Watershed Prioritization and Decision Making Based On Weighted Sum Analysis, Feature Ranking and Machine Learning Techniques.

Kishanlal Ramlal Darji
Pandit Deendayal Energy University
Dhruvesh Prehladbhai Patel (dhruvesh1301@gmail.com)
Pandit Deendayal Energy University https://orcid.org/0000-0002-2074-7158
Vinay Vakharia
Pandit Deendayal Energy University
Jaimin Panchal
Pandit Deendayal Energy University
Amit Kumar Dubey
ISRO: Indian Space Research Organisation
Praveen Gupta
ISRO: Indian Space Research Organisation
Raghavendra P Singh
ISRO: Indian Space Research Organisation

Research Article

Keywords: Morphometric analysis, RS, GIS, ML, Prioritization

Posted Date: May 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-359657/v1

Title: Watershed Prioritization and Decision Making based on Weighted Sum Analysis, Feature Ranking and Machine Learning Techniques

The full names of the authors: Kishanlal Darji[1], Dhruvesh Patel[1*], Vinay Vakharia[2], Jaimin Panchal[2], Amit Kumar Dubey[3], Praveen Gupta[3], Raghavendra P Singh[3]

Affiliations:

[1]Department of Civil Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

[2]Department of Mechanical Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

[3]Space Applications Center, ISRO, Ahmedabad, Gujarat, India

*__Correspondence:__ Dhruvesh Patel, Email: dhruvesh.patel@sot.pdpu.ac.in, Mobile Number: +919408479792

Watershed Prioritization and Decision Making based on Weighted Sum Analysis, Feature Ranking and Machine Learning Techniques

Kishanlal Darji[1], Dhruvesh Patel[1*], Vinay Vakharia[2], Jaimin Panchal[2], Amit Kumar Dubey[3], Praveen Gupta[3], Raghavendra P Singh[3]

[1]Department of Civil Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

[2]Department of Mechanical Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

[3]Space Applications Center, ISRO, Ahmedabad, Gujarat, India

[*]FULL CONTACT OF CORRESPONDING AUTHOR:

Dhruvesh Patel

e-mail: dhruvesh.patel@sot.pdpu.ac.in

Mobile Number: +919408479792

Abstract

Prediction and validation of Compound factors for prioritization of watersheds is an essential application using Machine Learning (ML) Techniques in water resources engineering. In the current paper, a method is proposed to derive 14 morphometric and 3 Topo-hydrological parameters using Remote Sensing (RS) and Geographical Information System (GIS), whereas prediction and validation of compound factor using ML techniques. Compound factor (CF) values are calculated using Weighted Sum Analysis (WSA), ReliefF, correlation coefficient techniques. A ten-fold cross-validation technique is applied to two machine learning models Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM). Predication accuracy of models has been further achieved by feature ranking. The accuracy of ML models is evaluated with three parameters, Mean Absolute Error (MEA), Correlation Coefficient (CC), and Root Mean Square Error (RMSE). With the ranked features and Ten-fold cross-validation, prediction results were found to be better. The methodology will be useful for the accurate prediction of CF values and to reduce the uncertainty in watershed prioritization for conservation techniques for soil and water.

Keywords: Morphometric analysis, RS, GIS, ML, Prioritization

1. Introduction

Watershed prioritization is crucial for the development of watershed management planning and better land management. It is essential to find out the watershed priority in the semi-arid and arid regions as it helps in water management in ungauged rivers. Spatial prioritization and watershed health help to understand watershed conditions, and it also helps in deriving better management strategies in the data-scarce region (Alilou et al. 2019).

Remote Sensing (RS) and Geographical Information System (GIS) is a promising tool to extract the important parametrical information from watersheds. The RS and GIS technique helps with the establishment of interrelationship with parameters and to decide the priority of watershed using the WSA technique (Malik et al. 2019). Many researchers have attempted to prioritize watershed (Kadam et al. 2017; Patel et al. 2012; Patel et al. 2015; Samal et al. 2015; Thakkar and Dhiman 2007) and utilized different analysis technique for watershed prioritization such as Multi-criteria decision analysis (MCDA) (Chowdary et al. 2013; Jaiswal et al. 2015; Memon et al. 2020; Samal et al. 2015; Vulević and Dragović 2017), Sediment Yield Index (Khan et al. 2001), Weighted Sum Analysis (WSA) (Aher et al. 2014; Memon et al. 2020) and Principal Component Analysis (PCA) (Farhan et al. 2016; Meshram and Sharma 2017). Out of this WSA technique is the most familiar and prominent method for prioritizing the watersheds in the present era, however, the recent development of ML techniques helps to classify, predict and forecast the laboratory and computer-simulated data for decision-making. Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest are ML models applied in various engineering applications.

ML is a probabilistic model frequently applied for pattern recognition applications. It has been used for applications like Motor current signature analysis (Singh et al. 2014), condition monitoring (Vakharia et al. 2015), fault diagnosis (Kankar et al. 2011), compressive strength prediction (Sonebi et al. 2016), EEG (Upadhyay et al. 2016), tool wear rate prediction (Vakharia et al. 2018) and many more applications, regardless of field.

For evaluating any type of ML model and for knowing how well the model predicts without biasedness, cross-validation is used. In K-fold cross-validation, the data is partitioned randomly into k number of approximately equal sets, from which model is trained on k-1 sets and tested on the remaining set. The procedure is repeated k number of times and the final result is the average of all the results obtained by repetition. The main benefit to apply this strategy is that each sample is used for training k-1 times and one time for testing and the average results are obtained. The regression model is used to develop a mathematical function based on the experimental data in which parameters act as features. It is observed that in a feature vector, prediction capability can be improved with identifications of relevant and irrelevant features. To discard irrelevant features, a feature selection strategy can be used to determine the utility of individual features in a feature vector (Vakharia et al. 2017). Other feature ranking techniques like Fisher score (Vakharia et al. 2016), Information Gain (Naseriparsa et al. 2014), ReliefF (Kononenko 1994) are also used to improve accuracy and prediction of classification and regression problems.

With the availability of data, the use of artificial intelligence techniques has been applied in a variety of civil engineering applications. According to literature research, it is found that in predicting and validating test results, machine learning strategies are helpful. Furthermore, for decision making a particular algorithm may not be suitable, in such situations detailed investigations are needed to assess the usefulness of ML algorithms.

