Optimization Enabled Neural Network for the Rainfall Prediction in India

Ananda R. Kumar Mukkala[1(✉)], S. Sai Satyanarayana Reddy[1], P. Praveen Raju[2], Mounica[2], Chiranjeevi Oguri[2], and Srinivasu Bhukya[2]

[1] Department of CSE, Sreyas Institute of Engineering and Technology, Hyderabad 500068, India
anandranjit@gmail.com
[2] Department of ECE, Sreyas Institute of Engineering and Technology, Hyderabad 500068, India

Abstract. Rainfall prediction plays a major role in ensuring the livelihood of many people especially, for the farmers. Heavy and irregular flow of rainfall can cause flood, landslide and much other destruction. To prevent this, rainfall should be predicted in a periodic manner. As a contribution, the proposed Spotted Hyena based nonlinear autoregressive model (SH-NARX) prediction model effectively predicts the rainfall in a yearly, monthly and quarterly manner using the Indian rainfall dataset. The data is collected and trained using the NARX neural network, which is a non linear autoregressive network that is optimized using the spotted hyena optimization for rainfall prediction. The performance of the prediction model is analyzed based on RMSE and PRD that are minimal, highlighting the higher accuracy rates.

Keywords: NARX neural network · Rainfall prediction · Spotted hyena · Indian rainfall data · RMSE

1 Introduction

Water plays a major role in the human life and no life can be pertained in the earth without water. One of the important sources of water is rainfall. Rainfall can be defined as a quasi-periodic signal with recurrent periodic fluctuations that can occurs at different levels of varying noises [8, 9, 3]. Rainfall is necessary to recharge the depleting ground sources as well as for agriculture. The agricultural productivity of the country entirely depends on the rainfall which is directly proportional to the growth of economy. The developing country such as India and Vietnam completely relies upon the monsoon rainfall for the agricultural productivity [10, 11, 12, 2]. The accurate prediction of the rainfall greatly helps in the futuristic planning of the agricultural activities as well as the severity of the natural disasters such as flood, landslide, and drought could be identified by predicting the rainfall [13, 14, 2]. It is quite difficult to forecast the rainfall due to the rapid change and complexity of the weather conditions.

Various researches are employed for the accurate detection of rainfall but in a broader perspective it is quite difficult to accurately predict the occurrence of rainfall at most of the times. Basically, the prediction of rainfall is performed using the empirical and

M. Singh et al. (Eds.): ICACDS 2022, CCIS 1614, pp. 12–23, 2022.
https://doi.org/10.1007/978-3-031-12641-3_2

dynamic methods. The empirical approach utilizes the previous information or historical prediction and this approach is most commonly used in the regression and artificial neural networks. The dynamic approach is utilized by the physical and statistical methods. The advancement of technology in recent years promotes the prediction of rainfall using the techniques regression, support vector machine (SVM), and K-nearest neighbor (KNN). Deep learning models are useful in examining large dataset and provide factual information and it is useful in computational applications [1]. Numerous methods are used in the prediction of rainfall and are categorized into three groups such as: statistical, dynamic and satellite based methods [15, 16] and the statistical methods are frequently used due to its inexpensive and time consuming nature [3].

The main objective of the research is to develop a rainfall prediction model to effectively predict the rainfall so that the destruction caused by the rainfall can be greatly avoided. As a contribution the data are collected from the rainfall predicted from the India dataset and then the SH-NARX model is used for the prediction which is optimized using the spotted hyena optimization. The NARX is a nonlinear autoregressive model that predicts the output based on the past values as well as current values. Then the final output are predicted and the main contribution of the proposed SH-NARX model are as follows.

