ANALYSIS OF SEDIMENT DEPOSITIONOF BRAHMANI RIVER BY NEURAL NETWORK AND MULTIPLE LINEAR REGRESSION

A Dissertation submitted in partial fulfilment of the requirement for the

Award of degree of

MASTER IN TECHNOLOGY

IN

HYDRAULICS AND WATER RESOURCE ENGINEERING

Submitted by: -

Rahul Sheokand

(2K19/HFE/09)

Under the guidance of

Dr. T. VIJAYA KUMAR

DEPARTMENT OF CIVIL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(FORMELY DELHI COLLEGE OF ENGINEERING)

DELHI-110042

AUGUST 2021

CANDIDATE'S DECLARATION

I, **RAHUL SHEOKAND,** Roll No. **2k19/HFE/09** of **M.Tech (HWRE),** hereby declare that the project Dissertation titled "**ANALYSIS OF SEDIMENT DEPOSITION OF BRAHMANI RIVER BY NEURAL NETWORK AND MULTIPLE LINEAR REGRESSION** " which is submitted by me to the department Hydraulics and Water Resource Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technological, is original and not copied from any source without proper citation.

I have not submitted the matter embodied in the report for the award of any other degree or Diploma.

Place: Delhi

Date: **(RAHUL SHEOKAND)**

CERTIFICATE

This is to certify that **Mr. Rahul Sheokand (Roll No. 2K19/HFE/09),** a student of M.Tech. Department of Civil Engineering, Delhi Technological University (FORMELY DELHI COLLEGE OF ENGINEERING) has worked under my guidance and completed his Dissertation entitled "**ANALYSIS OF SEDIMENT DEPOSITION OF BRAHMANI RIVER BY NEURAL NETWORK AND MULTIPLE LINEAR REGRESSION**" in partial fulfillment of the requirement for award of degree of M.tech in Civil Engineering from Delhi Technological University (FORMELY DELHI COLLEGE OF ENGINEERING).

This work has not been submitted for any other degree, award, or distinction elsewhere to the best of my knowledge and belief His solely responsible for the technical data and information provided in this work.

Place: Delhi
Date:

(Dr. T. Vijaya Kumar)
Associate Professor
Department of Civil Engineering
Delhi Technological University

ACKNOWLEDGEMENT

I take this opportunity to express my sincere regards and deep gratitude to **Dr. T. VIJAYA KUMAR** (associate professor, civil engineering department, DTU) for his consistent guidance, monitoring and constant encouragement throughout the course of this project work. Also, I express my gratitude to **Department of Civil Engineering, DTU** for giving me such an opportunity to commence this project in the first instance.

Professors and faculties of the civil engineering, DTU have been very supportive and cooperative. They have been always present for their kind opinions and suggestions regarding this project work and therefore I am deeply obliged to them.

Last but not the least, I would like to thank my family and my colleagues from the department who encouraged me to bring work to a successful close.

Place: Delhi (RAHULSHEOKAND)

Date:

Contents

List of figures

LIST OF ABBREVIATIONS

1. ANN – Artificial Neural Network
2. MLR – Multiple Linear Regression
3. SSC – Suspended Sediment Concentration
4. CWC – Central Water Commission
5. IMD – Indian Metrological Department
6. RMSE – Root Mean Square Error
7. R^2– Determination Coefficient
8. BPNN – Back Propagation Neural Network
9. GA – Genetic Algorithm
10. SBP – Standard Back Propagation
11. RBNN – Radial Basis Neural Network
12. FL– Fuzzy Logic
13. SRC– Sediment Rating Curve
14. ORG – Ordinary Rain Gauge
15. MLP – Multi Layer Perceptron

ABSTRACT

Estimating the settled sediment production is critical for water resource planning and management, as well as environmental protection. Sediment creation, transit, and the resulting sediment load in rivers are all influenced by environmental change. The natural sediment flow in the river is assessed in terms of sediment concentration using an artificial neural network. This is accomplished by teaching the network how to gather natural stream data from reputable sources. Choosing the right neural network structure the performance results based on two sorts of indicators, viz. correlation coefficient (R^2), root mean square error (RMSE). Internal suspicions are not explicitly built-in traditional techniques of estimating sediment production (e.g. regression models). This model, on the other hand, is unable to augment intellectual capacity of the underlying connections between the data gathered or to estimate the influence of each sediment yield component. It is easier and fewer expensive to develop artificial neural networks for sedimentation prediction and predicting factors. Environmental change has an impact on sediment production, transit, and the consequent sediment load in rivers. The Brahmani river basin was contrast using multiple linear regression and artificial neural networks in this study. This study used a four-year data collecting period and the back-propagation approach. The finest model findings will be used as an estimate in future hydrological structural blueprint studies.

1 CHAPTER-1

1.1 INTRODUCTION

Sedimentation is the process by which soil particles are eroded and deposited on a solid particle layer in the next reservoir and river-like body of water carried by flowing water or other transport medium. It's a complicated process that varies Depending on how much sediment there is in the watershed, the velocity of movement, and the mechanism of deposition. Reservoir storage capacity and life duration, as well as river flows, are all affected by sedimentation. Straight measuring is a time-consuming technique that is not carried out at every observing location. Artificial Neural Network (ANN) modeling, on the other hand, is a technique for detecting complex relationships between input and output data sets. Different techniques are utilized to optimise multidimensional non-linear objectives functions to check watershed features due to unpredictable weather patterns.

SSC is a significant influence on water course ecosystem and geomorphology. The forecast of suspended deposits in Hyperconcentrated Rivers aids in the distribution of water resource glitches and hydraulic projects such as barriers and basins.

Water quality is a global concern, and pollution caused by human activity or natural occurrences has an impact on people's lives. Sediments made consisting of silt and clay elements pollute numerous waterways. River sedimentation is a significant issue that must be considered when managing water resources.

Stream ecology and morphology are influenced by SSC. Water resource problems and distribution of hydraulic constructions such as barriers are supported by predictions of floating and in highly concentrated streams Suspended sediments are generally carried inside the fluid and at the same speed as it (water flow or air). The flow intensity and particle fineness determine the quantity of sediment suspended by turbulence. Only the finer fraction of the suspended sediments (typically silt and clay component) may be kept suspended indefinitely by flow turbulence. This percentage is known as "wash load," and it is seldom seen in large numbers near the bed surface. This percentage is generally linked to the residue source and is tough to calculate hypothetical; for example, the larger the expulsion, the greater the residue.

Water eminence is a global concern, and pollution caused by social activity or ordinary occurrences has an impact on people's lives. Sediments made consisting of silt and clay elements pollute numerous waterways. Sedimentation in rivers is a significant issue that must be considered when managing water supplies. The sediment load progressively fills a portion of the reservoir concentration (dead volume).Bed load and suspended materials make up the total sediment load, while wash load transports a major portion of the sediment load in propagation. As a result, the particle concentration of sediment is an essential component of the overall quantity of suspended sediment.

Because of the silt in the rivers, the lifecycle of barriers and basins might be shortened. When a stream is prevented and a basin is formed, relatively fast-moving rivers and flowing sediments are deposited in the reservoir

Each river takes all sorts of material, such as rock, soil, and clays, with it wherever it goes. When a river runs through any of these items, such as suspended silt, it may take them up and carry them downstream. The quantity of material transported by a river's flow is recognized as its sediment load. The geology and ecology of the river and its watershed, as well as the influence of social activities such as growth of industry and insect killer usage, are all factors in the suspended sediment capacity of the river and its watershed. The sediment load may be determined by multiplying the sediment concentration by the stream flow.

Modeling approaches have been used to comprehend hydrological processes throughout the last few decades. A model is characterized as either a physically based model or a systems theory model depending on how it is grounded in the management process. Models based on physics include descriptions of many physical processes that affect a system's hydrologic behavior. System theoretic models, on the other hand, do not take into account the physical features of the parameters Instead; they translate the data from input to output using transfer functions. Models based on artificial neural networks (ANNs) are system theory models for simulating rainfall–runoff–sediment cycles that have gained a lot of interest in recent years. Without a comprehensive grasp of the complicated physical rules controlling the process under study, ANN models are able to offer relatively precise models for the process under research. In instances when we can't examine every useful physical parameter, it's unsurprising those black box models Artificial neural networks are a sort of black box modeling tool that has lately gained popularity in disciplines such as hydrology and environmental engineering (ANN)

Since the 1930s, a variety of linear and nonlinear models have been created to replicate and foretell numerous hydrological processes and other variables. Hydrologic simulation models are continual improving as computer techniques evolve, making it easier for them to connect with future technology and give more sophisticated tools for operational applications. Multiple linear regression (MLR)is a statistical approach that predicts the outcome of a dependent variable by combining several independent variables. Regression models have been used to model a wide range of hydrologic phenomena with great effectiveness in recent years, including soil temperature, flood flows, and sediment prediction.