In the present paper, RS and GIS techniques are utilized to derive Morphometric parameters and Topo-hydrological parameters. The values and ranks to all 14 morphometric and 3 Topo-hydrological parameters are assigned. Initially, the CF values are calculated using well-known WSA techniques. After that, two feature ranking techniques ReliefF and Correlation coefficient are applied to calculate the weightage of morphometric parameters. To predict CF, a Ten-fold cross-validation technique is applied on Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM). The accuracy of ML models is examined with mean absolute error (MAE), Correlation coefficient (C_r), and root mean square error (RMSE). Based on better prediction capability, decision-making is done for watershed prioritization. As per the literature survey, most of the authors applied ML techniques to the variety of applications in RS and GIS techniques. To the author's best knowledge, decision-making for watershed prioritization using feature ranking and ML techniques is not explored till now. Therefore, to fill the gap, the novel approach for calculating CF value for watersheds prioritization for the decision-making activity is executed in the present research. It helps to provide an integrated solution to the decision-maker to restore the watersheds against any uncertain critical conditions.

2. Regression using machine learning

2.1 Support Vector Machine

Support Vector Machine (SVM) (Vapnik 2013) is an ML algorithm, used for classification and regression. This paper follows SVM-r (regression) which uses a set of training data x_m for m

number of observations to predict the Compound Factor (CF) as a response value. Linear SVM-r formulated as a convex optimization problem in which the goal is to minimize the error J as shown in (**Fig.1**).

$$J = \left[\frac{1}{2}\|v\|^2\right] + \left[C \sum_{j=1}^{m}(\xi_j - \xi_j^*)\right] \quad (1)$$

Where, (ξ_j) and (ξ_j^*) are the slack variables introduced to deal with infeasible constraints.

2.2 Multi-layer perceptron (MLP)

Multi-layer Perceptron (MLP) is an ML algorithm that used a function $F: X^n - X^m$ where n represents dimensions of input parameters and m represents the dimensions of parameters to be predicted. It is generally used for classification and regression which uses a set of features $g = g_1, g_2, g_3, \cdots, g_n$as an input for a response value y, which can be a linear function or a non-linear function. For regression, MLP uses back propagation for training without any activation function in the output layer. Hence for output, it gives a set of continuous values and mean square error as a loss function. (Pedregosa et al. 2011). (**Fig.2)** shows the architecture of MLP with one hidden layer.

3. Materials and methods

To determine the CF for any catchment area of the River, a Rel-River catchment, which is situated in Banaskantha district of Gujarat, India is considered in the present study as shown in (**Fig.3a)**. Basin lies between 24^0 50'N to 24^0 75' N latitude and 72^0 00'E to 72^0 45' E longitude and has an area of 431 km^2 (Memon et al. 2020). The basin area is partitioned into 51 micro-watersheds and

Cartosat DEM 10 m resolution is used to calculate slope and watershed boundary. ArcHydro toolset is used for calculating and deriving the boundary of the watershed. Drainage boundaries are derived using SOI Toposheets (42D02, 42D03, 42D06) and the ordering of the drainages are given according to the proposed technique by (Strahler 1964) (**Fig.3b).** ArcGIS 10.5 software was used to generate watersheds boundary, digitize drainages, and derive different morphometric parameters. Morphometric parameters such as linear aspects, areal aspects, relief aspects, (Faniran 1968; Melton 1957; Miller 1953; Ratnam et al. 2005; RE 1932; RE 1945; Schumm 1956; Strahler 1997) and Topo-hydrological parameters such as Stream Power Index (SPI) (Whipple and Tucker 1999), Sediment Transport Index (STI) (Moore and Burch 1986), and Topographic Wetness Index (TWI) (Beven and Kirkby 1979) are considered for the prioritization of watersheds, (**Table 1**). The ranking of the watersheds is calculated based on the WSA technique established by (Aher et al. 2014) (**Table2-4**). Every micro-watershed has different and unique characteristics, and watershed priority is carried out to identify the vulnerable areas for erosion. The weights have been calculated based on the correlation between parameters and their values. Its relation does a ranking of the parameters to soil erosion. Based on the literature, the linear factor showed a positive correlation with soil erosion, so that maximum priority is given to the higher value i.e., rank 1, while shape factors displayed a negative correlation for soil erosion and the ranking for these parameters given in reverse order.

The final ranking and prioritization of watersheds have been done using CF values. The CF values are calculated with help of WSA, ReliefF, and correlation coefficient methods (**Fig.4**).

3.1 WSA Techniques**:**

The WSA is a statistical approach which applied for the calculation of CF values. The CF values are estimated from morphometric parameters and weightage obtained using cross-correction analysis. The statistical expression for CF is written as follows (Aher et al. 2014).

$$CF\ (WSA) = R_{MP}\ \boldsymbol{X}\ W_{MP} \tag{2}$$

Where *CF (WSA)* represents compound factor, R_{MP} represents the rank (preliminary priority) estimated from morphometric and topo-hydrological parameters, and W_{MP} represents the weight of morphometric and Topo-hydrological parameters obtained using cross-correlation study. Weights are estimated with a ratio of morphometric and topo-hydrological parameters with a sum of the correlation coefficient value of each parameter (**Table 5**). Based on the weightage of parameters, a model is constructed for sorting of watershed prioritization, which is computed as follows:

$$CF\ (WSA) = (0.055447) \times R_b + (0.087678) \times D_d + (0.114486) \times F_u + (0.111822) \times R_t + (-0.08424) \times L_o + (0.025944) \times R_f + (-0.01748) \times B_s + (0.028237) \times R_e + (-0.00423) \times C_c + (0.112984) \times I_f + (0.007042) \times R_c + (\ 0.089087) \times C + (0.047331) \times R_h + (0.115931) \times R_n + (0.090761) \times STI + (0.105501) \times SPI + (0.113691) \times TWI \tag{3}$$

As observed from (**Table 6**) and equation 3, the highest CF value is 51.13 and it is for watershed 49, the second-highest value 47.48 and it is for watershed 48 and so on for other watersheds as shown in (**Fig.5a**).