- ***Spotted Hyena optimization:*** The spotted hyena optimization consists of faster convergence, and the search of the solution will be presented in a globalized search manner. Due to its simplicity and ease of use this optimization is widely used.
- ***SH-NARX model:*** The SH-NARX model has the capability to train the samples with less number of training cycles. The prediction error is also greatly reduced using this model.
- The PRD and RMSE values are improved to prove the significance of the proposed SH-NARX model

The alignment of the manuscript is as follows: The existing experimentations along with its advantages and disadvantages are enumerated in Sect. 2, the proposed optimization enabled SH-NARX model along with its architecture are depicted in Sect. 3. The brief discussions of the results achieved are enumerated in Sect. 4 band finally the rainfall prediction model is concluded in Sect. 5.

2 Motivation

In this context, the review of the existing prediction models with the challenges of the research is presented, which motivated the researcher in designing a prediction model based on ANN and optimization. R. Venkatesh *et al.* [1] introduced a rainfall prediction system using generative adversarial networks which performs well in both monthly and annual averages but the drawback is that it consumes more time and there is a need for the GPU-based computational resources. Duong Tran Anh *et al.* [3] developed a novel hybrid models for monthly rainfall prediction which predicts the rainfall in more accurate manner compared with conventional methods but it consists of a complex structure. Joao Trevizoli Esteves *et al.* [4] established an automatic ANN modeling that has the capability to find the global optimum solution but the results obtained has an average accuracy. Hatem Abdul-Kader et al. [6] predicted the rainfall using long short term memory (LSTM) which is strong enough to predict the rainfall in the Sidoarjo but the prediction mainly concentrated on single place. S. Dhamodharavadhani and R. Rathipriya [7] presented a Map Reduce-based exponential smoothing for the region wise rainfall prediction that showed efficient runtime improvement but the smoothing methods represented can be furthermore optimized. The major challenges of the research is enumerated as follows:

- The meteorological parameters considered for the prediction of rainfall is stochastic in nature which is a challenging task to be resolved [2].
- When there is a necessity for local-scale projections, the modeling of the variabilites present in the rainfall events are difficult to be modeled [3].
- It is a challenging task to obtain reliable and accurate prediction models because of the uncertainity and variability present in the prediction models that spatially forecast the rainfall for a short period of time [4].

