# Feature Selection with Enhanced Bat Algorithm and Modified Recursive Bayesian Deep Neural Network (MRBDNN) for Temperature Prediction

**R.Rajkumar, A. James Albert, S.P. Siddique Ibrahim**

***Abstract:*** ***Weather forecasting is major problemin ecological science. Existing statistical and Climate models are ineffective prediction tools of the long run temperature. Exact weather forecasting is tedious tasks that deal with huge amount of data. In this paper, an Enhanced Bat Algorithm (EBA) is proposed for selection of features from the temperature dataset. High dimensionality of data based on DNN with RMBLR are attempted in this work. Based on analysis of monthly high, average and low temperatures data sets, a novel Recursive Modified Bayesian Linear Regression (RMBLR) algorithm based on Deep Neural Network (DNN) is presented in this study.***

***Keywords:*** ***Feature Selection, BAT Algorithm, Recursive Bayesian, Temperature Prediction, Deep Neural Network.***

## I. INTRODUCTION

Weather prediction is the challenging task the world is facing today. It demands the expertise knowledge for the accurate prediction of weather [2] [3]. The applications that are great beneficiary of the weather forecasting are air traffic, agriculture and production, pollution dispersal, aviation industry and communication.The various methods such as linear regression, multi-layer perception and auto regression have been performed for the prediction. The dynamic nature of the weather makes it difficult for accurate prediction.

Temperature is one of the important factors for change in climate. It has been witnessed that the increase in temperature has changed the phonological order by changing the season for blooming of flowers and so on. The changes of season are also a proof for increase in temperature.

The tools like climate model and statistical time series forecasting method are found to be ineffective for long range temperature prediction.The non-linear involves the effective method named Recurrent Neural Network for complex systems to provide high accuracy

The Echo State Network (ESN) [7] is a type of recurrent network used for temperature prediction. The dataset is collected based on the analysis of monthly, mean and minimum temperature.In this study the long term temperature prediction is performed using the modified RMBLR- ESN [1]. This method is used to provide the maximum, mean and minimum temperature prediction for next 12 months with high accuracy.

**Revised Version Manuscript Received on 25 November, 2018.**

**R.Rajkumar**, Department of Mathematics, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.

**A. James Albert,** Department of Mathematics, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India

**S.P. Siddique Ibrahim,** Assistant Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.

## II. LITERATUR REVIEW

The author of [1] has proposed a method called Recurrent Neural Network (RNN) for the accurate prediction of temperature. The dataset has been collected on the analysis of maximum, mean and minimum temperature. The Echo State Network (ESN), type of neural network is applied in this system for the temperature forecasting. Radhika et.al [2] attempted ESN and RBLR with inflation factor. The Experimental explorations in Central England temperature time series predict monthly high, mean and low temperatures for the upcoming 12 months with high accuracy level.

Support Vector Machines (SVMs) application for weather prediction is presented in [2]. The maximum temperature of n days is provided as input based on the time series data to predict the temperature of the next day. By utilising the optimal value of the kernel system, the performance is calculated over the span of 2 to 10 days. The non-linear regression is employed in this method and the results are correlated with Multi-Layer Perception (MLP) with the back-prorogation algorithm.

The authors of [3] aims to predict the ambient atmospheric temperature of Indian coastal cities. A Fuzzy Knowledge – Rule base technique is used for prediction in this study. In order to estimate the temperature it utilises the historical data as well as the various meteorological parameters to develop the prediction process. The temperature prediction for the mean sea level pressure and the relative humidity has been provided by analysing the mean sea level, relative humidity and temperature for all 3 seasons. His model produces the low Root Mean Square for the temperature.

System for weather prediction using support vector regression model is introduced in [4]. The Support Vector Machine (SVM) is a novel approach of neural network technique. The comparison has been performed between the SVM and Multi-Layer Perception learning algorithm in this model. Thus this model provides the better results on weather forecasting.