Physical and applied models require data from point-by-point topography, ecological hydro climatology and geophysics. Preparing such data for sediment loads in any river, such as the Ganga or the Brahmani, will be difficult and expensive. Various scholars have successfully employed artificial intelligence approaches to address complicated non-linear global issues in hydrology and other fields.

1.2 OBJECTIVE

The primary goal of this research is forecast sediment concentrations in Brahmani Stream with the next goal in mind:

1 To create a model for predicting the amount of suspended silt in a highly concentrated environment river using feed forward backpropagation neural network

2 To access the concert of the model and indorse the finest model among those that have been produced.

3 In comparison to other commonly used discharge formulas, the novel approach produces a better outcome.

1.3 SIGNIFICANCE OF PROJECT

For the design and operation of many water resources, hydroelectric, and environmental engineering projects, suspended sediment calculations are critical. This research will present a novel numerical modelling approach for predicting suspended silt levels in rivers. Furthermore, further research and procedures may be used to ensure that the suspended sediment flux declines and the hydro structure is protected.

1.4 SCOPE OF STUDY

The Brahmani River's flow, temperature, rainfall, and sediment data will be used to predict suspended sediment under hyper-concentrated conditions. Using data from the Brahmani River, the study is confined to developing feed forward backpropagation Function Neural Network models for suspended sediment prediction. Two statistic metrics, determination coefficient (R square) and Root Mean Square Error (RMSE) are used to evaluate the performance of the prediction model.

1.5 FEASIBILITY OF THE PROJECT

The essential data like sediment, discharge, temperature and rainfall will be obtained from the Department of Central Water Commission (CWC) and Indian Metrological Department (IMD) in this study no fieldwork is necessary, decreasing the amount of time spent collecting data and allowing more time to be allocated to data analysis and prediction model building. The arranged data will run on matlab and spreadsheet and the result are comparing of both method.

2 CHAPTER-2

2.1 LITERATURE REVIEW

This chapter examines previously published literatures relevant to the study. Other books and articles were studied and used to aid in this investigation. Several studies address the capabilities of numerical modelling in estimating suspended silt concentration in various river sites.

Many researchers have done their work on the estimate and determination of residue absorption on different rivers.

Dillip k et al.(2018) done his research on modeling sediment concentration using BackPropagation Neural Network (BPNN) and Regression model simultaneously with genetic algorithm. He found that two dissimilar nonlinear methods including GA show not linear relationship. Production of this model displays three separate segments with temperature. In initial stage, residue concentration rises after that declines as discharge surges and eventually rises.

Avinash Agarwal et al. (2005) conducted research for the Vamsadhara river using ANN-based sediment yield models. He evaluated performance using numerous metrics, including RMSE, CC, and CE. He created a model using BPANN and LTF. In generalised model development, he concluded that quick and high convergence is not required.

Neural Network (Artificial) **Ajai Singh et al. (2013)** developed sediment yield forecast models at a single sample site in an Eastern Indian watershed. Four scenarios were tested to determine the kind and number of inputs for the ANN model. Incorporating regular rainwater and usual discharge variables better the ANN model's concert in calculating monthly sediment output in Nagwa's small and forested watershed. The only point of the Nagwa watershed in eastern India was related using SBP and RBNN models.

Using a neural network model for determine silt load concentration in rivers**, H.M. Nagy et al. (2002)** leverages significant features of ANN to address problems. It illustrates that, since sediment movement is inherently unpredictable and stochastic, when conventional methods fail, the neural networks model can be utilized to successfully move silt. The research entails modifying the collected field data in order to feed the network's memory and broaden its experience. It also entails utilizing the theoretical foundations of sediment transport to determine the problem's main factors, as well as estimating concentration based on available data and prior experience.

Yun-Mei Zhu et al. (2007) did research on suspended sediment flux modeling using artificial neural networks on the Longchuanjiang river in China's upper Yangtze basin. The normal rainfall, temperature, precipitation intensity, and water outflow are all linked together, he utilizes ANN to estimate the monthly flow of suspended material in the Longchuanjiang basin. The same data requirements apply, the power relation model and multiple regression models are unable to fit the observed suspended sediment flow as well as ANN. As of the dispersed data dealing out system and the non linear changes that must be performed, the most noticeable attribute of ANN be capable of deliver additional realistic results. Anomaly high or low numbers are to be expected.

The method of utilizing fuzzy logic (FL) may used to estimate the sediments content using river current data from the United States Geological Survey's Sacremento Freeport Station, according to **Demirci, M. and Baltaci, A. (2012)** The findings are compared using multi-linear regression (MLR) and the Sediment Rating Curve (SRC).The degree to which suspended silt in a hyperconcentration river can be anticipated with moderate accuracy could be greatly improved using a fuzzy rule-based model. To compare performance, the MSE, MAE, and R metrics got from the testing statistics are employed. The FL model, rather than MLR, had the highest accuracy in estimating total sediment load, while the SRC model had the lowest estimation result.

Shreya N and Prevandra K et al. (2017) compared multiple linear regression and artificial neural networks to study suspended loads in the Vamsadhra river basin (ANN). The chain rule of calculus is used to derive the back propagation computation. It entails adjusting the weight to reduce mistake. The log-sigmoid activation function was used in this investigation. The software MATLAB was used to create the model suspended sediment load. Hidden layer of neuron network is increased and different result are noted by hit and trial to develop model.

We are calculating suspended sediment loads and lacking data, like **Asli Ulke et al. (2009)** did on the Gediz River in Turkey. He used AI (ANN, ANFIS) and regression-based models to forecast bimonthly sediment loads since everyday observed river discharge and rainfall data in his research. Using two independent data sets, the interruption and extrapolation abilities of the models remained evaluated. Artificial intelligence methods proved to be superior to regression-based models in terms of interpolation. He came to the conclusion that including precipitation data in the input vector, together with flow rate data, progresses the SSL prediction by the created models.

3 CHAPTER-3

3.1 STUDY AREA

3.1.1 BASIC DISCRIPTION

The Brahmani is a major inter-state peninsular river in India that drains in the Bay of Bengal. The Brahmani stream geographical coordinates are around 200-28' to 230-35' north latitude and 830-52' to 870-03' east longitude. The Chhota Nagpur plateau borders the basin taking place the north, the Mahanadi stream on the westward and south, and the Bay of Bengal on the eastward. Before emptying into the Bay of Bengal, the river drifts through the states of Jharkhand, Chhattisgarh, and Odisha, draining a entire zone of 39,033 square kilometers.

The Brahmani River, also known as the South Koel, begins near Nagri hamlet in Jharkhand's Ranchi region at an elevation of about 600 metres in its upper parts. After a 310-kilometer voyage from its source, the river meets Sankh, a main Right Bank branch, in Panposh near Rourkela. Beneath this point, the stream is recognized as Brahmani. The stream has a whole span of 799 kilometers, with 260 kilometres in Jharkhand and the remainder in Odisha. At Dhamra's common mouth, where the Baitarani joins the Brahmani, the major river drains into the Bay of Bengal. The main tributaries include the Karo, the Sankh, the Tikra, the Sanakoi, the Telkoi, and the Ramiyala.

Water Year ranges from June 1st of one calendar year to May 31st of the next calendar year and covers one complete hydrological cycle.