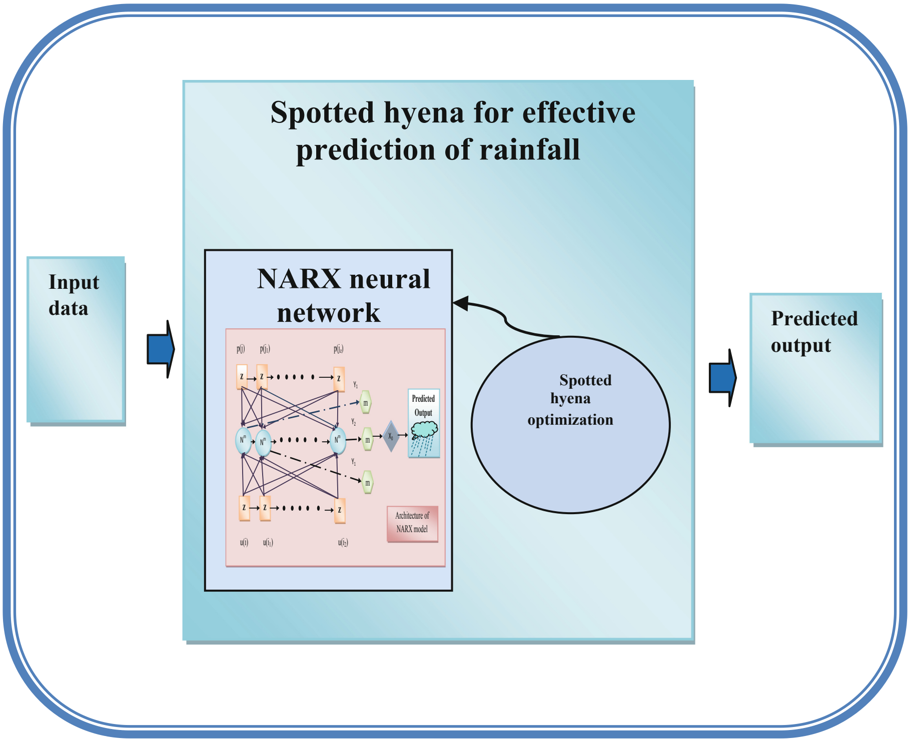

Fig. 1. Systematic representation of rainfall prediction model

3 Proposed Prediction Model Using the Optimization Enabled Neural Network Classifier

The main aim of the research is to develop a rainfall prediction model that effectively predicts the rain in a particular region in a periodic manner either monthly or yearly. The input data obtained from a particular region on a specific time is trained using the NARX neural network and the optimization of the neural network is performed using the spotted hyena optimization. The NARX neural network has the capability to predict the error between the predicted data and the ground output. Finally the probability of occurrence of the amount of rainfall is predicted using this network. The representation of the proposed model is shown in Fig. 1

3.1 Hybrid Optimization for Classifier Training

Spotted hyenas are skillfull animals that perform hunting in a well-organized preplanned manner. The spotted hyena characteristics are observed as the feature, which renders the optimal solution. The optimization features of the spotted hyena are used for declaring the best solution for the classifier, which is used for tuning of the internal model parameters. The social relation and the adaptability of the spotted hyenas are taken into account for the optimization problems, which is mathematically expressed in different phases.

Surrounding Phase: Initially the hyena start searching for the prey and once the prey is detected then it starts updating the information to different hyenas. The hyena which locates near to the location of the prey is considered as the current best solution and the information updated by the hyena can be mathematically expressed as follows

$$\vec{G_a} = | \vec{P} \,.\, \vec{M_b}(v) - \vec{M}(v) \tag{1}$$

$$\vec{M}(v+1) = \vec{M_b}(v) - \vec{L} \,.\, \vec{G_a} \tag{2}$$

Here $\vec{G_a}$ represents the distance between the target prey and the spotted hyena, v denotes the current iteration. $\vec{P}$ and $\vec{L}$ are the coefficient vectors, $\vec{M_b}$ indicates the position of the prey and $\vec{P}$ indicates rthe position of the spotted hyena. || and $\cdot$ absolute value and multiplication with respect to the vectors. The coefficient vectors $\vec{P}$ and $\vec{L}$ are given as follows

$$\vec{P} = 2.h\vec{w_1} \tag{3}$$

$$\vec{L} = 2\vec{q}\,.h\vec{w_2} - \vec{q} \tag{4}$$

$$\vec{q} = 5 - \left(iteration * {}^{5}\Big/ \max_{iteration} \right) \tag{3}$$

Here $iteration = 1, 2, 3 \ldots\ldots \max_{iteration}$. While performing the maximum number of iterations the $\vec{q}$ linearly decreases from the value 5 to 0. $h\vec{w_1}$ and $h\vec{w_2}$ are the random vectors has the value [0, 1] in the 2-dimensional environment.

Chasing Phase: The spotted hyenas always hunt in group with trusted members. The hyenas know the location of the prey where the prey is present. The hyena which knows the location of the prey is termed as optimal solution and all the other solutions are also updated. The hyenas started updating their positions can be represented by

$$\vec{G_a} = | \vec{P} . \vec{M_a} - \vec{M_g} | \tag{4}$$

$$\vec{M_g} = \vec{M_a} - \vec{L} . \vec{G_a} \tag{5}$$

$$\vec{E_a} = \vec{M_g} + \vec{M_{g+1}} + ... + \vec{M_{g+k}} \tag{6}$$

$\vec{M_a}$ defines the position of the first hyena who spotted the prey, $\vec{E_a}$ indicates the other spotted hyenas. Here K indicates the number of spotted hyenas are computed as follows

$$k = count_{numbers}\left(\vec{M_a}, \vec{M_{a+1}}, \vec{M_{a+2}} + ...\left(\vec{M_a} + \vec{J}\right)\right) \tag{7}$$

Here $\vec{J}$ refers to the random vector that has the values ranges from [0.5, 1], $count_{numbers}$ denotes the number of solutions and the count of all candidate solution after summing with $\vec{J}$, which is similar to the far best solution in the provided search space. $\vec{E_a}$ consists of number of optimal solutions in the group of clusters.