3.2 ReliefF

ReliefF is a feature ranking method for selecting important features for classification and regression problems. In this method, predictors that give different values for the same response values are penalized and predictors that give different values for different response values are rewarded. Final predictor weights are computed using intermediate weights.

$$P_w = \frac{W_{Rp}}{W_R} - \frac{p_R - W_{Rp}}{n - W_R} \tag{4}$$

Where *n* is several instances, W_{RP} is weighted with different response values and different values for predictor, W_R is weighted with different values for response *R* and P_R is weights with different values for predictor *p* (Robnik-Šikonja and Kononenko 1997).

The ReliefF method is used to assign the ranks to the morphometric parameters. I_f ranked as first with the weightage of 0.058 whereas F_u is ranked second with the weightage of 0.056. Other subsequent values, ranks, and weights are mentioned in (**Table 7**). The final equation of compound factor obtained using the ReliefF method is as follows:

$$CF\ (RF) = (0.00532) \times R_b + (0.02866) \times D_d + (0.05694) \times F_u + (0.04566) \times R_t + (0.02452) \times L_o + (-0.03597) \times R_f + (-0.03246) \times B_s + (-0.03282) \times R_e + (-0.0182) \times C_c + (0.05803) \times I_f + (-0.0198) \times R_c + (0.02878) \times C + (0.01382) \times R_h + (0.05113) \times R_n + (0.02103) \times STI + (0.02622) \times SPI + (0.03788) \times TWI \tag{5}$$

As observed from (**Table 8**) and equation 5, the highest CF value is 14.37 corresponding to watershed 49 and second-highest value 12.78 corresponding to watershed 48, likewise other subsequent CF values are calculated as shown in (**Table 8** and **Fig.5b**).

3.3 Correlation Coefficient (C_r)

The correlation coefficient (C_r) is generally used to measure and rank the relationship between the calculated and predicted values. A value near *+1* indicates a perfect correlation from the model output, whereas, the value of *-1* indicates negative correlations. If a value is closer to *0*, it means that there is no direct correlation between the calculated and predicted parameters (Gibbons and Chakraborti 2020).

$$\overline{X_p} = \sum_{i=1}^{m} \frac{X_{(p,i)}}{m}; \overline{Y_q} = \sum_{j=1}^{m} \frac{Y_{(q,j)}}{m} \tag{6}$$

$$\rho_{(p,q)} = \frac{\sum_{i=1}^{m}\left(X_{(p,i)}-\overline{X_p}\right)\left(Y_{(q,i)}-\overline{Y_q}\right)}{\sqrt{\sum_{i=1}^{m}\left(X_{(p,i)}-\overline{X_p}\right)^2 \sum_{j=1}^{m}\left(Y_{(q,j)}-\overline{Y_q}\right)^2}} \tag{7}$$

Here, $\rho_{(p,q)}$ gives the value of Pearson's linear correlation coefficient, *m* is the length of parameters, $\overline{X_p}$ and $\overline{Y_q}$ gives the values of a mean of each parameter.

The correlation Co-efficient method was used to assign the ranks to the morphometric parameters. I_f having the 1 rank with the weightage of 0.93 whereas F_u is 2 with a weightage of 0.9197. Other subsequent values, ranks, and weights are cited in (**Table 7**). Equation of compound factor based on the weightage of Correlation coefficient ranked features is as follows:

$$CF\ (CC) = (0.3689) \times R_b + (0.8069) \times D_d + (0.9197) \times F_u + (0.8659) \times R_t + (-0.7855) \times L_o + (0.1328) \times R_f + (-0.071) \times B_s + (0.1487) \times R_e + (-0.0267) \times C_c + (0.9373) \times I_f + (0.0484) \times R_c + (\ 0.8161) \times C + (0.2701) \times R_h + (0.9104) \times R_n + (0.7203) \times STI + (0.7744) \times SPI + (0.8488) \times TWI \qquad (8)$$

As observed from (**Table 9**) and equation 8, the highest CF value is 405.67 and it is for watershed 49, the second-highest value 375.3 and it is for watershed 48 as shown in (**Fig.5c**).

4. Results and discussion

In the present paper, the basic morphometric parameters are calculated and derived using RS and GIS Techniques. Various parameters are mentioned in (**Table 2 and 3**) and the ranks are assigned based on the relation of parameters with soil erodibility, which is shown in (**Table 4**). In the present study, the prioritization of watershed was decided based on CF values calculated through WSA, ReliefF, and correlation coefficient method. Validation and Comparison of CF values have been done using MLP and SVM.

4.1 Watershed priority based on CF values

The watersheds CF values are calculated using WSA, ReliefF, and correlation coefficient method. For prioritization, the lowest value of CF is given the priority rank 1 and the next lower value is given priority rank of 2, and so on for all the 51 micro-watersheds calculated through three methods mentioned above. Furthermore, the prioritized watersheds are categorized in five different categories i.e., very high, high, moderate, low, and very low. Finally, the categories watershed maps are utilized for the decision-making system.

As observed from (**Table 6**), the lowest CF values based on WSA are 1.34, hence it is assigned rank 1 and same ranking is assigned which will be useful for prioritizing the watersheds. (**Table 6, 8, 9**) shows the CF values calculated and the prioritized watersheds based on WSA, ReliefF, and correlation coefficient method. For CF values obtained through ReliefF, all the 51 micro-watersheds of Rel River catchment are classified into five priority categories (Aher et al. 2014) such as (i) very high (-1.355 to ≤ 1.789), (ii) high (1.789 to ≤ 4.933), (iii) medium (4.933 to ≤ 8.078), (iv) low (8.078 to ≤11.223), and (v) very low (> 11.223) as given in (**Table 10**). It is observed through (**Table 10)**, that the 8 micro-watersheds belongs to very high category (Ws no 1,3,4,5,7,8,9 and 11), 8 micro-watersheds under a high category (Ws no 2,6,10,12,14,15,17 and 19), 14 micro-watersheds under medium (Ws no 16,18,20,21,23,26,30,34,35,39,41,42,44 and 45), 12 micro-watersheds under a low category (Ws no 22,24,25,27,31,32,33,38,40,46,47 and 51), and 9 micro-watershed under the very low category (Ws no 13,28,29,36,37,43,48,49 and 50). The final priority category map of 51 micro-watersheds is shown in (**Fig.6**). It is observed that the percentage area of micro-watersheds under the very high category is 13.45%, for high category it is 15.72%, medium category is 30.24%, low category is 18.81%, and for very low category is 21.78%. This information is very helpful in the implementation of water management strategies in terms of soil and water conservation measures. The result also shows its vulnerability to erosion and runoff potential. This watershed is mostly situated in the upstream part of the catchment or basin, and they are influenced by high slopes and elongated basins. The priority map for soil erosion potential is shown in (**Fig.6)**.

