water

MDPI

Article

Hybridizing Artificial Intelligence Algorithms for Forecasting of Sediment Load with Multi-Objective Optimization

Arvind Yadav [1], Marwan Ali Albahar [2], Premkumar Chithaluru [3,4,*], Aman Singh [5,6], Abdullah Alammari [7], Gogulamudi Vijay Kumar [8] and Yini Miro [4,9]

1 Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522302, India
2 Department of Computer Science, Umm Al-Qura University, Mecca P.O. Box 715, Saudi Arabia
3 Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad 500075, India
4 Department of Project Management, Universidad Internacional Iberoamericana, Campeche 24560, Mexico
5 Higher Polytechnic School, Universidad Europea del Atlántico, 39011 Santander, Spain
6 Uttaranchal Institute of Technology, Uttaranchal University, Dehradun 248007, India
7 Faculty of Education, Curriculums and Teaching Department, Umm Al-Qura University, Makkah P.O. Box 734, Saudi Arabia
8 Department of CSE, Amrita Sai Institute of Science & Technology, Paritala 521180, India
9 Department of Project Management, Universidad Internacional Iberoamericana, Arecibo, PR 00613, USA
* Correspondence: bharathkumar30@gmail.com

Abstract: Forecasting of sediment load (SL) is essential for reservoir operations, design of water resource structures, risk management, water resource planning and for preventing natural disasters in the river basin systems. Direct measurement of SL is difficult, labour intensive, and expensive. The development of an accurate and reliable model for forecasting the SL is required. Sediment transport is highly non-linear and is influenced by a variety of factors. Forecasting of the SL using various conventional methods is not highly accurate because of the association of various complex phenomena. In this study, major key factors such as rock type (RT), relief (R), rainfall (RF), water discharge (WD), temperature (T), catchment area (CA), and SL are recognized in developing the one-step-ahead SL forecasting model in the Mahanadi River (MR), which is among India's largest rivers. Artificial neural networks (ANN) in conjunction with multi-objective genetic algorithm (ANN-MOGA)-based forecasting models were developed for forecasting the SL in the MR. The ANN-MOGA model was employed to optimize the two competing objective functions (bias and error variance) with simultaneous optimization of all associated ANN parameters. The performances of the proposed novel model were finally compared to other existing methods to verify the forecasting capability of the model. The ANN-MOGA model improved the performance by 12.81% and 10.19% compared to traditional AR and MAR regression models, respectively. The results suggested that hybrid ANN-MOGA models outperform traditional autoregressive and multivariate autoregressive forecasting models. Overall, hybrid ANN-MOGA intelligent techniques are recommended for the forecasting of SL in rivers because of their relatively better performance as compared to other existing models and simplicity of application.

Keywords: multi-objective-based genetic algorithm; water discharge; artificial neural network; sediment load; Mahanadi River

Citation: Yadav, A.; Ali Albahar, M.; Chithaluru, P.; Singh, A.; Alammari, A.; Kumar, G.V.; Miro, Y. Hybridizing Artificial Intelligence Algorithms for Forecasting of Sediment Load with Multi-Objective Optimization. *Water* **2023**, *15*, 522. https://doi.org/10.3390/w15030522

Academic Editor: Ian Prosser

Received: 25 November 2022
Revised: 5 January 2023
Accepted: 16 January 2023
Published: 28 January 2023

1. Introduction

Hydrology deals with the efficient utilization of water resources and their management through the use of hydrological forecasting. Forecasting sediment load (SL) is an important concern in water resource management, and it is essential to know information about reservoir operations, water resource planning, flooding, water pollution control, reservoir design in rivers, and risk management as well as for preventing natural disasters [1,2].

Moreover, SL forecasts are required for decision making and policy formulation in a variety of sectors such as hydropower, flood control, reservoir sedimentation, and reservoir operations [3]. Knowledge of the amount of SL in a river at a particular time can lead to better knowledge of flood potential and consequently help control over-bane flooding. The amount of accumulated sediment in a reservoir is a crucial factor in determining the reservoir's service life [4]. As a result, it is becoming increasingly important to measure SL, but it is challenging due to the complexity and non-linearity of the interactions of the various controlling factors with SL.

Jansson [5] demonstrated the effect of river basin features such as geology, storage capacity, soil, and relief (R) on SL. Syvitski et al. [6] revealed that long-term sediment discharge is associated with basin area and basin R, which was employed to examine the effects of climate change globally on the flux of sediment of the world's rivers. The SL in the river is caused by the physical as well as chemical weathering of the soil and rock within the basin. The catchment area (CA) affects SL due to variations in the properties of the catchment such as the capacity of storage and gradient [7]. In most of the world's major rivers, a good non-linear association between SL and water discharge (WD) has been found [8,9]. River sediment transportation and generation are significantly influenced by climatic variables such as temperature (T) and rainfall (RF) [10]. To design the numerous SL forecasting models based on prior research, temporal data such as WD, RF, SL, and T, as well as spatial data such as rock type (RT), R, and CA were taken into consideration.

The data-driven and physically based models are used for forecasting water resources [11]. The physically based models require a lot of data and use complicated mathematical equations to incorporate the physical process. Traditional time series regression models such as Autoregressive (AR), Autoregressive Moving Averages (ARMA), and Autoregressive Integrated Moving Averages (ARIMA) have been extensively used for hydrological time series forecasting [12]. These models have a lot of required input parameters, and most traditional modeling techniques suppose the data are linear and stationary, which makes them incapable of handling non-stationary and non-linearity processes associated with hydrology [13]. There are various non-linear forecasting methods that are also used for forecasting SL but those are also not capable of forecasting the SL accurately and efficiently [13]. Artificial intelligence (AI) data-driven methods have shown promise in modeling and forecasting non-stationary and non-linear processes in water resources, as well as in dealing with huge amounts of dynamicity and noise hidden in datasets. The drawbacks of linear AR, ARMA, and ARIMA and non-linear autoregressive networks with exogenous inputs forecasting models of SL can be overcome by using an AI-based non-linear system [13,14]. Long- and short-term forecasts are critical in reservoir operational processes, which are typically planned every month. Many researchers developed AI-based SL forecasting models using various controlling factors based on monthly data [15–17].

The goal of this research is not to describe the superiority of one technique over others, but rather to demonstrate that different modeling parameters must be chosen judiciously to generate a generalized, accurate, and reliable model. The ANN is chosen as a non-linear approach amongst AI techniques because it is currently among the most popular known AI techniques. The ANN works on the principle of the biological brain and the nervous system that goes along with it. Through proper learning, the ANN is capable of identifying the complex non-linear or linear relationship between outputs and inputs data without detailed knowledge of the character of the internal structure of physical processes. The ANN can establish non-linear links among outputs and inputs and makes them flexible and useful techniques for modeling the phenomena of hydrology [18]. It is useful for modeling when the physical presence of a process is unsure, there is no mathematical form for a description of the process, and reasonable experimental data are available [18,19]. The ANN is widely used in hydrology for forecasting the RF, runoff, flood, river discharge, and sediment yield modeling, which provide better results than

the traditional regression-based forecasting models [14,20–22]. Nevertheless, the ANN models possess some drawbacks such as overfitting and underfitting and local minima problems due to inaccurately choosing the ANN model's parameters (hidden node size, network topologies, number of hidden layers, nodes in hidden layers, initial weights, etc.) using greed search or trial-and-error approaches [23,24]. These ANN parameter selection approaches require a huge amount of computing time to calculate the parameter value and do not guarantee to provide an optimum solution [24]. It is possible to implement the genetic algorithm (GA) to solve the issues with ANN. The GA algorithm is a population-based global optimization algorithm that is based on Darwin's theory of evolution and is used to find the best parameters for ANN models. It generates diversity in a population of individuals (chromosomes) by employing various genetic operators such as mutation, selection, and crossover, and then provides the best solutions. Nowadays, the GA is among the most popular algorithms for global search optimization which is hybridized with ANN to overcome ANN's drawbacks [24,25]. The concurrent optimization of related parameters in ANN models using GA has been used to overcome the shortcomings of trial-and-error procedures as shown by a variety of applications of AI models. Numerous studies demonstrate that this strategy not only reduces the amount of computation required but also yields better results [26–28].

The ANN model parameters are optimized using GA by optimizing a single objective, namely Mean Square Error (MSE), as a criterion for the evaluation of performance, which is found in various studies [29,30]. Furthermore, the performance of feed-forward back-propagation ANN is also determined by achieving the lowest possible MSE [31]. The drawbacks of using a single objective (MSE) for the optimization of parameters of ANN have been well documented in past research [31,32]. The MSE is a summation of error variance and squared mean error (bias). MSE minimization does not guarantee that both error variance and bias are minimized. There is a conflict between bias (mean error) and variance, both of which affect ANN performance (MSE) [32]. In influencing MSE of estimation or forecasting, bias and variance contradict each other, resulting in significant uncertainty in evaluation. If one of them is reduced, then the other is increased [33]. Thus, the model selection issue can be viewed as a multi-objective optimization issue. As a result, minimizing both components is critical for attending generalization to avoid overfitting or underfitting with greater accuracy. Underfitting is caused by bias, whereas overfitting is caused by variance, which limits the model's generalization ability and contributes to its poor performance [31]. The optimization and compromise of bias and variance multi-objectives result in a reliable and accurate model [34,35]. The GA is a well-suited population-based search method for problems involving multiple objectives [19,36]. The Non-Dominated Sorting GA (NSGA), Controlled Elitist GA (CEGA), and Elitist Non-Dominated Sorting GA (NSGA-II) are more familiar with GA-based multi-objective algorithms [37]. Controlled Elitist Gas are more capable of maintaining population diversity for getting to the best Pareto front by controlling the number of elites than NSGA-II and NSGA [35]. Therefore, the Controlled Elitist MOGA-based ANN model has been used to optimize these two contradicting responses (variance and bias). Recently, many works of literature are available in various fields for prediction and forecasting using multi-objective optimization for ANN and it was found that the system's accuracy could be increased using this hybrid approach [38–40]. Rosales-Pérez [33] applied multi-objective GA for the optimization of AI models such as Support Vector Machine (SVM) parameters by optimizing bias and variance. Recently, there are few studies that have been conducted on the multi-objective optimization-based model to estimate the sediment load or suspended sediment concentration [41,42].

In this study, forecasting is a term that is used when inputs are past monthly time series of temporal data such as Q, RF, and T, as well as spatial data such as RT, R, and CA, and output data are the SL. Thus, this paper deals with the forecasting of the SL using the past observed SL data and other observed variables (RF, WD, and T) with spatial data (RT, CA, and R). This study was conducted in the MR system. Various researchers

have used monthly one-step-ahead forecasting studies in hydrology [43–45]. In this study, multi-objective genetic-based ANN forecasting models are developed for one-step-ahead of forecasting the SL in the MR basin with optimization of all ANN parameters (hidden layer neurons, combination coefficient, transfer function, inputs, bias weights, and connection weights) simultaneously.

Flood forecasting in the MR is studied by Kant et al. [46] with the use of a multi-objective evolutionary Neural Network (NN) and bootstrap NN. As per the author's knowledge, no researcher has yet attempted to develop a fully automatic, highly generalized, globally single hybrid AI-based forecasting model. In this study, a fully automated parameter tuning and highly generalized AI forecasting model is developed for forecasting of SL in the Mahanadi River, which reduced the need for human intervention. The proposed model would replace the use of multiple models to predict the SL, which stands for suspended sediment load. This single model was applied at each gauging station to forecast the SL in the MR Basin. In this study, a single hybrid ANN-MOGA model was developed for effectively forecasting the SL at individual stations in the MR using a huge number of combinations of temporal (SL, T, RF, and WD) and spatial (RT, CA and R) data of 11 gauge stations. All parameters for the ANN model were optimized concurrently using the multi-objective GA, which included bias and variance objectives. These approaches d not only improve the model's performance, but also significantly reduce computational time by eliminating grid searches and trial-and-error exercises. The forecasting capability of hybrid models was tested by comparing their performances to traditional Multivariate Autoregressive (MAR) and Autoregressive (AR) methods. It was revealed that the best accuracy was provided by the multi-objective GA-based ANN model with more generalization and it is the most suitable substituent among other comparative methods for forecasting the SL. If SL measurement is not possible, then approaches for multi-objective GA-based ANN modeling can be recommended for forecasting SL due to their ease of implementation and relatively better performance than other existing methods.

2. Study Area

To make the SL forecast, the MR basin was chosen. Flowing to the east, this river is a major waterway in the Indian peninsula (Figure 1). It is the fourth largest river in India and drains an area of 141,589 km^2 or approximately 4.3% of India's total land area [46]. Odisha receives 53% of the river's basin area, while Chhattisgarh receives 46% and Maharashtra, Madhya Pradesh, and Jharkhand share the remaining 1% [46]. Until it enters the Bay of Bengal, the river flows for a total of 851 km. Thr MR was located between 19°20′ and 23°35′ north, and 80°30′ and 86°50′ east. The MR contains the Hirakud dam which is the world's largest earthen dam. In terms of current sediment load, the MR is second among Indian peninsular rivers. Figure 1 shows the MR basin elevation map and the locations of all 11 hydro-climatological sites. The average annual RF was between 1200 and 1400 mm [47]. Approximately 90% of the yearly RF that the MR basin receives occurs during the monsoons. The MR basin has a dispersed pattern of RF strength. In the MR basin, the warmest months are April and May, with summer temperatures of 39 to 45 °C, and the coldest months are December and January, with winter temperatures of 4 to 12 °C [47]. The two largest bodies of water in the MR are Lake Chilka and the Hirakud Dam.

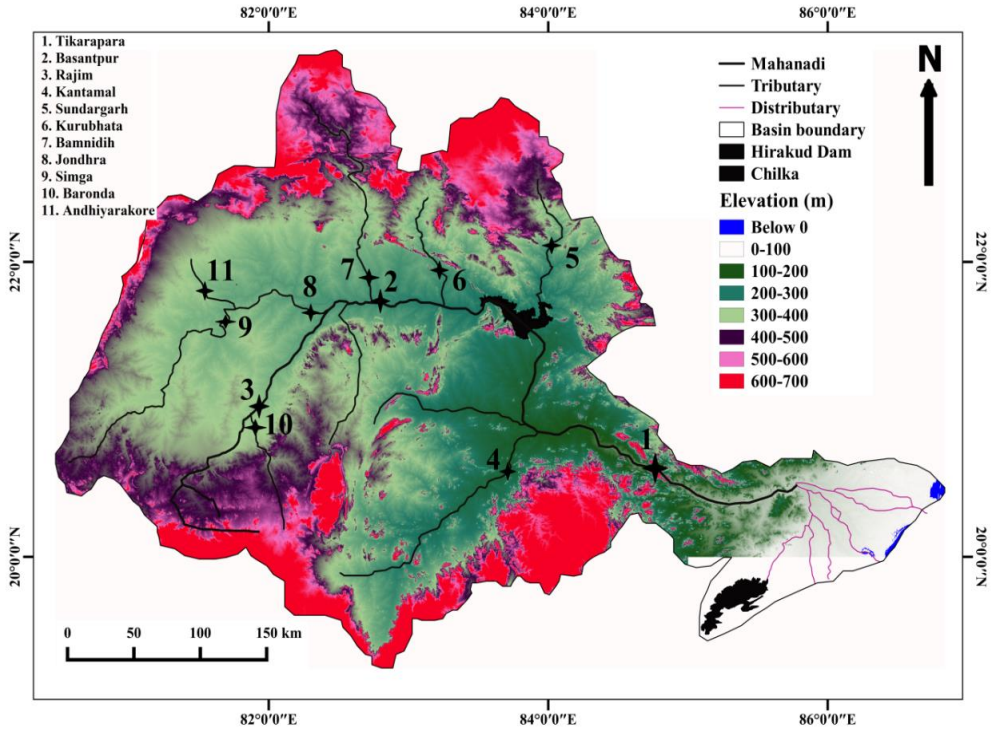


Figure 1. The elevation map of the Mahanadi River basin with the geographical location of 11 gauging stations [48].

3. Methodology and Data

The preliminary processing includes data normalization that was proposed before developing the models. To maintain uniqueness while developing a model, data normalization is used to remove the ranges between the datasets. It revealed convergence and fast processing throughout training and minimized forecasting errors [49].

Data standardization is another name for the process of normalizing numerical data, which results in much more highly accurate network training. The range of all variables is fixed at 0 to 1 by the normalization process which is described briefly in different literature [24,48].

Data normalization's primary goal is to remove the various ranges and dimensions of the variables included in the dataset. The normalization process of the data in the range of a and b is performed using Equation (1):

$$C_{norm} = a + \frac{C_i - C_{min}}{C_{max} - C_{min}} \times (b - a) \quad (1)$$

where C_i is the ith actual value, C_{norm} is the normalized value of C_i, C_{max} is the highest value, and C_{min} is the lowest value of the dataset. In this case, "a" represents the lowest value, and "b" represents the highest value of normalized data.

The used data consist of monthly RF, T, WD, and SL during the years 1990–2010 and spatial variables such as R, CA, and RT of eleven gauging stations in the MR for developing the proposed models. Figure 1 depicts the locations of all these stations. The dataset from the individual station is partitioned: training data (70%) are used to develop the models; validation data (15%) are used to avoid model overfitting, while testing data (15%) are used to evaluate the model's performance in a testing phase. Data from tests are regarded as "unseen" and "not used" in the process of modeling. Single testing, training, and validation for the MR basin were eventually produced by combining the data from all 11 stations. In this study, the forecasting of SL is performed using an ANN with a Multi-Layer perceptron (MLP) feed-forward using a Levenberg–Marquardt (LM) backpropagation algorithm. A wide variety of weight optimization strategies can be utilized during the training process

of the MLP-based ANN; however, the LM training algorithm trains more quickly than the gradient descent training algorithm and achieves convergence more quickly [50]. The weight updating rule of ANN is presented as [24,50]:

$$W_{k+1} = W_k - \left(J^T J\right) + \mu I)^{-1} J^T e \quad (2)$$

in which J is a Jacobian matrix, e is an error matrix, I represents an identity matrix, W is the weight of the ANN, and μ represents the combinational coefficient of LM, which plays an important role in the learning process of LM in an ANN. The flow chart of the ANN-MOGA method that has been proposed is shown in Figure 2.

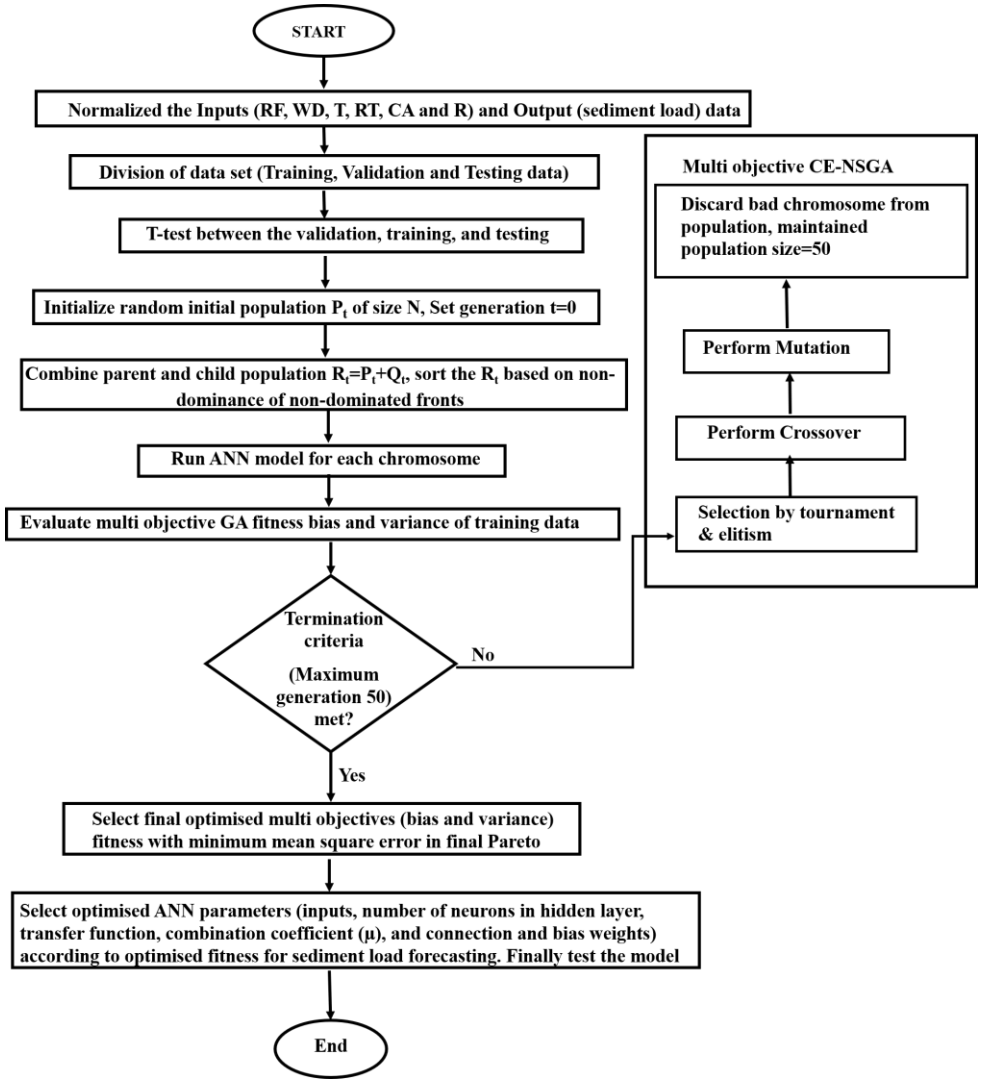


Figure 2. Flow chart of proposed hybrid ANN-MOGA model for SL forecasting.

Numerous artificial intelligence models have been successfully applied by various researchers for the prediction and forecasting in water resources [51–55]. The details description of MLP and LM training algorithms of ANN are discussed by various researchers [48,56]. Numerous factors affect the effectiveness of MLP-based ANN models, including the transfer function and number of nodes in the hidden layer, and the initial weights. If any of these factors is chosen incorrectly, the ANN will be poor, and the likelihood of the solution reaching the global optimum will be low. The GA has proven to be effective in resolving the issues of the ANN [27,57]. So, this research demonstrates the use of GA in ANN to overcome the drawbacks of ANN for forecasting SL with a selection of

all ANN model parameters optimally. In this study, the output demonstrates the value at time frame t in the time series models of SL forecasting if the forecasting model's inputs cover values corresponding to points of time $t-1$, $t-2$, $t-3$,..., $t-n$. In this study, the GA is used to choose the transfer function, inputs, neurons, combinational coefficient of LM (μ), bias, and connection weights. These are the five most important ANN parameters for SL forecasting. These ANN model parameters can be found in the chromosome, which is a binary string. The input parameters are found in the first part of the chromosome.

In the second part of the chromosome, a 3-bit binary number shows the transfer function for the hidden and output layers. In this section, the transfer functions in the output layer and the hidden layer are shown. Transfer functions are available in three different types: linear, log sigmoidal and tan sigmoidal. There are nine different ways to use transfer functions for the hidden and output layers. In the third part of the chromosome, 5 bits stand for the neurons in the hidden layer. During modeling, this binary number is turned into a decimal number. This is performed to make the hidden layer neurons. Due to the complexity and cost of processing the model, hidden neurons are limited to 32. All of the decimal numbers from 1 to 32 can be shown with 5-bit chromosomes. The 4th part of the chromosome denotes the μ, which is an 8-bit binary number. The μ showed decimal values from 0 to 255, which is normalized between 0.0010 and 9×10^9 [24]. In the fifth section, the biases and weights of the connections of ANN models are shown. The length of the chromosome changes because the number of hidden neurons and the number of inputs change. The ANN-MOGA forecasting models are designed with GA parameters such as the number of generations, the size of the population, the rate of mutation, and the probability of a crossover. In this study, a uniform crossover with a high probability value (0.6) and a low probability of constant mutation (0.05) was used. The values of each chromosome's fitness are estimated using the fitness function (RMSE) of the training dataset. The maximum generations (50) were considered as stopping criteria.