Bombarding Phase: When the value of $\vec{q}$ decreases the attacking of the prey is initiated. The vector $\vec{L}$ also gets decreased from 5 to 0. If the condition $L < 1$ is satisfied the hyena starts attacking the prey and is mathematically expressed as

$$\vec{M_{v+1}} = \frac{\vec{E_a}}{K} \tag{8}$$

$\vec{M_{v+1}}$ saves the best solution and updates the position of other search agents according to their positions.

Scouting Phase: The search of the prey for the hyenas always depends on the $\vec{E_a}$. They diverge from each other to search the prey and unite to attack the prey depending upon the vector $\vec{L}$ ranges between -1 and 1. $\vec{P}$ is a random vector that assigns weight to the prey which makes the searching more efficient.

3.2 Architecture of the Proposed Hybrid Optimization Enabled NN Classifier

The significant network used in the prediction for non-linear time series is NARX neural network. It is a recurrent feed forward network that has effective learning rate, and converges faster to the solution. It also enchants multi-layer network, recurrent loop and time delay in the non-linear prediction. The prediction is based on the past values and the present input values. The architecture of the NARX network consists of the input layer, hidden layer and the output layers. The output obtained from the NARX rainfall prediction is mathematically expressed as,

$$p(j+1) = F\left[(p(j), \ldots\ldots, p(j_{z1}); \;\; u(i), \ldots\ldots u(i_{z2}))\right] \tag{9}$$

where, $p(j)$ is the raindrop data in the j^{th} time series and $p(j_z)$ represents the rainfall data at j_z time series. z_1 and z_2 are the delay factors employed for the prediction (Fig. 2).

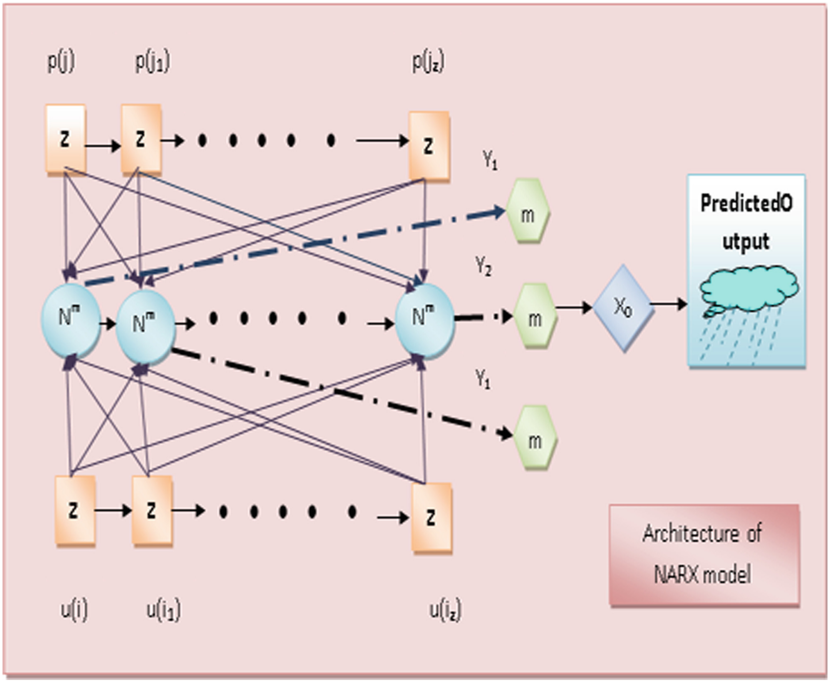

Fig. 2. Architecture of NARX rainfall prediction model

4 Results and Discussion

The results obtained using the proposed spotted hyena based optimization in the NARX model is summarized and evaluated in the following section as follows:

4.1 Experimental Setup

The research is accomplished using the software MATLAB in windows 10 operating system with 8 gb RAM. The dataset used for the execution is Rainfall in India to predict the rainfall.