4.2 Discussion

To predict accurate CF values for watersheds prioritization, two ML algorithms such as MLP and SVM have been utilized. The training and cross-validation performed all the 51 micro-watersheds and feature vector is normalized in the range [-1 to +1] to minimize the biasing error. To ensure the suitability of ML algorithms three performance metrics: mean absolute error, correlation coefficient, and root mean square error are assessed.

4.2.1 Correlation Coefficient

The correlation coefficient (R) is calculated to find the relation between the dependent variable and the independent variable. When R is observed as +1, it indicates a perfect correlation between the two variables while -1 indicated a negative correlation between two variables. The η shows the total number of observations, $\sum a$ shows the mean value of all the first variables in the data set, $\sum b$ represents the mean value of all the second variables in the data set. It is mathematically calculated by:

$$R = \frac{\eta \sum ab - \sum a \sum b}{\sqrt{\eta \sum a^2} - \sum a^2 [\eta \sum b^2 - (\sum b)^2} \tag{9}$$

4.2.2 Mean absolute error

Mean absolute error (MAE) has been used to analyze the performance of the ML model. The difference between the predicted value and the actual value is calculated by:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\emptyset_i - \emptyset| \tag{10}$$

Here, $\emptyset_i$ shows the predicted value by the machine learning model, $\emptyset$ shows the actual value from experiments.

4.2.3 Root mean square error

Root Mean Square Error (RMSE) represents the standard deviation of the residuals. RMSE is commonly used in Hydrology, forecasting, and regression analysis to verify experimental results.

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (\times_i - \times)^2} \tag{11}$$

Where i is variable, n represents data points, $\times_i$ represent actual observations, and × represent predicted observation.

(**Table. 11 and 12**) shows the results of SVM Ten-fold and MLP Ten-fold for CF (WSA), CF (RF), and CF (CC) values. The comparison was made between the correlation coefficient, MAE, and RMSE values. At first instant, the comparison was made through the correlation coefficient algorithm, however, it was observed that the values through SVM and MLP doesn't have any significant deviation for identifying a suitable CF. All the values come nearer to 1 which shows a good correlation, hence not included for made further comparison.

Now, (**Fig.7 a and b**) show the SVM Ten Fold and MLP Ten Fold results for comparison of CF values using MAE and RMSE. From (**Fig.7a**) it is clear that the minimum MAE is 0.0236 for CF(RF) using an SVM Ten-fold compare to minimum MAE is (0.2856) for CF (RF) using MLP

Ten-fold. Furthermore, minimum RMSE is 0.032 for CF (RF) using an SVM Ten-fold compare to minimum RMSE is 0.440 for CF (RF) using MLP Ten-fold. The maximum MAE is 0.3684 for CF (CC) using an SVM Ten-fold compare to the maximum MAE is 7.6514 for CF (CC) using MLP Ten-fold. Also, the maximum RMSE is 0.4361 for CF(CC) using an SVM Ten-fold compare to the maximum RMSE is 12.26 for CF (RF) using MLP Ten-fold. The present result shows that the SVM Ten-fold gives better results compare to MLP Ten-fold for the prediction of MAE and RMSE for CF (RF), CF(WSA), and CF (CC).

(**Fig.8 a, b, and c**) shows the variation in MAE and RMSE in all the three CF(R), CF (WSA), and CF (CC) using SVM Ten-fold and MLT Ten-fold method. The minimum MAE is 0.0236 for predicting CF (RF) compare to 0.3684 in CF (CC) using SVM Ten-fold. The maximum MAE is 0.2856 for predicting CF (RF) compare to 7.6514 in CF (CC) using MLP Ten-fold. Furthermore, the minimum RMSE for predicting CF (RF) is 0.032 compared to 0.4361 in CF (CC) using SVM Ten-fold. The maximum RMSE for predicting CF (RF) is 0.4402 compared to 12.2689 in CF (CC) using MLP Ten-fold. Similarly, the prediction of the comparison of MAE and RMSE for CF (WSA) using SVM Ten-fold and MLP Ten-fold is included in (**Fig.8b**).

5. Conclusion

In the methodology proposed, the utility of ML algorithms explored in detail for watershed prioritization and decision making system. For the prediction of CF values ML regression techniques such as SVM and MLP were used. Finally, the performance of models estimated after comparing the three parameters i.e., mean absolute error, Correlation Coefficient, and root mean square error. The findings are mentioned below:

1. ReliefF and Coefficient correlation found to be an efficient method to calculate the CF values for prioritization of watersheds in addition to WSA techniques.
2. SVM is an efficient algorithm for predicting CF-RF, CF-WSA, and CF-CC as compare to MLP.
3. The comparison of MAE and RMSE for predicting CF-RF, CF-WSA, and CF-CC shows that CF-RF is the best model for the prediction of CF values as compared to other calculations, hence it is utilized for prioritization of watersheds.
4. It is suggested that watersheds 1,3,4,5,7,8,9, and 11 have a very high vulnerability to soil erosion followed by the rest of the watersheds in the study region. Hence, appropriate soil and water conservation measures would be adopted as a protection against degradation.

The integrated framework of RS, GIS, feature ranking, and machine learning is an efficient technique for watersheds ranking and prioritization in water resource management. Ten-fold cross-validation and feature ranking is a novel approach for accurate prediction of CF for watershed ranking and prioritization. The utility of machine learning techniques relies on the data availability and calculated morphometric parameters.

6. Acknowledgment

The authors would like to thanks, various Government agencies and SAC-ISRO for providing useful data for research. Corresponding author is also thankful to Mr. Raviraj Dave formal JRF of

Civil Engineering Department, PDEU and research scholar, IIT Gandhinagar for helping in initial set up.