Both objective functions assessed the fitness values of each chromosome for the initial population. The chromosomes were sorted using a controlled non-dominating sorting strategy. In non-dominating sorting, the population was arranged according to the non-dominance level on various fronts (referred to as levels) [36]. To determine whether the solution does not predominate in the population, the following guidelines are used:

$$\text{Bias } [i] > \text{Bias } [j] \text{ and Variance } [i] \geq \text{Variance } [j] \tag{3}$$

or

$$\text{Bias } [i] \geq \text{Bias } [j] \text{ and Variance } [i] > \text{Variance } [j],\ i \neq j \tag{4}$$

where the chromosome numbers i and j are used. The solutions of the same non-dominated front are compared to establish the solution's overall ranking after the non-dominated fronts are obtained. To compare solutions from the same non-dominated front, a crowding distance is used [36]. The final step was to obtain the solutions' overall rank using the crowded-comparison operator, which combines the crowding distance and the measure of non-dominated rank. The multi-objective GA framework iteratively improves the beginning chromosomes according to their overall rank through various genetic operational processes, such as mutation, crossover, and selection. The tournament selection approach was used to make the decision [58]. The chosen chromosomes were subjected to crossover and mutation operations. Every generation also known as an iteration was followed by the crossover operation, which involved swapping out a portion of the binary strings of the available solutions to produce better individual solutions. According to the user selected mutation rate, the mutation operation is carried out by randomly flipping bits (0–1 or 1–0) of the chromosomes in order to diversify the existing solutions and avoid trapping at local minima. A child population Q_0 of size N is produced by the mutation, selection, and crossover operators at iteration 0. The overall number of chromosomal solutions for any iteration t following the genetic operations is $R_t = P_t \cup Q_t$ becomes twice (2N). P and Q represent the parent and child populations, respectively. R represents the total number

of chromosomal solutions after the genetic operations. The objective functions of each solution (R_t) were determined by calculating, and the solutions have been ranked using the previously discussed Non-Dominated Sorting criteria and crowding distance. The rest solutions were eliminated from the solution space, and the top N solutions determined by their whole rank were chosen (referred to as elitism) for the following generation. The maximum number of individuals permitted from the i^{th} non-dominated front as shown in Equation (5) is provided diversity in the new population, based on the geometric distribution [36].

$$N \times \left(\frac{1-r}{1-r^K} \right) \times r^{i-1} \tag{5}$$

where K indicates the number of undominated fronts and r represents the reduction rate, which has a value lower than 1.

The proposed study combined the Controlled Elitist NSGA (CE-NSGA) and ANN models. The CE-NSGA is a more familiar GA-based multi-objective algorithm. The CE-NSGA has more capability to maintain the diversity of population for convergence to an optimal Pareto front by controlling elite numbers. The ANN model was used to find the two objectives such as error variance and mean error in CE-NSGA for ranking of non-dominated. The ideal parameters for the ANN model were chosen based on the most optimal solution from the final CE-NSGA generation. With these best ANN parameters, the ANN model can be applied to forecast the river system's SL with given known input parameters. Figure 2 shows the flow chart for the multi-objective GA-based ANN model. In this study, AR forecasting models are designed by the linear combination of previous data of the variable (SL). The AR model is a fundamental class of time series model.

The AR model's equation is presented below [12]:

$$SL_{t+1} = a_0 SL_t + a_1 SL_{t-1} + a_2 SL_{t-2} + \ldots + a_n SL_{t-n} \tag{6}$$

where n is the number of orders for the AR model, and a_i (i = 0, 1, 2,..., n) represents the regression model's coefficients. The MAR was designed using training datasets and the linear combination of the Autoregressive of multiple variables (WD, T, RF, and SL and spatial variables (R, CA, and RT). The MAR formula is shown below

$$\begin{aligned} SL_{t+1} = a_1 WD_t + b_1 RF_t + c_1 T_t + d_1 SL_t, + a_2 WD_{t-1} + b_2 RF_{t-1} + c_2 T_{t-1} + \\ d_2 SL_{t-1} + \ldots + a_n WD_{t-n} + b_n RF_{t-n} + c_n T_{t-n} + d_n SL_{t-n} \ + eRT + fR + gCA \end{aligned} \tag{7}$$

The linear MAR forecasting model is represented by this equation up to n lags. The a_i, b_i, c_i, d_i, e, f, and g (i = 1,2, 3, ..., n) represent the coefficients of the MAR model. The a_i, b_i, c_i, and d_i represent the coefficients of WD, RF, T and SL, respectively. The coefficients of the RT, R, and CA are represented by the values e, f, and g, respectively. The maximum lag (n) in AR and MAR model for the SL forecasting is 12, after which the cyclicity begins due to seasonal behaviour of the data. There are four temporal variables: WD, SL, RF, and T. The RT, CA and R are the spatial variables.

4. Results and Discussion

4.1. Data Analysis

The non-linear comparison between different parameters is found using Spearman rank correlation coefficients which are presented in Table 1. The Spearman rank correlation of T and SL, RF and SL, and WD and SL are represented by r3, r2, and r1, respectively. The WD and SL have a significant and high Spearman rank correlation coefficient value which is represented by r1. It is found that SL has comparatively smallest and greatest values of the Spearman rank correlation with the WD at the Kntanmal and Rajim, respectively among all gauging stations (Table 1). The Pearson rank correlation coefficient between RF and SL has a significant value (greater than 0.5 at all eleven gauging stations) which is represented by r2. It shows that RF is significantly correlated with the SL.

Table 1. Spearman rank correlation coefficient ® of hydro-climatolic data from eleven stations for the MR.

Stations	r1 (WD-SL)	r2 (SL-RF)	r3 (SL-T)
Tikarapara	0.891951579	0.667566099	0.167164223
Sundargarh	0.933162643	0.719490012	0.083275757
Simga	0.912171157	0.669144968	0.016117111
Jondhara	0.953615062	0.634788084	0.024002708
Andhiyarakhore	0.930440984	0.679483679	0.172300271
Kurubhata	0.914790531	0.739541786	0.09457768
Bamnidih	0.792975574	0.673800075	0.19294038
Rajim	0.933515452	0.652771319	−0.053703323
Kantamal	0.784858255	0.653582343	0.045884228
Baronda	0.896128553	0.718865603	0.170015607
Basantpur	0.900261237	0.717722236	0.120971863

Furthermore, it is observed from Table 1 that the Spearman rank correlation coefficient between T and SL is small and insignificant which is represented by r3. This indicates that T has an indirect effect on SL and did not directly contribute significantly to SL. It is found that the SL data have the greatest coefficient of variation value, maximum/mean, skewness, and Kurtosis among all hydro-climatic data (WD, RF, T, and SL) in the MR which indicates that the SL is more dispersed than the other parameters (WD, RF, and T), as well as extremely erratic and complicated, with a non-normal distribution in the MR basin [24]. Thus, forecasting SL by the model is very difficult as compared to the other variables.

There were significant temporal as well as spatial variations in the SL. Variations of monthly average WD, RF, T, and SL data over 20 years with spatial variation (R, RT, and CA) of each gauge station in the logarithm scale are shown in Figure 3. Due to huge variations in the dataset, we have followed the logarithm scale so that maximum data values are shown in Figure 3. It is seen in Figure 3 that the pattern of decrease in WDs and corresponding sediment load is the same sediment load, except for Kantamal station. Further, Tikarapara station indicates the highest WD, CA, R, RF, and SL, whereas Andhiyarakore station shows the lowest values of it. Furthermore, it has been noted that SL changes proportionally with WD, with the Tikarapara gauge station recording the highest values due to the highest WD amongst all gauge stations. The highest CA is also observed at this gauge station. Lowest R is found at the Bamnidih. The lowest rock value is found at Kantamal. Variations in RF, WD, T, and SL during the decades of 1990–2000 and 2000–2010 at different locations such as upstream, midstream, and downstream in the MR are given in Figure 4. The highest decrement in sediment yield is found at Tikarapara station among all stations. This may have been caused by the SL trap at the Hirakud dam, which is upstream of the Tikarapara. Higher amount of SL is also decreased at Basantpur gauging, which may be due to traping of SL at large Minimata Bango dam nearest to this station.

Since the SL, WD, RF, and T data are seasonal and available monthly, they are influenced by the data from the previous month. The maximum lag for the forecasting model must be selected. To evaluate the temporal correlations of SL, WD, RF, and T, an Autocorrelation Function (ACF) is employed. Figure 5 depicts the ACF of the SL with different time lags. The highest correlation was found at lag 1, and it decreases as the time lag increases (Figure 5). Additionally, it was noticed that ACF exhibited the next highest peak value at lag 12, supporting the seasonality of the SL dataset. The ACF plot illustrates that the highest lag is 12, after which seasonal data behavior indiciate cyclicity.

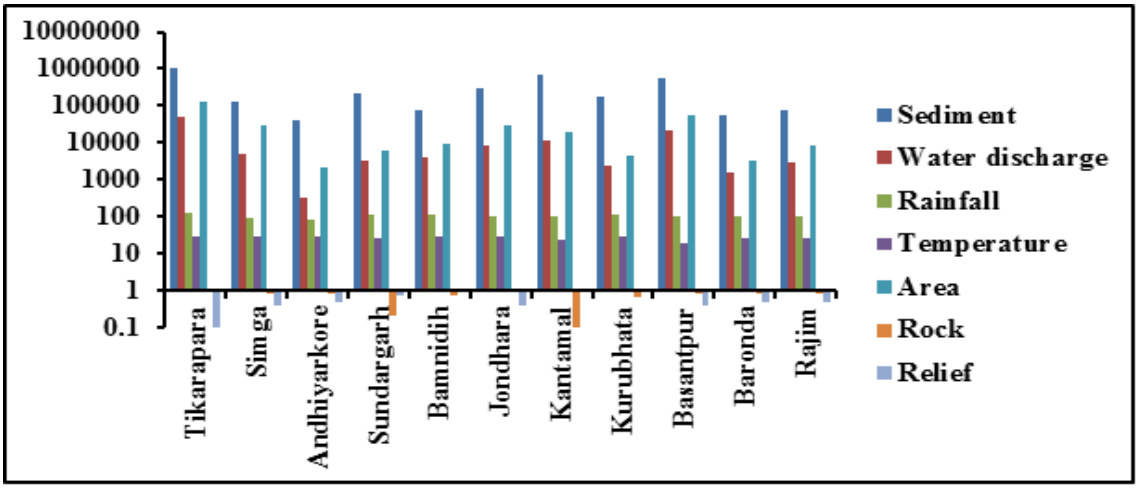


Figure 3. Spatial variation of monthly average hydro-climatic and geomorphological data such as water discharge (cms) (m^3/s), temperature (°C), rainfall (mm) and sediment load (t/ month), relief (m), and catchment area (km^2).

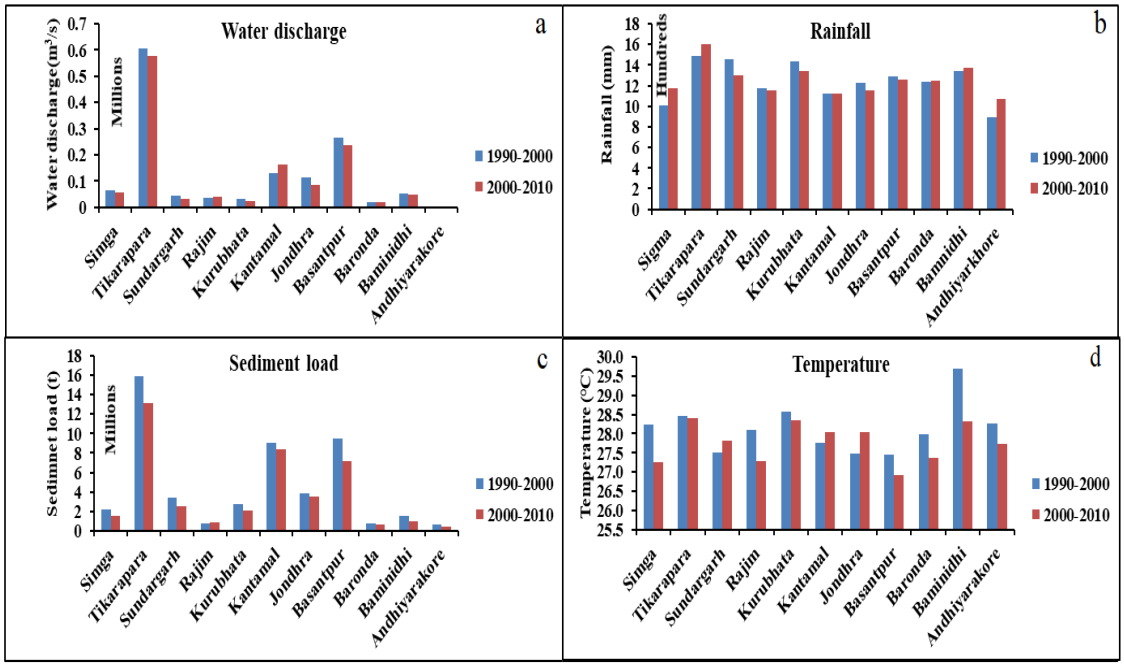


Figure 4. Comparative analysis of long-term variation over the past two decades of (**a**) WD, (**b**) RF, (**c**) SL, and (**d**) T.

4.2. ANN-MOGA Forecasting Model

A set of final solutions in the form of Pareto solutions have been provided by the model at predetermined stopping criteria that correspond to the highest number of generations (50). Figure 6a depicts the variation in bias and variance during the training stage. In the training stage, MSE (0.00352) and its subsequent variance value (0.000651) and bias (0.00424) were found to be the best Pareto solution. Figure 6b shows the variation in crowding distance among individuals. Figure 6c displays the rankings of the individual. Figure 6d illustrates the Pareto average spread variation with generation.

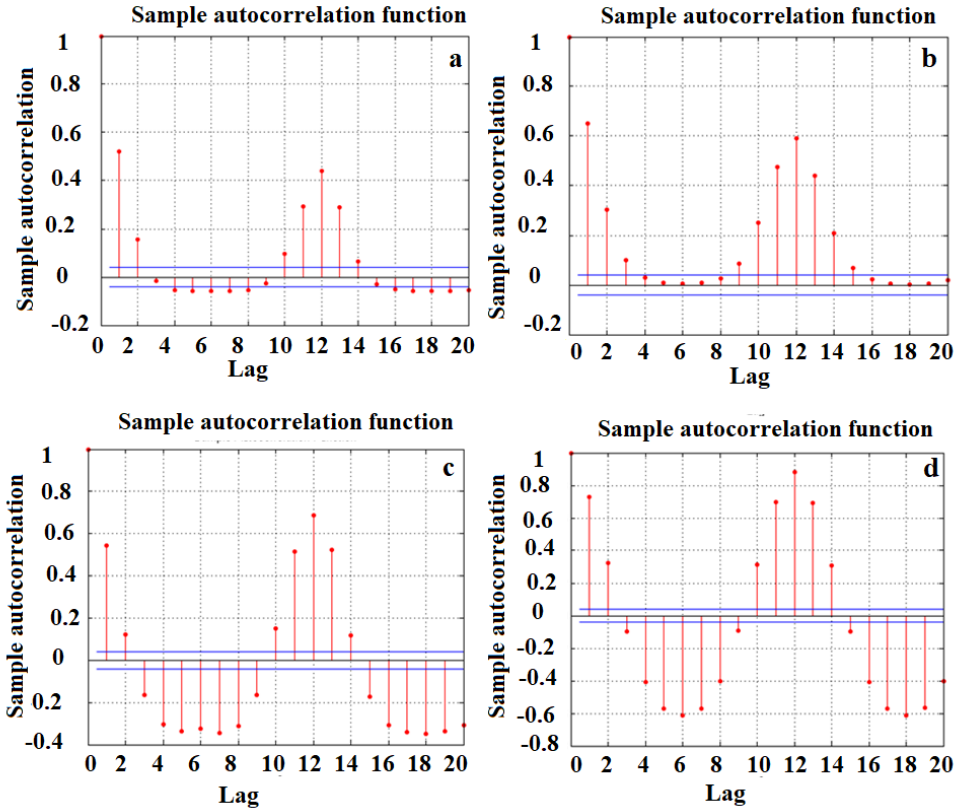


Figure 5. The autocorrelation function (ACF) for monthly hydro-climatical data (**a**) SL, (**b**) WD, (**c**) RF, (**d**) T.

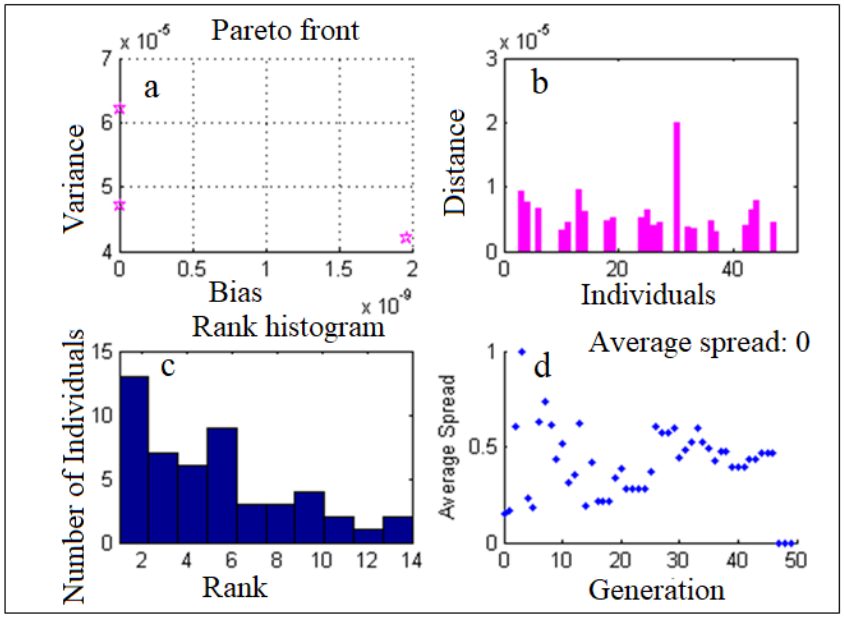


Figure 6. (**a**) Dilemma between the bias and variance objectives. (**b**) Variation of individuals and distance. (**c**) Variation of individuals and corresponding ranking. (**d**) Generation-wise variation of average spreads of Pareto.

Table 2 lists the various types of models with lag input variables. The ANN-MOGA-51 models are built using previous time series of temporal variables (WD, R, SL, and T) with 12 lags for each variable and 3 spatial variables (CA, RT, and R). The ANN-MOGA-48 models are built on prior time series temporal data (WD, R, SL, and T) with 12 lag for each variable and without considering the spatial data. The ANN-MOGA-15 model is developed by 15 input variables (12 lag time series inputs of SL and 3 spatial factors) for forecasting one-step-ahead of SL.

Table 2. Various multi-objective GA-based ANN models using different input variables with lags.

Models	Number of Initial Inputs	Input Parameters
ANN-MOGA-51	51	SL, WD, RF, T, RT, R and CA
ANN-MOGA-48	48	SL, WD, RF and T
ANN-MOGA-15	15	SL, RT, R and CA
ANN-MOGA-12	12	SL

The ANN-MOGA-12 models are established using previous time series SL only data with 12 lags. The lag for each temporal data (RF, T, WD, and SL) is twelve. All of these models' performances are compared to assess the forecasting capability of models as per statistical error analysis. The statistical error is used as an evaluation metric, and the performances of the different models are compared to one another to determine which model performs the best. Validation, training, and testing dataset are assessed to evaluate error statistics such as the Mean Square Error (MSE), the correlation coefficient (r), the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the error variance (VAR) for the forecasting models.

$$\mathrm{RMSE} = \sqrt{\left(\frac{\sum_{i=1}^{N}(O_i - F_i)}{N}\right)^2} \tag{8}$$

$$\mathrm{MSE} = \left(\frac{\sum_{i=1}^{N}(O_i - F_i)}{N}\right)^2 \tag{9}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}|O_i - F_i|}{N} \tag{10}$$

$$\mathrm{r} = \frac{\sum_{i=1}^{N}\left(O_i - \overline{O}_i\right)\left(F - \overline{F}_i\right)}{\sum_{i=1}^{N}\left(O_i - \overline{O}_i\right)\sum_{i=1}^{N}\left(F_i - \overline{F}_i\right)} \tag{11}$$

$$\mathrm{VAR} = \frac{\sum_{i=1}^{N}\left(E - \overline{E}\right)^2}{N} \tag{12}$$

where, O_i, $\overline{O_i}$, F_i and $\overline{F_i}$ are measured, measured mean, forecasted and forecasted mean values, respectively. The N value represents the number of samples. The $\overline{E}$ and E represent the mean error and error values.

All hybrid multi-objective GA-based ANN hybrid forecasting models use statistical errors obtained from testing, validation, and training datasets for one-step-ahead forecasting of the SL value, which is presented in Table 3. The statistics of all errors of testing, validation, and training data for the ANN-MOGA-12, ANN-MOGA-48, ANN-MOGA-15, and ANN-MOGA-51 models reveal that r is relatively high, while error variance, RMSE, MAE, and MSE are all low. It can be implied that these models are highly accurate at predicting SL. High levels of consistency across the three datasets show that generalized forecasting models were produced and that neither overfitting nor underfitting occurred. Among all comparative models, the ANN-MOGA-51 model stands out as the best. Table 4 lists the GA-optimized parameters chosen for the designed ANN-MOGA models. Two

and twenty-nine are the optimum combinational coefficient (μ) and neurons, respectively. Pure linear and log sigmoidal transfer functions are optimally selected at the output layer and hidden layer, respectively, in the ANN-MOGA-51 model. Twenty-nine neurons are the best number for an ANN model's hidden layer. After the evolution run, the best-fit chromosomes were used to determine the best solution for the ANN- MOGA-51 model.

Table 3. Performance comparison of the hybrid multi-objective GA-based ANN forecasting models in the testing phase.

Models	RMSE	Initially Inputs No.	MSE	MAE	VAR	r
ANN-MOGA-51	0.011639	51	0.000135	0.003802	0.000136	0.643313
ANN-MOGA-48	0.013343	48	0.000178	0.00381194	0.0001783	0.5674853
ANN-MOGA-15	0.013637	15	0.000186	0.0044	0.000186	0.513217
ANN-MOGA-12	0.01181	12	0.000139	0.003626	0.00014	0.623344

Table 4. Optimally selected parameters of ANN-MOG models.

Models	Transfer Function	Neurons	Inputs	μ
ANN-MOGA-51	Log-sigmoid, pure linear	29	22	2
ANN-MOGA-48	Tan-sigmoid, pure linear	19	21	9
ANN-MOGA-15	Pure linear, tan-sigmoid	15	10	10
ANN-MOGA-12	Tan-sigmoid, pure linear	4	6	6

Table 5 displays the error statistics for the 11 gauge stations determined with the best ANN-MOGA-51 model. It is also clear from Table 5 that there is a wide range of performance between the stations, with some stations providing good performance and others showing poor performance. The unpredictable and complex non-linear nature of SL delivers the forecasting model incapable of providing accurate forecasts of SL at the Andhiyarakhore and Baronda stations. The forecasting model's inaccuracy is a direct result of the low CA and the flat land. Located in the Raipur district of Chhattisgarh, India, Simga is the first gauging station of the MR after its origin near Nagri town and Pharsiya village. Both WD and SL are low at this station. At the Simga station, the forecasting model is providing worse results. Due to the high coefficient of variation (COV), max/mean ratio, skewness, and kurtosis value of influential parameters, many AI models were not performing well at some MR stations [36]. Forecasted and actual SL have a high degree of correlation at the Tikarapara, Kurubhata, and Jondhra gauging stations (r values greater than 0.7). This reveals the proposed model's high performance at these locations. The r value of 0.4780 for the Basantpur gauge station is not significant. Poor correlation between the WD and SL, caused by the large Minimata Bango dam upstream of this station, contributes to the proposed model's bad performance at this location. The remaining gauging stations have a significant r value between 0.5 and 0.7, indicating a moderate correlation [59].

Table 5. Error statistics of ANN-MOGA-51 forecasting model during validation, testing, and training phase.