4.2 Dataset Description

The dataset utilized for the experiment is collected from the National Data Sharing and Accessibility Policy (NDSAP) and the rainfall predicted from the India are taken into account from the year 1901 to 2015. For the detailed enumeration the dataset is classified into yearly, monthly and quarterly basis.

4.2.1 Performance Metrics

RMSE: Root mean square error is used to measure the variation occurred between the original image and the segmented image. When the value of RMSE is small then the performance of the segmented image will be high and is mathematically expressed as follows

$$RMSE = \frac{\sqrt{\sum_{U=1}^{X} \sum_{V=1}^{Y} [H[U, V] - I[U, V]]^2}}{U \times V} \tag{10}$$

Here X and Y represents the size of the image U and V represents the position of the pixel in the image. $H(U, V)$ is the image that is segmented and $I(U, V)$ is the original image subjected to the model.

PRD: Percentage root mean square difference is a measure used to calculate the difference between the original and reconstructed image.

$$PRD = \sqrt{\frac{\sum_{z=0}^{t-1} \left[I(U, V) - \hat{I}(U, V) \right]^2}{\sum_{Z=0}^{t-1} [I(U, V)]^2}} \tag{11}$$

Here $I(U, V)$ represents the original image and $\hat{I}(U, V)$ is the reconstructed image and t represents the number of samples.

4.3 Comparative Methods

The methods used to compare the proposed SH-NARX network model are LM-NARX, Linear regression modeling, and RBF-HPSOGA. The values are measured for varying sizes but for the simplified version several values are enumerated in the below sections.

4.3.1 Comparative Analysis

Analysis Based on RMSE Values: Initially the RMSE values are measured based on the size of the training data in an yearly basis and are shown in Fig. 3a.The RMSE value for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training size of 0.65 are 0.054645613, 0.131040967, 0.053178225, 0.006395029 respectively. Similarly for the training size of 0.7 for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model are 0.054386483, 0.128716118, 0.03784572, and 0.005845776 respectively. Furthermore the RMSE values of the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training percentage 0.8 are 0.05399651, 0.119272397, 0.02033044, and 0.003127474 respectively.

The RMSE values for the methods in a monthly basis is measured and shown in Fig. 3b. The RMSE value for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training size of 0.55 are 1.567359981, 18.07018917, 17.05821744, and 0.108860232 respectively. Similarly the RMSE value of the data under the training size 0.45 for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model is 2.520388226, 31.2543726, 30.70843581, 0.138727057 respectively. Correspondingly the values of the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training size 0.45 are 2.071838111, 26.768377, 23.2582918, and 0.126290269 respectively.

The RMSE values in a quarterly manner are measured and exhibited in the Fig. 3c. The RMSE value of the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training size 0.5 are 0.110449739, 0.242172466, 0.556730604, and 0.013896454 respectively. Similarly for the training size of 0.65, the RMSE values are measured for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model and are enumerate as 0.074979355, 0.155438121, 0.220683831, and 0.011657799 respectively. Likewise the RMSE values for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training size 0.8 are listed as 0.057285533, 0.127414519, 0.107858025, and 0.009971462 respectively.

Analysis Based on PRD Values: The PRD values are also measured for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model and are revealed in Fig. 4a. Initially the values are measured for the yearly basis for varying training sizes of data and the values are enumerated. The PRD values for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training size 0.7 are 5.272694952, 9.619634919, 3.917220226, and 1.535235903 respectively. In a similar manner the values of the training size 0.6 for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model are 5.296588236, 9.765313608, 5.145555559, and 1.65364854 respectively. Similarly the PRD values for the training size 0.45 are 5.331479133, 10.01870928, 9.247074244, and 1.869263315 respectively.

