaning.

In the state of Texas, Ali Danandeh Mehr et al. used a model based on Wavelet-Linear Genetic Programming (WLGP) for drought forecasting (with different lead times such as 3, 6, and 12 months) [9]. The WLGP was set to optimize spectral band of predictors and use them to predict the drought index or the predictand directly. It uses the NINO 3.4 Index and the Palmer's modified drought index (PMDI) as predictors. The authors ended up showing that the model using standard linear genetic programming was not able to learn the nonlinear structure due to the nature of drought for than three months in lead time.

A recursive multistep neural network (RMSNN) , a direct multistep neural network (DMSNN) and an autoregressive integrated moving average (ARIMA), were used in [10] for drought forecasting in the Kansabati Riverbasin in India. The paper uses SPI as the drought quantification parameter. The three models are compared, it is observed that ARIMA models provide good results for up to 2-months lead time but not as accurate as the direct multi-step approach. Due to the error build-up between the predicted and observed values every time step, the performance of both ARIMA and recursive multi-step approach deteriorates over longer time periods.

Various machine learning methods such as support vector regression, artificial neural networks and coupled wavelet-artificial neural networks in [11] were used for longterm prediction of drought in the region of the Awash River Basin in Ethiopia. The results of their forecast showed that the wavelet-Artificial Neural Network model gave the best performance for long-term drought prediction. Artificial neural networks (ANN) have shown good performance in different forms complex time series forecasting [12]. The authors in [12], clearly depict the scope for ANN in time-series forecasting and their drawbacks which state that it is impossible to create a universal model that predicts everything. They also discuss the lengthy and tedious process of training the model which includes extensive experimenting and tinkering.

In [13], the authors have used RNNs with LSTM units to predict stock prices. They predicted the stock returns of NIFTY 50 by collecting 5 years of historical data. They concluded that RNN and LSTM are one of the most precise methods available to forecast time-series data.

Poornima.S et al. in [14] had carried out different statistical methods such as Holt-winters and moving average methods and analyzed weather data using forecasting algorithms.

Table 2: Analysis of existing models

Author	**Method**	**Drought Quantification Parameter**	**Drawbacks**
Madadgar and Moradkhani	Probabilistic approach with Bayesian framework	Standardized Streamflow Index(SSI)	The need for dependent variables in conditional forecasting
Danandeh Mehr et al.	Wavelength-Linear Genetic Programming	NINO 3.4 Index and Palmer's modified drought index (PMDI)	Standard linear genetic programming's inability to learn nonlinear structure of drought for than three months in lead time.
AK Mishra and VR Desai	ARIMA	Standardized Precipitation Index(SPI)	Provide good results for up to 2-months lead time but not as accurate as the direct multi-step approach.
AK Mishra and VR Desai	RMSNN, DMSNN	Standardized Precipitation Index(SPI)	Build-up of error between the predicted and observed values at each time step affects performance over long periods.
A Belayneh et al.	Machine learning and ANN	Standardized Precipitation Index(SPI)	Impossibility to create a universal model, lengthy and tedious process of training the model which includes extensive experimenting and tinkering.

It was shown that the performance of the technique was more dependent on the pattern of the data as opposed to the size of the data. It also explained the difficulty faced due the non-linear and unpredictable nature of the time series data as it shows abrupt changes at times.

III. PROPOSED MODEL

The proposed model looks to make predictions on rainfall data using a Recurrent Neural Network with LSTM units and use this data to calculate SPI values to predict the risk of drought.