ANN-MOGA-51	MSE	RMSE	r	Error Variance	MAE
Training	0.000241	0.015526	0.668938	0.000241	0.004677
Validation	5.25×10^{-5}	0.007243	0.7730	5.26×10^{-5}	0.002867
Testing	0.000135	0.011639	0.643313	0.000136	0.003802
Tikarapara	0.000396	0.019905	0.731051	0.000407	0.010517
Simga	1.17×10^{-5}	0.003422	0.5930	9.81×10^{-6}	0.001707
Andhiyarakhore	1.10×10^{-5}	0.003319	0.4001	9.06×10^{-5}	0.001749
Sundargarh	2.60×10^{-5}	0.005097	0.635	2.67×10^{-5}	0.002466
Bamnidih	3.33×10^{-5}	0.005769	0.695	3.27×10^{-5}	0.002765
Jondhara	3.02×10^{-5}	0.005499	0.737	2.91×10^{-5}	0.002534
Kurubhata	2.06×10^{-5}	0.001434	0.914	1.97×10^{-6}	0.000865
Basantpur	0.000229	0.015121	0.478143	0.000222	0.006185
Baronda	5.66×10^{-6}	0.002379	0.495	5.23×10^{-6}	0.001319
Rajim	2.10×10^{-6}	0.001448	0.669	2.15×10^{-6}	0.000722
Kantamal	0.000806	0.028395	0.659518	0.000801	0.01198

Figures 7 and 8 illustrate the ANN-MOGA-51 model's hydrologic graph and scatter plot, respectively. The hydrologic graph revealed that, except for the Kantamal, Andhiyarakore, Simga, and Bamnidih stations, the predicted SL corresponds to the observed SL data's variability. Andhiyarakhore is a small tributary that has a relatively small CA; however, despite its size, it carries a relatively small SL. The reason for this is that relatively small CA basins are unable to store SL and allow for the complete removal of all material that has been eroded. The presence of a large dam named Minimata Bango at Bamnidih is the primary factor that contributes to the modeled output not being accurate. Although Simga has a topography that is almost entirely composed of limestone and a relatively large catchment area, the area is relatively flat. Because of this, the sediment yield and water discharge are relatively low in comparison to those of other tributaries, such as Seonath and Tel, which have a smaller catchment area. Further, its complex non-linear erosion and transportation process of sedimentation resulted in the poor performance of the proposed model at some gauging stations. The high skewness, Kurtosis, coefficient of variation (COV) and maximum/mean value of suspended sediment load indicate its complex and highly erratic behaviours as compared to other variables (water discharge, rainfall and temperature). Thus, estimation of suspended sediment load through mathematical models is very difficult comparative to other variables [24]. The highest coefficient of variation of rainfall is found at Kantamal. This could also explain the corresponding water discharge and the SL dataset's wide variation and non-normal distribution. Similarly, the ANN-MOGA-51 results are nearer to the bisector line which is also known as the 45° line, except for the four gauging stations that were mentioned earlier (Figure 8). The scatter plots and hydrographs show that the magnitudes and medium, high, and low SL forecasted values generated by the best ANN-MOGA-51 forecasting model are also fairly close to the corresponding actual SL values. The ANN-MOGA-51 model displayed a positive value SL at each of the 11 gauge stations, although the SL output was either 0 or very near zero (Figures 7 and 8). Based on these findings, it was determined that the application of ANN in conjunction with GA is the method that yields the most accurate results for calculating SL in the MR basin system. The forecasting model provides the highest level of accuracy at the Tikarapara gauging station compared to any other station gauging station. This may be due to the location of Tikarapara, which is at the most downstream portion of the MR basin and possesses the highest WD, CA, RF, and SL of any of the gauging stations [24,47].

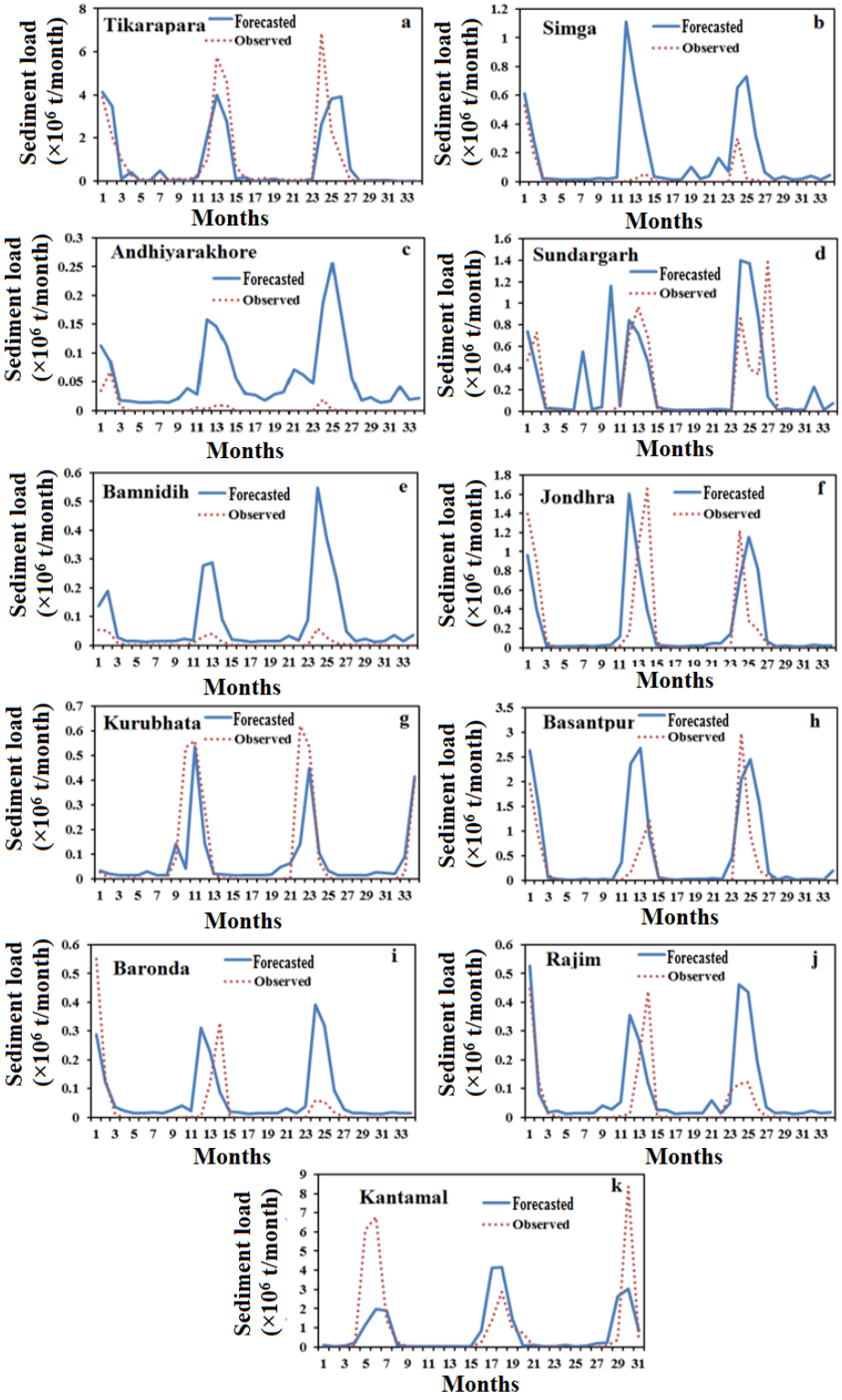


Figure 7. Comparison of the actual and forecasted SL during the testing phase of the ANN-MOGA-51 forecasting model (**a**–**k**).

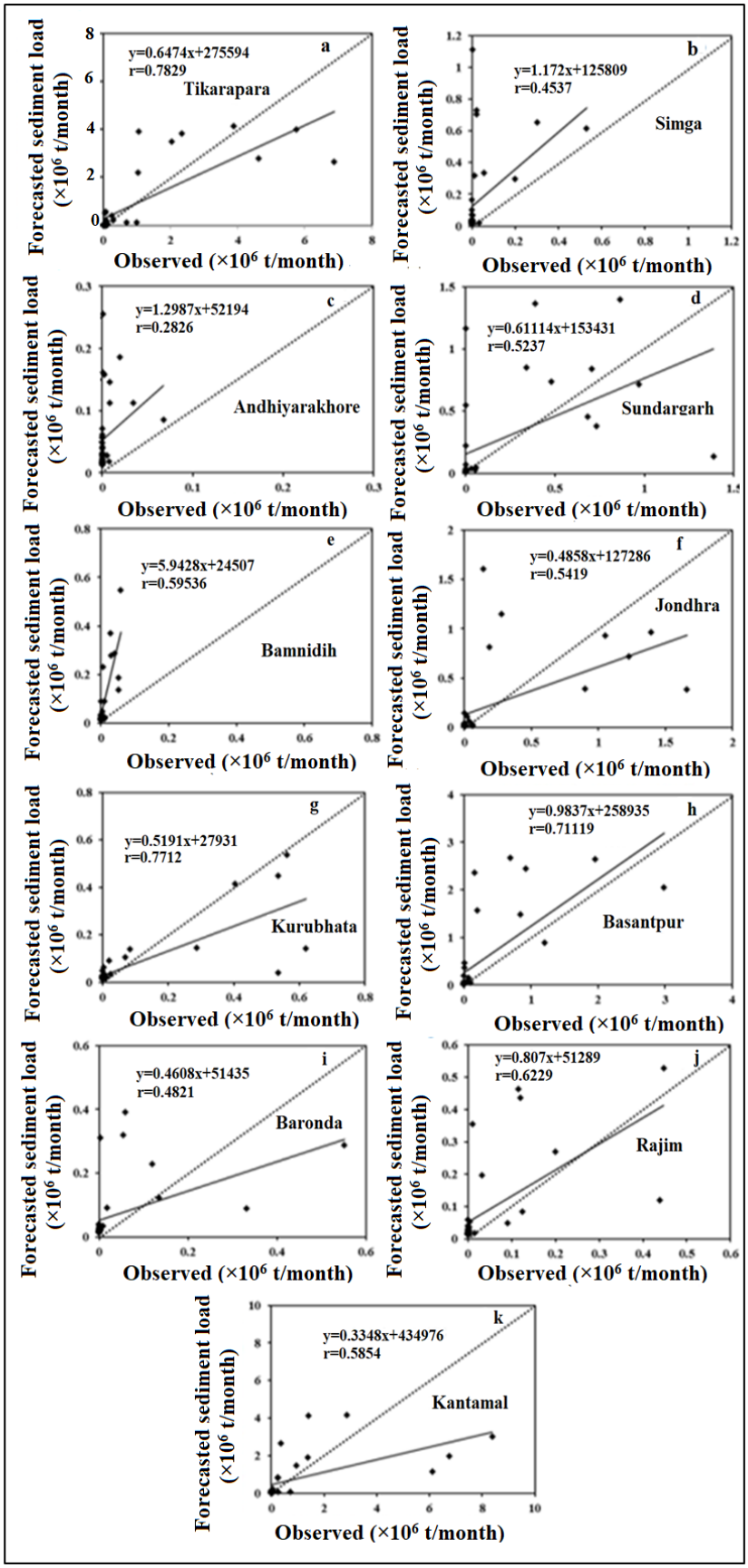


Figure 8. Scatter plot of the actual and forecasted SL of the ANN-MOGA-51 forecasting model during the testing phase (**a**–**k**).

4.3. AR Forecasting Model

In an autoregression model, the SL forecast is generated through a linear combination of the SL time series data. In this study, AR models are developed by the linear combination of previous data of the variable (sediment yield). The AR model is a fundamental class of time series model. It predicts future values by adding up the weighted sums of lagged past data. Various AR models were developed by different input parameter selections with different autoregression of the variable and compared the performances of each other. In this method, the maximum lag selection is considered 12 due to seasonal variation of data and using ACF.

The error statistics of the AR model are given in Table 6. The test dataset's low MAE and RMSE, as well as its high *r* value, revealed that the AR forecasting model could justifiably fit the data. The training dataset revealed a similar pattern of behavior, which is not surprising given that the linear model will never be overfitted to the training data. It reveals that the RMSE, MSE, and MAE of training and testing data are trending in the same direction because these are related to one another in a direct proportion, as expected for the linear model. It is observed that this model does not provide satisfactory performance at various gauging stations. The r values are not significant at various gauging stations. Poor correlation is found in all gauging stations except Tikarapara, Kurubhata, and Bamnidih. It is seen in the hydrograph (Figure 9) and scatters plot (Figure 10) that the AR model generates a greater percentage of negative SL values at low SL values as compared to other models. The AR provided the best result at Tikarapara similar to MAR and ANN-MOGA-51 models.

Table 6. Error statistics of the autoregressive (AR) forecasting model at different gauging stations.

AR	RMSE	MSE	MAE	VAR	r
Training	0.01640	0.00027	0.00473	0.00027	0.61268
Testing	0.01335	0.00018	0.00365	0.00018	0.49670
Tikarapara	0.02576	0.00066	0.01119	0.00066	0.51087
Simga	0.00185	3.42×10^{-6}	0.00069	3.49×10^{-06}	0.19322
Andhiyarakhore	0.00020	4.07×10^{-8}	9.30×10^{-5}	4.15×10^{-08}	0.40166
Sundargarh	0.00555	3.08×10^{-5}	0.00303	3.02×10^{-05}	0.44791
Bamnidih	0.00024	5.56×10^{-8}	0.00012	5.46×10^{-08}	0.58226
Jondhara	0.00668	4.46×10^{-5}	0.00323	4.48×10^{-05}	0.48917
Kurubhata	0.00222	4.92×10^{-6}	0.00103	4.72×10^{-06}	0.74650
Basantpur	0.01059	0.00011	0.00411	0.00011	0.32857
Baronda	0.00187	3.49×10^{-6}	0.00073	3.56×10^{-6}	0.25906
Rajim	0.00177	3.14×10^{-6}	0.00079	3.18×10^{-6}	0.37205
Kantamal	0.03386	0.00115	0.01481	0.00116	0.36473

4.4. The Multivariate Autoregressive (MAR) Forecasting Model

The MAR model was designed using training datasets and the combination of the autoregression of multiple factors (WD, RF, SL, and T) as well as spatial variables (CA, RT, and R). The MAR model forecasted a one-step-ahead of the SL value using 51 input parameters (12 from each of the 4 temporal parameters and 3 from the spatial parameters). There is no need for a validation dataset because the linear model does not overfit.

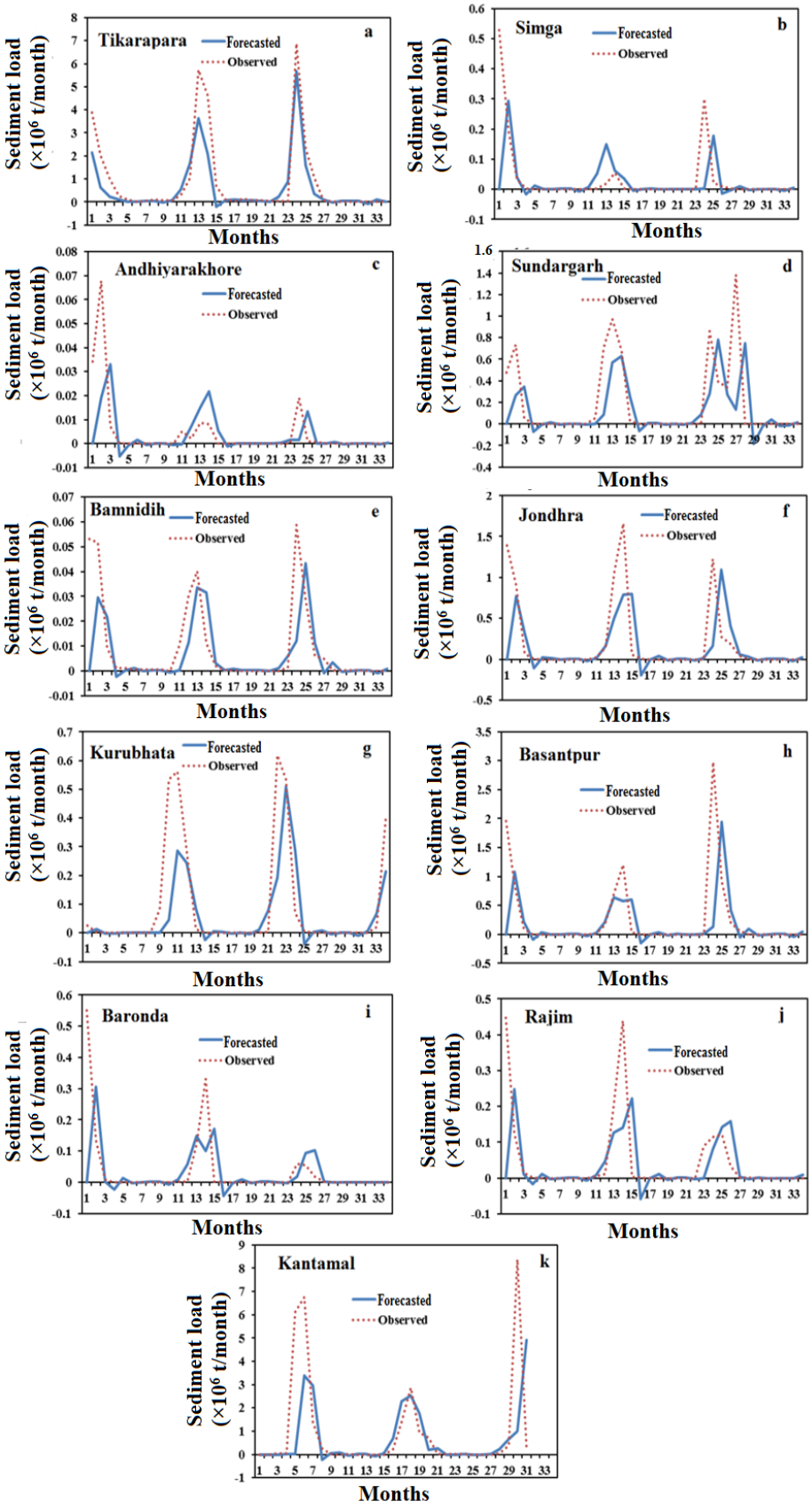


Figure 9. Comparison of the actual and forecasted SL during the testing phase of the AR forecasting model (**a**–**k**).

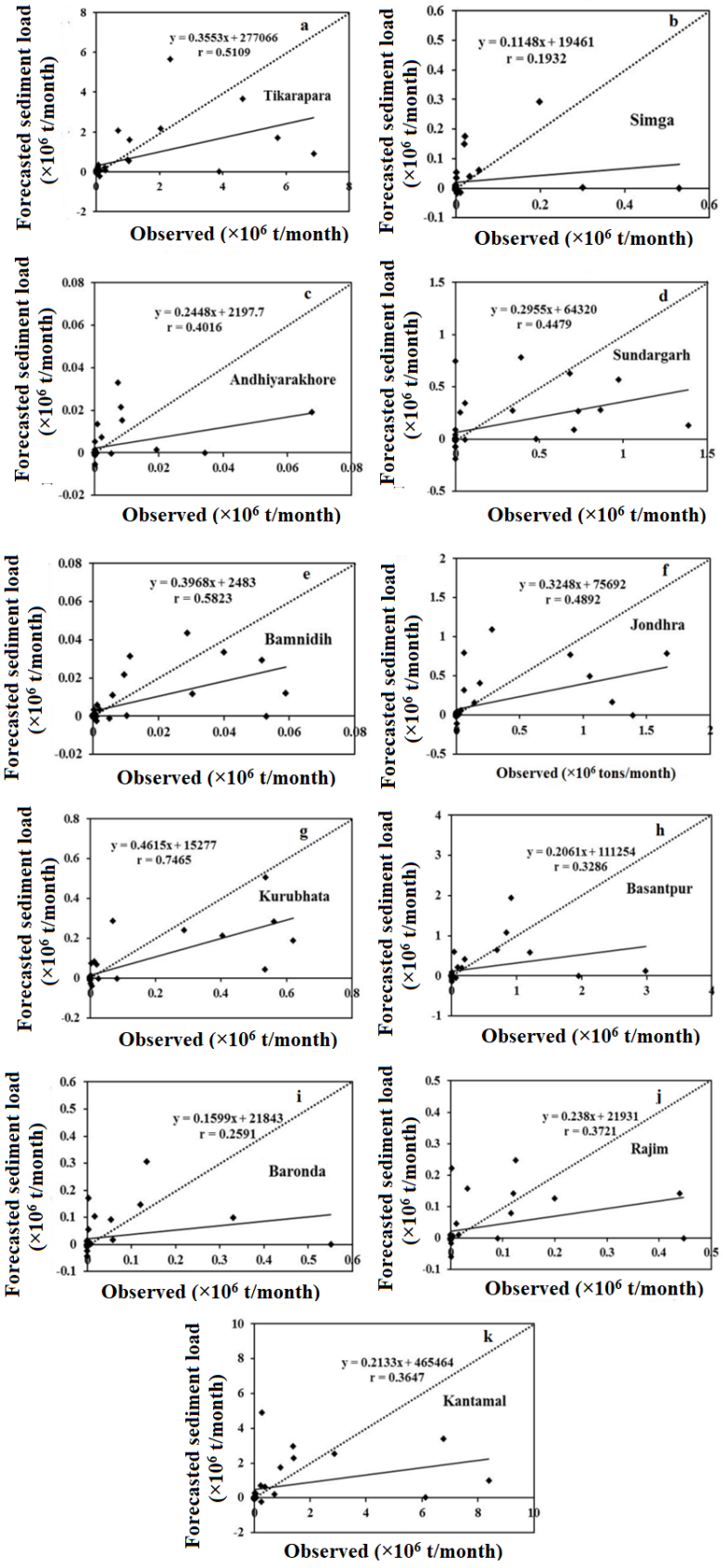


Figure 10. Scatter plot of the actual and forecasted SL of the AR forecasting model during the testing phase (**a**–**k**).

For testing, the same testing data that were applied to the ANN models were used. Table 7 displays the MAR model's error statistics. During the phases of training and testing, it was observed that the RMSE and MAE are very low, and r is high, demonstrating that the MAR forecasting model can reasonably fit the data. It is observed that Tikarapara shows the highest coefficient of correlation between the actual and forecasted SL values, and Andhiyarakhore station has the lowest correlation coefficient. The proposed model offered the best level of accuracy at the Tikarapara site and the lowest accuracy at Andhiyarakhore.

Table 7. Error statistics of single forecasting MAR model at each station.

MAR	MAE	VAR	r	MSE	RMSE
Training	0.00640	0.00023	0.68010	0.00023	0.01516
Testing	0.00562	0.00017	0.55620	0.00017	0.01296
Tikarapara	0.01306	0.00054	0.62380	0.00052	0.02284
Simga	0.00395	2.65×10^{-5}	0.49380	2.99×10^{-5}	0.00547
Andhiyarakhore	0.00260	1.20×10^{-5}	0.39130	1.16×10^{-5}	0.00341
Sundargarh	0.00486	4.91×10^{-5}	0.45670	4.79×10^{-5}	0.00692
Bamnidih	0.00272	1.53×10^{-5}	0.61440	1.66×10^{-5}	0.00407
Jondhara	0.00458	5.44×10^{-5}	0.63700	5.36×10^{-5}	0.00732
Kurubhata	0.00228	1.01×10^{-5}	0.74210	9.80×10^{-6}	0.00313
Basantpur	0.00610	0.00010	0.62500	0.00011	0.01025
Baronda	0.00306	1.75×10^{-5}	0.44590	1.70×10^{-5}	0.00412
Rajim	0.00309	1.64×10^{-5}	0.43390	1.59×10^{-5}	0.00398
Kantamal	0.01652	0.00111	0.39780	0.00108	0.03292

Figures 11 and 12 show the MAR model's hydrograph and scatter plot. The figure clearly shows that the MAR model presented a negative number where the SL is zero or near zero. It was also realized that the model could not capture low SL as evidenced by the scatter plot, which predicts a negative value in the case of low SL data including all stations. The modeled SL at Bamnidih and Andhiyarakore varies greatly at all peaks and during small SL. However, SL cannot be negative in actuality.

In the scatter plot, it is found that the MAR results are not closer to the bisector line at the Andhiyarakhore and Bamnidih gauging stations. It was noticed that the linear MAR model is unable to handle non-linearity behavior and, as a result, some negative forecasted values of SL are obtained.

4.5. Comparison Results of Forecasting Models

The best hybrid model (ANN-MOGA-51), traditional MAR, and AR models were compared using the same test data. Table 8 shows that the ANN-MOGA-51 model has the least RMSE, variance, MSE, and highest r score when compared to all other comparative models (MAR and AR). This statistical study indicates that the ANN-MOGA-51 model is the best. As a result, when the optimized input variables and associated elements are taken into account, the ANN-MOGA-51 model outperforms both the AR and MAR models. This advantage is due to the selection of optimal of all ANN parameters using the GA. The good performance of the ANN-MOGA-51 model may be attributed to the utilization of time series lag data of SL, RF, T, and WD with spatial data (CA, R and RT), which are more informative by ANN in conjunction with multi-objective GA method and lagged input variable selection using multi-objective GA.