Secondly the PRD values are measured in a monthly basis and are shown in Fig. 4b. The PRD values for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training size of 0.8 are 3.197236148, 17.74855856, 17.20523995, and 0.20707524 respectively. For the training size of 0.4 of the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model is given by 3.464095582, 185.7834588,

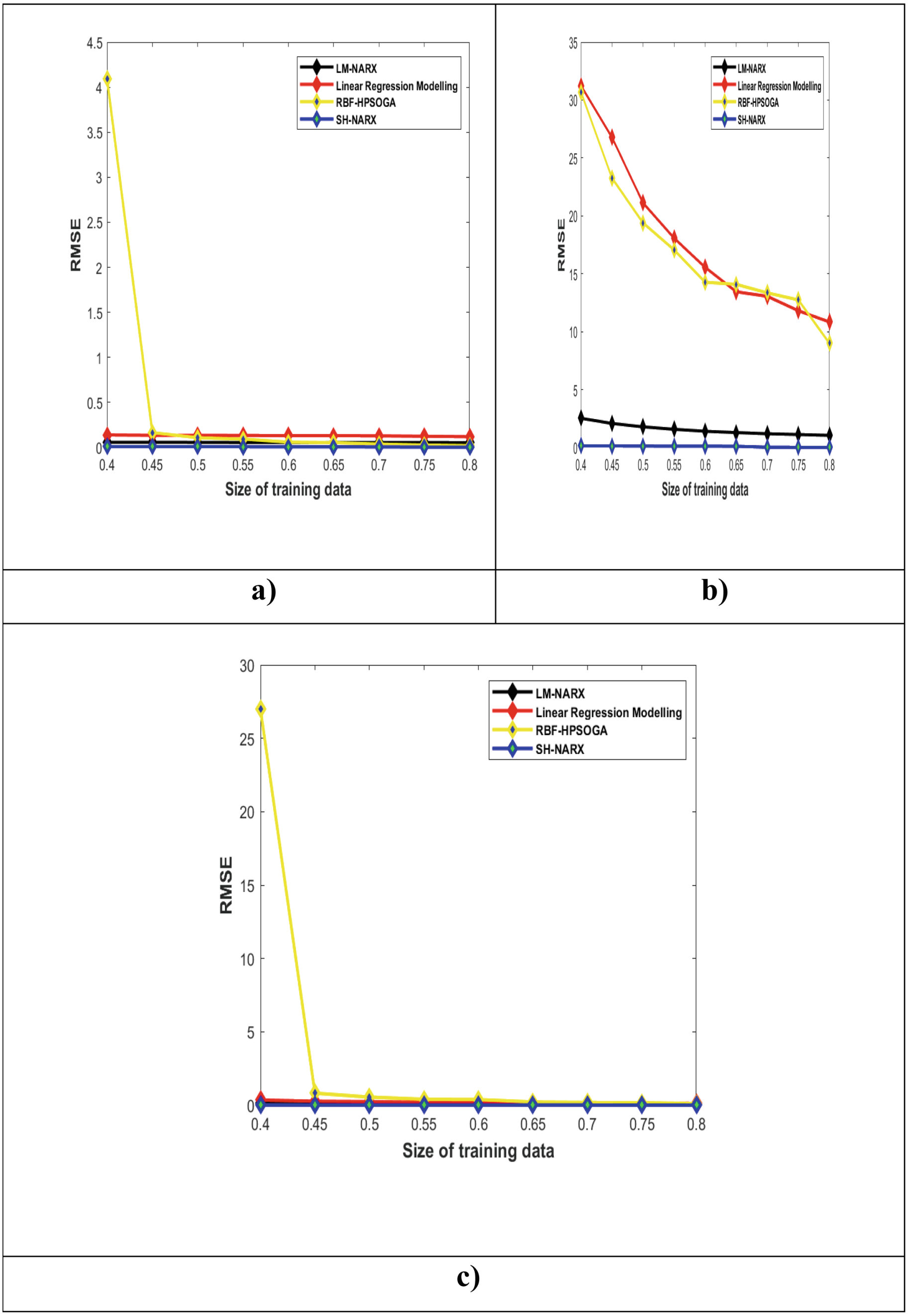

Fig. 3. Analysis based on the RMSE values in the rainfall prediction model a) yearly basis b) monthly basis c) quareterly basis

138.234133, and 0.954183783 respectively. For the training percentage 0.6 for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model the values are enumerated as 3.312602169, 41.12819639, 33.25213038, and 0.868705142 respectively.