In [5] Amartya Raj Gurung has proposed the Artificial Neural Network for weather prediction. The ANN consists of Back Propagation Neural Network which can properly approximate the large class of functions. The real time data set has been used for processing. The result obtained has been compared with data of metrological department for the potential outcome.

The authors of [6] predicted the rainfall using the Artificial Neural Network (ANN) which is found to be the accurate methodology for prediction. The ANN comprises of various techniques of prediction. In his paper the methods like BPN, MLP, RBFN, SOM and SVM are discussed which are found to be suitable for rainfall prediction. However back propagation network (BPN) is the most suitable technique for the accurate prediction of rainfall. This study helps the different authors for the accurate prediction of rainfall.

The non-linear system, echo state network (ESN) to learn the mechanism in biological brains is used in [7]. This method is effective and easy to use. It is found that the usage of chaotic time series has improved the accuracy by a factor of 2400 over the other techniques. The error rate of signal is improved by two orders of magnitude by equalizing the communication channel in this methodology.

An Artificial Neural Network (ANN) with Back propagation learning for the prediction of average monsoon temperature of India is introduced in [8]. The input matrix for the neural network has been found using the six predictors. It is also found to produce the weather forecast with small prediction error. It is figured this method produce the better results.

The author of [9] has proposed integration of back propagation technique with the genetic algorithm technique for the time series based temperature prediction model. This hybrid method has been developed to analyse the effects of both under training and over training the models. Thus this study provides the compensability between time series based on hybrid method and back propagation method.

Pankaj Kumar [10] has strived adaptiveneuro-fuzzy inference system (ANFIS) to model the relationship between maximum and minimum temperature. The mean weekly maximum and mean weekly minimum temperature dataset of a decade from the year 1997 to 2006 has been obtained from regional centre of Meterological department, Dehradun, India. Time n maximum temperature data are provided as input to analyse the maximum temperature of upcoming week.

## III. PROPOSED SYSTEM

This section presents the detailed description for the feature selection with enhanced bat algorithm and modified recursive Bayesian deep neural network (MRBDNN) for temperature prediction. The figure 1 shows the process flow of the proposed algorithm.In this work, an Enhanced Bat Algorithm (EBA) is introduced for selection of features from the temperature dataset. For prediction, a novel Recursive Modified Bayesian Linear Regression (RMBLR) algorithm based on Deep Neural Network (DNN) is presented in this paper.

### A. BAT Algorithm (BA)

BAT algorithm simulates the behaviour of bats. Each bat is considered with the parameters like position, velocity, pulse rate, loudness and frequency. The performance of Bat algorithm was better compared to PSO and GA. The implementation of GA is more complicated than other algorithms. This type of interaction influences the quality of a solution and the time [11].

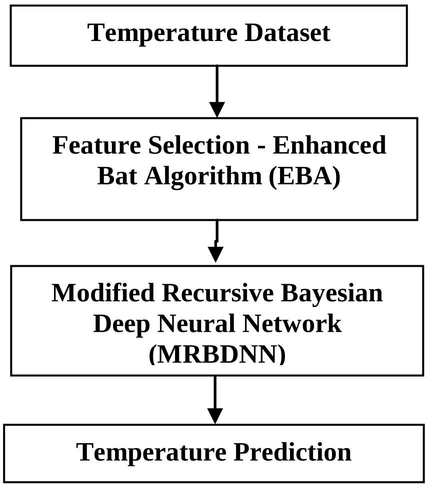


**Figure 1: Proposed System – Process Flow**

**<u>BAT Algorithm</u>**

1. Objective function: $g(y)$, $y = (y_1, \dots, y_d)$
2. Consider bat population $x_i$ and velocity $v_i$, $i = 1,2, \dots, n$
3. Define frequency $g_i$at $y_i$
4. Initialize pulse emission rate $r_i$ and loudness $A_i$
5. While (t<maximum number of iterations)
6. Generate new solutions by adjusting frequency, and updating velocities and location/solutions.
7. If (rand $>r_i$)
8. Select a solution among the best solutions
9. Generate a local solution around the selected best solution
10. End If
11. If (rand< $A_i$ and g(yi)<g(y*))
12. Accept new solutions
13. Increase $r_i$reduce $A_i$
14. End If
15. Ranks the bats and find current best y*
16. End While
17. Display results