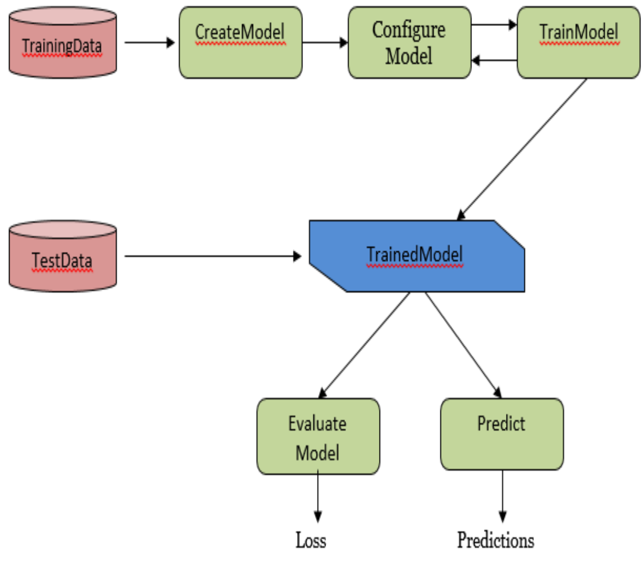


Figure 1: Workflow

a. Recurrent Neural Networks

Recurrent Neural Networks(RNN) are becoming increasingly popular. They follow the basic working of conventional ANN, they are especially useful when the problem involves sequential data. This is useful in gaining better understanding of language and highly efficient in predicting time-series data. In an RNN model, each neuron or unit possesses some internal memory that it uses to maintain information about previous inputs.

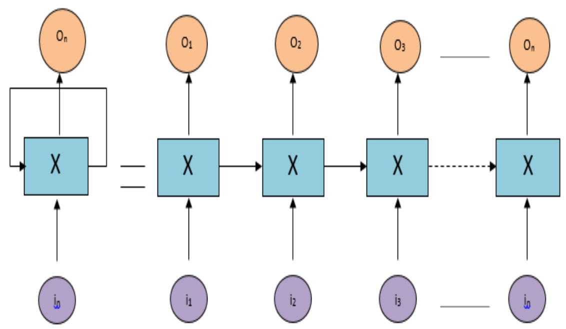


Figure 2: Unrolled RNN

b. Long Short Term Memory

When RNNs are trained on gradient-based learning methods and backpropagation, the vanishing and exploding gradient problems are observed. In these methods, the weights of the neural networks receives updates according to the derivative of the error function related to the current weight in each iteration. The gradient therefore gets vanishingly small and the weights stop changing eventually, posing a problem during the training phase of the model. In order to avoid this, the Long short term memory or LSTM was introduced. They are units in an RNN network, they contain a cell, an input gate, a forget gate and an output gate. LSTM networks are best fit for processing, classifying and making predictions on time-series data, due to the possibility of lags of variable duration between important events in time series data.

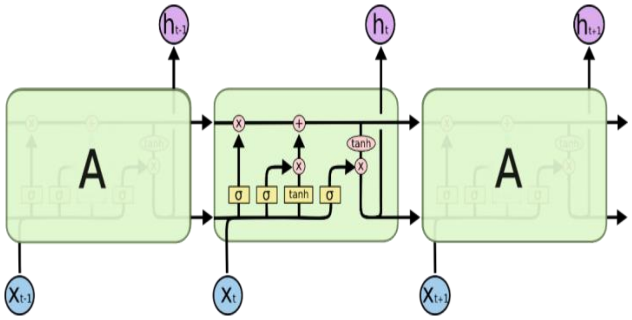


Figure 3: Illustration of LSTM unit

Equations for LSTM units with forget gate

$$f_t = \sigma\,(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma\,(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$o_t = \sigma\,(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t$$

$$h_t = o_t * \tanh\,(C_t)$$

The main element in LSTM is the cell state that runs through the cell horizontally. There are gates that control whether information can be added to or removed from the cell state. The first step involves the forget gate deciding whether the information is important and if it should be kept or lost. Sigmoid function gives a value between 0 and 1 deciding how much a component must be allowed where 1 completely allows the component and 0 completely rejects it. The next step involves the input gate that decides what value is going to be stored in the cell state, it is a combination of sigmoid and tanh layer usually. These are then combined and sent into the cell state. The old cell state C_{t-1} is then updated into a new one C_t. The output gate o_t decides the output h_t that is formed by combining a sigmoid function of the input with the tanh function of the cell state. Finally an output for the layer and the updated cell state are obtained to continue to the next step.