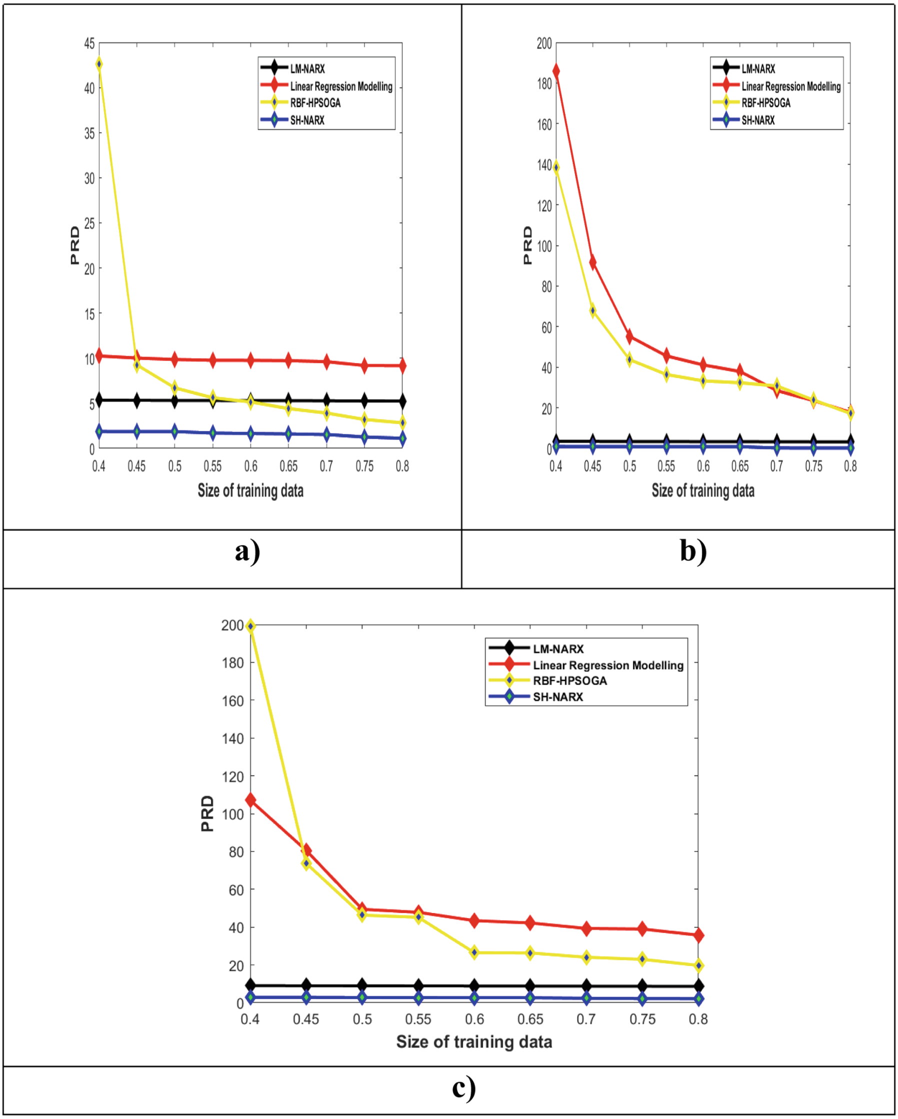

Fig. 4. Analysis based on the PRD values in the rainfall prediction model a) yearly basis b) monthly basis c) quarterly basis

Finally the PRD values are measured in a quarterly basis and are depicted in Fig. 4c. The PRD values for the training size of 0.8 for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX methods are 8.694244079, 35.76247282, 19.7700456, and 2.201340922 respectively. Similarly the PRD values for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model for the training size 0.65 are given by 8.824927992, 42.2069493, 26.35117132, and 2.660116499 respectively. Similarly the PRD values for the training size of 0.7 are given by 8.781661461, 39.29613037, 24.07781106, 2.387528763 for the methods LM-NARX, Linear regression modeling, RBF-HPSOGA and the proposed SH-NARX prediction model respectively.