### B. Enhanced BAT Algorithm (EBA)

Bat algorithm is improved by the integration of Chi-Square feature selection for selecting the best features in a random manner. Modified Bat optimization algorithm is introduced for the selection of appropriate features from the dataset. In the modified bat algorithm, the temperature dataset is considered as the initial Bat population and each data has a certain frequency $gr_i$ and velocity $ve_i$. The frequency and velocity are estimated and updated after the every iteration. The features are grouped together based on the similarity and ranked by resampling method. The current solution is compared with the ranked solutions and they are sorted in the best order.

**<u>Enhanced BAT Algorithm</u>**

1. Initialization of $ds_i$, we, Where ds=Dataset, 1<=i<=n


*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*

2. Initialization of $gr_i$, $ve_i$, Where frequency$gr_i$, velocity$ve_i$
3. While (t<dsi) Do
4. Update $gr_i$, $ve_i$
5. Generate the first set of solution
6. Update $ds_i$
7. Generate random number r
8. Select random instance from $ds_i$
9. Calculate similarity
10. Select the location among the current best solution
11. If (a better solution is found)
12. Update the current best
13. Generate a new solution
14. If (rand<$ds_i$& (g($x_i$)<g(gbest)))
15. Accept the new solution
16. Sketch the global best solution
17. Sort the features based on gbest
18. Stop iteration if target is reached.

### C. RBLRWAIF-ESN

In [1], the recursive Bayesianlinear regression (RBLR) algorithm based on ESNRBLRWAIF-ESN method was proposed. Single input single output systems was considered, it is show that ESN has single input and single output

**RBLRWAIF-ESN Algorithm**

Step 1.Start the input $W^{in}$ and the reservoir $W^r$inconstantly and give starting values $w_t$ and $p_t$. The mean $w_0$,=0 and its covariance matrix $p_0$=unit matrix.

Step 2.Renew the ESN state to derive the new reservoir state $x_t$.

Step 3.Renew the parameters $S_t$, $k_t$, $w_t$ and $p_t$.

Step 4. Making predictions is straightforward. The average of the prediction is$y_t' = \varphi_t * w_t$.

Step 5. Compute the covariance matrix $p_t = (1 + 0.2 * \propto_t) * p_t$, where $\propto_t$ is an adaptive inflation factor.

Step 6. Go to step 2.

### D. Recursive Bayesian Deep Neural Network (MRBDNN)

A novel Recursive Modified Bayesian Linear Regression (RMBLR) algorithm based on Deep Neural Network (DNN) is presented to analyse monthly temperatures statistics. The algorithm consists of two main components: DNN and a RMBLR algorithm. Weights are marginalized. Then, we could now applyapproximations to evaluate this integral.

**MRBDNN Algorithm**

1. Prediction of single continuous target t from vector x of inputs
2. Assumption of p(t|x) is Gaussian with precision β
3. Output of neural network y(x,w) gives mean
$p(t \mid x,w,\beta) = N(t \mid y(x,w),\beta^{-1})$
4. Prior probability function is
$p(w \mid \alpha) = N(w \mid 0,\alpha^{-1}I)$
5. Likelihood probability function is
$p(D \mid w,\beta) = N(t_n \mid y(x_n ,w),\beta^{-1})$
6. Posterior probability function is
$p(w \mid D,\alpha,\beta)$ α $p(w \mid \alpha)p(D \mid w, \beta)$
7. The output distribution is given by
$P (C1|x, D) = \int y ( x; w ) p ( w|D ) d w$

Here,
The output activation function ->y(x; w)
Interpret as P(C1|x, w))

## IV. EXPERIMENTAL RESULTS

The relative experiments are conducted to examinethe performance of the proposed algorithm. For comparison,seasonalautoregressive integrated moving average (SARIMA), a three-layer FFNN-BPand RBLRWAIF-ESN Algorithm is also appliedfor temperature prediction.