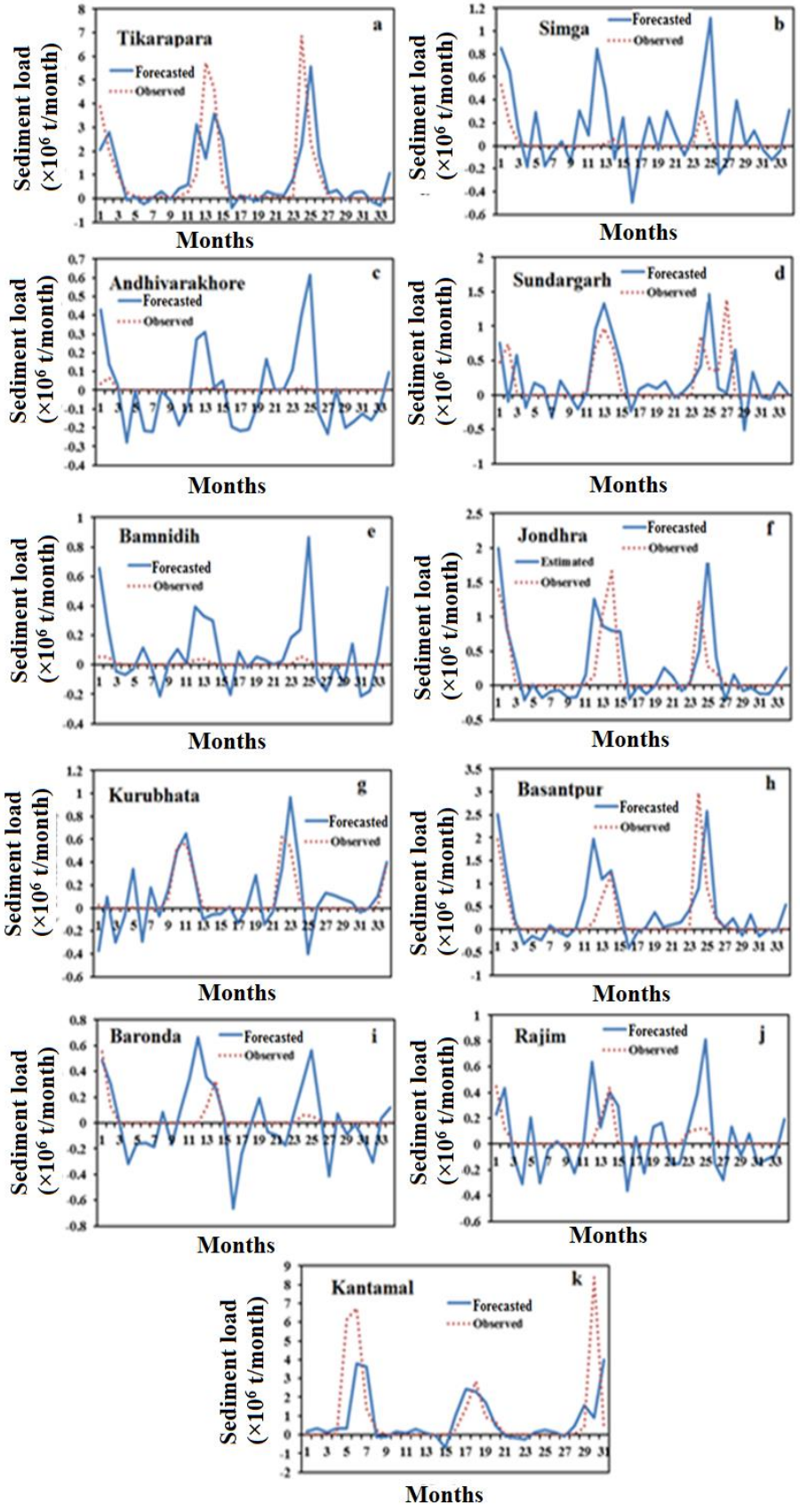


Figure 11. Comparison of the actual and forecasted SL during the testing phase of the MAR forecasting model (**a**–**k**).

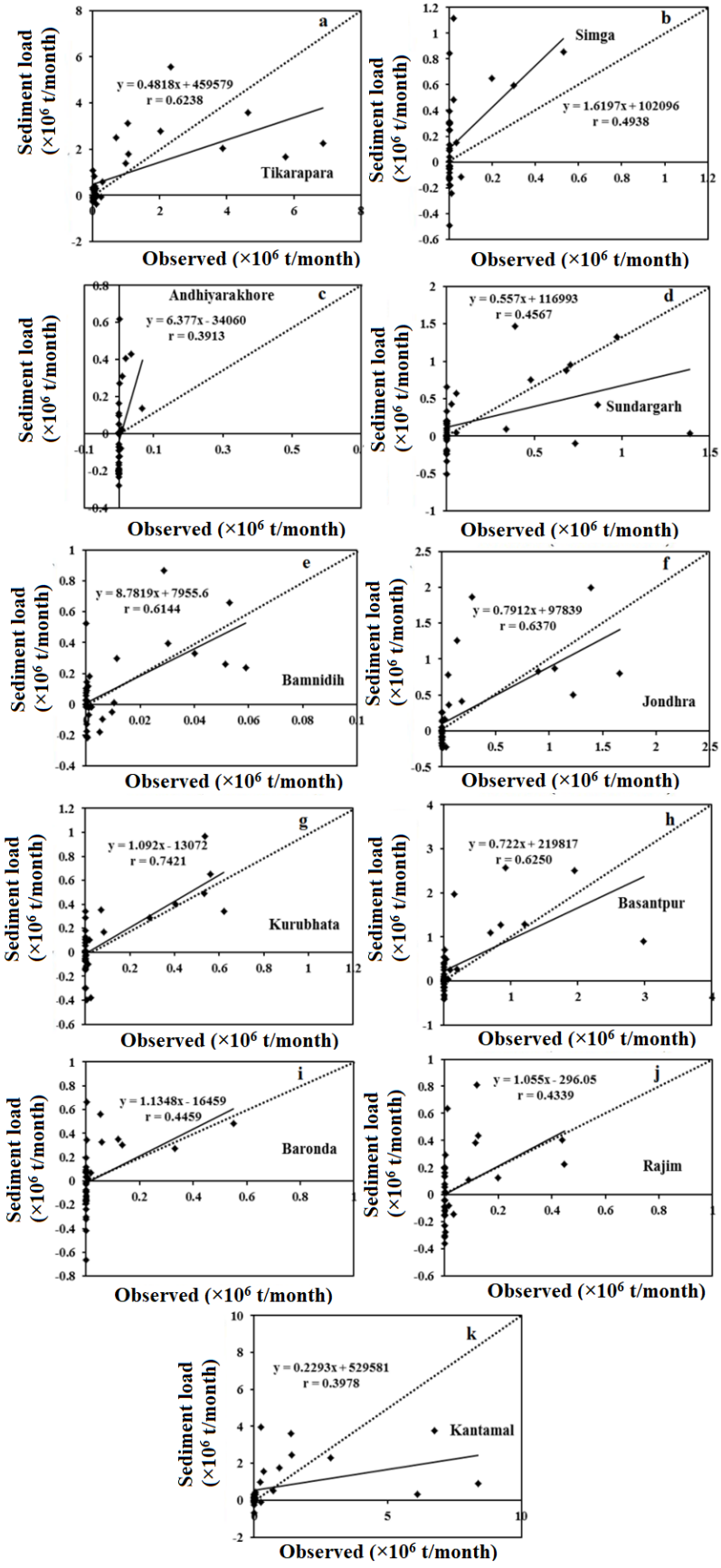


Figure 12. Scatter plot of the actual and forecasted SL of the MAR forecasting model during the testing phase (**a**–**k**).

Table 8. Statistical performance evaluation of ANN-MOGA-51, AR, and MAR models along with all gauging stations in a testing phase.

Models	ANN-MOGA-51		MAR		AR	
Statistics	**RMSE**	**r**	**RMSE**	**r**	**RMSE**	**r**
Testing	0.01164	0.6433	0.01296	0.5562	0.01335	0.4967
Tikarapara	0.01991	0.7311	0.02284	0.6238	0.02576	0.5109
Simga	0.00342	0.5930	0.00547	0.4938	0.00185	0.1932
Andhiyarakhore	0.00332	0.4001	0.00341	0.3913	0.00020	0.4017
Sundargarh	0.00510	0.6350	0.00692	0.4567	0.00555	0.4479
Bamnidih	0.00577	0.6950	0.00407	0.6144	0.00024	0.5821
Jondhara	0.00550	0.7370	0.00732	0.6370	0.00668	0.4892
Kurubhata	0.00143	0.9140	0.00313	0.7421	0.00222	0.7465
Basantpur	0.01512	0.4781	0.01025	0.6250	0.01059	0.3286
Baronda	0.00238	0.4950	0.00412	0.4459	0.00187	0.2591
Rajim	0.00145	0.6690	0.00398	0.4339	0.00177	0.3721
Kantamal	0.02840	0.6595	0.03292	0.3978	0.03386	0.3647

Table 8 demonstrates that the AR model performs the worst due to the highest RMSE, MSE, variance, and lowest *r* as compared to other models. This worst performance in the AR model is also caused due to the consideration of only SL as the lag time input variable and the exclusion of temporal (WD, T, and RF) and spatial (R, RT, and CA) data. The ANN-MOGA-51 model improved the performance by 12.81% and 10.19% from traditional AR and MAR regression models, respectively. It is also observed from Tables 3 and 8 that all hybrid intelligence-based models (ANN-MOGA-12, ANN-MOGA-15, ANN-MOGA-48 and ANN-MOGA-51) are providing better results than the traditional regression models (MAR and AR) models on the basis of RMSE and r as performance evaluation criteria.

5. Conclusions

This study revealed the forecasting of SL by the ANN-MOGA-51, MAR, and AR models with a time lag at eleven stations in the MR using various inputs of hydro-climatic factors (RT, RF, T, WD, R, and CA). The input parameters of the SL in the MR were found to be the primary governing factors. The results showed that the ANN-MOGA-51 models performed well and had a higher generalization capability, which was obtained by concurrently optimizing all ANN parameters using the MOGA. As a result, simultaneously optimizing all ANN parameters and input subsets with the MOGA is a better method with satisfactory performance and less computation cost than the traditional grid search and trial-and-error methods. The hydrograph and scatter plots of the ANN-MOGA-51 model also show that the magnitude of the proposed model's medium, low, and high SL forecasting was closer to the observed values. The best ANN-MOGA-51 model forecasted a positive sediment value even when SL was zero or near zero at all 11 sites in the MR, which is an interesting finding from the hydrograph and scatter plots. On the other hand, MAR and AR models provided negative SL values where SL is low or close to zero. This demonstrated that the data, particularly small, valued samples, exhibit significant non-linear behavior which is not captured by traditional MAR and AR forecasting models.

The results revealed that the hybrid ANN-MOGA model performed significantly better than other traditional MAR and AR models in terms of performance. This is the most appropriate approach because of the relatively better performance and ease of implementation. Thus, the proposed forecasting models are of great assistance to water resource planners and managers because they allow for a better understanding of the problems caused by sedimentation and allow for the finding of alternative solutions to manage the issues in the future by utilizing prior knowledge of forecasting SL. The RF intensity is also an important factor of the SL that is not incorporated in this research due to its unavailability for the improvement of modeling performance but will be considered in future research.

Author Contributions: Conceptualization, A.Y., P.C. and G.V.K.; methodology, A.Y., P.C., A.S. and A.A.; software, G.V.K., P.C. and M.A.A.; validation, A.Y., P.C. and Y.M.; analysis, A.Y., P.C., A.S. and A.A.; investigation, Y.M.; resources, A.S. and M.A.A.; writing—original draft preparation, A.Y., P.C. and A.A.; writing—review and editing, G.V.K., P.C., A.S. and M.A.A.; supervision, A.S., A.A. and Y.M.; project administration, Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4330898DSR02).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study's data were obtained after a non-disclosure agreement had been signed with the Central Water Commission. The author also thanks the National Institute of Technology, Rourkela, India, for providing the necessary facilities for this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Frémion, F.; Bordas, F.; Mourier, B.; Lenain, J.; Kestens, T.; Courtin-Nomade, A. Influence of dams on sediment continuity: A study case of a natural metallic contamination. *Sci. Total Environ.* **2016**, *547*, 282–294. [CrossRef] [PubMed]
2. Xia, X.; Dong, J.; Wang, M.; Xie, H.; Xia, N.; Li, H.; Zhang, X.; Mou, X.; Wen, J.; Bao, Y. Effect of water-sediment regulation of the Xiaolangdi reservoir on the concentrations, characteristics, and fluxes of suspended sediment and organic carbon in the Yellow River. *Sci. Total Environ.* **2016**, *571*, 487–497. [CrossRef] [PubMed]
3. Honorato, A.G.D.S.M.; Silva, G.B.L.D.; Guimarães Santos, C.A. Monthly streamflow forecasting using neuro-wavelet techniques and input analysis. *Hydrol. Sci. J.* **2018**, *63*, 2060–2075. [CrossRef]
4. Dutta, S. Soil erosion, sediment yield and sedimentation of reservoir: A review. *Model. Earth Syst. Environ.* **2016**, *2*, 123. [CrossRef]
5. Jansson, M.B. *Land Erosion by Water in Different Climates*; UNGI Report No. 57; Department of Physical Geography, University of Uppsala: Uppsala, Sweden, 1982.
6. Syvitski, J.P.M.; Peckham, S.D.; Hilberman, R.; Mulder, T. Predicting the Terrestrial Flux of Sediment to the Global Ocean: A Planetary Perspective. *Sediment. Geol.* **2003**, *162*, 5–24. [CrossRef]
7. Gupta, H.; Chakrapani, G.J. Temporal and spatial variations in water flow and sediment load in Narmada River Basin, India: Natural and man-made factors. *Environ. Geol.* **2005**, *48*, 579–589. [CrossRef]
8. Ramesh, R.; Subramanian, V. Temporal, spatial and size variation in the sediment transport in the Krishna River basin, India. *J. Hydrol.* **1988**, *98*, 53–65. [CrossRef]
9. Bastia, F.; Equeenuddin, S.M. Spatio-temporal variation of water flow and sediment discharge in the Mahanadi River, India. *Glob. Planet. Chang.* **2016**, *144*, 51–66. [CrossRef]
10. Thodsen, H.; Hasholt, B.; Kjarsgaard, J.H. The influence of climate change on suspended sediment transport in Danish rivers. *Hydrol. Process.* **2008**, *22*, 764–774. [CrossRef]
11. Merritt, W.S.; Letcher, R.A.; Jakeman, A.J. A review of erosion and sediment transport models. *Environ. Model. Softw.* **2003**, *18*, 761–799. [CrossRef]
12. Salas, J.D.; Delleur, J.W.; Yevjevich, V.; Lane, W.L. *Applied Modeling of Hydrologic Time Series*; Water Resources Publications: Littleton, CO, USA, 1980; 484p.
13. Adamowski, J.; Chan, H.F.; Prasher, S.O.; Ozga-Zielinski, B.; Sliusarieva, A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Wat. Resour. Res.* **2012**, *48*, W01528. [CrossRef]
14. Pektas, A.O.; Cigizoglu, H.K. Long-range forecasting of suspended sediment. *Hydrol. Sci. J.* **2017**, *62*, 2415–2425. [CrossRef]
15. Partal, T.; Cigizoglu, H.K. Estimation and forecasting of daily suspended sediment data using wavelet-neural networks. *J. Hydrol.* **2008**, *358*, 317–331. [CrossRef]
16. Nourani, V.; Alizadeh, F.; Roushangar, K. Evaluation of a two-stage SVM and spatial statistics methods for modeling monthly river suspended sediment load. *Water Resour. Manag.* **2016**, *30*, 393–407. [CrossRef]
17. Meshram, S.G.; Ghorbani, M.A.; Deo, R.C.; Kashani, M.H.; Meshram, C.; Karimi, V. New approach for sediment yield forecasting with a two-phase feedforward neuron network-particle swarm optimization model integrated with the gravitational search algorithm. *Water Resour. Manag.* **2019**, *33*, 2335–2356. [CrossRef]
18. ASCE. Task Committee on Application of Artificial Neural Networks in Hydrology Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **2000**, *5*, 115–123. [CrossRef]
19. Reddy, P.V.B. Modelling and Optimization of Wire Electrical Discharge Machining of Cr-Mo-V Special Alloy Steel Using Neuro Genetic Approach. Ph.D. Thesis, Jawaharlal Nehru Technological University, Anantapur, India, 2014.
20. Tokar, A.S.; Johnson, P.A. Rainfall-Runoff Modelling using Artificial Neural Networks. *J. Hydrol. Eng.* **1999**, *4*, 232–239. [CrossRef]

21. Dawson, C.W.; Harpham, C.; Wilby, R.L.; Chen, Y. An evaluation of artificial neural network techniques for flow forecasting in the river Yangtze, China. *Hydrol. Earth Syst. Sci.* **2002**, *6*, 619–626. [CrossRef]
22. Kar, A.K.; Lohani, A.K.; Goel, N.K.; Roy, G.P. Development of Flood Forecasting System Using Statistical and ANN Techniques in the Downstream Catchment of Mahanadi Basin, India. *J. Water Resour. Prot.* **2010**, *2*, 880–887.
23. Bishop, M. *Neural Networks for Pattern Recognition*; Clarendon Press: Oxford, UK, 1998.
24. Yadav, A.; Chatterjee, S.; Equeenuddin, S.M. Suspended Sediment Yield Estimation using Genetic Algorithm-based Artificial Intelligence Models in Mahanadi River. *Hydrol. Sci. J.* **2018**, *63*, 1162–1182. [CrossRef]
25. Holland, J. *Adaptation in Natural and Artificial Systems*; The University of Michigan Press: Ann Arbor, MI, USA, 1975.
26. Hosseini, S.A.; Abbaszadeh Shahri, A.; Asheghi, R. Prediction of bedload transport rate using a block combined network structure. *Hydrol. Sci. J.* **2022**, *67*, 117–128. [CrossRef]
27. Chatterjee, S.; Bandopadhyay, S. Reliability estimation using a genetic algorithm-based artificial neural network: An application to a load-haul-dump machine. *Expert Syst. Appl.* **2012**, *39*, 10943–10951. [CrossRef]
28. Asheghi, R.; Hosseini, S.A.; Saneie, M.; Shahri, A.A. Updating the neural network sediment load models using different sensitivity analysis methods: A regional application. *J. Hydroinformatics* **2020**, *22*, 562–577. [CrossRef]
29. Adib, A.; Mahmoodi, A. Prediction of Suspended Sediment Load using ANN GA Conjunction Model with Markov Chain Approach at Flood Conditions. *KSCE J. Civ. Eng.* **2016**, *1*, 447–457. [CrossRef]
30. Chatterjee, S.; Bandopadhyay, S. Goodnews Bay Platinum Resource Estimation Using Least Squares Support Vector Regression with Selection of Input Space Dimension and Hyperparameters. *Nat. Resour. Res.* **2016**, *20*, 117–129. [CrossRef]
31. Samarasinghe, S. *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*; CRC Press: Boca Raton, FL, USA, 2016; p. 555.
32. Geman, S.; Bienenstock, E.; Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **1992**, *4*, 1–58. [CrossRef]
33. Rosales-Perez, A.; Escalante, H.J.; Gonzalez, J.A.; Reyes-Garcia, C.A. Bias and variance optimization for SVMs model selection. In Proceedings of the Twenty-Sixth International FLAIRS Conference, St. Pete Beach, FL, USA, 22–24 May 2013.
34. Kulasiri, D.; Verwoerd, V. *Stochastic Dynamics: Modeling Solute Transport in Porous Media*; North Holland Series in Applied Mathematics and Mechanics; Elsevier: Amsterdam, The Netherlands, 2002; p. 44.
35. Levin, S.A. Population dynamics in models in heterogeneous environments. *Annu. Rev. Ecol. Syst.* **1976**, *7*, 287. [CrossRef]
36. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 181–197. [CrossRef]
37. Bharti, P.S.; Maheshwari, S.; Sharma, C. Multi-objective optimization of electric-discharge machining process using controlled elitist NSGA-II. *J. Mech. Sci. Technol.* **2012**, *26*, 1875–1883. [CrossRef]
38. Behzadian, K.; Kapelan, Z.; Savic, D.; Ardeshir, A. Stochastic sampling design using multi-objective genetic algorithm and adaptive neural network. *Environ. Model. Softw.* **2009**, *24*, 530–541. [CrossRef]
39. Zhou, C.C.; Yin, G.F.; Hu, X.B. Multi-objective optimization of material selection for sustainable products: Artificial neural networks and genetic algorithm approach. *Mater. Des.* **2009**, *30*, 1209–1215. [CrossRef]
40. Yadav, A.; Chatterjee, S.; Equeenuddin, S.M. Suspended sediment yield modeling in Mahanadi River, India by multi-objective optimization hybridizing artificial intelligence algorithms. *Int. J. Sediment Res.* **2021**, *36*, 76–91. [CrossRef]
41. Sakai, K.; Osawa, K.; Yoshinaga, A. Development of suspended sediment concentration analysis model and its application with multi-objective optimization. *Paddy Water Environ.* **2005**, *3*, 201–209. [CrossRef]
42. Peng, Y.; Ji, C.; Gu, R. Multiobjective optimization model for coordinatec regulation of water flow and sediment in cascade reservoirs. *Water Resour. Manag.* **2014**, *28*, 4019–4033. [CrossRef]
43. Cigizoglu, H.K.; Kisi, O. Methods to improve the neural network performance in suspended sediment estimation. *J. Hydrol.* **2006**, *317*, 221–238. [CrossRef]
44. Cigizoglu, H.K.; Alp, M. Generalized regression neural network in modelling river sediment yield. *Adv. Eng. Softw.* **2006**, *37*, 63–68. [CrossRef]
45. Piotrowski, A.P.; Napiorkowski, M.J.; Napiorkowski, J.J.; Osuch, M. Comparing various artificial neural network types for water temperature prediction in rivers. *J. Hydrol.* **2015**, *529*, 302–315. [CrossRef]
46. Kant, A.; Suman, P.K.; Giri, B.K.; Tiwari, M.K.; Chatterjee, C.; Nayak, P.C.; Kumar, S. Comparison of multi-objective evolutionary neural network, adaptive neuro-inference system and bootstrap-based neural network for flood forecasting. *Neural Comput. Appl.* **2013**, *23*, S231–S246. [CrossRef]
47. India-WRIS. Water Resources Information System of India. Available online: http://india-wris.nrsc.gov.in/wrpinfo/index.php?title=Mahanadi (accessed on 8 August 2016).
48. Yadav, A.; Chatterjee, S.; Equeenuddin, S.M. Prediction of Suspended Sediment Yield by Artificial Neural Network and Traditional Mathematical Model in Mahanadi River Basin, India. *J. Sustain. Water Resour. Manag.* **2017**, *4*, 745–759. [CrossRef]
49. Rojas, R. *Neural Network: A Systematic Introduction*; Springer: Berlin, Germany, 1996; pp. 151–184.
50. Hagan, M.T.; Menhaj, M.B. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* **1994**, *5*, 989–993. [CrossRef]
51. Ebtehaj, I.; Bonakdari, H.; Safari, M.J.S.; Gharabaghi, B.; Zaji, A.H.; Madavar, H.R.; Khozani, Z.S.; Es-haghi, M.S.; Shishegaran, A.; Mehr, A.D. Combination of sensitivity and uncertainty analyses for sediment transport modeling in sewer pipes. *Int. J. Sediment Res.* **2020**, *35*, 157–170. [CrossRef]

52. Riahi-Madvar, H.; Gharabaghi, B. Pre-processing and Input Vector Selection Techniques in Computational Soft Computing Models of Water Engineering. In *Computational Intelligence for Water and Environmental Sciences*; Springer: Singapore, 2022; pp. 429–447.
53. Dehghani, M.; Seifi, A.; Riahi-Madvar, H. Novel forecasting models for immediate-short-term to long-term influent flow prediction by combining ANFIS and grey wolf optimization. *J. Hydrol.* **2019**, *576*, 698–725. [CrossRef]
54. Riahi-Madvar, H.; Dehghani, M.; Memarzadeh, R.; Gharabaghi, B. Short to long-term forecasting of river flows by heuristic optimization algorithms hybridized with ANFIS. *Water Resour. Manag.* **2021**, *35*, 1149–1166. [CrossRef]
55. Riahi-Madvar, H.; Seifi, A. Uncertainty analysis in bed load transport prediction of gravel-bed rivers by ANN and ANFIS. *Arab. J. Geosci.* **2018**, *11*, 1–20. [CrossRef]
56. Gowda, C.C.; Mayya, S.G. Comparison of back propagation neural network and genetic algorithm neural network for stream flow prediction. *J. Comput. Environ. Sci.* **2014**, 290127. [CrossRef]
57. Senthil Kumar, A.R.; Sudheer, K.P.; Jain, S.K.; Agarwal, P.K. Rainfall-runoff modelling using artificial neural networks: Comparison of network types. *Hydrol. Process. Int. J.* **2005**, *19*, 1277–1291. [CrossRef]
58. Ghosh, A.; Das, M.K. Non-dominated rank-based sorting genetic algorithms. *Fundam. Inform.* **2008**, *83*, 231–252.
59. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]

Measurement: Sensors 27 (2023) 100744

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Reliable cluster based data aggregation scheme for IoT network using hybrid deep learning techniques

Guguloth Ravi [a,*], M. Swamy Das [b], Karthik Karmakonda [b]

[a] Department of CSE, University College of Engineering (UCE), Osmania University(OU), Mall Reddy College of Engineering & Technology, Hyderabad, Telangana, India
[b] Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

ARTICLE INFO

Keywords:
IoT
Cluster formation
CH selection
Optimization algorithm
Deep neural network
Data aggregation

ABSTRACT

Background: Several IoT nodes are deployed in the monitoring environment to ensure reliability. In both sensor and sink nodes, the same data is sensed and forwarded. While redundant data maintains reliability, sink nodes waste energy processing the redundant data.
Objective: In order to upholdthe compromiseamong energy ingesting and reliability, necessary to eliminate the redundancies in sensed data up to an appropriate level. Data aggregation algorithms currently assign time slots based on data sensing period and program rate, disregarding packet loss and latency.
Methodology: In this paper, we suggest a cluster based reliable data aggregation (CRDA) scheme for IoTnetwork which ensures data collection and aggregation in energy efficient manner and transfer to another end very effectively. We first introduce a monarch and sine-cosine (MSC) algorithm to form clusters by grouping the IoT sensors which ensures the effective data transferring. In data aggregation phase, we utilize the multiple design metrics to compute the trust degree of each IoT sensors and design an improved sunflower optimization (ISFO) algorithm to optimize the design constraints. The highest trust degree owned swelling is act as cluster head (CH) of the cluster which ensures data aggregation. A reformative optimal–learning-based deep neural network (ROL-DNN) is then used to compute routes between IoT sensors which ensures reliabledata aggregation and transferring.
Results: and analysis: Finally, we validate our proposed routing with the different simulation scenario and their results are compared with the existing routing protocols to prove the effectiveness.