7. Declarations

Funding

This work was supported by SAC-ISRO (Grant no; SAC/EPSA/GHCAG/LHD/SARITA/01/19)

Conflicts of interest/Competing interests

None

Availability of data and material

The authors confirm that the data supporting the findings of this study are available within the article or could be requested from the corresponding author, upon reasonable request.

Code availability

Not applicable

Authors' contributions

Conceptualization: [Dhruvesh Patel]; Methodology: [Dhruvesh Patel, Vinay Vakharia]; Analysis and investigation: [Dhruvesh Patel, Vinay Vakharia, Kishanlal Darji, Jaimin Panchal]; Writing-original draft preparation: [Dhruvesh Patel, Vinay Vakharia]; Data curation: [Dhruvesh Patel, Kishanlal Darji]; Review, editing and supervision [Dhruvesh Patel, Vinay Vakharia, Amit Kumar Dubey, Praveen Gupta and Raghavendra P Singh]; Funding acquisition: [Dhruvesh Patel]

Ethical approval

Not applicable

Consent to participant

Not applicable

Consent to publication

Not applicable

8. References

Aher P, Adinarayana J, Gorantiwar SJJoH (2014) Quantification of morphometric characterization and prioritization for management planning in semi-arid tropics of India: a remote sensing and GIS approach 511:850-860 doi:https://doi.org/10.1016/j.jhydrol.2014.02.028

Alilou H et al. (2019) Evaluation of watershed health using Fuzzy-ANP approach considering geo-environmental and topo-hydrological criteria Journal of environmental management 232:22-36 doi:https://doi.org/10.1016/j.jenvman.2018.11.019

Beven KJ, Kirkby MJJHSJ (1979) A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant Hydrological Sciences Journal 24:43-69 doi:https://doi.org/10.1080/02626667909491834

Chowdary V, Chakraborthy D, Jeyaram A, Murthy YK, Sharma J, Dadhwal V (2013) Multi-criteria decision making approach for watershed prioritization using analytic hierarchy process technique and GIS Water resources management 27:3555-3571 doi:https://doi.org/10.1007/s11269-013-0364-6

Faniran AJAJS (1968) The index of drainage intensity: a provisional new drainage factor Aust J Sci 31:326-330

Farhan Y, Anbar A, Al-Shaikh N, Mousa R (2016) Prioritization of semi-arid agricultural watershed using morphometric and principal component analysis, remote sensing, and GIS techniques, the Zerqa River Watershed, Northern Jordan Agricultural Sciences 8:113-148 doi:https://doi.org/10.4236/as.2017.81009

Gibbons JD, Chakraborti S (2020) Nonparametric statistical inference. CRC press,

Jaiswal R, Ghosh N, Galkate R, Thomas T (2015) Multi criteria decision analysis (MCDA) for watershed prioritization. Aquat Proc 4: 1553–1560. doi:https://doi.org/10.1016/j.aqpro.2015.02.201

Kadam AK, Jaweed TH, Umrikar BN, Hussain K, Sankhua RN (2017) Morphometric prioritization of semi-arid watershed for plant growth potential using GIS technique Modeling Earth Systems and Environment 3:1663-1673 doi:https://doi.org/10.1007/s40808-017-0386-9

Kankar P, Sharma SC, Harsha SJJoV, Control (2011) Rolling element bearing fault diagnosis using autocorrelation and continuous wavelet transform Journal of Vibration 17:2081-2094

Khan M, Gupta V, Moharana P (2001) Watershed prioritization using remote sensing and geographical information system: a case study from Guhiya, India Journal of Arid Environments 49:465-475 doi:https://doi.org/10.1006/jare.2001.0797

Kononenko I Estimating attributes: Analysis and extensions of RELIEF. In: European conference on machine learning, 1994. Springer, pp 171-182. doi:https://doi.org/10.1007/3-540-57868-4_57

Malik A, Kumar A, Kandpal HJAJoG (2019) Morphometric analysis and prioritization of sub-watersheds in a hilly watershed using weighted sum approach Arabian Journal of Geosciences 12:118 doi:https://doi.org/10.1007/s12517-019-4310-7

Melton MA (1957) An analysis of the relations among elements of climate, surface properties, and geomorphology. Columbia Univ New York,

Memon N, Patel DP, Bhatt N, Patel SBJNH (2020) Integrated framework for flood relief package (FRP) allocation in semiarid region: a case of Rel River flood, Gujarat, India 100:279-311 doi:https://doi.org/10.1007/s11069-019-03812-z

Meshram SG, Sharma S (2017) Prioritization of watershed through morphometric parameters: a PCA-based approach Applied Water Science 7:1505-1519 doi:https://doi.org/10.1007/s13201-015-0332-9

Miller VC (1953) A quantitative geomorphic study of drainage basin characteristics in the clinch mountain area virginia and tennessee. Columbia univ new york,

Moore I, Burch GJWRR (1986) Sediment transport capacity of sheet and rill flow: application of unit stream power theory Water Resources Research 22:1350-1360 doi:https://doi.org/10.1029/WR022i008p01350

Naseriparsa M, Bidgoli A-M, Varaee TJapa (2014) A hybrid feature selection method to improve performance of a group of classification algorithms arXiv preprint arXiv doi:https://arxiv.org/ct?url=https%3A%2F%2Fdx.doi.org%2F10.5120%2F12065-8172&v=abba9fe4

Patel DP, Dholakia MB, Naresh N, Srivastava PK (2012) Water harvesting structure positioning by using geo-visualization concept and prioritization of mini-watersheds through morphometric analysis in the Lower Tapi Basin Journal of the Indian Society of Remote Sensing 40:299-312

Patel DP, Srivastava PK, Gupta M, Nandhakumar N (2015) Decision Support System integrated with Geographic Information System to target restoration actions in watersheds of arid environment: A case study of Hathmati watershed, Sabarkantha district, Gujarat Journal of Earth System Science 124:71-86

Pedregosa F et al. (2011) Scikit-learn: Machine learning in Python the Journal of machine Learning research 12:2825-2830

Ratnam KN, Srivastava Y, Rao VV, Amminedu E, Murthy KSRJJotISoRS (2005) Check dam positioning by prioritization of micro-watersheds using SYI model and morphometric analysis@ Remote sensing and GIS perspective Journal of the Indian Society of Remote Sensing 33:25-38 doi:https://doi.org/10.1007/BF02989988