2.3 Implementation

The dataset was split into training and testing data. Data from 1st January 1958 till 31st December 2013 was taken to train the model and the model was tested on data from 1st January to 31st December 2014.The predictions for this data was done with different activation functions and number of layers and the results were compared. The results of the prediction are given below:

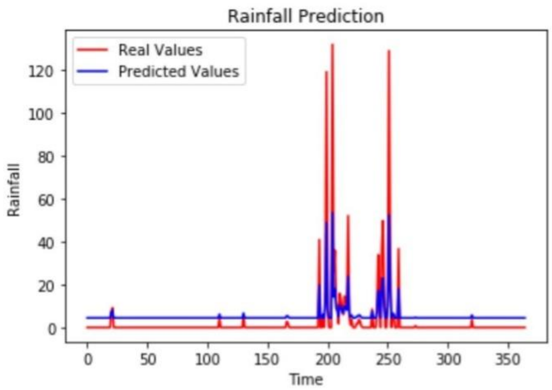


Figure 4: 1 Layer(50 units), Sigmoid

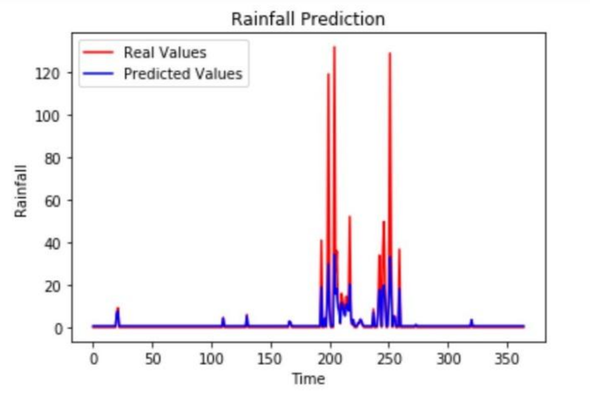
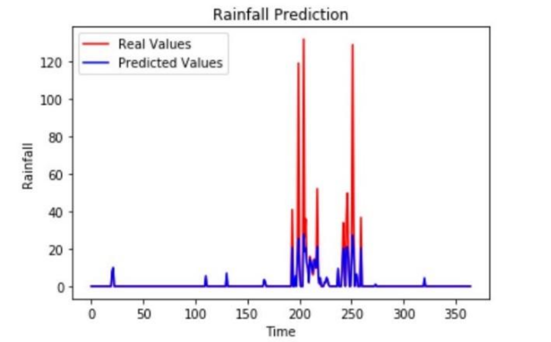