1. Introduction

Internetof Things (IoT) permits brilliant articles and savvy frameworks to collect and share information worldwide to allow shrewd climate [1]. There is a developing interest in the utilization of remote detecting advancements in different IoT situations.Given the enormous development of objects and their applications, gathering and examining their item information is becoming one of the significant difficulties. As sensor hubs are controlled by batteries, energy productive activities are critical. For this reason, it is attractive for the sensor hub to de-copy the information got from the adjoining hubs prior to communicating the last information to the focal station. Information accumulation [2,3] is one of the powerful strategies to wipe out information overt repetitiveness and further develop energy productivity; expanding the lifetime of remote sensor organizations (WSNs). A difficult issue for data management is effectively deliver data to relevant users. It uses economical techniques such as efficient flow distribution systems for IoT [4]. The system collects integrated data streams generated from different collectors and transmits relevant data to relevant users based on user queries entered into the system [5]. Create two new data structures to meet the requirements of high efficiency data flow propagation in two conditions, such as point-to-point systems and flow propagation in wireless transmission systems. Assessment of approaches using real-world datasets shows that they can transmit connected data streams more efficiently than current technology [6].

In IoT advances, multiple data configurations are recommended for efficient data processing and minimal data recovery. This includes storing centralized data, such as the cloud system, on nearby distribution systems. Smart cities are the practical implementation of IoT, which aims to provide people with efficient, reliable and secure applications such as water, electricity and transportation through rational management [7]. The database-based IoT platform is required to implement

* Corresponding author.
E-mail addresses: g.raviraja@gmail.com (G. Ravi), msdas_cse@cbit.ac.in (M. Swamy Das), Karthikk_cse@cbit.ac.in (K. Karmakonda).

https://doi.org/10.1016/j.measen.2023.100744
Received 12 November 2022; Received in revised form 2 February 2023; Accepted 2 March 2023
Available online 6 March 2023

manageable and sustainable smart applications and build new applications on them. A service-based architecture developed using key features of smart application are web resource management, end-to-end data, and IoT application platform [8]. Raising the inefficiency of health infrastructure and biomedical systems is one of the least ambitious aspects of modern society. Recent advantage in IoT knowledge design is used to control the development of intelligent organizations to support and improve health and biological processes. Possible examples include automation, population monitoring, biomechanical facilities in hospitals, affiliated medical associations, and direct medical monitoring of physical effects to predict medical disorders [9]. RFID low-power devices using passive and/or passive devices use ability to transfer data when using suspicious electromagnetic field. Passive RFID symbols do not require resource energy to work, so their lifespan calculated in decades, allowing RFID technology to adapt to a variety of applications, including healthcare. IoT devices can remotely access the Internet, but can use various data collection and new management features provided by IOT to create applications [10] those health system solutions. However, the combination of these technologies has not yet been used for security purposes.

Portable edge processing worked with protection saving information rundown for IoT in light of the homomorphic properties of the Boneh-Goh-Nissim cryptosystem [11].A correspondence productive information total tree (AT) [12] is proposed for complex necessities in IoT. Changed gain of total capability considering arrangement limit and collection cost mutually. Area based secure rethought conglomeration (LBOA) [13] utilizes steering chain, request security encryption, and other cryptographic capability to amplify the worth of detected information that sticks to the area system.A gadget based unknown protection saving (APPA) [14] confirmation conspire is proposed for information combination in haze improved IoT frameworks that upholds numerous specialists to oversee savvy gadgets and mist hubs locally.A consistency-ensured and energy effective rest booking calculation (CG-E2S2) [15] is utilized for information that not entirely set in stone by a Markov dynamic cycle thinking about the energy proficiency of IoT gadgets.Brilliant meters compute their portions by information encoding (FHE) in a haphazardly created polynomial (secure MPC) [16]. Encrypted information shares are collected in a succession to the organization entryway utilizing an aggregator without uncovering the genuine worth of the measurement. An unquestionable secure conglomeration conspire [17] executes an entrusted collection hub to perform information accumulation from source hubs without unveiling information. TinyECC gives start to finish protection and empowers early assault location through jump by-bounce assessment, in this way lessening the need to depend altogether on synchronization for check [18].Information incorporation actually settles on the best choices consequently over Low-power and lossy organization (LA-RPL) [19] for IoT. Protection saving Heath information mix programming safely gathers wellbeing information from different sources and gives fair motivators to patients. Boneh-Goh-Nissim cryptosystem and Shamir secret sharing are utilized to keep up with information obscurity security and adaptation to non-critical failure [20].

Our contributions.A cluster based reliable data aggregation (CRDA) arrangements proposed for IoT network which ensures data collection and aggregation as energy efficient manner and transfer to another end very effectively. The key contributions of plannedCRDA arrangement are abridged as follows.

1. A monarch and sine-cosine (MSC) algorithm is used to form clusters by grouping the IoT sensors which ensures the effective data transferring.
2. Then, we utilize the multiple design metrics to compute the trust degree of each IoT sensors and design an improved sunflower optimization (ISFO) algorithm to optimize the design constraints.
3. ROL-DNN is used to compute routes between IoT sensors which ensures reliable data aggregation and transferring.
4. Finally, we validate our proposed routing with the different simulation scenario and their results are associated with the prevailing routing protocols to prove the effectiveness.

The rest of this paper is prearranged as follows: Sect. 2 designates the recent works connected tothe reliable data aggregationscheme. Sect. 3 illustrates the problem methodology and system perfect of our planned CRDA scheme. The thorough employed process of our plannedCRDA scheme is explained in Sect. 4 with the proper mathematical model. The imitation outcomes and qualified analysis are deliberated in Sect. 5. Finally, this paper is decided in Sect. 6.

2. Related works

Today, the Internet of Things (IoT) is receiving significant courtesy in various theoretical and real-world explore areas. The data-intensive countryside of IoT makes collecting data in such organizations very difficult. The primary objective of data combination is to reach large amount of Quality of service (QoS) such as optimal data broadcast latency, dependability and energy ingesting.

Chandnani et al. [21] have proposed the ANT-PSO-AODV is a secure data aggregation method that uses trust-based techniques and ANT particle swarm optimization to detect and secure data transmission from CH to the sink. Clusters are formed based on proximity within each zone and secret sharing scheme is used for secure data transmission. The performance of ANT-PSO-AODV is analyzed in terms of end-to-end delay, throughput, packet delivery ratio, energy consumption, and routing overhead by varying node mobility, number of sensor nodes, and input data rate. Amarlingam et al. [22] have proposed compressed sensing-aided reduced-complexity, which divides the network into limited overlapping groups to allow an optimal trade-off between power consumption, computational complexity and error recovery. The metric matrix used for bounded overlap clustering corresponds to ISO properties that guarantee cluster detection. Laplacianeigen-values are based on weighted laterality matrices to find sparse representations of unhurried data from arbitrarily convoluted systems. This enables high accuracy accumulated data to be read at the sink end. A graphical Laplacian eigen base in view of the weight nearness grid is utilized to track down a meager portrayal of the deliberate information from haphazardly requested networks, which allows recovering the accumulated data at the sink node with high reliability. Badiger et al. [23] have proposed the EDAS is an efficient data aggregation scheme for IoT-based wireless sensor networks that uses the improved low energy adaptive clustering algorithm (I-LEACH) to form an optimal number of cluster heads by considering node residual energy and average network energy. Data redundancy is removed using network coding, which integrates linear XOR operations and ensures non-repeated data transmissions. EDAS is evaluated in terms of network parameters and its performance is compared to existing schemes.

Sajidi et al. [24] have developed the Fuzzy data integration for IoT (F-LEACH) project, which enables healthcare applications to extend network life. Fuzzy membership function optimized for the changing network situation, and the multi-run average was chosen as the best parameter. The main objective of data aggregation schemes is to combine and combine information parcels productively to decrease energy utilization and gridlock and further develop information precision. This study presents a multi-strategy similar investigation of IoT information assortment connected with energy squander, network lifetime, energy utilization, deferral, and number of sensor hubs. Wu et al. [25] (2022) have proposed a secure and efficient multi-functional data combination. As IoT data becomes increasingly important to potential privacy effectiveness. A powerful, feature-rich, and data collection model is used to extend SEMDA which support heterogeneous privacy. SEMDA security analysis was performed using a simple but attractive model. And this shows that SEMDA not guarantees privacy and confidentiality. But also guarantees its completeness and accuracy. Khan

et al. [26] (2022) have designed IoT uses a beta-dominated set-based cluster aggregation mechanism (βDSC2DAM), It is improvement of traditional cluster data integration mechanism. It compares to a traditional clustering algorithm during data collection. The average hour is estimated as the maximum hour of the package. According to certain complexity, time parameters, βDSC2DAM performs better than traditional IoT clustering algorithms.

Saleem et al. [27] have plannedfog-enabled privacy-preserving data aggregation (FESDA) scheme. This plan opposes malevolent information infusion assaults by separating added values from outer assailants [34]. To accomplish security, the shrewd meter utilizes a changed variant of the Paillier crypto-framework to scramble clients' utilization information.FESDA is shortcoming lenient, implying that regardless of whether few brilliant meters fall flat, information assortment from different gadgets won't be impacted [35]. To consider its exhibition in contrast to three other contending plans as far as accumulation, encryption and correspondence costs. The discoveries show that PPFA lessens the correspondence cost by half contrasted with the coordination conspire. Zeng et al. [28] have proposed a privacy-preserving multi-dimensional and directional data aggregation (MMDA) scheme for edge computing-enhanced IoT communications. MMDA empowers edge gadget m complex information of IoT gadgets to perform line accumulation and section collection in two headings. Such information is utilized to ascertain the amount of the information in each line and every segment to safeguard security.MMDA gives the capacity to give more experiences to the IoT control place for investigation and handling. MMDA takes on sequential confirmation to decrease approval cost. Nonetheless, these current plans are hard to carry out because of complicated estimations or high correspondence prerequisites. Onesimu et al. [29] have proposed a privacy-preserving data collection scheme for IoT-based healthcare systems. An anonymization model in view of bunching is utilized to plan a compelling security assurance plan to meet protection prerequisites and keep medical services IoT from different security assaults. UPGMA is utilized to decrease correspondence cost and accomplish better security. Aziz et al. [30] have proposed an efficient multi-hop cluster-based coordination scheme for IoT using hybrid CS (EMCA-CS).This really builds the organization lifetime and lessens the reconfiguration blunder among CS and directing conventions. EMCA-CS incorporates calculation to parcel the field into various hexagonal cells and select a CH hub from each bunch in view of a few measures. Each CH sums up its group information utilizing the half breed CS strategy [36]. Dark Wolf-based algorithm used to make an ideal way for CHs to send compacted information to the BS, and a CSMO-GWO calculation was introduced to further develop the CS grid development process. Reproduction results demonstrate that the exhibition beats the standard strategy as far as expanding WSN lifetime, diminishing energy utilization, and lessening standardized mean square mistake.

3. Problem methodology and network model

3.1. Research gaps

There are some technical challenges in integrating wireless sensors into existing 5G networks, including channel features, route planning for wireless sensors, power charging capabilities, and integration of wireless sensors and IoT systems [37,38]. Coverage estimation and modeling are required to determine the number of wireless sensors, the optimal height of wireless sensors, and the neutrality of message among wireless sensors and ground users. The use of wireless sensors in urban environments requires consideration of road damage. Wireless sensors often run on limited battery power, so the performance of sensors is limited by their limited internal power. Therefore, it is important to use energy wisely to complete sensors and transportation. In IoT networks, optimal routing is required for data collection. However, in order to provide a means of communication to ground users, the IoT network must be considered. Limited energy is a barrier to IoT networks. Also, energy harvesting can be used to enhance IoT sensors using green energy [39]. Energy-efficient methods such as wireless power transfer can be smart solutions to improve charging efficiency. The unique features of IoT such as fast mobility, fast deployment, easy programmability and scalability and applications for future IoT systems make it an excellent solution for understanding the architecture of future IoT systems.

With the rapid rise in the number of healthcare sensors, researchers are increasingly interested in supporting a wide range of exciting new applications and scenarios. IoT consists of many devices connected with different types of sensor support, especially for the detection of medical health standards [21]. More commitment means less power consumption and longer waiting time for packets in the queue. Delays increase in the process of collecting high-quality data. The discarded packets are sent before expiration, which leads to increased power consumption [21–24]. Also, many uncoordinated beams are likely to drop due to channel issues such as high beam and mid-air impacts.Broadcast aggregation improves power ingesting and quality of provision during the data collection process. Aggregation sends multiple packets with maximum aggregation achieved by QoS constraints [25]. In data aggregation, the maximum number of packets is combined into a single packet. Security is key factor in enabling widespread adoption of IoT technologies and applications [26]. Without privacy, reliability, and privacy, IoT solutions are unlikely to gain mass adoption. IoT is characterized by diversity and connectivity with limited resources, making it a powerful Internet. Blockchain with IoT devices are essential to provide an end-to-end secure connection between IoT entities [27,28,40]. One of the main challenges resulting from the quick spread of the network of things is the node density that contains large amounts of data in different networks, which affects the probability of collision and network overcrowding [29]. Current clustering techniques have solved these problems by ignoring the characteristics and types of IoT traffic [21–30]. Existing solutions suffer from additional storage, communications and energy costs [41]. Recently, many data integration projects have introduced the data integration process to increase its efficiency and makes process more accurate. However, these methodsare used for specific application environments, which make design difficult. The main objectives of our CRDA scheme are given as follows.

1. New data combination scheme is used to improve the energy efficiency, data transmission rate, increase network lifetime with minimum delay and transmission of data efficiently
2. The network cluster integration model ensures QoSand reduces traffic congestion.To increase the network lifetime using reliable transmission ofdata
3. To validate proposed scheme using different real-time IoT requests such as smart city, E-healthcare, E-agriculture, E-education, and etc.

3.1.1. Network model

Fig. 1 displays the network model of our proposed CRDA system with the sample network model which used to increase network lifetime, accomplish traffic information load balance and decrease reproduction mistake. First, an MSC algorithm is used to form the efficient balanced clustering geometry based on the IoT nodes information, such as node location and distance between base station to corresponding target node.In each group, the proposed plot considers CH determination as a dynamic issue formed in view of numerous standards, for example, energy consumption, link quality, path loss, distance from base station and delay. Then, the multiple design metrics are optimized through ISFO algorithmto find solution to the conclusion problem. Besides, ROL-DNN technique is then used to calculate bestoptimum path between IoT nodes which ensures reliable data aggregation and transferring. Our reproduction results show that the proposed strategy beats existing gauge methods in decreasing energy utilization and expanding network lifetime.

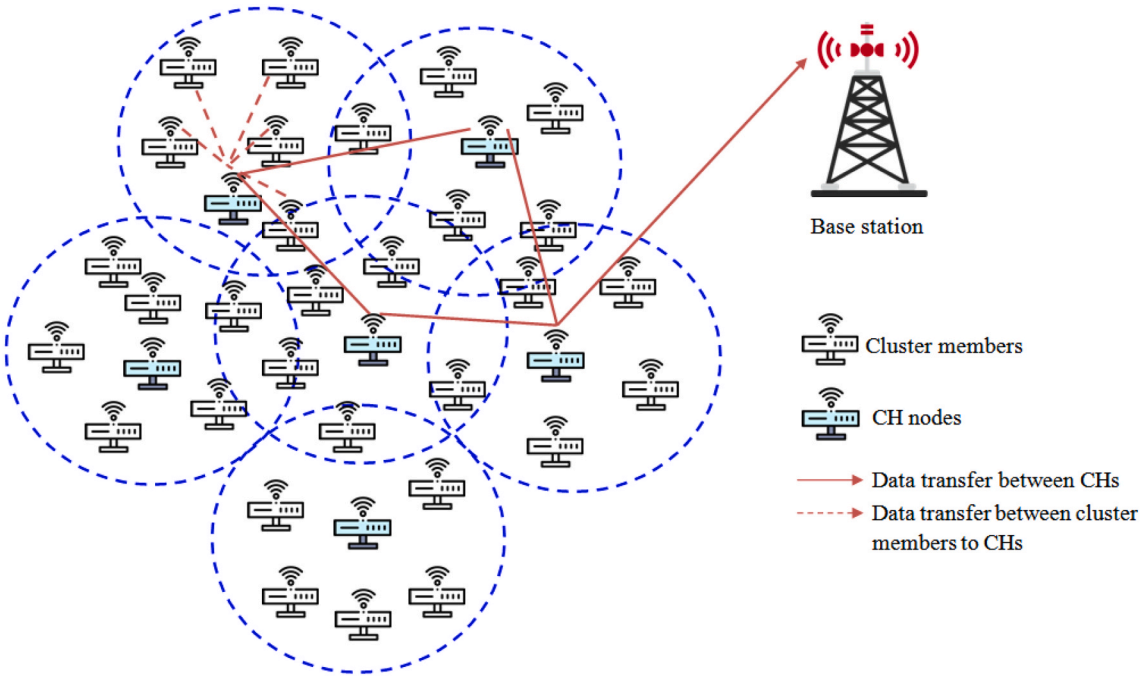


Fig. 1. Network model of proposed CRDA scheme.

4. Proposed methodology

In this section, we designate the working processing of our proposed CRDA structure with proper mathematical model which consist three-fold process are clustering, CH selection and optimal path computation.

4.1. Cluster formation

The creation of energy-efficient becomes crucial since IoT nodes are energy-constrained and run on a small internal battery. In order to prepare for impending demand, energy conscious IoT networks must simultaneously forecast their energy use. A collection of sensor nodes that can sense, calculate, and transmit make up the network. Energy conservation in IoT becomes a major concern to increase network lifetime. Since clustering is regarded as an efficient and suitable way for transmitting the data without any issues, multiple efforts have been made to improve the routing protocols in the network to date. For clustering purpose, in this work, we utilized a monarch and sine-cosine (MSC) algorithm for cluster formation which groups IoT sensors to ensure the energy efficient data aggregation and transfer. The probability q ∈ [0, 1] of the MSC algorithm that a migrant encounters a life-threatening situation depends on two main factors: fatigue, peripheral navies (storms, predators, individuals, birds and etc.) equal probability. For optimal problematic, the fitness of the answer is relative to the value of the impartial function. Founded on these rules the main steps of MSC are shown in the procedure. Time-compensated sun compass $y_j^{(s+1)}$ information is used to create a new position,

$$y_j^{(s+1)} = y_j^{(s)} + \alpha \times \delta_s \tag{1}$$

The δ_s step size is the light strength used in a sun range or the attractive field strength used in attractive scope. After a dangerous condition occurs, a portion of the migratory populace dies. With the existence of the fittest in mind, we select the worst people as victims. New immigrants are created to fill the void in the population. A new migrated $y_j(New)$ with strong magnetic susceptibility is produced from randomly selected individual y_j,

$$y_j(New) = y_j \oplus \gamma \otimes \eta \tag{2}$$

The size of the leap η relies upon the force of the attractive field. α and $\gamma > 0$ mean the inquiry size of step size or bounce size, contingent upon the issue we are thinking about. Taking into account the earth as an attractive dipole with attractive second M and exiled person in wide position, the brilliant force (J) and attractive power not set in stone.

$$J = J_o E^{-Vr} \tag{3}$$

$$A = \frac{\mu_0}{4\pi}\frac{m}{D^3} \tag{4}$$

where is the underlying light power, V is the light ingestion coefficient, and R is the separation from the source; and vacuum penetrability, D is the upward separation from the attractive dipole.

$$J(v) = U \sim \exp(R) \tag{5}$$

$$A(D) = V \sim lo\ \max(\lambda, 2) \tag{6}$$

The J(V) tracks an exponential distribution and A(D) follows a Lomax or Pareto type II circulation. Therefore, the final form of the optimization reckoning is given as follows:

$$y_j^{(s+1)} = \begin{cases} y_j^{(s)} + \alpha \times u, if\ rand \le \beta, \\ y_j^{(s)} + \alpha \times v, otherwise; \end{cases} \tag{7}$$

The above reckoning (Equation (5)) shows that the steering of monarch refugees is a Markov chain (MC) process where the next state ($y_j^{(s+1)}$) depends only on the current state ($y_j^{(s)}$) and the step size is finite. Sine-cosine calculation (SCA) is a gathering of calculations as of late presented by Mirza, and he likewise proposed a few multitude procedures like dark wolf streamlining and whale improvement. SCA follows the very idea as other help strategies in that it comprises of an underlying populace of irregular possible arrangements, and afterward this underlying populace is refreshed in view of a bunch of conditions until it meets a halting basis. SCA uses the sine equation or the cosine equation to update the individuals (possible solution). This modified state provides the learning and exploitation of global optimization depending on the current state and some random numbers.

$$e_S = \frac{1}{N_S}\sum_{j=1}^{N}\left(l_j^r - l_j^c\right)^2 \tag{8}$$

where e_S is the average error between the true l_j^r and l_j^c the estimated and is the true and j-th peak load, respectively, and N is the total number of data in the set. In SC-DA, the state of each agent is initialized as follows. In the next iteration, the agents' states are updated with the following two equations with some probabilities:

$$Y_j^{s+1} = \begin{cases} Y_j^s + D_1 \sin(D_2)\left|D_3 w_j^s - Y_j^s\right|, D_4 < D \\ Y_j^s + D_1 \cos(D_2)\left|D_3 w_j^s - Y_j^s\right|, D_4 \geq D \end{cases} \tag{9}$$

where Y_j^{s+1} is the ongoing place of the jth layered arrangement at thedsth iteration; D_2 random number from 0 to 2π; D_3 and D_4 are two uniformly distributed random numbers in the range [0, 1]; D is the threshold, which is set to 0.5; Excellent solution level; The equation D_1 is shown as follows

$$D_1 = b - b\frac{s}{S} \tag{10}$$

where a is constant (0, 2); s characterizes the current number of repetitions and S represents the maximum number of iterations. The steps complicated in the MSC based cluster formation are described in Algorithm 1.