5 Conclusion

Rainfall prediction is an important factor because it has the capability to cause destruction. To prevent this, the rainfall is predicted using the proposed SH-NARX prediction model which will be useful in the early prediction of the rain fall. The NARX model effectively predicts the rainfall based on the previous output from the previous month or year and improves the recognition accuracy. The effectiveness of the model is enhanced by the spotted hyena optimization and the experimental analysis using PSD and RMSE shows that the proposed method is more efficient. For more effectiveness the model can be tested on not only quarterly data but also on fortnight data also.

References

1. Venkatesh, R., Balasubramanian, C., Kaliappan, M.: Rainfall prediction using generative adversarial networks with convolution neural network. Soft. Comput. **25**(6), 4725–4738 (2021). https://doi.org/10.1007/s00500-020-05480-9
2. Pham, B.T., et al.: Development of advanced artificial intelligence models for daily rainfall prediction. Atmos. Res. **237**, p. 104845 (2020)
3. Tran Anh, D., Duc Dang, T., Pham Van, S.: Improved rainfall prediction using combined pre-processing methods and feed-forward neural networks. J **2**(1), 65–83 (2019)
4. Esteves, J.T., de Souza Rolim, G., Ferraudo, A.S.: Rainfall prediction methodology with binary multilayer perceptron neural networks. Clim. Dyn. **52**(3), 2319–2331 (2019)
5. Haq, D.Z., et al. Long short-term memory algorithm for rainfall prediction based on El-Nino and IOD data. Procedia Comput. Sci. **179**, 829–837 (2021)
6. Abdul-Kader, H., Mohamed, M.: Hybrid machine learning model for rainfall forecasting. J. Intell. Syst. Internet Things **1**(1), 5–12 (2021)
7. Dhamodharavadhani, S., Rathipriya, R.: Region-wise rainfall prediction using MapReduce-based exponential smoothing techniques. In: Peter, J., Alavi, A., Javadi, B. (eds.) Advances in Big Data and Cloud Computing. AISC, vol. 750, pp. 229–239. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1882-5_21
8. Wu, C.L., Chau, K.W., Li, Y.S.: Methods to improve neural network performance in daily flows prediction. J. Hydrol. **372**, 80–93 (2009)
9. Sang, Y.F.: A review on the applications of wavelet transform in hydrology time series analysis. Atmos. Res. **122**, 8–15 (2013)
10. Sahai, A.K., Soman, M.K., Satyan, V.: All India summer monsoon rainfall prediction using an artificial neural network. Clim. Dyn. **16**(4), 291–302 (2000)

11. Trinh, T.A.: The impact of climate change on agriculture: findings from households in Vietnam. Environ. Resour. Econ. **71**(4), 897–921 (2018)
12. Le, L.M., et al.: Development and identification of working parameters for a lychee peeling machine combining rollers and a pressing belt. AgriEngineering **1**(4), 550–566 (2019)
13. Abbot, J., Marohasy, J.: Skilful rainfall forecasts from artificial neural networks with long duration series and single-month optimization. Atmos. Res. **197**, 289–299 (2017)
14. Navone, H.D., Ceccatto, H.A.: Predicting Indian monsoon rainfall: a neural network approach. Clim. Dyn. **10**(6–7), 305–312 (1994)
15. Davolio, S., Miglietta, M.M., Diomede, T., Marsigli, C., Morgillo, A., Moscatello, A.: A meteo-hydrologicalprediction system based on a multi-model approach for precipitation forecasting. Nat. Hazards Earth Syst. Sci. **8**, 143–159 (2008)
16. Diomede, T., et al.: Discharge prediction based on multi-model precipitation forecasts. Meteorol. Atmos. Phys. **101**, 245–265 (2008)