### A. Dataset Collection:

The Climate Change: Earth Surface Temperature Data is considered from the Kaggle. The taken dataset covers the Global Land and Ocean-and-Land Temperatures. A subset of data from Kaggle, featuring 7 countries and 16 cities is used for analysis. The dataset consists of 48470 records. The subset temperature dataset consist of attributes like record_id, month, day, year, AverageTemperatureFahr, AverageTemperatureUncertaintyFahr, City, country_id, Country, Latitude and Longitude.

### B. Tool:

R programming is used to measure the performance of the proposed algorithm and comparision. RStudio IDE is used for implementation of the proposed system. The dataset is loaded, and features are selected using EBA. Then MRBDNN algorithm applied to get the output.

### C. Performance Metrics:

**Mean Absolute Error (MSE)**

The MSE is

$$MSE = \sum (f(x_i) - y_i)^2 / N \quad \text{------- 1}$$

**Root Mean Squared Error (RMSE)**

The difference between sample, population values and observed values are speculated by model or an estimator.

The RMSE is given by

$$RMSE = \sqrt{MSE} = \sqrt{\sum (f(x_i) - y_i)^2 / N} \quad \text{------ 2}$$

**Relative Absolute Error (RSE)**

Calculated as the Mean absolute error divided by the error of the classifier.

The RSE is given by


*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*

$$RSE = \sum (f(x_i) - y_i)^2 / \sum (\bar{y}_i - y_i)^2 \quad \text{------ 3}$$

**Root Relative Squared Error (RRSE)**

The RRSE is given by

$$RRSE = \sqrt{RSE} = \sqrt{\sum (f(x_i) - y_i)^2 / \sum (\bar{y}_i) - y_i)^2} \quad \text{----- 4}$$

## V. RESULTS AND DISCUSSIONS

The original dataset is given as input to the EBA algorithm. It selects the best features for the prediction. Table 1 gives the performance analysis of 4 algorithms namely SARIMA, FFNN-BP, RBLRWAIF-ESN and MRBDNN. Four performance metrics are used for evaluating the performance. MSE, RMSE, RSE and RRSE are used in evaluation.

**Table 1: Performance Analysis**

| **Algorithm** | *MSE* | *RMSE* | *RSE* | *RRSE* |
|---|---|---|---|---|
| *SARIMA* | 0.0467 | 0.2133 | 10.4978 % | 45.2448 % |
| *FFNN-BP* | 0.0529 | 0.1692 | 11.9007 % | 35.9002 % |
| *RBLRWAIF-ESN* | 0.035 | 0.1586 | 7.8705 % | 33.6353 % |
| *MRBDNN* | 0.0327 | 0.1291 | 7.3555 % | 27.3796 % |

The results shows that SARIMA performs well compared to FFNN-BP in terms of MSE and RSE. RBLRWAIF-ESN gives good results compared to SARIMA and FFNN-BP in all four performance metrics. But our proposed algorithm MRBDNN outperforms all three algorithms with all 4 metrics.