$$Tn_{r-energy} = Tn_{B-energy} - Tn_{c-energy} \tag{11}$$

where, $Tn_{r-energy}$ presents the senor node residual energy, $Tn_{B-energy}$ average energy and C-energy consumed energy.

$$Tn_{c-energy} = SA_b \times Tn_{s-energy} + rA_a \times Tn_{r-energy} \tag{12}$$

where A_b , rA_a are transmitted and received bits in sensors also, $Tn_{s-energy}$ mean the all out worth of received and transmitted energy.

$$Tn_{S-energy} = Tn_{Sn-energy} + Tn_{SB-energy} \times dis \tag{13}$$

$$Tn_{r-energy} = Tn_{Sn-energy} \tag{14}$$

$Tn_{Sn-energy}$ and $Tn_{SB-energy}$ provide the transmitter with the necessary energy. 'd' is the coldness among the next forwarder and the source node, and energy refers to the transmitter amplifier with the necessary Sn-energy and SB-energy. The connection quality is the next characteristic to be calculated after the power level, and it is used to gauge how well data is transferred.

$$lP = \begin{cases} 1 & f \quad lqa_i > lqa_{optimal} \\ 0 & \dfrac{lqa_{\max} - lqa_i}{lqa_{\max}} \quad if \quad lqa_{worst} > lqa_i < lqa_{optimal} \\ & if \quad lqa_i > lqa_{worst} \end{cases} \tag{15}$$

$Lqa_{optimal}$ and Lqa_{worst} are utilized for associated and detached states. Besides, these limits additionally ordered into connect I for separated. Whenever the beneficiary (rv_j) sensor hub gets a bundle (lqa_i) or not exactly Lqaworst or this is grouping into associated connect (lqa_i) when

Input	: $y_j^{(s+1)}$ parameters (location, distance between cluster members to base station)
Output	: D and w_j^s value
1	Generate initial population of parameters
2	The time compensated sun compass $y_j^{(s+1)} = y_j^{(s)} + \alpha \times \delta_s$
3	The η is the jump size be contingent upon the magnetic field strength
4	For j=1, i=0
5	Update position $y_j^{(s+1)} = \begin{cases} y_j^{(s)} + \alpha \times u, & if\ rand \leq \beta, \\ y_j^{(s)} + \alpha \times v, & otherwise; \end{cases}$
6	Exploration and exploitation for global optimization $e_S = \frac{1}{N_S}\sum_{j=1}^{N}(l_j^r - l_j^c)^2$
7	Update the final values of b
8	end if
9	End

4.2. CH selection and data aggregation

4.2.1. Design constraints for trust degree computation

Next, we utilize multiple design constraints for node trust degree computation, energy efficiency, link quality, path loss, distance between target node to base station and aggregation delay. Energy consumption is a major concern in the design of routing protocol. Energy ingesting and reduction in sensor nodes can be caused by a variety of factors, including routing, making bad or complicated path choices, meddling, path loss, and bit error rate. It is challenging to refresh or substitute the series of the sensor nodes in these networks because of the internal placement of the sensor nodes in the human body. Minor or complicated surgery is the only option for changing the internal sensor matrix.The sensor node with the highest power will be the next forward data propagator for determining the energy level. The energy level of a sensor node is described as follows:

worth is higher than $Lqa_{optimal}$ or, this is characterizing into momentary for (lqa_i) range among $Lqa_{optimal}$ and Lqa_{worst}. At the point when remote correspondence happens then there is path loss.

$$Ql_{(F,d)} = Ql_o + 10N\log_{10}\frac{d}{do} + T \tag{16}$$

Ql_o is the way misfortune at a reference distance and is numerically characterized as

$$Ql_o = 10\log_{10}\left[\frac{4\pi dF}{C}\right]^2 \tag{17}$$

Update the fitness value of Ql_o.

$$Ql_{(F,d)} = 10\log_{10}\left[\frac{4\pi dF}{C}\right]^2 + 10N\log_{10}\frac{d}{do} + T \tag{18}$$

Table 1
Summary of Research gaps.

Ref.	Year	Methodology	Technique used	Parameter improved	Research gaps
[21]	2022	Secure data aggregation	ANT-PSO-AODV	Energy consumption, routing overhead	However, it fails to handle data redundancy and has lower throughput.
[22]	2019	Data aggregation for energy constrained IoT	Laplacian eigen basis for weight adjacency matrix in data aggregation	Throughput, packet delivery ratio	Ignore the computationally complex energy efficiency aspect of the node.
[23]	2022	Efficient data aggregation scheme	I-LEACH	Throughput, end-to-end delay	High computational complexity due to LEACH
[24]	2021	Fuzzy-based data aggregation	Fuzzifcated LEACH	Energy consumption, aggregation ratio	Real-time implementation becomes expensive.
[25]	2022	Secure and efficient multifunctional data aggregation	Lightweight cryptographic technique	Energy consumption, overhead, aggregation ratio	Several aggregates determine the energy efficiency of data storage.
[26]	2022	Cluster-based data aggregation	βDSC^{2}DAM	Data aggregation time, average latency, end-to-end delay	This scheme is built using point-to-point bilinear mapping and hash function operations, so it is not efficient.
[27]	2019	FESDA	Paillier crypto-system	Energy consumption, aggregation ratio	It depends solely on data protection principles
[28]	2020	MMDA-IoT	CNN	Data aggregation time	Not suitable for high density nodes.
[29]	2021	Privacy preserving data collection (PPDC)	Cluster based k-anonymity	Execution time	Heavy computational complexity and poor data utility
[30]	2020	EMCA-CS	CSMO-GWO	Lifetime, energy consumption	Affected by data dimensionality issues

In the distance boundary with numerical model, $F^{dist}(q)$ indicates the distance among two ordinary hubs. $F^{dist}(p)$ shows the distance between target node and CH of organization. The best maximum optimal solution of $F^{dist}(p)$ ought to be inside the range [0, 1].

$$F^{dist} = \frac{F^{dist}(p)}{F^{dist}(q)} \tag{19}$$

$$F^{dist}(p) = \sum_{j=1}^{l}\sum_{i=1}^{m}\left\|d_j - CH_i\right\| + \left\|CH_i - A_i\right\| \tag{20}$$

$$F^{dist}(q) = \sum_{j=1}^{l}\sum_{i=1}^{m}\left\|d_j - d_i\right\| \tag{21a}$$

Specifies the data aggregation delay of the nodes, where the delay value should be in the range [0, 1]. Reducing the number of nodes in cluster also reduces the resulting latency. CH represents the numerator value and each node range represents the denominator value.

$$F^{delay} = \frac{\max\limits_{i=1}^{M}(CH_i)}{l} \tag{22a}$$

4.2.2. Design metric optimization

In data aggregation phase, we utilize the multiple design metrics to compute the trust degree of each IoT sensors and design an improved sunflower optimization (ISFO) algorithm to optimize the design constraints. The highest trust degree owned node is act as CH of the cluster which ensures data aggregation. Based on our objective, we formulate the following optimal function for decision making.

$$TD = \min\left(Tn_{energy}, LP, E^{dist}, F^{delay}\right) \cup \max\left(QI_{F,d}\right) \tag{23a}$$

In general, sunflower optimization (SFO) procedure is used to find the best orientation to the sun, taking into account the special nature of sunflowers. In SFO procedure, the fertilization process is replicated based on random seed production considering the minimum coldness among flower j and flower J+1. Though, each natural flower has millions of pollen gametes, and in order to obtain a fast answer in the optimization, the algorithm considers only one pollen gamete, which is reproduced individually for each sunflower. SFO algorithm is one of the latest evolutions of nature-inspired soft computing algorithms. The cycle of sunflowers is always the same: they wake up every day and come with the sun like the hands of a clock. In this study, we make changes in the fitness optimizations to update the basic SFO algorithm into improved sunflower optimization (ISFO) algorithm. In ISFO algorithm, define the travel in the opposite direction during the night and leave the next morning. The law of radiation governs the sunflower cycle.

$$P_Y = \frac{Q_Y}{4\pi R_Y^2} \tag{24a}$$

where, P_Y is the heat intensity received by each sunflower (y); Q_Y is the ideal individual R_Y is the distance between each individual solar energy and the current population. An inverse quadratic relationship between radiant heat and distance defines,

$$\overrightarrow{T} = \frac{Y^* - Y_Y}{\left\|Y^* - Y_Y\right\|}, y = 1, 2, \ldots, nq \tag{25a}$$

where Y^* is the ideal individual in the current population, Y_Y denotes each solution, and nq is the proposed population size. The movement of the sunflower on the sun is represented as follows.

$$D_y = \lambda.Q_Y(\left\|Y_Y + Y_{Y-1}\right\|)\cdot\left\|Y_Y + Y_{Y-1}\right\| \tag{26a}$$

Here, λ is a finite constant associated with the inertial displacement of each sunflower; $Q_Y(\left\|Y_Y + Y_{Y-1}\right\|)$ is the sunflowers are prone to pollination. In ISFO algorithm, pollination takes place randomly, with short distances between each flower and the outgoing flower. As a result, sunflower pollination takes place in a new state, where sunflowers closer to the sun make small movements to develop local pollination, while other sunflowers move normally. Based on the above, the mechanism for updating the position of each sunflower is implemented by moving the sunflowers (D_Y) and their orientation to the sun (T_Y) as follows:

$$Y_{Y+1} = Y_Y + D_Y \times T_Y \tag{27a}$$

ISFO algorithm is highly sensitive to two defined parameters: pollination rate (QR) and mortality rate (M). This high sensitivity does not provide the best ability to detect sunflowers. Also, the inertial displacement of each sunflower is fixedly defined, which limits the search behavior of ISFO algorithm. Both of these controls do not support ISFO algorithm search checks. In this article, two modifications are proposed to overcome the limitations mentioned above. The first change suggests changing the pollination rate (QR) from fixed defined value to an adaptive value.

$$QR = 0.5 \times \left(1 - \frac{W}{\max_w}\right) \tag{28a}$$

It describes the coefficient vector (*QR*) decreasing linearly from 0.5 to 0 iterations. The second variable describes the passive displacement (λ) corresponding to each sunflower.

$$\lambda = (u_a - l_a) \times \left(1 - \frac{W}{\max_w}\right) \tag{29a}$$

Here u_a and l_a are the upper and lower bounds of the decision variables. W is the current iteration. In this case, to establish equilibrium between search and misuse in the algorithm, the search starts with a high deviation search, and in the final step it explorations locally in the search space. This progression is designed as shadows

$$\vec{z}_{j+1}^{New} = \begin{cases} \vec{z}_{j+1} + \gamma \times T_j \times \vec{d}_j \times F\left(\vec{z}_{j+1}\right), rand > 0.5 \\ \vec{z}_{j+1} - \gamma \times T_j \times \vec{d}_j \times F\left(\vec{z}_{j+1}\right), rand \le 0.5 \end{cases} \tag{30a}$$

Applying this system to the total quantity of plants, the subsequent improved version of fitness is describe as follows:

$$n_{q+1}^{New} = n_q^{New} + \beta_j \times n_q^{New} \tag{31}$$

$$\beta_{j+1} = 4 \times \left(\beta_j - \beta_j^2\right) \tag{32}$$

where, β_j designates the value for the j-th chaotic repetition, and the Initial value β_j designates a random value in the range 0 and 1. The working steps involved in the design constraints optimization and CH selection is given in Algorithm 2.

with respect to quality of service (QoS), which happens because of inefficient data aggregations. They also fail to extend the network lifetime and reduce energy depletion.In order to ensure optimal data transfer between CHs, we introduce reformative optimal–learning-based deep neural network (ROL-DNN) technique which compute optimal path between source and destination.The ROL-DNN technique considers the solution population as a group of learners who learn in two main stages: learning between teachers; and student-generated learning from peer interactions. The difference between the means for each student is calculated using the initial population created with the best study variables. This helps update the underlying populace of the instructor matrix

$$Diff \quad in \quad mean_{i,g,j} = K_j \times \left(Q_{i,kbest,}j - M_f - M_f \times s_{i,j}\right) \tag{21b}$$

A K_j student's best performance in subject j is a random number between 0 and 1 and $Q_{i,kbest,}$ TF is a teaching factor between 1 and 2 that can be calculated using.

$$F_f = round\left[1 + rand(0,1)\right. \tag{22b}$$

For specimen, if the accidental value is 0.5, then Sf = 2. Likewise, if random worth is 0 then Sf = 1

$$Q_{i,g,j}^{'} = Q_{i,g,j} + diff \quad in \quad mean_{i,g,j} \tag{23b}$$

where $Q_{i,g,j}^{'}$ is the effective worth of $Q_{i,g,j}$. It is utilized to refresh the arrangement got from the underlying populace. New wellness values were gotten and contrasted and the wellness upsides of the underlying populace. The best wellness values are passed to the student stage as starting values.This finishes the instructor phase.Consider two haphazardly chosen students P and Q, separately,

Input	: P_Y parameters
Output	: Initial value β_j
1	Generate initial population of parameters
2	The law of radiation manages the cycle of a sunflower $P_Y = \frac{Q_Y}{4\pi R_Y^2}$
3	While Do
4	For j=1, i=0
5	The sunflowers move across the sun $D_y = \lambda . Q_Y \left(\| Y_Y + Y_{Y-1} \|\right) \cdot \| Y_Y + Y_{Y-1} \|$
6	Define local search space $\vec{z}_{j+1}^{New} = \begin{cases} \vec{z}_{j+1} + \gamma \times T_j \times \vec{d}_j \times F\left(\vec{z}_{j+1}\right), & rand > 0.5 \\ \vec{z}_{j+1} - \gamma \times T_j \times \vec{d}_j \times F\left(\vec{z}_{j+1}\right), & rand \le 0.5 \end{cases}$
7	Find inertial displacement $\lambda = \left(u_a - l_a\right) \times \left(1 - \frac{W}{\max_w}\right)$
8	Update fitness using $n_{q+1}^{New} = n_q^{New} + \beta_j \times n_q^{New}$
9	Update the final values of b
10	End

4.3. 4.3 optimal path computation

After data aggregated in CH nodes, the data should transfers between two IoT nodes which need energy efficient and high network lifetime. Traditional sensor network platforms were created with a focus on energy consumption at the expense of communication throughput.To collect aggregated data, this has high transmission throughput and low delay to reach destination end.According to the investigation, conventional path selection algorithm fails to deliver improved data transfer

$$Q_{i,q,j}^{''} = Q_{i,q,j}^{'} + K_j \times \left(Q_{i,q,j}^{'} - Q_{i,q,j}^{'}\right) \; if \; Q_{total\ q,j}^{'} < Q_{total\ P,j}^{'} \tag{24b}$$

$$Q_{i,q,j}^{''} = Q_{i,q,j}^{'} + K_j \times \left(Q_{i,q,j}^{'} + Q_{i,q,j}^{'}\right) \; if \; Q_{total\ q,j}^{'} < Q_{total\ P,j}^{'} \tag{25b}$$

DNN can use their internal state to process variable-length input sequences. ROL-DNNcan be thought of as multiple replicas of the same network, each sending a message to its successor. They can relate

previous information to the current task. However, when this gap increases, DNN cannot learn to integrate information. The short-term memory problem of DNNis that short-term memory has high influence, but long-term memory has little influence. The main reason is that the structure of ROL-DNN allows information to be passed each time g_s.

$$k_s^{(m)} = \sigma\left(Z_s^{(m)} \cdot \left[k_{s-1}^{(m)}, y_s^{(m)}\right] + V_s^{(m)} k_{s-1}^{(m)} + A_f^{(m)}\right) \quad (26b)$$

However, one of the problems with simple ROL-DNNis that end nodes are too weak to extract information from time nodes at large intervals. ROL-DNN offers a solution to the long-term bias problem by using a gating system to control information flow and loss. The network is trained using dynamic back propagation. It is a gradient descent learning algorithm based on the assumption that the initial state of the network is independent of the initial weights. Let's $d(m+1)$ specify the target value of the network $(s+1)$. Defined as the network error at that time $(s+1)$.

$$E(s+1) = d(m+1) - x(s+1) \quad (27b)$$

The network cost function is the squared error between the actual values and the predicted values given by the following equation:

$$I(m+1) = \frac{1}{2}[E(m+1)]^2 \quad (28b)$$

For all nodes in the current stack, the cumulative output error time is compute using iteration as follows:

$$e_K = {}^1/_{Mtr} \sum_{j=1}^{Mtr} o_j - z_{Kj} \quad (29b)$$

where z_{Kj} the kth node w represents the output with respect to j-th data value. M_{tr} is a training set. Calculate weight changes and update the weight value.

Table 2
Simulation setup.

Parameters	Values
Simulation Area	1000 m × 1000 m
Number of IoT sensors	200–1000
Data size	4000 bits
Control packet size	200 bits
Senor sensing range	80 m
Initial energy of sensor nodes	2J
MAC protocol	CSMA/CA
Bandwidth	250 kb/s
Payload size	30 bytes
Transmission range	100 m
Avg. energy consumption of transmitting node	24.92 mJ per byte
Avg. energy consumption of receiving node	19.72 mJ per 1 byte
Simulation time	30 times

current state. Algorithm 3 describes the steps involved in the optimal path computation for data transfer using ROL-DNN technique.

5. Results and comparative analysis

In this segment, we validate our future cluster based reliable data aggregation (CRDA) scheme with the different simulation scenarios are impact of node density and simulation rounds. The entire simulation is carried in Network Simulator (NS2) tool. The simulation results of proposed CRDA plot is contrasted and the current condition of-workmanship plans, upgraded edge delicate stable political race convention (ETSSEP) [33], stable energy productive bunching convention (SEECP) [32], dependability Improved-Drain (SILEACH) [31] and effective multi-jump cluster based collection (EMCA-CS) [30]. The presentation of proposed and it are investigated through the different to exist plans measures are energy consumption, aggregation delay, network lifetime, throughput, overhead and data transfer rate (see Table 1).

Input	:number of nodes, CHs, TD, threshold condition
Output	:W_s moment
1	Generate initial population of parameters
2	This helps in updating the initial population of the teacher phase $Differencein\ \ mean_{i,k,j} = R_j \times (Y_{i,kbest,}j - S_f - S_f \times m_{i,j})$
3	For j=0, i=1 do
4	Consider two randomly selected learners P and Q, respectively, $Y''_{i,q,j} = Y'_{i,q,j} + R_j \times \left(Y'_{i,q,j} - Y'_{i,q,j}\right) if\ \ Y'_{total\ q,j} < Y'_{total\ P,j}$
5	Defined as the network error at that time $(s+1)$ $E(s+1) = D(s+1) - x(s+1)$
6	The efforts to minimize the formal impartial as follows: $obj(\theta) = \sum_j l(\hat{x}_j, x_j) + \sum_K \Omega(F_K),\ \ F_K \in f$
7	Update the final value
8	End

The purpose of ROL in DNN is to determine and optimize the loss function, i.e. the objective function, by adding weak learners using gradient descent optimization method and arbitrary differentiable loss functions. ROL seeks to reduce its formal scope to:

$$obj(\theta) = \sum_j l\left(\hat{x}_j, x_j\right) + \sum_K \Omega(F_K), F_K \in f \quad (30b)$$

where, W_s is the information about previous iterations which sent to the

5.1. Simulation setup

The sink is situated in the centre of a terrain measuring 1000 m by 1000 m, where the nodes are situated. A grid of 100 nodes is created, and the remaining 900 are distributed at random. The sensor node's transmission and reception power consumption are 24.92 and 19.72 mJ per byte, respectively, according to the simulation requirements. IoT sensors come in a variety of numbers, ranging from 200 to 1000. Each sensor node and each gateway are assumed to have a starting energy of

Table 3
Comparative analysis of proposed and existing data aggregation schemes with node density.

Data aggregation schemes	Energy consumption (J)					Aggregation delay (s)					Network lifetime (%)				
	200	400	600	800	1000	200	400	600	800	1000	200	400	600	800	1000
ETSSEP	589.25	748.23	769.48	847.98	897.36	15.26	20.58	31.45	45.89	59.78	56.23	45.24	42.78	40.15	38.45
SEECP	465.80	624.78	646.03	724.53	773.91	13.81	19.13	30.00	44.44	58.33	67.12	56.13	53.67	51.04	49.34
SILEACH	342.35	501.33	522.58	601.08	650.46	12.36	17.68	28.55	42.99	56.88	78.01	67.02	64.56	61.93	60.23
EMCA-CS	218.90	377.88	399.13	477.63	527.01	9.91	15.23	26.10	40.54	54.43	88.90	77.91	75.45	72.82	71.12
CRDA	95.45	254.43	275.68	354.18	403.56	4.46	9.78	20.65	35.09	48.98	99.79	88.80	86.34	83.71	82.01

Data aggregation schemes	Throughput (Mbps)					Overhead (%)					Data transfer rate (%)				
	200	400	600	800	1000	200	400	600	800	1000	200	400	600	800	1000
ETSSEP	7456	6458	5632	4785	4247	60.25	65.45	69.78	72.34	75.03	72.15	71.45	70.35	68.79	67.32
SEECP	7811	6813	5987	5140	4602	57.65	62.85	67.18	69.74	72.43	78.04	77.34	76.24	74.68	73.21
SILEACH	8166	7168	6342	5495	4957	55.05	60.25	64.58	67.14	69.83	85.93	85.23	84.13	82.57	81.1
EMCA-CS	8521	7523	6697	5850	5312	44.46	49.66	53.99	56.55	59.24	92.82	92.12	91.02	89.46	87.99
CRDA	8876	7878	7052	6205	5667	27.86	33.06	37.39	39.95	42.64	99.71	99.01	97.91	96.35	94.88

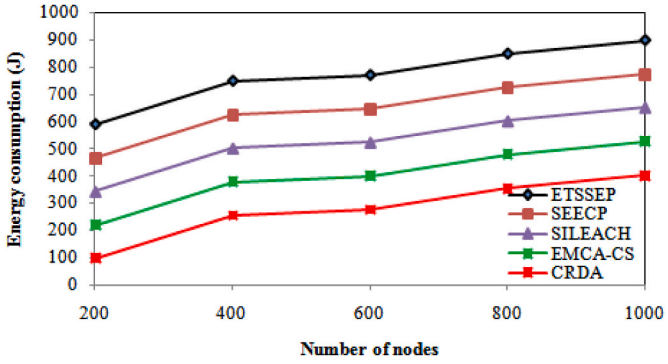


Fig. 2. Results of energy consumption with node density.

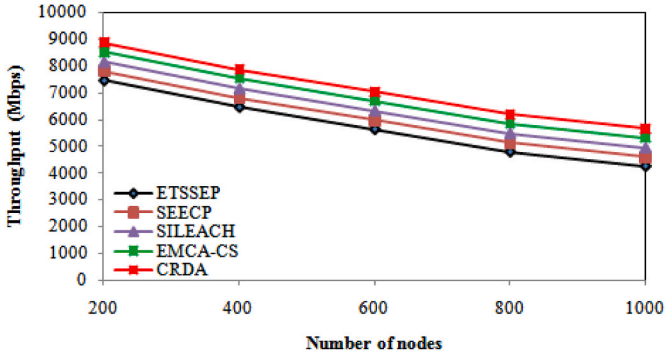


Fig. 5. Results of throughput with node density.

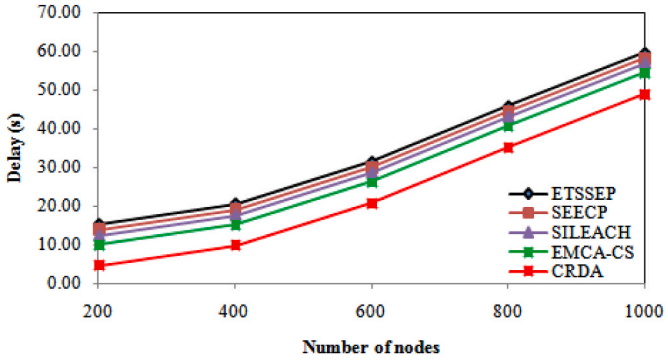


Fig. 3. Results of aggregation delay with node density.

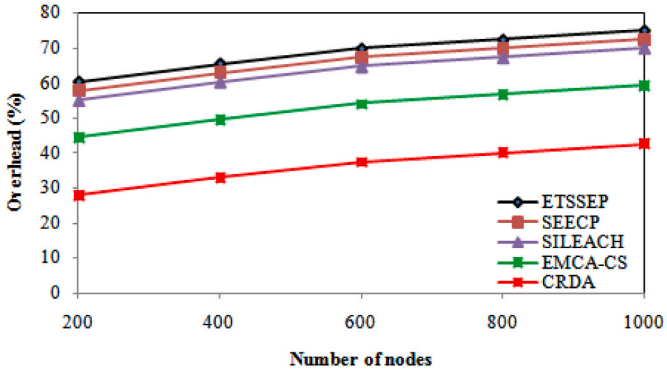


Fig. 6. Results of overhead with node density.