RE H (1932) Drainage-basin characteristics Eos, transactions american geophysical union 13:350-361 doi:https://doi.org/10.1029/TR013i001p00350

RE H (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology GSA Bulletin 56:275-370 doi:10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2 %J GSA Bulletin

Robnik-Šikonja M, Kononenko I An adaptation of Relief for attribute estimation in regression. In: Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97), 1997. pp 296-304

Samal DR, Gedam SS, Nagarajan R (2015) GIS based drainage morphometry and its influence on hydrology in parts of Western Ghats region, Maharashtra, India Geocarto International 30:755-778 doi:https://doi.org/10.1080/10106049.2014.978903

Schumm SAJGSoAB (1956) Evolution of Drainage Systems and Slopes in Badlands at Perth Amboy, New Jersey Geological Society of America Bulletin 67:597 doi:10.1130/0016-7606(1956)67[597:Eodsas]2.0.Co;2

Singh S, Kumar A, Kumar NJPMS (2014) Motor current signature analysis for bearing fault detection in mechanical systems Procedia Materials Science 6:171-177

Sonebi M, Cevik A, Grünewald S, Walraven JJC, materials B (2016) Modelling the fresh properties of self-compacting concrete using support vector machine approach Construction Building materials 106:55-64 doi:https://doi.org/10.1016/j.conbuildmat.2015.12.035

Strahler A (1964) Quantitative geomorphology of drainage basins and channel networks In: Chow Ven Te (Ed) Handbook of applied hydro McGraw Hill Book Company New York

Strahler AN (1997) Quantitative geomorphology. In: Geomorphology. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 898-912. doi:10.1007/3-540-31060-6_304

Thakkar AK, Dhiman S (2007) Morphometric analysis and prioritization of miniwatersheds in Mohr watershed, Gujarat using remote sensing and GIS techniques Journal of the Indian society of remote sensing 35:313-321 doi:https://doi.org/10.1007/BF02990787

Upadhyay R, Padhy P, Kankar PJJoBS (2016) Application of S-transform for automated detection of vigilance level using EEG signals Journal of Biological Systems 24:1-27 doi:https://doi.org/10.1142/S0218339016500017

Vakharia V, Gupta V, Kankar PJIJAV (2015) Ball bearing fault diagnosis using supervised and unsupervised machine learning methods Int J Acoust Vib 20:244-250

Vakharia V, Gupta V, Kankar PJJotBSoMS, Engineering (2017) Efficient fault diagnosis of ball bearing using ReliefF and Random Forest classifier Journal of the Brazilian Society of Mechanical Sciences Engineering 39:2969-2982 doi:https://doi.org/10.1007/s40430-017-0717-9

Vakharia V, Gupta V, Kankar PJSC (2016) A comparison of feature ranking techniques for fault diagnosis of ball bearing Soft Computing 20:1601-1619 doi:https://doi.org/10.1007/s00500-015-1608-6

Vakharia V, Pandya S, Patel PJLCR, Engineering S (2018) Tool wear rate prediction using discrete wavelet transform and K-Star algorithm Life Cycle Reliability Safety Engineering 7:115-125 doi:https://doi.org/10.1007/s41872-018-0057-5

Vapnik V (2013) The nature of statistical learning theory. Springer science & business media,

Vulević T, Dragović N (2017) Multi-criteria decision analysis for sub-watersheds ranking via the PROMETHEE method International Soil and Water Conservation Research 5:50-55 doi:https://doi.org/10.1016/j.iswcr.2017.01.003

Whipple KX, Tucker GEJJoGRSE (1999) Dynamics of the stream-power river incision model: Implications for height limits of mountain ranges, landscape response timescales, and research needs Journal of Geophysical Research: Solid Earth 104:17661-17674