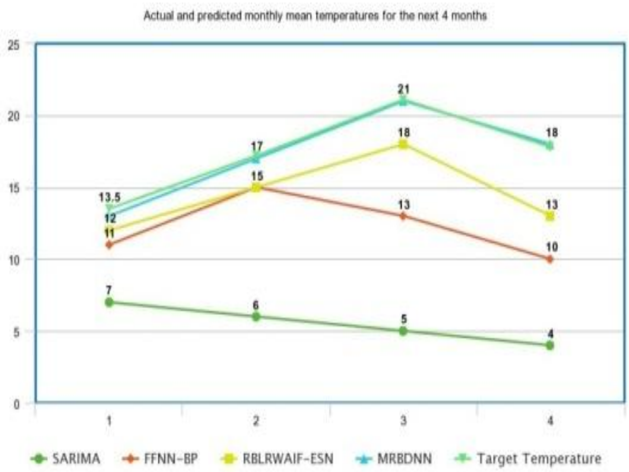


**Figure 2: Actual and Predicted Monthly Mean Temperatures for the Next 4 Months**

The figure 2 represents the actual and predicted monthly mean temperatures for the next 4 months. The actual temperature is compared with the prediction of 4 algorithms SARIMA, FFNN-BP, RBLRWAIF-ESN and MRBDNN. SARIMA prediction highly deviate from the actual. The experiment (Figure 2) showed that the MRBDNN model could track the actual temperature values.

## VI. CONCLUSION

Temperature and climate prediction is one of the recent research field that gains lot of importance.High dimensionality is one of the issue in temperature dataset. Enhanced Bat Algorithm (EBA) is proposed for selection of features from the temperature dataset. A novel Recursive Modified Bayesian Linear Regression (RMBLR) algorithm based on Deep Neural Network (DNN) is presented in this study for prediction. In this paper, basic statistical measures like MSE, RMSE, RSE and RRSE are used for evaluation of the proposed algorithm. The performance of MRBDNNwas correlate with RBLRWAIF-ESN, SARIMA andFFNN-BP. The results indicate that MRBDNNperforms better than other algorithms. It can beconcluded that MRBDNN can be used as aneffective method for temperature prediction.

## REFERENCES


1. Biaobing Huang, Guihe Qin, Rui Zhao, Qiong Wu, AlirezaShahriari, “Recursive Bayesian echo state network with an adaptive inflation factor for temperature prediction” , Neural Computer and Application, November 12, 2016.
2. Y.Radhika, M.Shashi, “Atmospheric temperature prediction using Support Vector Machine”, International Journal of Computer Theory and Engineering, Vol.1, No.1, April 2009.
3. Dipi A. Patel, R.A.Christian, “Ambient atmospheric temperature prediction using fuzzy knowledge-rule based for inland cities in India”, World Applied Sciences Journal, 2012.
4. TarunRao, N.Rajasekhar, Dr.T.V.Rajinikanth, “An efficient approach for weather forecasting using Support Vector Machine”, International Conference on Computer Technology and Science, Vol.47, 2012.
5. Amartya Raj Gurung, “Forecasting weather system using Artificial Neural Network (ANN): A survey paper”, International journal of Latest Engineering Research and applications (IJLERA), Vol.02, October 2017.
6. Deepak RanjanNayak, AmitavMahapatra, Pranati Mishra, “A survey on rainfall prediction using Artificial Neural Network”, International Journal of Computer Applications, Vol.72, No.16, June 2013.
7. Herbert Jaeger, Harald Hass, “Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication”, Science, Vol.304, April 2, 2004.
8. S.S.De, A.Debnath, “Artificial neural network based prediction of maximum and minimum temperature in summer monsoon months over India”, Applied Physics Research, Vol.1, No.2, November 2009.
9. Shaminder Singh, PankajBhambri, Jasmeen Gill, “Time series based temperature prediction using back propagation with genetic algorithm technique”, International Journal of Computer Science Issues, Vol.8, Issue 5, No.3 September 2011.
10. Pankaj Kumar, “Minimum weekly temperature forecasting using ANFIS”, Computer Engineering and Intelligent Systems, Vol.3, No.5, 2012
11. C. Zhen, Z. Yongquan, and L. U. Mindi, “A simplified Adaptive Bat Algorithm Based on Frequency”, J.Comput. Inf. Syst., vol. 9: 16, pp. 6451–6458, 2013.