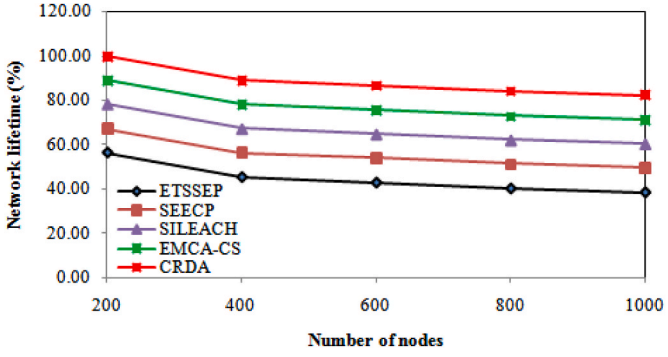


Fig. 4. Results of network lifetime with node density.

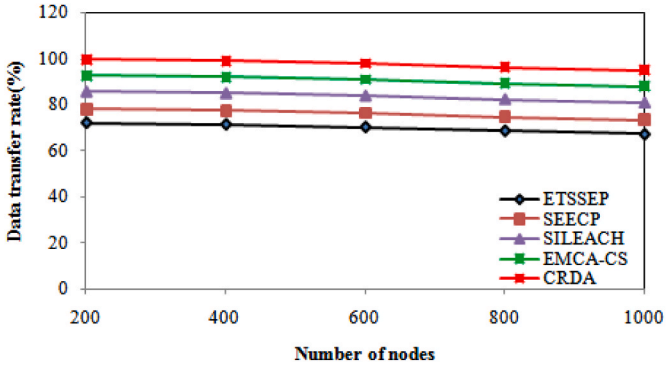


Fig. 7. Results of data transfer rate with node density.

Table 4
Comparative analysis of propose and existing data aggregation schemes with simulation rounds.

Data aggregation schemes	Energy consumption (J)					Aggregation delay (s)					Network lifetime (%)				
	500	1000	1500	2000	2500	500	1000	1500	2000	2500	500	1000	1500	2000	2500
ETSSEP	578.69	737.67	758.92	837.42	886.80	9.70	15.02	25.89	40.33	54.22	45.67	34.68	32.22	29.59	27.89
SEECP	455.24	614.22	635.47	713.97	763.35	8.25	13.57	24.44	38.88	52.77	56.56	45.57	43.11	40.48	38.78
SILEACH	331.79	490.77	512.02	590.52	639.90	6.80	12.12	22.99	37.43	51.32	67.45	56.46	54.00	51.37	49.67
EMCA-CS	208.34	367.32	388.57	467.07	516.45	4.35	9.67	20.54	34.98	48.87	78.34	67.35	64.89	62.26	60.56
CRDA	84.89	243.87	265.12	343.62	393.00	1.90	4.22	15.09	29.53	43.42	89.23	78.24	75.78	73.15	71.45

Data aggregation schemes	Throughput (Mbps)					Overhead (%)					Data transfer rate (%)				
	500	1000	1500	2000	2500	500	1000	1500	2000	2500	500	1000	1500	2000	2500
ETSSEP	7446	6448	5622	4775	4237	49.69	54.89	59.22	61.78	64.47	61.59	60.89	59.79	58.23	56.76
SEECP	7801	6803	5977	5130	4592	47.09	52.29	56.62	59.18	61.87	67.48	66.78	65.68	64.12	62.65
SILEACH	8156	7158	6332	5485	4947	44.49	49.69	54.02	56.58	59.27	75.37	74.67	73.57	72.01	70.54
EMCA-CS	8511	7513	6687	5840	5302	33.90	39.10	43.43	45.99	48.68	82.26	81.56	80.46	78.90	77.43
CRDA	8866	7868	7042	6195	5657	17.30	22.50	26.83	29.39	32.08	89.15	88.45	87.35	85.79	84.32

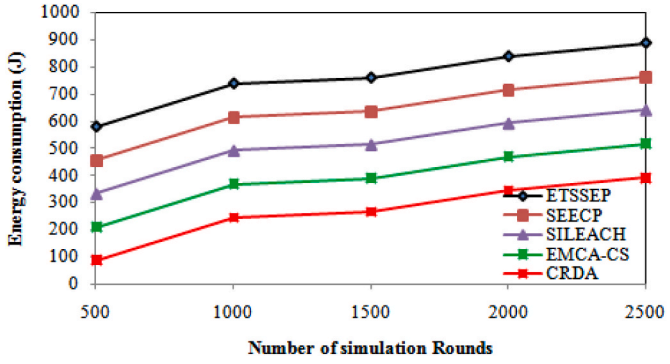


Fig. 8. Results of energy consumption with simulation rounds.

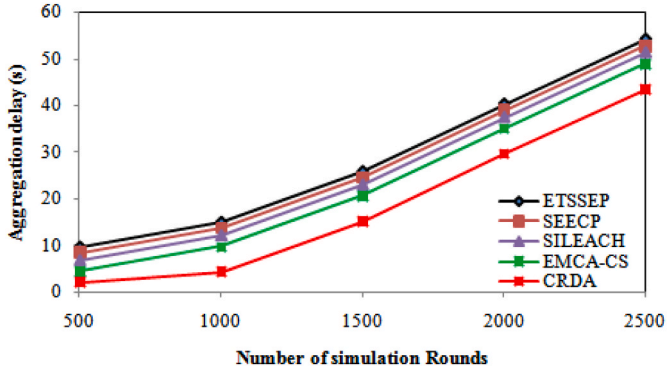


Fig. 9. Results of aggregation delay with simulation rounds.

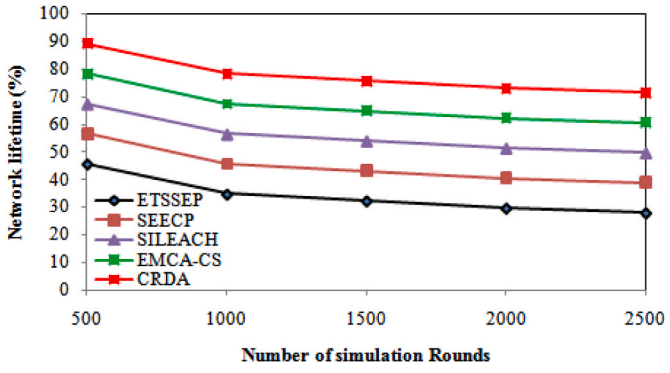


Fig. 10. Results of network lifetime with simulation rounds.

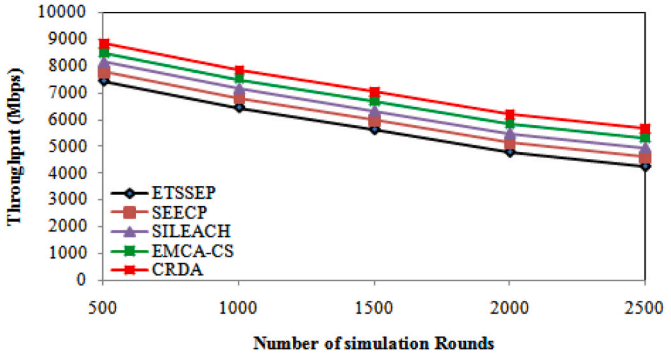


Fig. 11. Results of throughput with simulation rounds.

2J. Each node communicates with the others via the CSMA/CA MAC layer protocol. To send the data packets to the gateways, the sensor nodes employ the TDMA that the BSs choose. The IoT sensor nodes' transmission range is 100 m, and the data packet size is 4000 bits. The simulations are run 30 times, and the accompanying graph shows the typical outcome of the runs. Table 2 describes the summary of simulation setup.

5.2. Comparative analysis

5.2.1. Impact of IoT nodes

The purpose of the section is to validate the performance of proposed and existing data aggregation schemes with the following simulation setup, node density of 200–1000 and a fixed network area of 1000 × 1000 m^2. Our CRDA arrangement is associated with existing state-of-the-art ETSSEP [33], SEECP [32], SILEACH [31] and EMCA-CS [30]

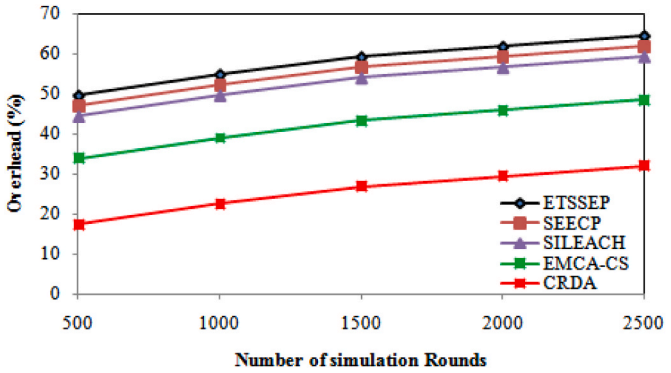


Fig. 12. Results of overhead with simulation rounds.

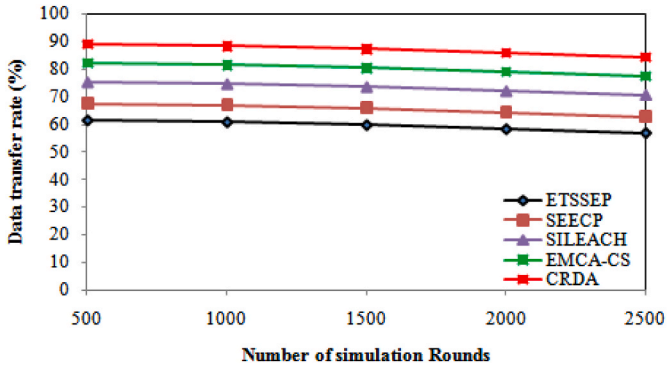


Fig. 13. Results of data transfer rate with simulation rounds.

schemes in Table 3. A comparison of existing and proposed aggregation schemes for energy consumption is shown in Fig. 2. In CRDA scheme, we achieve 64.092%, 57.24%, 47.158%, and 30.854% lowerenergy consumption than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for aggregation delay is shown in Fig. 3. In CRDA scheme, we achieve31.221%, 28.212%, 24.927% and 18.638% loweraggregation delay than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for network lifetimeis shown in Fig. 4.

In our CRDA scheme, we achieve49.427%, 37.07%, 24.713%, and 12.357% higher network lifetime than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for throughput is shown in Fig. 5. In our CRDA scheme, we achieve 19.9%, 14.925%, 9.95% and 4.975% higher throughput than those existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for overhead is shown in Fig. 6. In our CRDA scheme, we achieve 47.239%, 45.162%, 42.913% and 31.45% minimum overhead than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for data transfer rates shown in Fig. 7. In our CRDA scheme, we achieve 28.246%, 22.209%, 14.123% and 7.061% maximum data transfer rate than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes.

Our CRDA scheme has been shown to improve performance with increasing node density in IoT networks. IoT nodes enable wireless connections between objects, but their small size, limited battery life, and resource constraints make handling the large amounts of data they generate a challenge. To address these constraints, many researchers have focused on optimizing energy consumption to extend network lifetime. However, these methods often result in reduced throughput and do not effectively handle data redundancy. Our CRDA scheme addresses these issues by eliminating data redundancy and conserving energy. From this comparative analysis, we conclude that the performance of our CRDA scheme is improved with respect to node density.

5.2.2. Impact of simulation rounds

The purpose of the section is to validate the performance of data aggregation schemes with the following simulation setup, simulation rounds as 500 to 2500 and a fixed number of nodes as 500. Our proposed CRDA arrangement is associated with existing state-of-the-art ETSSEP [33], SEECP [32], SILEACH [31] and EMCA-CS [30] schemes in Table 4. A comparison of existing and proposed aggregation schemes for energy consumption is shown in Fig. 8. In our CRDA scheme, we achieve 67.813%, 61.243%, 51.301% and 34.5% lower energy consumption than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for aggregation delay is shown in Fig. 9. In CRDA scheme, we achieve 44.25%, 40.44%, 36.07% and 27.35% lower aggregation delay than those of existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for network lifetime is shown in Fig. 10.

In our CRDA scheme, we achieve 55.07%, 41.302%, 27.535% and 13.767% higher network lifetime than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for throughput is shown in Fig. 11. In our CRDA scheme, we achieve 18.952%, 14.214%, 9.476% and 4.738% higher throughput than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for overhead is shown in Fig. 12. In our CRDA scheme, we achieve 57.434%, 55.377%, 53.111% and 40.882% minimum overhead than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. A comparison of existing and proposed aggregation schemes for data transfer rate is shown in Fig. 13. In our CRDA scheme, we achieve 31.431%, 24.713%, 15.715% and 7.858% maximum data transfer rate than those of the existing ETSSEP [33], SEECP [32], SILEACH [31], and EMCA-CS [30] schemes. To alleviate the growing network traffic in IoT applications, there is a need for systems that can reduce data transmission flux due to inherent sensor constraints. Prediction and data aggregation techniques, as described in Refs. [30–34], offer promising solutions for meeting this requirement, especially for forecasting processes that require large amounts of data collection, transmission, and recording. However, these solutions face challenges with communication overhead when integrating IoT with other technologies. Our CRDA scheme has been shown to effectively address this issue by reducing communication overhead in simulation.

6. Conclusion

We have planned a cluster based reliable data aggregation (CRDA) scheme for IoT network which ensures data collection and aggregation as energy efficient manner and transfer to another end very effectively. An MSC algorithm is used for clustering which groups the IoT sensor which ensures the effective data transferring. The multiple design metrics are used to compute the trust degree of each IoT sensors and ISFO algorithm is used to optimize design constraints and compute CHs. ROL-DNN technique is then introduced to compute routes between IoT sensors which ensures reliable data aggregation and transferring.From the simulation results, we observed that the energy ingesting of our CRDAsystem is 49.836% and 51.872% efficient than the existing data aggregation schemes with respect to impact of nodes and simulation rounds, respectively. The aggregation delay of our CRDA scheme is 25.75% and 37.283% efficient than the existing data aggregation schemes with respect to impact of nodes and simulation rounds, respectively. The network lifetime of our CRDA scheme is 30.892% and 34.712% efficient than the existing data aggregation schemes with respect to impact of nodes and simulation rounds, respectively. The throughput of proposedCRDA scheme is 12.438% and 15.889% efficient than the existing data aggregation schemes with respect to impact of nodes and simulation rounds, respectively.The overhead of proposedCRDA scheme is 41.691% and 51.115% efficient than the existing data aggregation schemes with respect to impact of nodes and simulation rounds, respectively. The data transfer rate-of proposedCRDA scheme is 17.91% and 22.34% efficient than the existing data aggregation schemes with respect to impact of nodes and simulation rounds, respectively.Our simulation results demonstrate that proposed CRDA-scheme extends network lifetime, decreases energy consumption, and improves quality of service. Going forward, we plan to extend our proposed CRDA scheme for real-time applications such as E-learning, smart agriculture, and smart healthcare. This will ensure reliable data transfer and we will also validate it with a new cryptographic technique.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] Y. Shi, J. Han, X. Wang, J. Gao, H. Fan, An obfuscatable aggregatable sign cryption scheme for unattended devices in IoT systems, IEEE Internet Things J. 4 (2017) 1067–1081.

[2] N. Kouzayha, M. Jaber, Z. Dawy, Measurement-based signaling management strategies for cellular IoT, IEEE Internet Things J. 4 (2017) 1434–1444.

[3] H. Yu, B. Ng, W.K. Seah, On-demand probabilistic polling for nanonetworks under dynamic IoT backhaul network conditions, IEEE Internet Things J. 4 (2017) 2217–2227, https://doi.org/10.1109/JIOT.2017.2751524.

[4] E. Fitzgerald, M. Pióro, A. Tomaszwski, Energy-optimal data aggregation and dissemination for the Internet of Things, IEEE Internet Things J. 5 (2018) 955–969, https://doi.org/10.1109/JIOT.2018.2803792.

[5] Z. Qin, D. Wu, Z. Xiao, B. Fu, Z. Qin, Modeling and analysis of data aggregation from convergecast in mobile sensor networks for industrial IoT, IEEE Trans. Ind. Inf. 14 (2018) 4457–4467.

[6] S. Badsha, I. Khalil, X. Yi, M. Atiquzzaman, Designing privacy-preserving protocols for content sharing and aggregation in content centric networking, IEEE Access 6 (2018) 42119–42130.

[7] J. Shi, H. Zhang, Y. Bai, G. Han, G. Jia, A novel data aggregation preprocessing algorithm in flash memory for IoT based power grid storage system, IEEE Access 6 (2018) 57279–57290, https://doi.org/10.1109/ACCESS.2018.2873685.

[8] S. Sanyal, P. Zhang, Improving quality of data: IoT data aggregation using device to device communications, IEEE Access 6 (2018) 67830–67840, https://doi.org/10.1109/ACCESS.2018.2878640.

[9] J. He, L. Cai, P. Cheng, J. Pan, L. Shi, Distributed privacy-preserving data aggregation against dishonest nodes in network systems, IEEE Internet Things J. 6 (2018) 1462–1470, https://doi.org/10.1109/JIOT.2018.2834544.

[10] R. Li, C. Sturtivant, J. Yu, X. Cheng, A novel secure and efficient data aggregation scheme for IoT, IEEE Internet Things J. 6 (2018) 1551–1560, https://doi.org/10.1109/JIOT.2018.2848962.

[11] X. Li, S. Liu, F. Wu, S. Kumari, J.J. Rodrigues, Privacy preserving data aggregation scheme for mobile edge computing assisted IoT applications, IEEE Internet Things J. 6 (2018) 4755–4763, https://doi.org/10.1109/JIOT.2018.2874473.

[12] B. Yin, X. Wei, Communication-efficient data aggregation tree construction for complex queries in IoT applications, IEEE Internet Things J. 6 (2018) 3352–3363, https://doi.org/10.1109/JIOT.2018.2882820.

[13] J. Zhang, Y. Zong, C. Yang, Y. Miao, J. Guo, LBOA: location-based secure outsourced aggregation in IoT, IEEE Access 7 (2019) 43869–43883, https://doi.org/10.1109/ACCESS.2019.2908429.

[14] Z. Guan, Y. Zhang, L. Wu, J. Wu, J. Li, Y. Ma, J. Hu, APPA: an anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT, J. Netw. Comput. Appl. 125 (2019) 82–92, https://doi.org/10.1016/j.jnca.2018.09.019.

[15] H. Ko, J. Lee, S. Pack, C.G.- E2S2, Consistency-guaranteed and energy-efficient sleep scheduling algorithm with data aggregation for IoT, Future Generat. Comput. Syst. 92 (2019) 1093–1102, https://doi.org/10.1016/j.future.2017.08.040.

[16] S. Tonyali, K. Akkaya, N. Saputro, A.S. Uluagac, M. Nojoumian, Privacy-preserving protocols for secure and reliable data aggregation in IoT-enabled smart metering systems, Future Generat. Comput. Syst. 78 (2018) 547–557, https://doi.org/10.1016/j.future.2017.04.031.

[17] T. Li, C. Gao, L. Jiang, W. Pedrycz, J. Shen, Publicly verifiable privacy-preserving aggregation and its application in IoT, J. Netw. Comput. Appl. 126 (2019) 39–44, https://doi.org/10.1016/j.jnca.2018.09.018.

[18] O.R. Merad Boudia, S.M. Senouci, M. Feham, Secure and efficient verification for data aggregation in wireless sensor networks, Int. J. Netw. Manag. 28 (2018) e2000, https://doi.org/10.1002/nem.2000.

[19] M.H. Homaei, E. Salwana, S. Shamshirband, An enhanced distributed data aggregation method in the Internet of Things, Sensors 19 (2019) 3173, https://doi.org/10.3390/s19143173.

[20] W. Tang, J. Ren, K. Deng, Y. Zhang, Secure data aggregation of lightweight E-healthcare IoT devices with fair incentives, IEEE Internet Things J. 6 (2019) 8714–8726, https://doi.org/10.1109/JIOT.2019.2923261.

[21] N. Chandnani, C.N. Khairnar, Bio-inspired multilevel security protocol for data aggregation and routing in IoT WSNs, Mobile Network. Appl. 27 (3) (2022) 1030–1049, https://doi.org/10.1007/s11036-021-01859-6.

[22] M. Amarlingam, K.D. Prasad, P. Rajalakshmi, S.S. Channappayya, C.S. Sastry, A novel low-complexity compressed data aggregation method for energy-constrained IoT networks, IEEE Transactions on Green Communications and Networking 4 (2020) 717–730, https://doi.org/10.1109/TGCN.2020.2966798.

[23] V.S. Badiger, T.S. Ganashree, Data aggregation scheme for IOT based wireless sensor network through optimal clustering method, Measurement: Sensors 24 (2022), 100538, https://doi.org/10.1016/j.measen.2022.100538.

[24] S.N. Sajedi, M. Maadani, M. NesariMoghadam, F.- Leach, A fuzzy-based data aggregation scheme for healthcare IoT systems, J. Supercomput. 78 (2022) 1030–1047, https://doi.org/10.1007/s11227-021-03890-6.

[25] Q. Wu, F. Zhou, J. Xu, Q. Wang, D. Feng, Secure and efficient multifunctional data aggregation without trusted authority in edge-enhanced IoT, J. Inf. Secur. Appl. 69 (2022), 103270, https://doi.org/10.1016/j.jisa.2022.103270.

[26] A.R. Khan, M.A. Chishti, βDSC2DAM: beta-dominating set centered cluster-based data aggregation mechanism for the Internet of things, J. Ambient Intell. Hum. Comput. (2022) 1–18, https://doi.org/10.1007/s12652-021-03692-x.

[27] A. Saleem, A. Khan, S.U.R. Malik, H. Pervaiz, H. Malik, M. Alam, A. Jindal, FESDA: fog-enabled secure data aggregation in smart grid IoT network, IEEE Internet Things J. 7 (2019) 6132–6142, https://doi.org/10.1109/JIOT.2019.2957314.

[28] P. Zeng, B. Pan, K.K.R. Choo, H. Liu, MMDA: multidimensional and multidirectional data aggregation for edge computing-enhanced IoT, J. Syst. Architect. 106 (2020), 101713, https://doi.org/10.1016/j.sysarc.2020.101713.

[29] J.A. Onesimu, J. Karthikeyan, Y. Sei, An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services, Peer-to-Peer Netw. Appl. 14 (2021) 1629–1649, https://doi.org/10.1007/s12083-021-01077-7.

[30] A. Aziz, W. Osamy, A.M. Khedr, A.A. El-Sawy, K. Singh, Grey Wolf based compressive sensing scheme for data gathering in IoT based heterogeneous WSNs, Wireless Network 26 (2020) 3395–3418, https://doi.org/10.1007/s11276-020-02265-8.

[31] A.S. Al-Zubaidi, A.A. Ariffin, A.K. Al-Qadhi, Enhancing the stability of the improved-LEACH routing protocol for WSNs, J. ICT Res. Appl. (2018) 12, https://doi.org/10.5614/itbj.ict.res.appl.2018.12.1.1.

[32] N. Mittal, U. Singh, B.S. Sohi, A stable energy efficient clustering protocol for wireless sensor networks, Wireless Network 23 (2017) 1809–1821, https://doi.org/10.1007/s11276-016-1255-6.

[33] S. Kumar, S.K. Verma, A. Kumar, Enhanced threshold sensitive stable election protocol for heterogeneous wireless sensor network, Wireless Pers. Commun. 85 (2015) 2643–2656, https://doi.org/10.1007/s11277-015-2925-x.

[34] M. Alshammari, M. Al-Smadi, O.A. Arqub, I. Hashim, M.A. Alias, Residual Series representation algorithm for solving fuzzy duffing oscillator equations, Symmetry 12 (2020) 572, https://doi.org/10.3390/sym12040572.

[35] O. Abu Arqub, J. Singh, B. Maayah, M. Alhodaly, Reproducing kernel approach for numerical solutions of fuzzy fractional initial value problems under the Mittag–Leffler kernel differential operator, Math. Methods Appl. Sci. (2021), https://doi.org/10.1002/mma.7305.