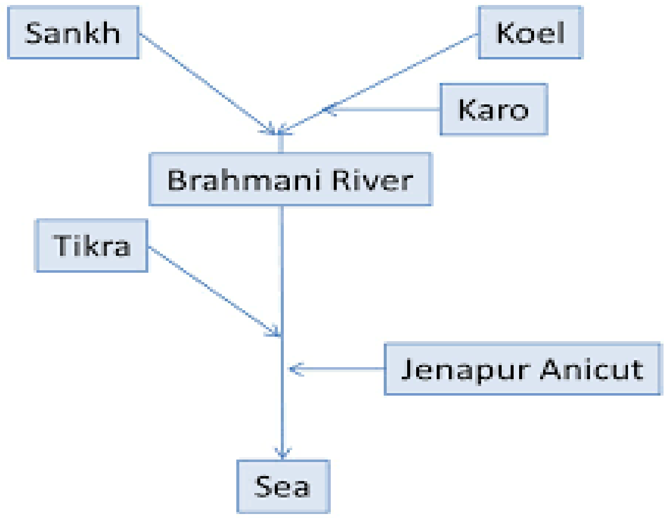

Figure 1 Schematic view of Brahmani River

Source: Policy and Advisory Technical Assistance 8089 IND Phase II

3.1.2 CLIMATE CHARATERSTIC

Tropical weather with scorching summertime and moderate wintertime characterizes the basin. The south-west monsoon as well as certain downpours in the lower levels induced by cyclonic depressions in the Bay of Bengal, have an impact on this basin from June to October. The regular yearly rainwater in this basin is 1460 mm. The greatest temperature ranges from 38 to 43 degrees Celsius, while the lowest ranges from 10 to 15 degrees Celsius.

3.1.3 GEOLOGY

Mineral resources abound in the basin. The basin's primary mineral resources include bauxite, chinaclay, chromite, coal, dolomite, fireclay, graphite, gemstones, iron ore, limestone, manganese ore, mineral sand, nickel ore, pyrophylite and quartz. Current explosion of the mineral industry has bowed the state into a hotspot, with entrepreneurs from all over the world crowding for their share of fortune.

3.2 STUDY MAP

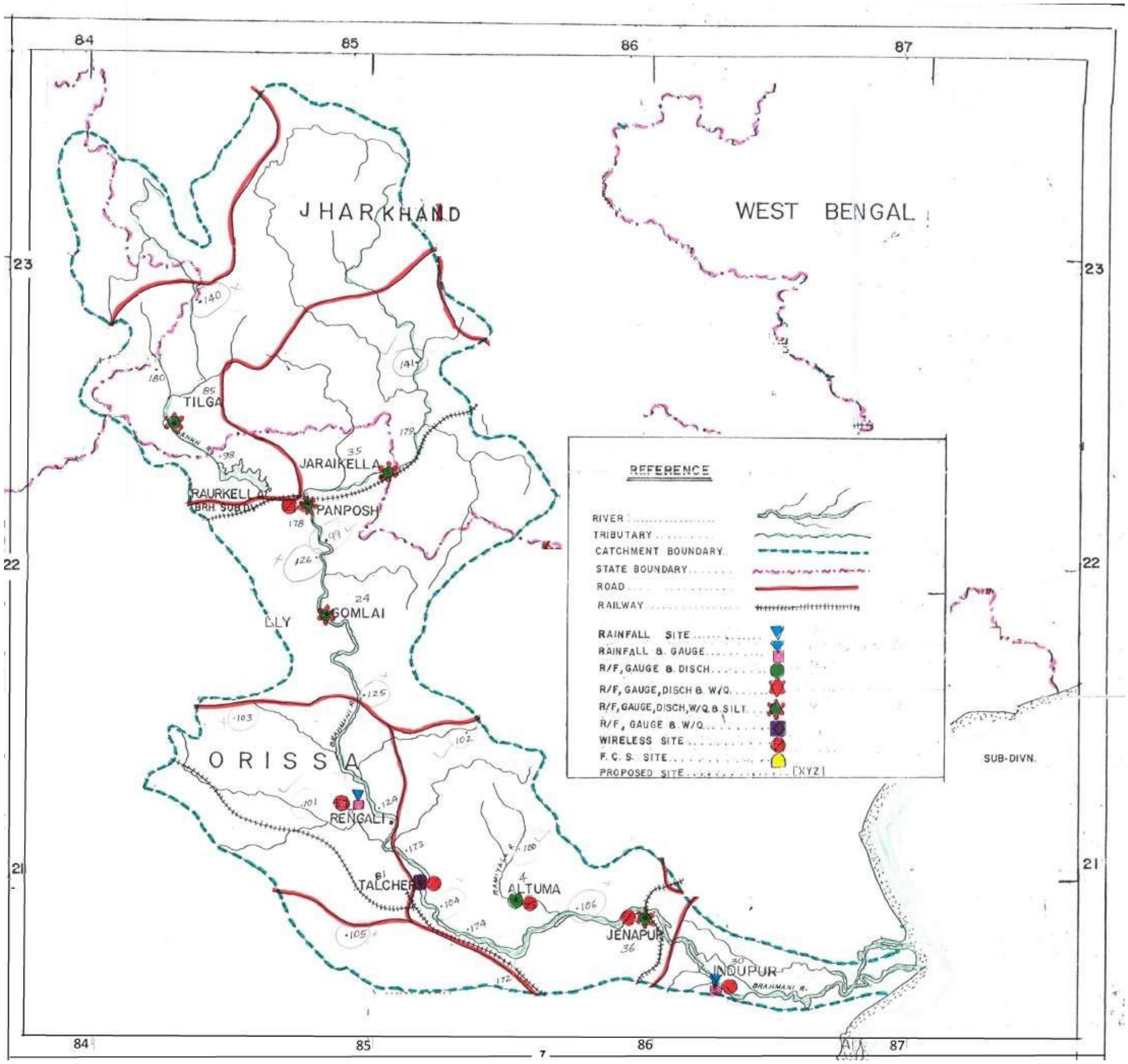

Figure 2 Map of River

Source: Ministry of water resource, river development and Ganga rejuvenation

3.3 OBSERVATION TECHNIQUE

Meteorological observations at GDS (Gauge, Discharge, sediment) site are recorded regularly at 08.30hrs. The Meteorological instruments are installed in the premises of the site offices. The information gathered at the sites is forwarded to the Division office for further processing. CWC specifies the observation strategy that will be used to obtain meteorological data.

3.3.1 RAINFALL

Rainfall is measured by Ordinary Rain-gauge (ORG) which is made of fiberglass reinforced polyester material as per Indian Standard Specification No. IS 5225: 1992

Rainfall is measured in terms of the depth of water collected on a level area of any size, assuming that the rain falls evenly over the area at the same rate as it falls uniformly over the area of the rain gauge's rim collector. One centimeter of rainfall, for example, signifies that if the rain falls on a flat surface that does not absorb it and from which it cannot run off or evaporate, it will total one centimeter, it would form a sheet of water one centimeter in depth. To measure the rainfall, the water collected in the receiving bottle is poured in to specially graduated glass cylinders manufactured according to Indian Standard Specification No. IS: 4849-4968. Care is taken to avoid spilling of the collected water.

3.3.2 MAXIMUM MINIMUM TEMPERATURE

A general-purpose maximum and minimum thermometer are used at sites for registering extremes of temperature during a day i.e., 24 hrs. The thermometer of ‘A’ pattern having range -40^0 C to +60^0 C with smallest scale divisions equivalent to 1.0^0 C (IS: 7000-1973) is generally used.

The thermometer is housed in wooden louvered enclosure known as Stevenson screen, which support the thermometers and shield it from direct solar radiation and from precipitation while allowing free circulation of air around it, and prevent accidental damage.

3.3.3 STREAM FLOW DATA (Discharge)

For assessing discharge at sites, the area-velocity approach is commonly used. The velocity of the flow is measured using a cup type current meter, and for depths up to 3 m, a sounding rod is used to determine depth. and a log line for depths greater than 3 m. The discharge is measured using the area velocity method once a day, at 0800 Hrs.

Flows on non-observed days are computed. For each season (monsoon and non-monsoon), the observed stage and discharge statistics are presented, as well as a mean Stage V/s. The discharge curve is constructed, paying special attention to the distributed spots in terms of area, velocity, and other factors. The variables that cause the curves to move are further addressed by examining the river cross section at regular intervals and superimposing prior years' Stage V/s. Discharge curves. As a result, the current curve's trend is complete. Finally, these Stage V/s. Discharge Curves are used to calculate the discharges of non-observed days.

3.3.4 SEDIMENT DATA

During the monsoon season, sediment observations are conducted daily, and once a week (on Monday) during the non-monsoon season. Non-observed days' data is estimated/ interpolated using a discharge vs. sediment load relationship calculated using observed sediment concentration and weighted mean discharge from the same year.

At 0.6 depths, sediment samples are collected from all verticals along the hydrological observation sections, where velocity is monitored for discharge computation using a Punjab type bottle sampler. For analysis, the collected samples from all segments are merged into three to seven groups, each with compartments or groups of equal or almost equal discharges. Suspended sediment load is graded into three categories: coarse, medium, and fine. The coarse and medium grades are separated using sieving, whereas the fine grade is separated through filtration of left-over samples after sieving through filter paper.