Figure 5: 2 Layers(100,50 units), Sigmoid



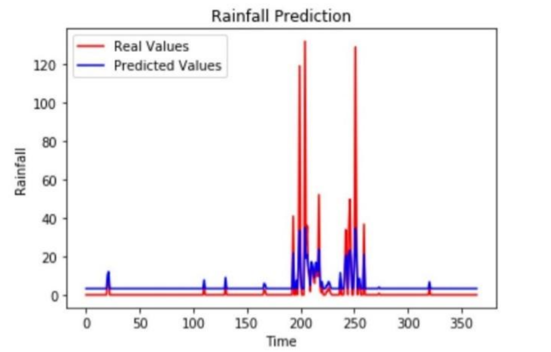


Figure 6: 2 Layers(100,50 units), Tanh

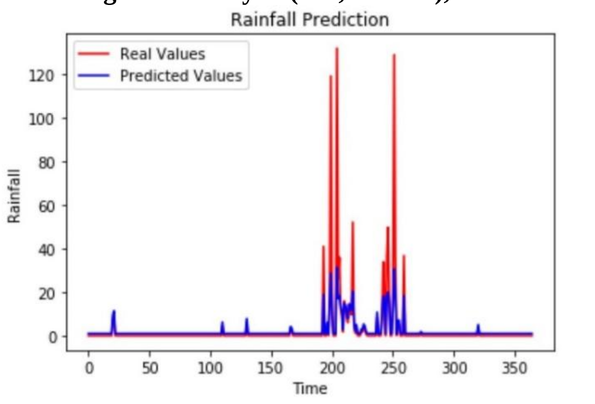


Figure 7: 2 Layers(100,50 units), ReLU

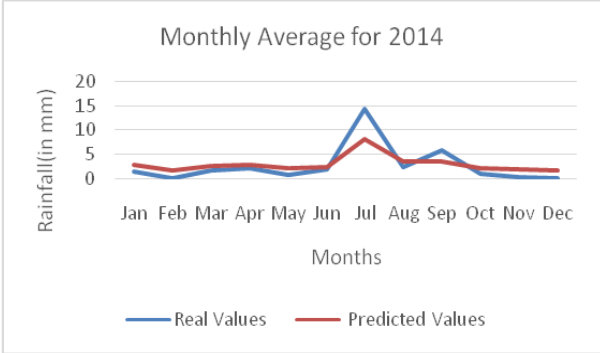


Figure 8: 3 Layers(100,50,50 units), Sigmoid

2.4 Training the model

The model used was a recurrent neural network with LSTM units. Data from 1958 to 2013 was given as training input. The data was shifted by a row to in order to obtain a model that works on time-series data. Being a recurrent neural network, the model promises good performance on time-series data. In an attempt to compare and improve the performance and efficiency of the model, various factors such as the number of layers in the neural network, the activation function and the optimizers were compared to obtain results.

Activation Functions:

Activation functions are used to introduce non-linearities in a neural network. A neural network with without an activation function would simply be a linear function. This type of model will show its inability to learn complex, non-linear structures of the training data.

Sigmoid – It is the most simple out of the lot. It is an S shaped curve and is not centered on zero. This can cause it to go into different extreme directions. However its derivative is simple and needs lower computational power.

Tanh – The function is a hyperbolic tangent, it has a range from -1 to 1 and this makes the function 0-centered. It is usually chosen over the sigmoid activation, however it still poses a vanishing gradient problem.

ReLU – Rectified linear units have been gaining increasing popularity. It uses a simple and effective mathematical form. It reduces the vanishing gradient problem much better than the other cases.

Optimizers:

Optimizers are used in order to update the parameters of the model such as bias and weights during the training process in order to give accurate predictions. Various alternatives of gradient descent have been used for this process. Several researchers have come with new, improved optimization algorithms but however they have proved to be efficient only for a specific case and suffered when it came to generalization.

RMSprop was proved to speed up gradient descent and was an optimizer with adaptive learning rate. It was similar to adagrad but it found a way to handle the diminishing learning rates.

Gradient descent with momentum had shown great promise, the Adaptive moment estimation or Adam optimizer combined the features of momentum and rmsprop. It has a high learning speed, is observed to converge quickly and preserves the learning rate. The adam and rmsprop optimizer are known to perform well on a wide range of models and are the most versatile.

Retrieval Number: B2485078219/2019©BEIESP
DOI:10.35940/ijrte.B2485.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

The comparison of the training process of the model with these parameters were obtained as shown below:

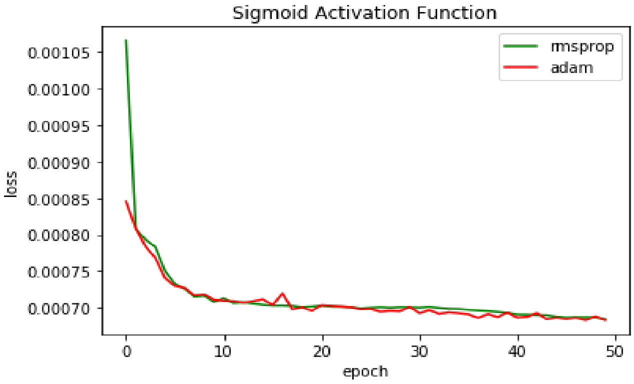


Figure 10: Comparison of loss with 2 Layers,

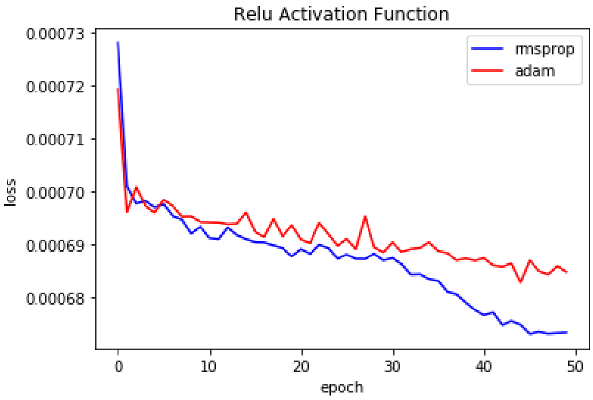
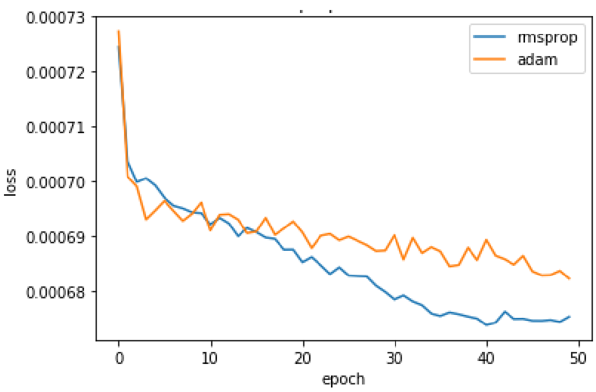