Figures

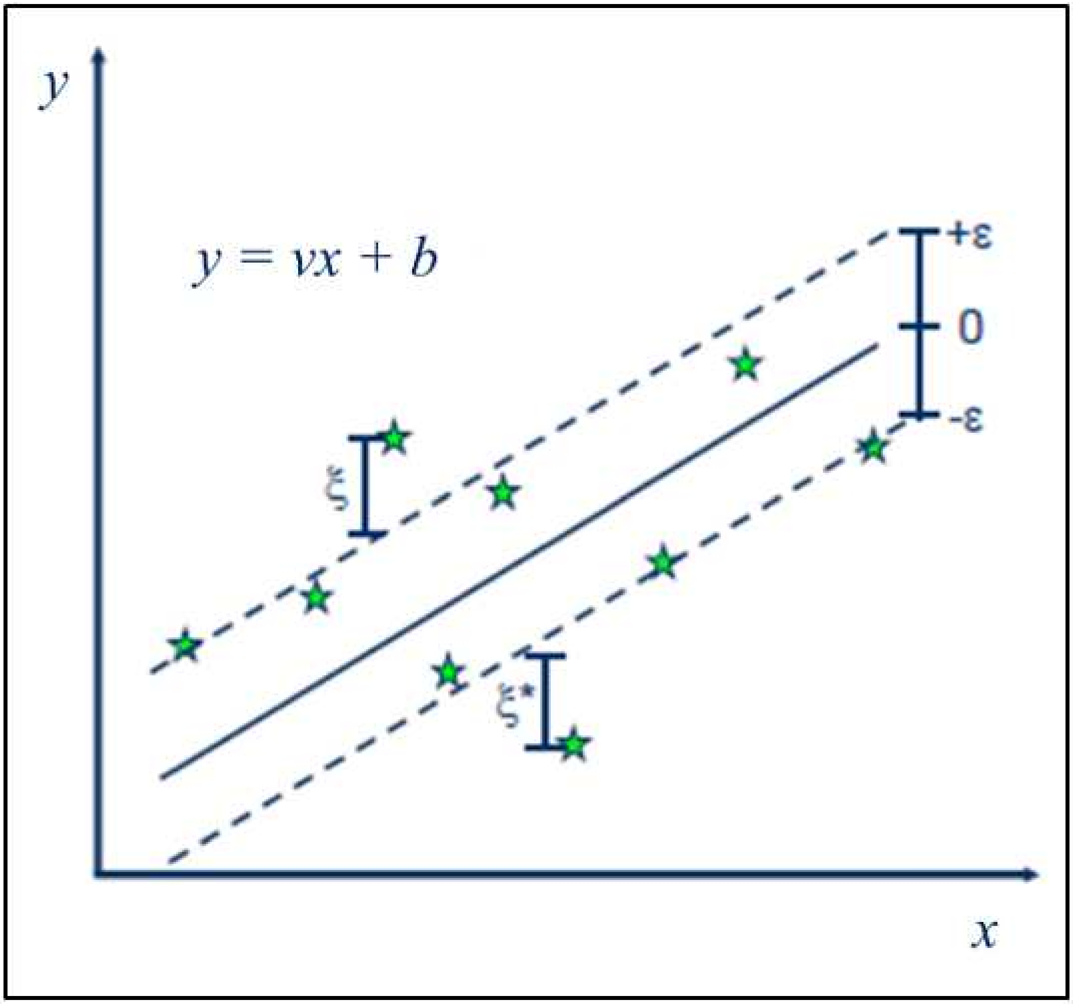


Figure 1

Support Vector Machine Regression

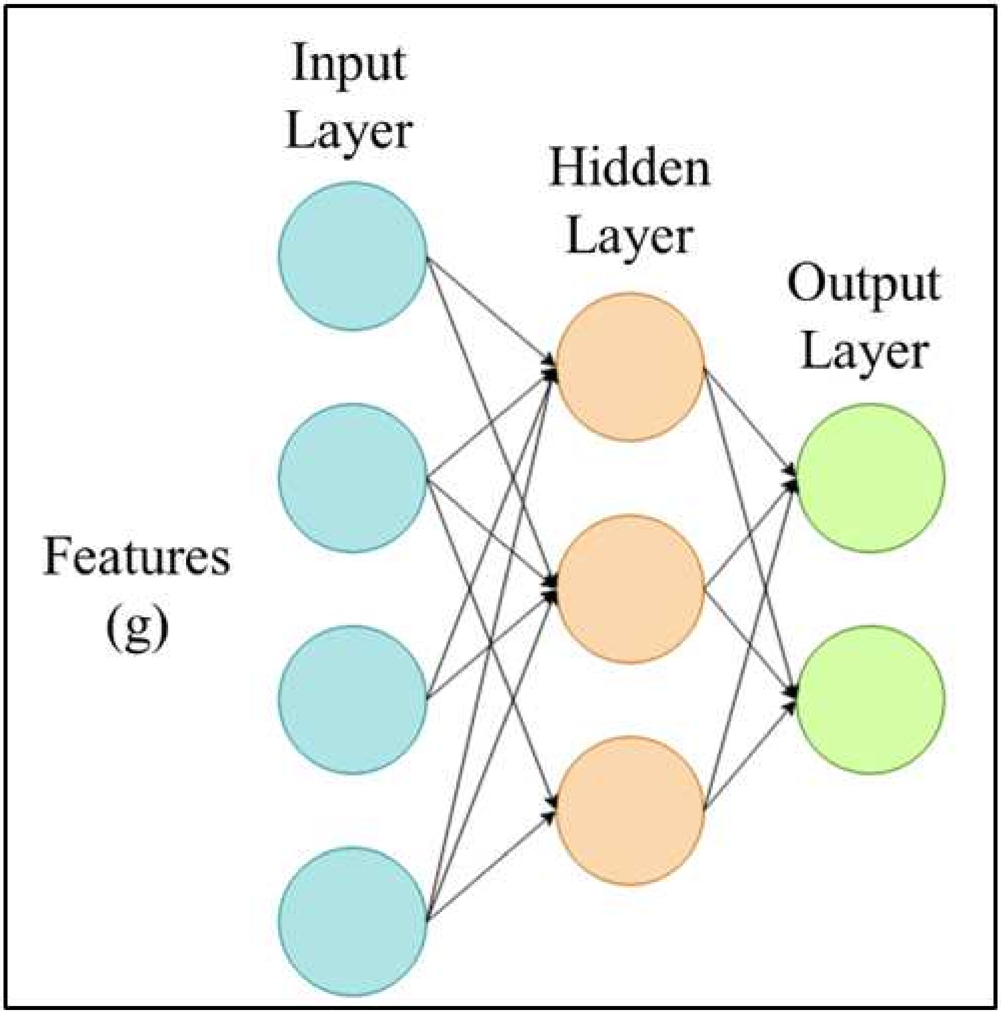


Figure 2

One hidden layer MLP

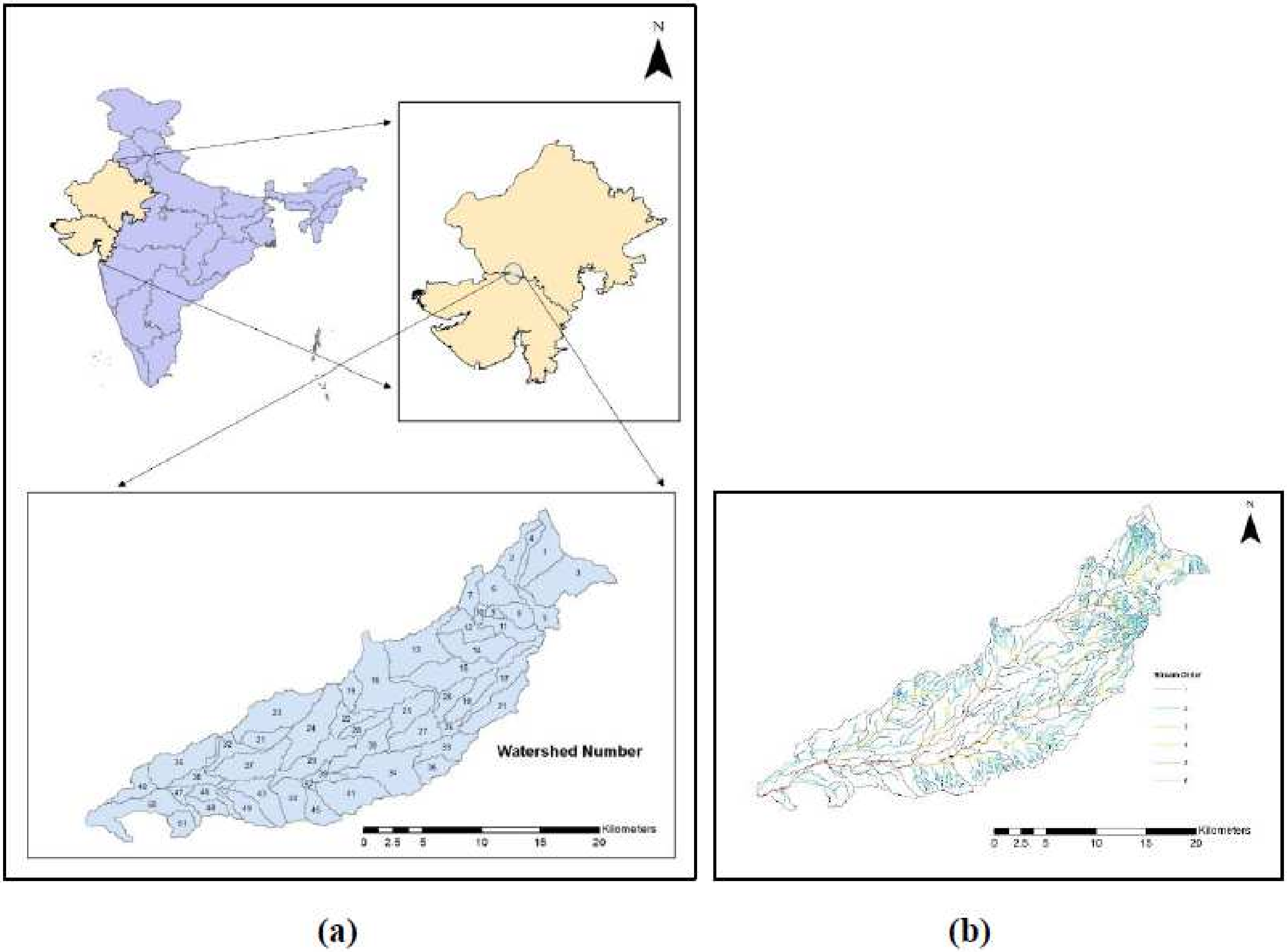