4 CHAPTER 4

4.1 METHDOLOGY

4.1.1 DATA COLLECTION

The research data is obtained from Central Water Commission (CWC) and Indian Metrological Department (IMD) of JENAPUR station of Brahmani River Orissa. The total drainage area of 33955 km.sq. The data for this river is taken from 2015 to 2019.

During the monsoon season, sediment observations are conducted daily, and once a week during the non-monsoon season. The total of daily observed suspended sediment (g/l) is used to determine the quantity of suspended sediment load in three grades: coarse, medium, and fine. Grade wise concentration is derived gravimetrically as per standard procedure. Daily 4-year data of Rainfall, temp and discharge data is also collected. In this research we take the average temperature.

4.1.2 DATA PROCESSING

The ranges of the original input and output data are very varied since they are made up of distinct parameters with varying physical meanings and units. To guarantee that each variable in the model is treated equally, data is generally rescaled to a certain interval, such as [-1-1]; [0-1]; [0.1-0.9]. The benefit of doing so is that it allows for possibly extreme values outside of the calibration data range to be accommodated. For this investigation, the input and output variables were normalized in the range [0-1].

$$Y_{istd} = 0.1 + 0.8*[(Yi - Y_{imin})/ (Yimax- Y_{imin})]$$

Where Y_{istd}= standard value of i,

Y_i= original value

Y_{imax}= max value

Y_{imim}= min value

The activities may be broken down into two categories. Essentially, the first half is devoted to the training phase, during which the information will be analyses in order to obtain the input information. The testing data, which will evaluate on MATLAB to locate the hidden neuron and estimate the output data, will be the second portion. These numbers will be used to see if the model can accurately forecast suspended sediment.

4.2 ARTIFICIAL NEURAL NETWORK (ANN)

A controlled learning structure made up of a huge quantity of basic units known as neurons or perceptron is known as an artificial neural network (ANN). Each neuron in the connected layer has the ability to make basic decisions and send them to other neurons. Three layers make up an artificial neural network (input layer, hidden layer, output layer).

The ANN examined the input and output datasets to determine if there was any association. The neuron accepts the series of inputs or indications and TRAINLM functions have been utilized as transfer functions to create sedimentation as an output. The ANN models improved the accuracy of suspended sediment estimate over conventional approaches.

Artificial neurons are employed in this study in a Feed-Forward Back-Propagation system that includes an input layer, a hidden layer, and an output layer. There is at least one hidden layer sandwiched between the input and output layers. The trainlm function is used as a training function, TRANSIG is used as a transfer function, and the backpropagation method is used as a training algorithm in this study.

The neural network's foundation is MLP. A multilayer perceptron (MLP) is a set of perceptron layers that can response composite problems correctly. Individually perceptron in the initial layer sends signals to the succeeding layer's perceptron, and so on.

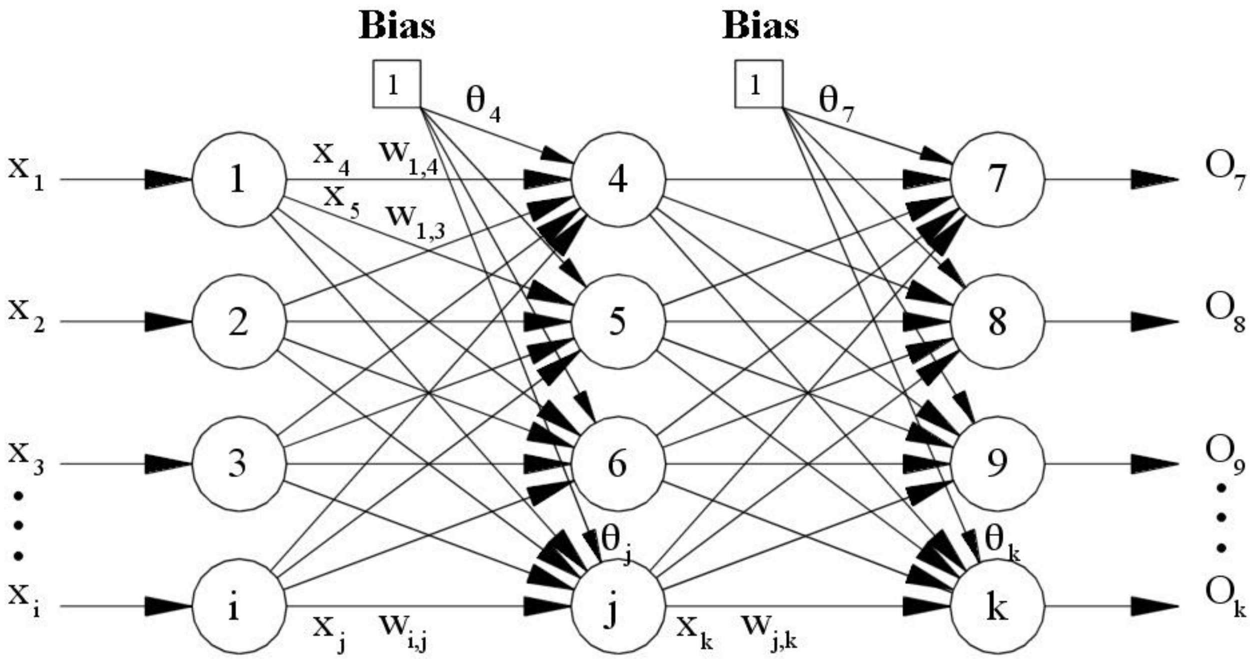

Figure 3 Back Propagation Neural Network

Source: https://www.cse.unsw.edu.au/~cs9417ml/MLP2/BackPropagation.html

4.2.1 BACK-PROPAGATION NEURAL NETWORK

The feed forward architecture of a back propagation neural network is made up of three layers: an input layer, a hidden layer, and an output layer. The point of this network is to train it to find a compromise between correctly reacting to training input patterns and convey the best replies to comparable input. The three phases of back propagation are training, testing, and validation. Backpropagation looks for the weights that decline the loss function to the smallest value based on the best prophecy by following the derivatives of the activation function in each successive neuron.

A hyperparameter is a surrounding that influences the structure or function of a neural network. The activation function, the number of hidden layers, and the number of training epochs are all typical hyper parameters.

1 Forward pass + 1 Backward pass = 1 epoch

The inputs are sent through the network in a forward pass, allowing individually neuron to respond to a part of the input. Neurons create output and permit it on to the succeeding layer, till the network produces its own output.

Each neuron is assigned a numerical weight in the weight space. Each neuron's output is determined by its weight and activation function.

4.2.2 ACTIVATION FUNCTION

The output of each joint in the neural network is determined by an activation function, which is a mathematical equation. It takes each neuron's input and converts it to an output, which is generally among1 and 0 or -1 and 1.

The activation function's role in a neural network input, which is generally a real value, is supplied into the network's neurons. Through the back-propagation process, the derivative of the activation function aids the network in learning.

4.2.3 TRANSFER AND TRAINING FUNCTION

The transfer function of an LTI (Linear and Time invariant) arrangement is the proportion of the system's Laplace transforms of output to the Laplace transform of input, presumptuous entirely preliminary circumstances are zero. Unsupervised and supervised training are the two sorts of training procedures. Supervised training includes either manually "grading" the network's performance or supplying the network with the information it needs to succeed. Unsupervised training is when the network is left to its own devices to make sense of the inputs.

4.2.4 THE PERCEPTRON LEARNING PROCESS: -

1. Computes the sum of the inputs that are supplied into the perceptron's in the input layer, multiplied by their weight.
2. Multiplies the numeral by a "bias weight" it is a methodological stage that allows each perceptron's output function (the activation function) to be moved up, down, left, and right on the numeral chart.
3. Uses the activation function to feed the total.
4. The output is the outcome of the step function.

4.2.5 ERROR – CORRECTION LEARNING: -

It's a method of differentiate the arrangement's output to the expected output value and by the difference to guide training. The error value can be utilised to unswervingly alter the tap weight using a technique like the back-propagation algorithm. At each training iteration, the error correction learning algorithm tries to reduce the error signal.

4.2.6 BIAS AND VARIANCE

The model's bias reflects how well it matches the data of training. A large partiality indicates that the neural network is unable to make accurate estimates, even for the example it was trained on.

The variance shows how well the model matches the validation set's unseen examples. A large variance indicates that the neural network is unable to anticipate new examples that it has not seen before. Both bias and variance should be minimal in a good model.