Figure 11: Comparison of loss 2 Layers,

Sigmoid activation ReLU activation



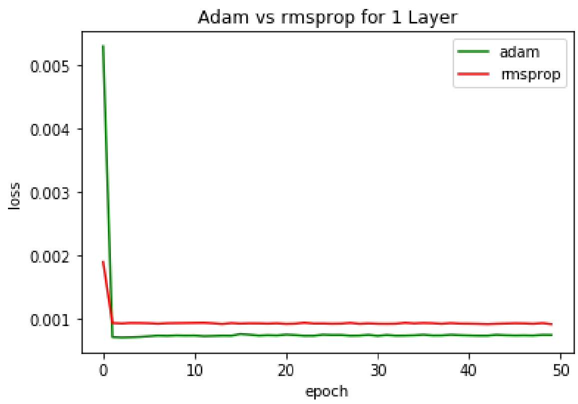


Figure 12: Comparison of loss 2 Layers, sigmoid activation

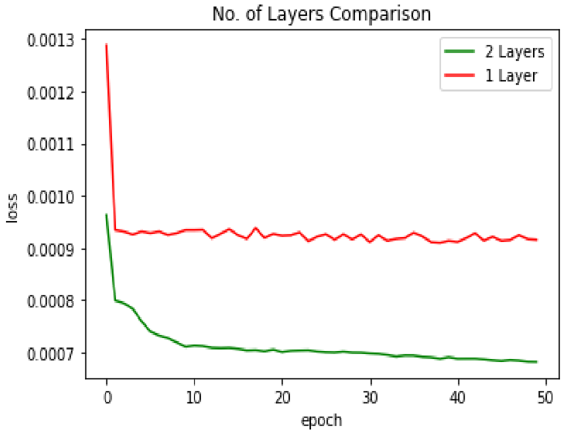


Figure 13: Comparison of loss 1 Layer, Tanh activation

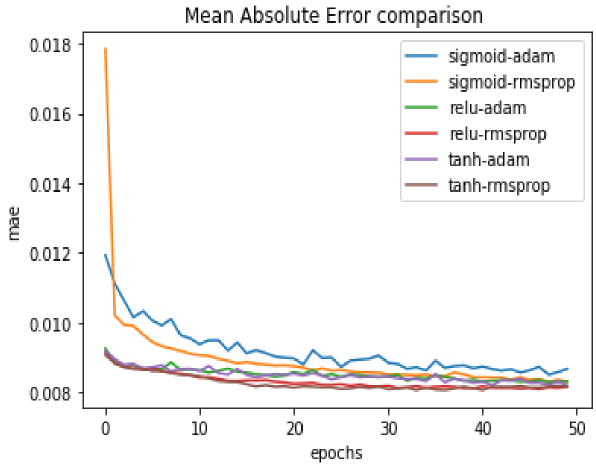


Figure 14: Comparison of loss with Sigmoid activation and rmsprop optimizer

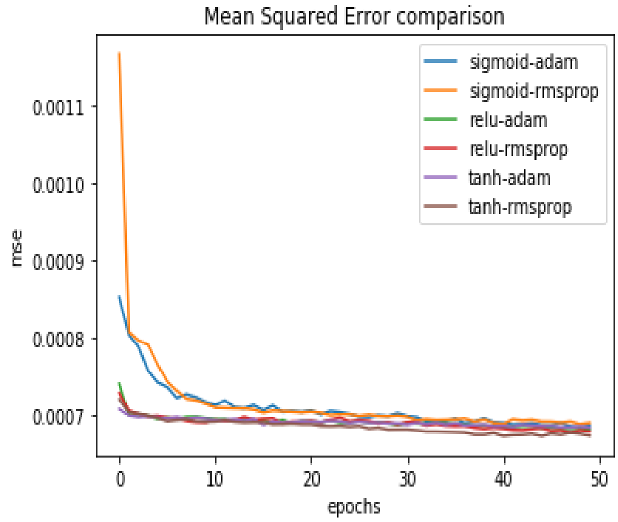


Figure 14: Comparison of MAE with 2 Layer

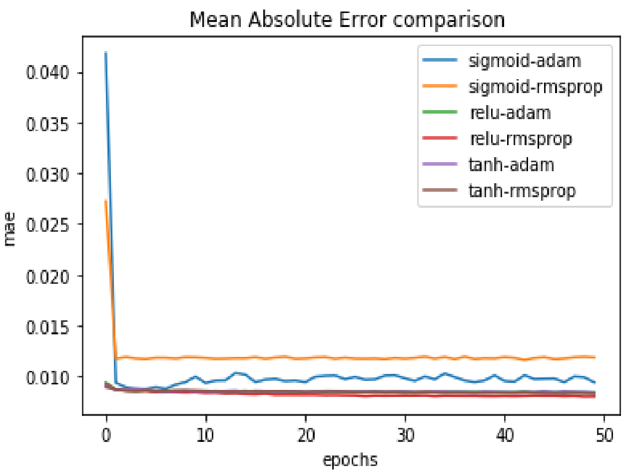


Figure 15: Comparison of MSE with 2 Layer for

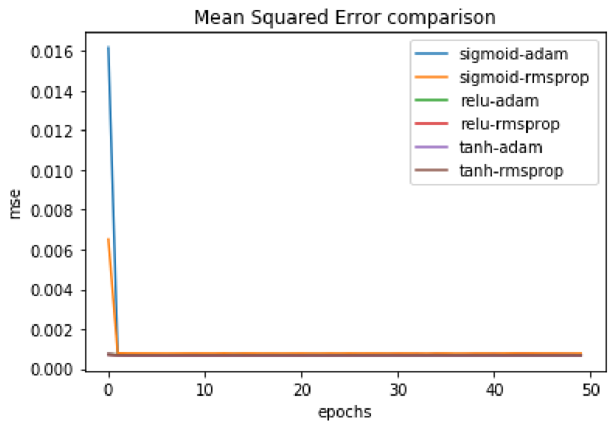


Figure 16: Comparison of MAE with 1 Layer

Table 3: Comparison of performance of 2 Layer Models after 50 epochs

		Loss(e-04)	RMSE	MAE
Sigmoid Activation	adam optimizer	6.8562	0.0261843	0.0087
	rmsprop optimiser	6.8923	0.0262532	0.0083
ReLU Activation	adam optimizer	6.8069	0.02609	0.0083
	rmsprop optimiser	6.7808	0.02604	0.0082
tanh Activation	adam optimizer	6.8311	0.0261364	0.0082
	rmsprop optimiser	6.7273	0.025937	0.0082

Table 4: Comparison of performance of 1 Layer Models after 50 epochs

		Loss(e-04)	RMSE	MAE
Sigmoid Activation	adam optimizer	7.0619	0.0265742	0.0094
	rmsprop optimiser	7.6318	0.0276257	0.0119
ReLU Activation	adam optimizer	6.8701	0.0262109	0.0084
	rmsprop optimiser	6.9463	0.0263558	0.008
tanh Activation	adam optimizer	6.8965	0.0262612	0.0084
	rmsprop optimiser	6.8911	0.0262509	0.0083