[36] O. Abu Arqub, J. Singh, M. Alhodaly, Adaptation of kernel functions based approach with Atangana–Baleanu–Caputo distributed order derivative for solutions of fuzzy fractional Volterra and Fredholm integrodifferential equations, Math. Methods Appl. Sci. (2021), https://doi.org/10.1002/mma.7228.

[37] M. Gheisari, M.S. Yaraziz, J.A. Alzubi, C. Fernández-Campusano, M.R. Feylizadeh, S. Pirasteh, A.A. Abbasi, Y. Liu, C.C. Lee, An efficient cluster head selection for wireless sensor network-based smart agriculture systems, Comput. Electron. Agric. 198 (2022), 107105, https://doi.org/10.1016/j.compag.2022.107105.

[38] A.A. Movassagh, J.A. Alzubi, M. Gheisari, M. Rahimi, S. Mohan, A.A. Abbasi, N. Nabipour, Artificial neural networks training algorithm integrating invasive weed optimization with differential evolutionary model, J. Ambient Intell. Hum. Comput. (2021) 1–9, https://doi.org/10.1007/s12652-020-02623-6.

[39] M. Gheisari, H.E. Najafabadi, J.A. Alzubi, J. Gao, G. Wang, A.A. Abbasi, A. Castiglione, OBPP: an ontology-based framework for privacy-preserving in IoT-based smart city, Future Generat. Comput. Syst. 123 (2021) 1–13, https://doi.org/10.1016/j.future.2021.01.028.

[40] O.A. Alzubi, J.A. Alzubi, K. Shankar, D. Gupta, Blockchain and artificial intelligence enabled privacy-preserving medical data transmission in Internet of Things, Trans. Emerg. Telecommun. Technol. 32 (2021) 4360, https://doi.org/10.1002/ett.4360.

[41] O.A. Alzubi, J.A. Alzubi, A.M. Al-Zoubi, M.A. Hassonah, U. Kose, An efficient malware detection approach with feature weighting based on Harris Hawks optimization, Cluster Comput. 25 (2022) 2369–2387, https://doi.org/10.1007/s10586-021-03459-1.

Fog Computing: Applications, Challenges, and Opportunities

Rajanikanth Aluvalu, Lakshmi Muddana, V Uma Maheswari, Krishna Keerthi Channam, Swapna Mudrakola, MD Sirajuddin, CVR Syavasya

Abstract

Cloud computing, is a widely accepted utility computing model. All the application processing takes place in the cloud data center managed by the cloud service provider. This includes network latency and delays in processing. Each time the application is executed, data has to be transported from node to the cloud. This will increase network traffic and is practically not feasible to transport data from node to remote cloud server and back. Fog computing, a new paradigm of cloud computing will help in overcoming this challenge. In fog computing technology, the data processing tasks are executed at the node level either completely or partially, which highly increases the speed of responses. Also, it reduces latency, processing costs, and bandwidth problems, and improves the efficiency of customer driver services with better response time. Fog is highly useful in locations where network connectivity is an issue because fog has a separate protocol suite that will support weak network connections. In this article, the various parameters of the fog computing paradigm such as challenges, application, and opportunities are

studied and presented.

Keywords

Fog Computing; Edge Computing; Customer-Driver Service; Cloud Computing

Full Text:
PDF

References

1. Saharan KP, Kumar A. Fog in comparison to cloud: A survey. International Journal of Computer Applications 2015; 122(3): 10-12.

2. Bonomi F, Milito R, Zhu J, Addepalli S. Fog computing and its role in the internet of things. In: MCC '12: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing; 2012 Aug 17; Helsinki, Finland. New York: Association for Computing Machinery; 2012. p. 13-16. doi: 10.1145/2342509.2342513.

3. Hao X, Yeoh PL, Ji Z, et al. Stochastic analysis of double blockchain architecture in IoT communication networks. IEEE Internet of Things Journal 2022; 9(12): 9700-9711. doi: 10.1109/JIOT.2022.3142761.

4. Macedo ELC, Delicato FC, de Moraes LFM, Fortino G. Assigning trust to devices in the context of consumer IoT applications. IEEE Consumer Electronics Magazine 2022. doi: 10.1109/MCE.2022.3154357.

5. Abid MA, Afaqui N, Khan MA, et al. Evolution towards smart and software-defined Internet of Things. AI 2022; 3(1): 100-123. doi: 10.3390/ai3010007.

6. Chavhan S, Gupta D, Gochhayat SP, et al. Edge computing AI-IoT integrated energy efficient intelligent transportation system for smart cities. ACM Transactions on Internet Technology (TOIT) 2022; 22(4): 1-18. doi: 10.1145/3507906.

7. Ali O, Ishak MK, Bhatti MKL, et al. A comprehensive review of internet of things:

Technology stack, middlewares, and fog/edge computing interface. Sensors 2022; 22(3): 995. doi: 10.3390/s22030995.

8. Talaat FM. Effective prediction and resource allocation method (EPRAM) in fog computing environment for smart healthcare system. Multimedia Tools and Applications 2022; 81: 8235-8258. doi: 10.1007/s11042-022-12223-5.

9. Farooqi AM, Alam MA, Hassan SI, Idrees SM. A fog computing model for VANET to reduce latency and delay using 5G network in smart city transportation. Applied Sciences 2022; 12(4): 2083. doi: 10.3390/app12042083.

10. Gardasu AK, Kotha RK. A fog computing solution for advanced security, storage techniques for platform infrastructure. EasyChair Prepring No. 7460. 2022.

11. Sun L, Xue G, Yu R. TAFS: A truthful auction for IoT application offloading in fog computing networks. IEEE Internet of Things Journal 2022; 10(4): 3252-3263. doi: 10.1109/JIOT.2022.3143101.

12. Almiani M, Razaque A, Alotaibi B, et al. An efficient data-balancing cyber-physical system paradigm for quality-of-service (QoS) provision over fog computing. Applied Sciences 2022; 12(1): 246. doi: 10.3390/app12010246.

13. Verma P, Tiwari R, Hong WC, et al. FETCH: A deep learning-based fog computing and IoT integrated environment for healthcare monitoring and diagnosis. IEEE Access 2022: 10: 12548-12563. doi: 10.1109/ACCESS.2022.3143793.

14. Laghari AA, He H, Halepoto IA, et al. Analysis of quality of experience frameworks for cloud computing. IJCSNS 2017: 17(12): 228-233.

15. Dubey H, Yang J, Constant N, et al. Fog data: Enhancing telehealth big data through fog computing. In: Proceedings of the ASE BigData & SocialInformatics; 2015 Oct 7-10; Kaohsiung Taiwan. New York: Association for Computing Machinery; 2015. p. 1-6. doi: 10.1145/2818869.2818889.

16. Kocabas O, Soyata T, Couderc JP, et al. Assessment of cloud-based health monitoring using

homomorphic encryption. 2013 IEEE 31st International Conference on Computer Design (ICCD); 2013 Oct 6-9; Asheville, NC, USA. New York: IEEE; 2013. p. 443-446. doi: 10.1109/ICCD.2013.6657078.

17. Wang Q, Guo S, Wang Y, Yang Y. Incentive mechanism for edge cloud profit maximization in mobile edge computing. ICC 2019-2019 IEEE International Conference on Communications (ICC); 2019 May 20-24; Shanghai, China. New York: IEEE; 2019. p. 1-6. doi: 10.1109/ICC.2019.8761241.

18. Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning 2011; 3(1): 1-122. doi: 10.1561/2200000016.

19. Wang L, An H, Chang Z. Security enhancement on a lightweight authentication scheme with anonymity fog computing architecture. IEEE Access 2000; 8: 97267-97278. doi: 10.1109/ACCESS.2020.2996264.

20. Hong J, Xue K, Li W. Comments on “DAC-MACS: Effective data access control for multiauthority cloud storage systems”/security analysis of attribute revocation in multiauthority data access control for cloud storage systems. IEEE Transactions on Information Forensics and Security 2015; 10(6): 1315-1317. doi: 10.1109/TIFS.2015.2407327.

21. Liang J, Zhang M, Leung VC. A reliable trust computing mechanism based on multisource feedback and fog computing in social sensor cloud. IEEE Internet of Things Journal 2020: 7(6): 5481-5490. doi: 10.1109/JIOT.2020.2981005.

22. Alwakeel AM. An overview of fog computing and edge computing security and privacy issues. Sensors 2021: 21(24): 8226. doi: 10.3390/s21248226.

23. Mukherjee M, Matam R, Shu L, et al. Security and privacy in fog computing: Challenges. IEEE Access 2017: 5: 19293-19304. doi: 10.1109/ACCESS.2017.2749422.

24. Priyadarshini R, Barik RK. A deep learning based

Information

For Readers

For Authors

For Librarians

intelligent framework to mitigate DDoS attack in fog environment. Journal of King Saud University-Computer and Information Sciences 2022; 34(3): 825-831. doi: 10.1016/j.jksuci.2019.04.010.

25. Yaseen Q, Jararweh Y, Al-Ayyoub M, AlDwairi M. Collusion attacks in Internet of Things: Detection and mitigation using a fog based model. 2017 IEEE Sensors Applications Symposium (SAS); 2017 Mar 13-15; Glassboro, NJ, USA. New York: IEEE; 2017. p. 1-5. doi: 10.1109/SAS.2017.7894031.

26. Aliyu F, Sheltami T, Shakshuki EM. A detection and prevention technique for man in the middle attack in fog computing. Procedia Computer Science 2018: 141: 24-31. doi: 10.1016/j.procs.2018.10.125.

27. Tu S, Waqas M, Rehman SU, et al. Security in fog computing: A novel technique to tackle an impersonation attack. IEEE Access 2018; 6: 74993-75001. doi: 10.1109/ACCESS.2018.2884672.

28. Jia B, Hu H, Zeng Y, et al. Double-matching resource allocation strategy in fog computing networks based on cost efficiency. Journal of Communications and Networks 2018; 20(3): 237-246. doi: 10.1109/JCN.2018.000036.

29. Alenizi F, Rana O. Minimising delay and energy in online dynamic fog systems. arXiv preprint arXiv:2012.12745. 2020. doi: 10.48550/arXiv.2012.12745.

30. da Silva RA, da Fonseca NL. Resource allocation mechanism for a fog-cloud infrastructure. 2018 IEEE International Conference on Communications (ICC); 2018 May 20-24; Kansas City, MO, USA. New York: IEEE; 2018. p. 1-6.

31. Mijuskovic A, Chiumento A, Bemthuis R, et al. Resource management techniques for cloud/fog and edge computing: An evaluation framework and classification. Sensors 2021; 21(5): 1832. doi: 10.3390/s21051832.

32. Taneja M, Davy A. Resource aware placement of IoT application modules in Fog-Cloud Computing Paradigm. 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM); 2017 May 8-12; Lisbon, Portugal. New York: IEEE; 2017. p.

1222-1228. doi: 10.23919/INM.2017.7987464.

33. Javaid S, Javaid N, Saba T, et al. Intelligent resource allocation in residential buildings using consumer to fog to cloud based framework. Energies 2019; 12(5): 815. doi: 10.3390/en12050815.

34. OpenFog Consortium Architecture Working Group. OpenFog reference architecture for fog computing. OPFRA001.020817. OpenFog Consortium Architecture Working Group; 2017.

35. Tuli S, Gill SS, Casale G, Jennings NR. iThermoFog: IoT-Fog based automatic thermal profile creation for cloud data centers using artificial intelligence techniques. Internet Technology Letters 2020; 3(5): e198. doi: 10.1002/itl2.198.

36. Zhang Q, Wang J, Kim M. Heterofuzz: Fuzz testing to detect platform dependent divergence for heterogeneous applications. In: Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering; 2021 Aug 23-28; Athens Greece. New York, United States: Association for Computing Machinery; 2021. p. 242-254. doi: 10.1145/3468264.3468610.

37. Wu HY, Lee CR. Energy efficient scheduling for heterogeneous fog computing architectures. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC); 2018 Jul 23-27; Tokyo, Japan. New York: IEEE; 2018. p. 555-560. doi: 10.1109/COMPSAC.2018.00085.

38. Oueis J, Strinati EC, Barbarossa S. The fog balancing: Load distribution for small cell cloud computing. 2015 IEEE 81st Vehicular Technology Conference (VTC spring); 2015 May 11-14; Glasgow, UK. New York: IEEE; 2021. p. 1-6. doi: 10.1109/VTCSpring.2015.7146129.

39. Intharawijitr K, Iida K, Koga H. Analysis of fog model considering computing and communication latency in 5G cellular networks. 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops); 2016 Mar 14-18; Sydney, NSW, Australia. New York: IEEE; 2016. p. 1-4. doi: 10.1109/PERCOMW.2016.7457059.

40. Machen A, Wang S, Leung KK, et al. Live service migration in mobile edge clouds. IEEE Wireless Communications 2017; 25(1): 140-147. doi: 10.1109/MWC.2017.1700011.

41. Hu P, Ning H, Qiu T, et al. Security and privacy preservation scheme of face identification and resolution framework using fog computing in internet of things. IEEE Internet of Things Journal 2017; 4(5): 1143-1155. doi: 10.1109/JIOT.2017.2659783.

42. Koo D, Hur J. Privacy-preserving deduplication of encrypted data with dynamic ownership management in fog computing. Future Generation Computer Systems 2018; 78: 739-752. doi: 10.1016/j.future.2017.01.024.

43. Shynu PG, Nadesh RK, Menon VG, et al. A secure data deduplication system for integrated cloud-edge networks. Journal of Cloud Computing 2020; 9(1): 1-12. doi: 10.1186/s13677-020-00214-6.

44. Du M, Wang K, Liu X, et al. A differential privacy-based query model for sustainable fog data centers. IEEE Transactions on Sustainable Computing 2017; 4(2): 145-155. doi: 10.1109/TSUSC.2017.2715038.

45. Huang C, Lu R, Zhu H, et al. EPPD: Efficient and privacy-preserving proximity testing with differential privacy techniques. 2016 IEEE International Conference on Communications (ICC); 2016 May 22-27; Kuala Lumpur, Malaysia. New York: IEEE; 2016. p. 1-6. doi: 10.1109/ICC.2016.7511194.

46. Yi S, Qin Z, Li Q. Security and privacy issues of fog computing: A survey. In: Xu K, Zhu H (editors). Wireless algorithms, systems, and applications. International Conference on Wireless Algorithms, Systems, and Applications; 2015 Aug 10-12; Qufu, China. Cham: Springer; 2015. p. 685-695. doi: 10.1007/978-3-319-21837-3_67.

47. Hatzivasilis G, Soultatos O, Ioannidis S, et al. Review of security and privacy for the Internet of Medical Things (IoMT). 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS); 2019 May 29-31; Santorini, Greece. New York: IEEE; 2019. p. 457-464. doi: 10.1109/DCOSS.2019.00091.

48. Chiang M. Fog networking: An overview on research opportunities. arXiv preprint arXiv:1601.00835. 2016. doi: 10.48550/arXiv.1601.00835.

49. Kang J, Yu R, Huang X, Zhang Y. Privacy-preserved pseudonym scheme for fog computing supported internet of vehicles. IEEE Transactions on Intelligent Transportation Systems 2017; 19(8): 2627-2637. doi: 10.1109/TITS.2017.2764095.

50. Margariti SV, Dimakopoulos VV, Tsoumanis G. Modeling and simulation tools for fog computing—A comprehensive survey from a cost perspective. Future Internet 2020; 12(5): 89. doi: 10.3390/fi12050089.

51. Shi W, Cao J, Zhang Q, et al. Edge computing: Vision and challenges. IEEE Internet of Things Journal 2016; 3(5): 637-646. doi: 10.1109/JIOT.2016.2579198.

52. Dubey H, Yang J, Constant N, et al. Fog data: Enhancing telehealth big data through fog computing. In: Proceedings of the ASE BigData & SocialInformatics; 2015 Oct 7-9; Kaohsiung Taiwan. New York: Association for Computing Machinery; 2015. p. 1-6. doi: 10.1145/2818869.2818889.

53. Singh S. Smart meters big data: Behavioral analytics via incremental data mining and visualization [PhD thesis]. Ottawa: University of Ottawa; 2016.

54. Gia TN, Jiang M, Rahmani AM, et al. Fog computing in healthcare internet of things: A case study on ecg feature extraction. 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing; 2015 Oct 26-28; Liverpool, UK. New York: IEEE; 2015. p. 356-363. doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.51.

55. Lynn R, Wescoat E, Han D, Kurfess T. Embedded fog computing for high-frequency MTConnect data analytics. Manufacturing Letters 2018; 15: 135-138. doi: 10.1016/j.mfglet.2017.11.002.

56. Wang X, Yang LT, Kuang L, et al. A tensor-based

big-data-driven routing recommendation approach for heterogeneous networks. IEEE Network 2019; 33(1): 64-69. doi: 10.1109/MNET.2018.1800192.

57. Gao L, Luan TH, Yu S, et al. FogRoute: DTN-based data dissemination model in fog computing. IEEE Internet of Things Journal 2016; 4(1): 225-235. doi: 10.1109/JIOT.2016.2645559.

58. Ahanger TA, Tariq U, Nusir M, et al. A novel IoT–fog–cloud-based healthcare system for monitoring and predicting COVID-19 outspread. The Journal of Supercomputing 2022; 78(2): 1783-1806. doi: 10.1007/s11227-021-03935-w.

59. Barik RK, Dubey H, Samaddar AB, et al. FogGIS: Fog Computing for geospatial big data analytics. 2016 IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering (UPCON); 2016 Dec 9-11; Varanasi, India. New York: IEEE; 2017. p. 613-618. doi: 10.1109/UPCON.2016.7894725.

60. Dhande R. Dynamic replica management in fog-enabled IoT using enhanced data mining technique [Master's thesis]. Dublin: National College of Ireland; 2020.

61. Pérez JL, Gutierrez-Torre A, Berral JL, Carrera D. A resilient and distributed near real-time traffic forecasting application for Fog computing environments. Future Generation Computer Systems 2018; 87: 198-212. doi: 10.1016/j.future.2018.05.013.

62. Ning Z, Dong P, Wang X, et al. Deep reinforcement learning for intelligent internet of vehicles: An energy-efficient computational offloading scheme. IEEE Transactions on Cognitive Communications and Networking 2019; 5(4): 1060-1072. doi: 10.1109/TCCN.2019.2930521.

63. Pandit MK, Mir RN, Chishti MA. Adaptive task scheduling in IoT using reinforcement learning. International Journal of Intelligent Computing and Cybernetics 2020; 13(3): 261-282. doi: 10.1108/IJICC-03-2020-0021.

64. Sami H, Mourad A. Dynamic on-demand fog formation offering on-the-fly IoT service deployment. IEEE Transactions on Network and

Service Management 2020; 17(2): 1026-1039. doi: 10.1109/TNSM.2019.2963643.

65. Wilkes MV. Artificial intelligence as the year 2000 approaches. Communications of the ACM 1992; 35(8): 17-23.

66. Zou Z, Jin Y, Nevalainen P, et al. Edge and fog computing enabled AI for IoT-an overview. 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS); 2019 Mar 18-20; Hsinchu, Taiwan. New York: IEEE; 2019. p. 51-56. doi: 10.1109/AICAS.2019.8771621.

67. Ji W, Liang B, Wang Y, et al. Crowd V-IoE: Visual internet of everything architecture in ai-driven fog computing. IEEE Wireless Communications 2020; 27(2): 51-57. doi: 10.1109/MWC.001.1900349.

68. Dastjerdi AV, Gupta H, Calheiros RN, et al. Fog computing: Principles, architectures, and applications. Internet of Things 2016; 61-75. doi: 10.1016/B978-0-12-805395-9.00004-6.

69. Cheng X, Dale C, Liu J. Statistics and social network of youtube videos. 2008 16th International Workshop on Quality of Service; 2008 Jun 2-4; Enschede, Netherlands. New York: IEEE; 2008. p. 229-238. doi: 10.1109/IWQOS.2008.32.

70. Chen N, Chen Y, You Y, et al. Dynamic urban surveillance video stream processing using fog computing. 2016 IEEE Second International Conference on Multimedia Big Data (BigMM); 2016 Apr 20-22; Taipei, Taiwan. New York: IEEE; 2016. p. 105-112. doi: 10.1109/BigMM.2016.53.

71. Jain S, Gupta S, Sreelakshmi KK, Rodrigues JJ. Fog computing in enabling 5G-driven emerging technologies for development of sustainable smart city infrastructures. Cluster Computing 2022; 25: 1111-1154. doi: 10.1007/s10586-021-03496-w.

DOI: https://doi.org/10.32629/jai.v5i2.545

Refbacks

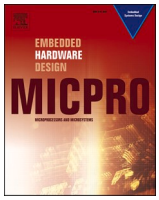
- There are currently no refbacks.

Microprocessors and Microsystems 98 (2023) 104794

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro



The novel emergency hospital services for patients using digital twins

Rajanikanth Aluvalu [a,*], Swapna Mudrakola [b], Uma Maheswari V [c], A.C. Kaladevi [d], M.V.S Sandhya [e], C. Rohith Bhat [f]

[a] Department of IT, Chaitanya Bharathi Institute of Technology, Hyderabad, India
[b] Matrusri Engineering College, Hyderabad, India
[c] Department of CSE, Chaitanya Bharathi Institute of Technology, Hyderabad, India
[d] Department of Computer Science and Engineering, Sona College of Technology, Salem, India
[e] Osmania University, Hyderabad, India
[f] Institute of Computer Science and Engineering, SIMATS School of Engineering, Chennai, India

ARTICLE INFO

Keywords:
Digital twins
Digital health record
Expert system
Sensors
Smart health care system

ABSTRACT

The Digital twins will duplicate the actual objects, create the virtual world and execute using IoT devices and Sensors. The Emergency Room Service (ERS) is a critical phase for patients in health condition evaluation, Digital Health records will help us in understanding the cause of illness, medical history will help us to start the treatment. The most challenging for ERS is anonymous person or unknown follow ups about patients. The proposed model (Emergency Service Room with Digital Twins), helps to treat a patient with fast-track service and reduce the length of Stay in ER. The risk factor of a patient's life by reviewing the medical history of the patients through digital health records. This novel method will help doctors in treating patients by Computing Image Processing in face recognition of patients. The biometric used for authentication to access the cloud for digital health records. The communication system used to acknowledge the family, Insurance Company and expert adviser. The empirical results successfully proved the novel proposed idea with above 80% of success rate. We can build an intelligent expert system to collaborate the Digital Health Record, E-H-S, Expert Adviser, and an expert system in the future. In treating Anonym patients in the Emergency department.

1. Introduction

The role of digital twins (DT) to Autonomous business medical applications through digital devices. The function of DT includes - Monitoring Digital Twins, Imaginary Digital Twins, Predictive Digital Twins, Prescriptive Digital Twins and Recollection Digital Twins are controlled by the life cycle of the control module in the framework. Smart farming is an advanced and most required approach in farming. Digital Twins help to study the environment virtually. Dynamic actions are taken based on the intelligent decision taken by the Digital Twin framework designed using IoT devices, twining the devices in data acquisition, and data analysis on the system. The applications of smart farming like Livestock Farming, Planthouse, Dairy houses, Chemical-Free Farming and Agriculture Farming. The Digital Twin collaborates with different modules in the farming life cycle through the virtual world. In this European IoF2020 project, six digital twins exist to analyze the current state, predict future actions, and monitor and some of application area are shown in Fig. 1 [1] and a list of medical applications using digital twins is discussed in Table 1.

Identifying the Challenges in Digital Twins will help exploit the research world's solutions. The first Challenge is that AI algorithms are used in digital twins and the block chain. Most of the AI methods are black-box in nature. The AI decision-making methods suffer from apprehensible logic clearness in the reasoning. The explained Artificial Intelligence is expected for trustworthiness [2]. It isn't easy to collect data about physics and chemical info about food raw materials in the digital twins' Challenge in the food safety industry. Adopting new technology in the existing system is also difficult. Much statistical calculation needs to be done to predict the values [3]. The Digital Twins challenge in Constructions is a comprehensive design model expected from design to execution in planning, model construction, monitoring, operation executions, and quality check management [4].

Recent Trends in Digital Twins: Digital twins are used in eliminating mistakes and defects (Poka Yoke) products in the manufacturing industry. The robot manufacturing products work in collaboration with digital twins and predict the failure of the equipment, failed execution,

* Corresponding author.
E-mail address: rajanikanth.aluvalu@ieee.org (R. Aluvalu).

https://doi.org/10.1016/j.micpro.2023.104794
Received 26 May 2022; Received in revised form 15 December 2022; Accepted 12 February 2023
Available online 15 February 2023