4.3 BACK-PROPAGATION WORKING

1. Weights are initialized, and training set inputs are fed into the network, giving the model generator its first calculation.
2. Error function: - calculated by determining how much the estimate deviates from the real value.
3. Gradient descent using back-propagation: - The back-propagation algorithm determines how much each model weight influences the output values. To do so, it calculates the partial derivative from the error.
4. Weight update: Back-propagation is used to the combined result after a bunch of samples is processed in one major forward pass. The number of batches used in training and the size of each batch, referred to as iteration, are critical hyperparameters that must be fine-tuned in order to get the optimum results.

The back-propagation technique states that the network's tap weights are changed repeatedly throughout training to approach the error function's minimum.

This is accomplished by using the following equation: -

$$W_{ij}^{l}[n] = W_{ij}^{l}[n-1] + \delta w_{ij}^{l}[n]$$

$$W_{ij}^{l-1}[n] = \eta \delta j^{l} X_{I}^{l-1}[n] + \mu \Delta w_{ij}^{l}[n-1]$$

X_{i}^{l-1}= output from the previous interlayer

W_{ij}^{l}= weight from the previous interlayer's I input to the current layer's j element

η= learning rate

μ= momentum parameter

4.4 MULTIPLE LINEAR REGRESSION

A linear connection between two or more variables is the subject of regression analysis. MLR is used to solve issues in which the dependent variable, y, and the independent variables, x1, x2, x3,...xn, have been recorded, and the goal is to examine the connection between the variables x1.... Xn.

Regression analysis is done by excel and we compared the value of correlation coefficient (R^2) and RMSE with Matlab.

4.4.1 GOODNESS OF FIT

The coefficient of determination (R^2) is a assess of an predictable regression line's overall goodness of fit, or the proportion or percentage of total variance in the dependent variable Y explained by all regressors. It tells: how great the model is.

4.4.2 ASSUMPTION OF MULTIPLE LINEAR REGRESSION

Homogeneity of variance (homoscedasticity) means that the amount of the error in our forecast doesn't change much when the independent variable's values fluctuate.

Observation independence the observations in this dataset were collected using statistically acceptable techniques, and there are no hidden relationships between variables.

Some of the independent variables in multiple linear regression are likely to be correlated, therefore double-check them before constructing the regression model. Only one of two independent variables should be utilised in the regression model if they are too closely linked ($r2 > 0.6$).

Normality The data is distributed normally.

The line of greatest fit across the data points is a straight line, rather than a curve or any other type of grouping factor.

4.4.3 RESIDUAL ANALYSIS

Non random residual indicate that the predicted value is biased you need to fix the model to produce unbiased predication

The modified R2 statistic is used to assess the fit of regression models with changing numbers of independent variables. The modified R2 value rises only when the additional term improves the model fit more than would be anticipated by chance alone.

The predicted R2 is used to assess how effectively a regression model predicts the future.

You can use regression analysis to do the following: -

1. Model multiple independent variables
2. Use polynomial term to model curvature
3. Include uninterpreted and absolute variables
4. Examine the interaction term to see if the effect of one independent variable is influenced by the value of another.

4.5 T00LS AND SOFTWARE

Microsoft Excel	For data analysis, a spreadsheet programme that includes graphing tools and pivot tables.
MATLAB R2019a	Numerical computation, visualization, and programming are all made easier with This dynamic environment and high-level language It is employed in the analysis of data, the development of algorithms, and the creation of models and applications. Instead of using spreadsheets or standard programming languages, language tools and built-in functions provide a faster way to explore alternative techniques.

4.6 MODEL PERFORMANCE

The goodness of fit of the ANN and MLR models to the testing data was assessed using two performance metrics.

1. Error in the root mean square (RMSE) It expresses the residue mistake as a mean square error, which looks like this:
   i. The given below formula is used in excel

$$RMSE = \sqrt{\frac{\sum_{1}^{n}(y_o - y_e)^2}{n}}$$

Y_o= Actual value of sediment

Y_e = predicted value of sediment

N = no of data

ii. The given below formula is used in matlab

$$RMSE = \mathrm{sqrt}\ [\mathrm{mean}\ (Y_t - Y_o)^2]$$

Y_t= target value

Y_o = output value

2. Correlation coefficient (r) it's a metric for how closely an estimated model's projected values match the real-world data. By using excel (regression) and matlab software we compare the value of $R^2$the value more close to 1 its mean our model is good.

4.7 MODEL DEVELOPMENT

For the present study MATLAB (R2019a) software was used to predict suspended sediment load. In this study we have taken four input data Rainfall(mm), Discharge(cumecs), Daily observed suspended sediment(g/l), average temperature of day Soft computing and traditional approaches were used to estimate sediment flux. This river's data spans the years 2016, 2017, 2018, and 2019, with the most recent data being from 2019. These years were chosen because they include the most up-to-date and full data. Data from the previous three years will be utilized for training, and data from the current year will be put to the test.

For rainfall, discharge daily observed sediment load, temperature, and suspended sediment concentration, there are 1461 data points available, of which 1096 will be used for training and the remaining 365 for testing. The data were standardized to a range of 0 to 1 in this investigation.

Different intervals of time series of rainfall, outflow, temperature, and daily observed suspended sediment remained used as input data and sediment flux (M/T) is used as output data to estimate daily SSC. Finding the optimum most effective input combination is crucial stage in any modeling process. Hidden neurons can be discovered utilizing the trial-and-error technique by selecting the best input data.

4.8 PROCEDURE

ANNs are characterised as feed forward networks based on the way of the input flow and dispensation. MLP stands for multi-layer perception and is a feed forward network. MLP has been proven to be a universal approximator in a lot of studies. Any finite nonlinear function may be successfully represented by an MLP with one hidden layer. The MLP contains three layers in this project: an input layer, a hidden layer, and an output layer. The neurons in each layer were linked to those in neighbouring layers, but data only flowed in one way: from input to output.

Neuron in the output layer represents the suspended sediment movement. A hit and trail method is utilized to determine the amount number of neurons in the buried layers. In the input layer, neurons represent input variables. Four inputs are utilised in this investigation, with two levels and a hidden layer of twenty layers providing the best predicted value.

The back-propagation (BP) approach is used to train the network. Input data feed forward and error data back-propagation are two directions in which BP incorporates information processing. The data is managed by the neurons in the input layer before being sent downcast to the succeeding layer by connections. Individual neuron analyses its net input as the weighted sum of total inputs before passing the information to the succeeding layer. The net input of each neuron is amplified or inhibited by a transformation function linked with it. In this example, TRANSIG is used as a transfer function. The training function and the transformation function were selected as a transig function, which is commonly used in current hydrological models. After the data processes reach the last layer and the last output is created, TRAINLM is used to calculate an error that shows the alteration among the expected and observed outputs.

Relevant physiographic data were first adjusted in the range of 0 to 1. The data was at odds into two training and testing groups while designing the artificial neural network. As a result, 70% and 30% of the data are allocated to training and testing, correspondingly. The data set was normalized using the equation below.

$$Y_{istd} = 0.1 + 0.8*[(Yi - Y_{imin})/ (Yimax - Y_{imin})]$$

Where Y_{istd} is the standard value of i,

Y_i= original value

Y_{imax}= max value

Y_{imim}= min value

The RMSE and R^2 between predictable and observed yearly sediment yield were utilized to evaluate performance in multiple linear regression and ANN techniques as follows:

$$RMSE = sqrt [mean (Y_t – Y_o)^2]$$

Y_t= target value

Y_o = output value

4.9 MULTIPLE LINEAR REGRESSION (MLP)

To re-create the connection multiple linear regression was used to assess the affiliation among the input and the sediment movement. The MLR model was donated as follows:

$$Y\hat{}= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

X_1, X_2, X_3 , X_4 are the independent variable

B_0 = regression intercept

B_j= coefficient of independent variable

Multiple linear regression calculates three things to get the best-fit line for each independent variable:

- The regression coefficients that result in the least amount of overall model error.
- The entire model's t-statistic.
- The p-value associated with it (How probable is it that the t-statistic occurred by accident if the null hypothesis of no connection between the independent and dependent variables is correct?)

The t-statistic and p-value for each of the regression coefficients in the model are then calculated.

Multiple linear regression can be used when you need to know the following.

The degree to which two or more independent variables are connected to a single dependent variable (For example how rainfall, temperature, and fertiliser application impact crop development).