IV. ANALYSIS

During the training the training process of the model it is observed that in the ReLU and tanh activation function, the loss or the mean squared error reduces at a faster rate i.e in lesser epochs while using the adam optimizer. However this value saturates and does not reduce beyond a point. Whereas while using the rmsprop optimizer the rate at which the loss and mse reduce are lower and take more epochs but they reduce the loss to lower rate overall. The sigmoid activation shows similar characteristics but the final values of loss are similar. This is probably due to the fact that the adam optimizer uses momentum and tries to get to the point of the minima quicker but misses the point as the learning rate increases. In models that are less complex like 1 Layer models, adam seems to perform significantly better. In the comparison different activation functions, sigmoid is marginally outperformed by tanh and ReLU functions in the number of epochs taken, while the final value after 50 epochs is somewhat similar in range. As expected, the 2 layer is much more efficient in learning the pattern and reducing the loss. The training accuracy of all the models were around 86% and did not pose any significant changes. The mean absolute error was found to be least while using ReLU activation function along with rmsprop optimizer, this is due to the ability of rmsprop to handle the diminishing learning rate problem and maintaining a steady learning rate. In simple models like the ones with one layer, the optimizers did not contribute to any noteworthy improvements. Tanh and ReLU performed similar to the sigmoid activation and the mse values saturated quite quickly as opposed to mae values which showed marginal improvement.

V. CONCLUSION

The proposed model therefore aims at applying the concepts of Recurrent Neural Networks along with LSTM units to predict time series data of rainfall values. The use of LSTM eliminates the vanishing gradient problem and improves the performance of the RNN. Therefore, the RNN with LSTM is expected to show promising results in predicting rainfall values and thus predicting the risk of drought. This can prove crucial in helping people be prepared and alerting the concerned authorities to take appropriate actions in order to minimize the damage induced by droughts. It is observed that with increase in number of layers the predictions are more accurate and that the change in activation functions do not have a major impact on the performance, however they affect the training process of the model. Future work would be to find ways to improve the LSTM units in the RNN by altering the structure of the gates and changing the activation functions.

REFERENCES

1. Antonino Cancelliere, G Di Mauro, B Bonaccorso, and G Rossi. Drought forecasting using the standardized precipitation index. Water resources management, 21(5):801–819, 2007.

2. Donald A Wilhite and Michael J Hayes. Drought planning in the united states: status and future directions. In The arid frontier, pages 33–54. Springer, 1998.
3. T. B. Mckee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," Jan. 1993. [Online].
4. A Ramachandra Rao and G Padmanabhan. Analysis and modeling of palmer's drought index series. Journal of hydrology, 68(1):211–229, 1984.
5. Shahrbanou Madadgar and Hamid Moradkhani. A bayesian framework for probabilistic seasonal drought forecasting. Journal of Hydrometeorology, 14(6):1685–1705, 2013.
6. AK Mishra and VR Desai. Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment, 19(5):326– 339, 2005.
7. Cristina Fern´andez, Jos´e A Vega, Teresa Fonturbel, and Enrique Jim´enez. Streamflow drought time series forecasting: a case study in a small watershed in north west spain. Stochastic Environmental Research and Risk Assessment, 23(8):1063–1070, 2009.
8. Mehmet O¨ zger, Ashok K Mishra, and Vijay P Singh. Long lead time drought forecasting using a wavelet and fuzzy logic combination model: a case study in texas. Journal of Hydrometeorology, 13(1):284–297, 2012.
9. Ali Danandeh Mehr, Ercan Kahya, and Mehmet O¨ zger. A gene–wavelet model for long lead time drought forecasting. Journal of Hydrology, 517:691–699, 2014.
10. AK Mishra and VR Desai. Drought forecasting using feed-forward recursive neural network. Ecological Modelling, 198(1):127–138, 2006.
11. A Belayneh, J Adamowski, B Khalil, and B Ozga-Zielinski. Long-term spi drought forecasting in the awash river basin in ethiopia using wavelet neural network and wavelet support vector regression models. Journal of Hydrology, 508:418–429, 2014.
12. Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu. Forecasting with artificial neural networks:: The state of the art. International journal of forecasting, 14(1):35–62, 1998.
13. Roondiwala, Murtaza & Patel, Harshal & Varma, Shraddha. (2017). Predicting Stock Prices Using LSTM. International Journal of Science and Research (IJSR). 6. 10.21275/ART20172755.