The value of the dependent variable at a particular value of the independent variables (for example a crop's anticipated yield at various rainfall, temperature, and fertiliser applications).

Also, the RMSE is computed as follows:

$$RMSE = \sqrt{\frac{\sum_{1}^{n}(y_o - y_e)^{2}}{n}}$$

Y_o= Actual value of sediment

Y_e = predicted value of sediment

N = no of data

The MLR model's performance was also assessed using RMSE and R2 and compared to that of the ANN model.

5 CHAPTER 5

5.1 RESULT AND DISCUSSION

5.1.1 ANN MODEL

Trial and error were conditioned calculate the amount of neurons in the buried layer, and the most optimal neural network architecture was chosen for each case based on R^2, R, and RMSE.

The input and output layers, as well as the hidden layer, were designed for the feed forward back-propagation artificial neural network model. The artificial neural network is trained using the Levenberg-Marquardt (LM) training method with Transig as an activation function and a maximum of 1000 iterations Initially, a traditional regression analysis was performed, which involved randomly selecting 70% of available data for model creation and the remaining 30% for testing, and corelating the output parameter to the input parameters.

4-20-1 indicates 4 input layer, 20 hidden layers, and 1 output layer in the L-M based MLP neural network architecture. The number of hidden layers ranges from 10 to 20, and using the hit and trail error approach, we discovered that hidden layer 20 has an RMSE of 0.0075 and an R^2 of 0.98, indicating that our model is good.

The scatter plots below show that the actual and predicted values of the ANN model anticipated that the more locations are equal to the 45-degree line where the sediment value was equivalent to the projected values in the learning, validation, and test datasets.

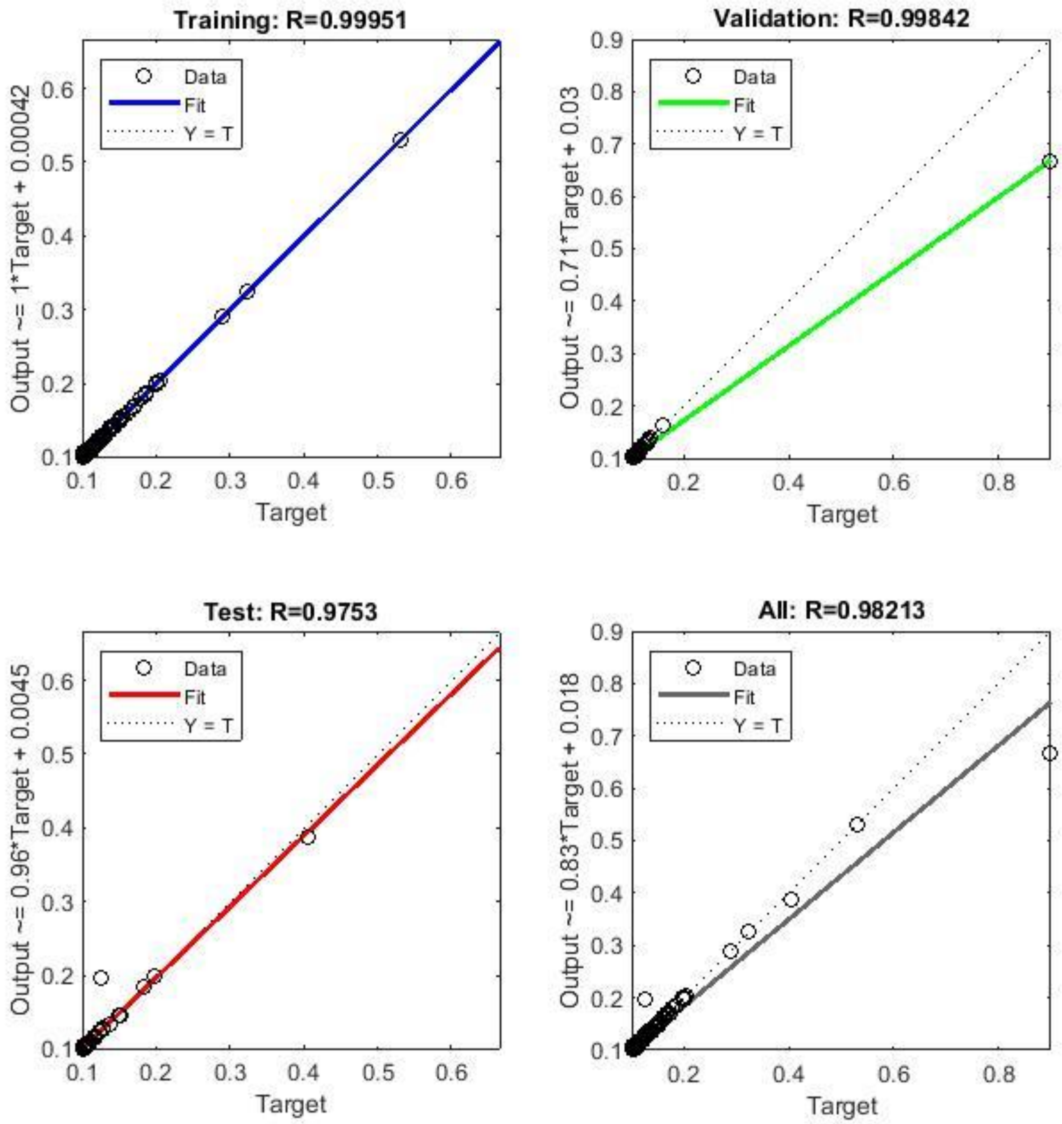

Figure 4Scatter plots of a suspended Sediment yield calculated using an ANN model.

The input variable choice is also crucial in influencing the network's accuracy. Rainfall, discharge, and daily observed suspended sediment were chosen from the initial set of climatic variables, along with temperature, since they were shown to be strongly connected to sediment deposition and could be used to reflect the effect of climatic conditions. Temperature is a metric of soil moisture, which has a major influence on soil erosion and sediment delivery to rivers, especially in the preceding month.

As indicated in the diagram below, the testing stage simply employed input data at the model, and there were no goal values in the testing stage. The bulk of the projected values appear to be very similar to the observed pattern.

These findings demonstrate that there are no outliers that might lower prediction accuracy for the time being. The results show that, based on a contrast of actual and predicted sediment deposit, modeling can be utilized to reliably forecast sediment flux rather than other modeling.

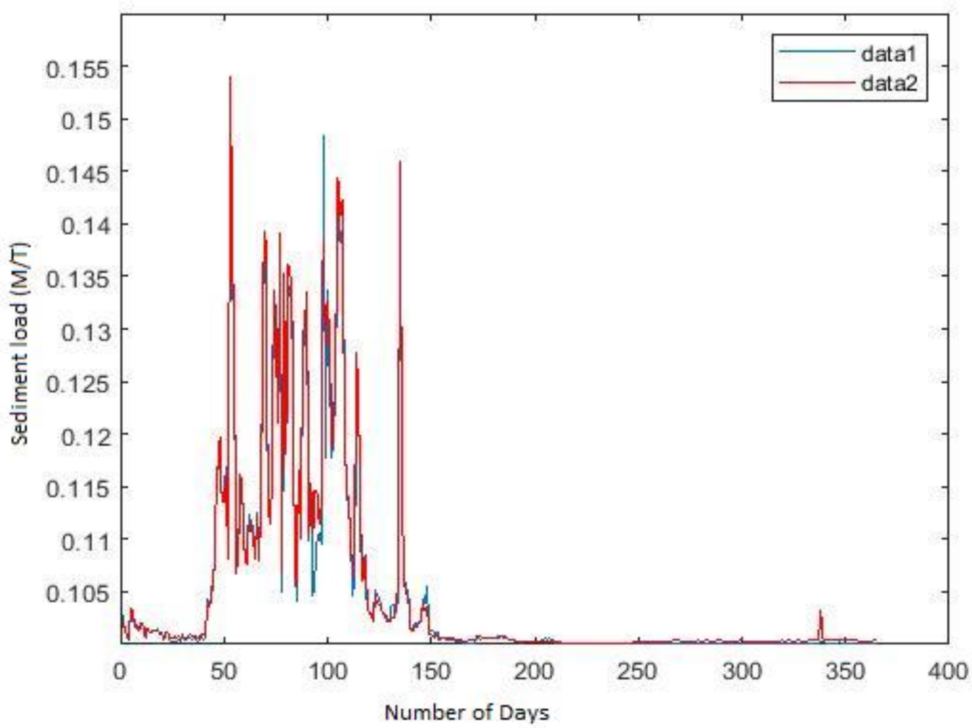

Figure 5 plot between the observed and predicted value of testing

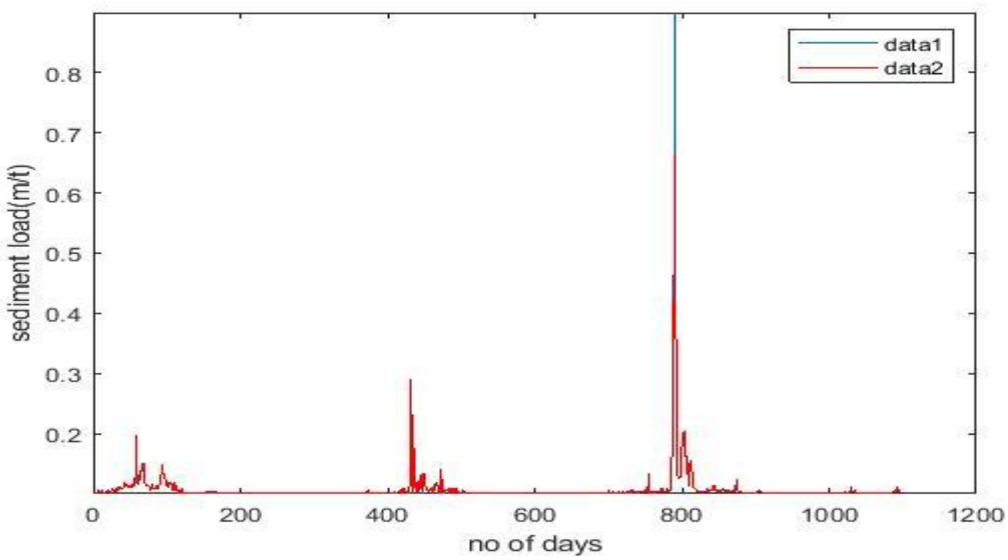

Figure 6 plot between the observed and predicted value of training

Figure 6 shows the ANN error, which is defined as the difference between observed and projected suspended sediment flow.

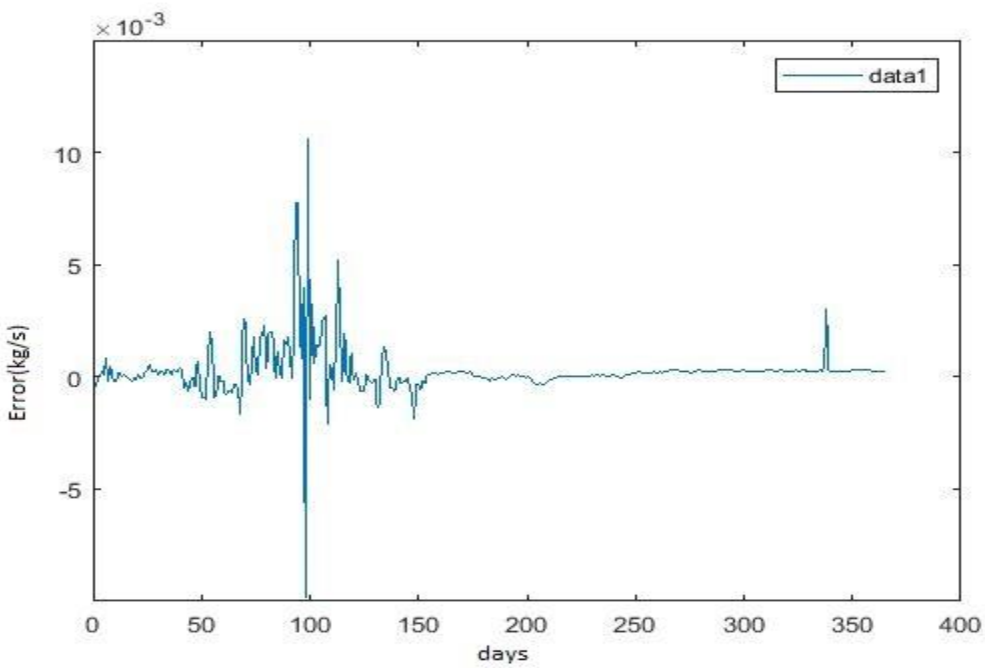

Figure 7 error plot of testing data

The below graph shows the best validation performance In general, the error decreases as the number of training epochs increases, but it may begin to grow on validation data that network starts to over fit the training data. In an epoch, we utilize all of the data exactly once to train the neural network using all of the training data for one cycle. A forward pass and a backward pass were combined to make one epoch.

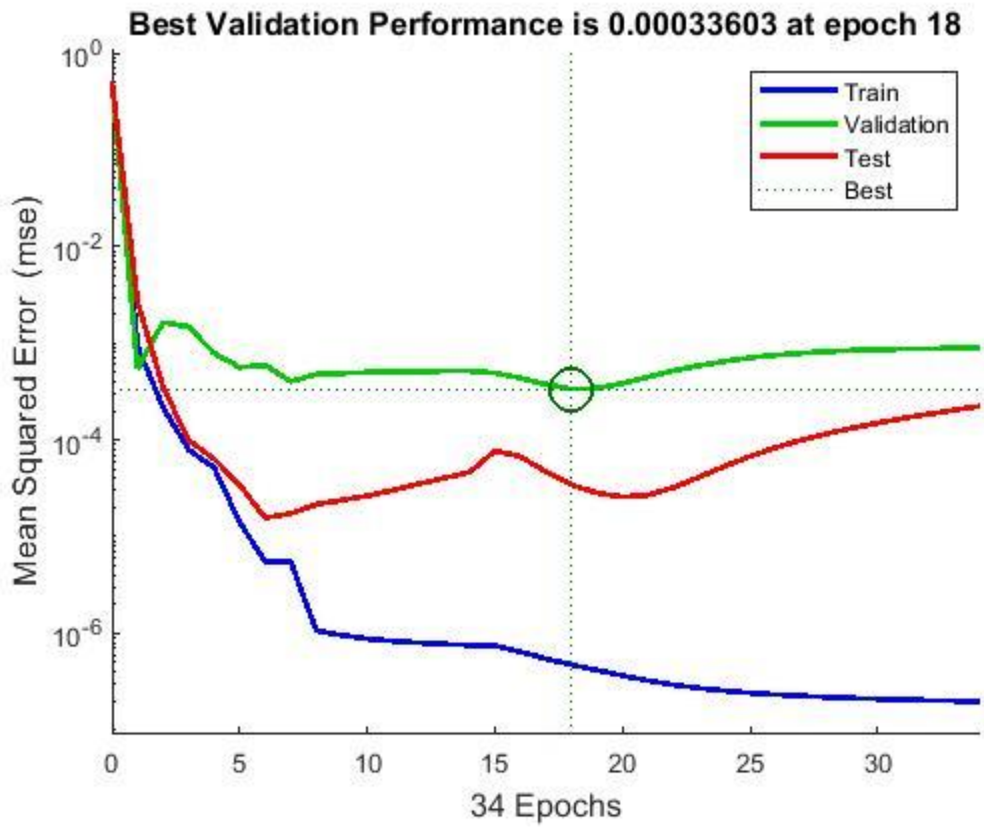

Figure 8 validation performance

5.1.2 MLR MODEL RESULT

The training regression model's significance F-value was 1.7773E-30, indicating that it is a viable MLR model. The p value is always less than α values which is 0.05 suggesting that the variable is relevant for MLR.

Table: - statistics of the MLR models on goodness-of-fit

PARAMETER	VALUE
R^2 of model	0.62
Adjusted R^2 of model	0.61
Standard error	0.017

$$\hat{Y} = -13666.5 - 186.65x_1 - 179.34x_2 + 16.89x_3 + 137014.5x_4$$

X_1, X_2, X_3, and X4 are the independent variable

B_0 = regression intercept

B_j= coefficient of independent variable

And the RMSE of MLR model is measured as 0.0556 and R^2 is 0.62

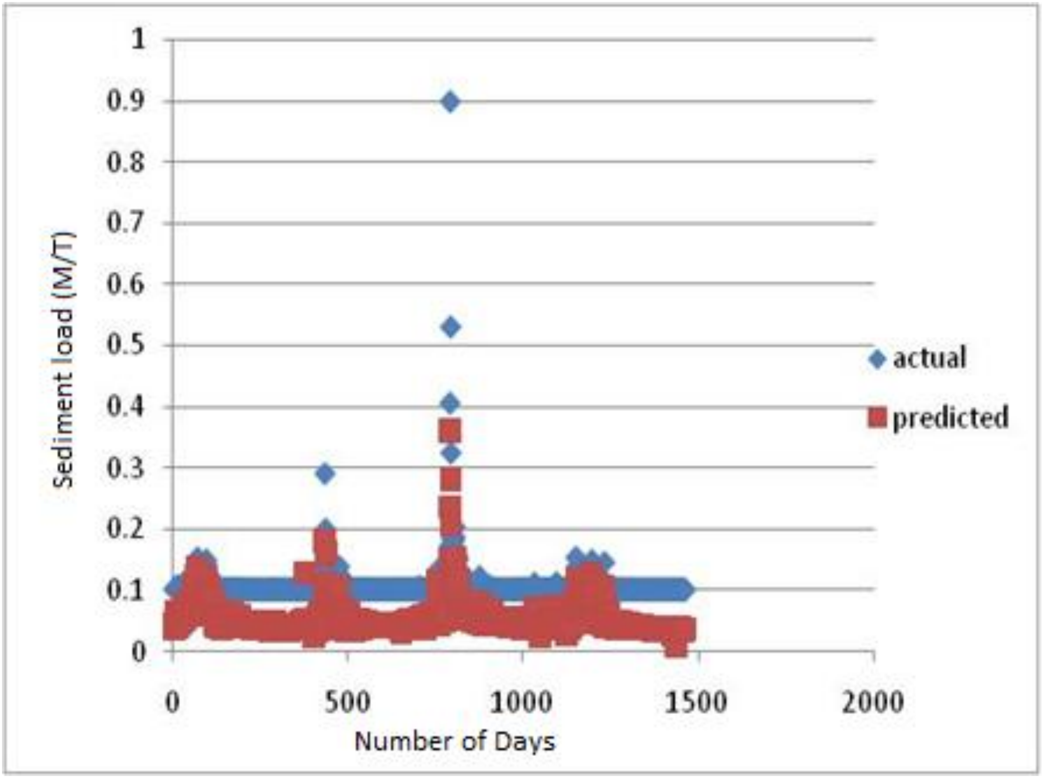

Figure 9 Actual and predicted suspended sediment yields are compared

Table: - Comparison of both model

MODEL	RMSE	R^2
ANN TRAINING	0.0075	0.98
ANN TESTING	0.0093	0.97
MLR	0.055	0.62

We can observe from the table above that statistical or conventional models are incapable of modeling complicated and nonlinear sediment yield processes, but the performance of ANN models is fairly good. The ANN model shows that the suspended sediment has the highest accuracy of prediction, with a value near to 1.

5.2 CONCLUSION

To predict yearly sediment flux yield in the BRAHMANI River, in this investigation, several linear regression models and artificial neural networks were utilized. The findings of collate the two techniques revealed that the artificial neural network had a greater perfection for forecasting the yearly residue load than the multiple linear regression method. The vast and parallel processing system, as well as the non-linear transformations, is the reasons behind this. In the multiple linear regression techniques, the regression tool and the linear equation were used to estimate the yearly sediment, and the final relationship was established with R2 and RMSE of 0.62 and 0.055, respectively. The best quantity of predicted sediment was reached with 4 inputs and 20 neurons in the hidden layer, with RMSE of 0.0075 and R2 = 0.98, according to the findings of the neural network implemented using back propagation algorithm. Also, The results of using the various functions revealed that the moving functions tansig as the artificial neural network's threshold function were more suited than the other functions. Using a neural network is preferable than other regression models since it is less sensitive to the input.

As a result, because the sediment flux phenomena are a non-linear process, ANN is a furtherdependable approach for monitoring sediment load across the river than the traditional regression method. The MLR models did not match the data set under investigation well.

6 REFRENCES

Agarwal, A., Singh, R.D., Mishra, S.K. and Bhunya, P.K., 2005. ANN-based sediment yield models for Vamsadhara river basin (India). Water Sa, 31(1), pp.85-100.

Ampomah, R., Hosseiny, H., Zhang, L., Smith, V. and Sample-Lord, K., 2020. A regression-based prediction model of suspended sediment yield in the cuyahoga river in Ohio using historical satellite images and precipitation data. Water, 12(3), p.881.

Anmala, J., Zhang, B. and Govindaraju, R.S., 2000. Comparison of ANNs and empirical approaches for predicting watershed runoff. Journal *of* water resources planning and management, 126(3), pp.156-166.

Arabkhedri, M., Lai, F.S., Ibrahim, N.A. and Mohamad-Kasim, M.R., 2010. Effect of adaptive cluster sampling design on accuracy of sediment rating curve estimation. Journal of Hydrologic Engineering, 15(2), pp.142-151.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000. Artificial neural networks in hydrology. II: Hydrologic applications. Journal of Hydrologic Engineering, 5(2), pp.124-137.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000. Artificial neural networks in hydrology. I: Preliminary concepts. Journal of Hydrologic Engineering, 5(2), pp.115-123.

Cigizoglu, H.K., 2004. Estimation and forecasting of daily suspended sediment data by multi-layer perceptrons. Advances in Water Resources, 27(2), pp.185-195.

Cobaner, M., Unal, B. and Kisi, O., 2009. Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. Journal of hydrology, 367(1-2), pp.52-61.

Demirci, M., & Baltaci, A. (2012). Prediction of suspended sediment in river using fuzzy logic and multilinear regression approaches. Neural Comput & Applic.

Eisazadeh, L.L., Sokouti, R., Homaee, M. and Pazira, E., 2013. Modeling sediment yield using artificial neural network and multiple linear regression methods. Int J Biosci, 3(9), pp.116-122.

Fu, L., 1994. Neural Networks in Computer Intelligence, McGraw-Ill. Inc., New york.

Ghose, D.K. and Samantaray, S., 2018. Modelling sediment concentration using back propagation neural network and regression coupled with genetic algorithm. Procedia Computer Science, 125, pp.85-92.

Gujarati, D.N. 2003. Basic Econometrics. New York: McGraw Hill Book Co.

Jain, S.K., 2001. Development of integrated sediment rating curves using ANNs. Journal of hydraulic engineering, 127(1), pp.30-37.

Karnin, E.D., 1990. A simple procedure for pruning back-propagation trained neural networks. IEEE transactions on neural networks, 1(2), pp.239-242.

Nagy, H.M., Watanabe, K.A.N.D. and Hirano, M., 2002. Prediction of sediment load concentration in rivers using artificial neural network model. Journal of Hydraulic Engineering, 128(6), pp.588-595.

Nivesh, S. and Kumar, P., 2017. Modelling river suspended sediment load using artificial neural network and multiple linear regression: Vamsadhara River Basin, India. Ijcs, 5(5), pp.337-344.

Nourani, V., Kalantari, O. and Baghanam, A.H., 2012. Two semidistributed ANN-based models for estimation of suspended sediment load. Journal of Hydrologic Engineering, 17(12), pp.1368-1380.

Raghuwanshi, N.S., Singh, R. and Reddy, L.S., 2006. Runoff and sediment yield modeling using artificial neural networks: Upper Siwane River, India. Journal of Hydrologic Engineering, 11(1), pp.71-79.

Sarangi, A. and Bhattacharya, A.K., 2005. Comparison of artificial neural network and regression models for sediment loss prediction from Banha watershed in India. Agricultural water management, 78(3), pp.195-208.

Singh, A., Imtiyaz, M., Isaac, R.K. and Denis, D.M., 2013. Comparison of artificial neural network models for sediment yield prediction at single gauging station of watershed in eastern India. Journal of Hydrologic Engineering, 18(1), pp.115-120.

Ulke, A., Tayfur, G. and Ozkul, S., 2009. Predicting suspended sediment loads and missing data for Gediz River, Turkey. Journal of Hydrologic Engineering, 14(9), pp.954-965.

Wang, P. and Linker, L.C., 2008. Improvement of regression simulation in fluvial sediment loads. Journal of Hydraulic Engineering, 134(10), pp.1527-1531.

Yeniguuml, K., Bilgehan, M., Gerger, R. and Mutlu, M., 2010. A comparative study on prediction of sediment yield in the Euphrates basin. International Journal of Physical Sciences, 5(5), pp.518-534.

Zhu, Y.M., Lu, X.X. and Zhou, Y., 2007. Suspended sediment flux modeling with artificial neural network: an example of the Longchuanjiang River in the Upper Yangtze Catchment, China. Geomorphology, 84(1-2), pp.111-125.