Figure 3

(a) Study area map of Rel River basin (b) drainage network map of Rel River Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

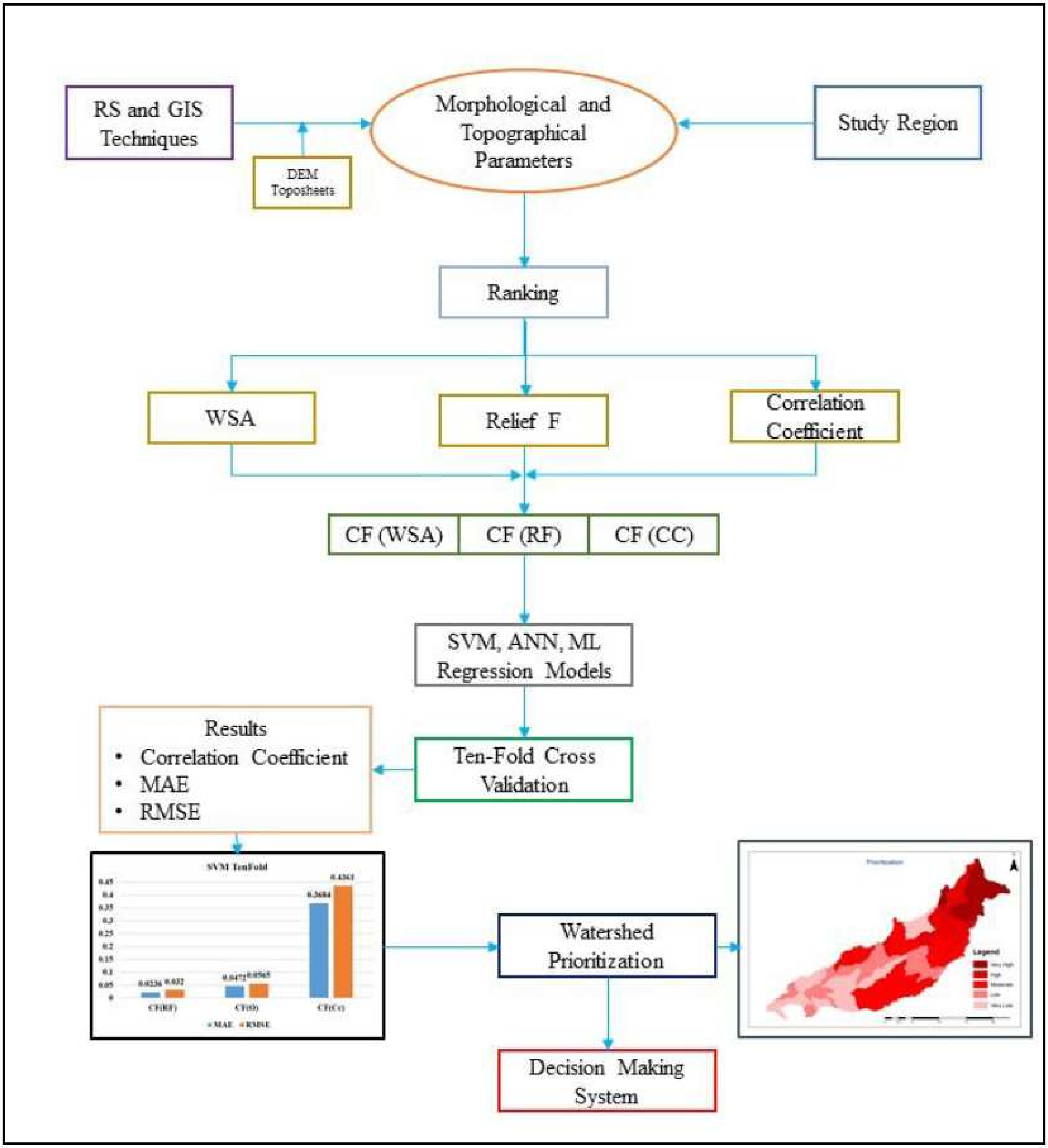


Figure 4

Methodology of the study

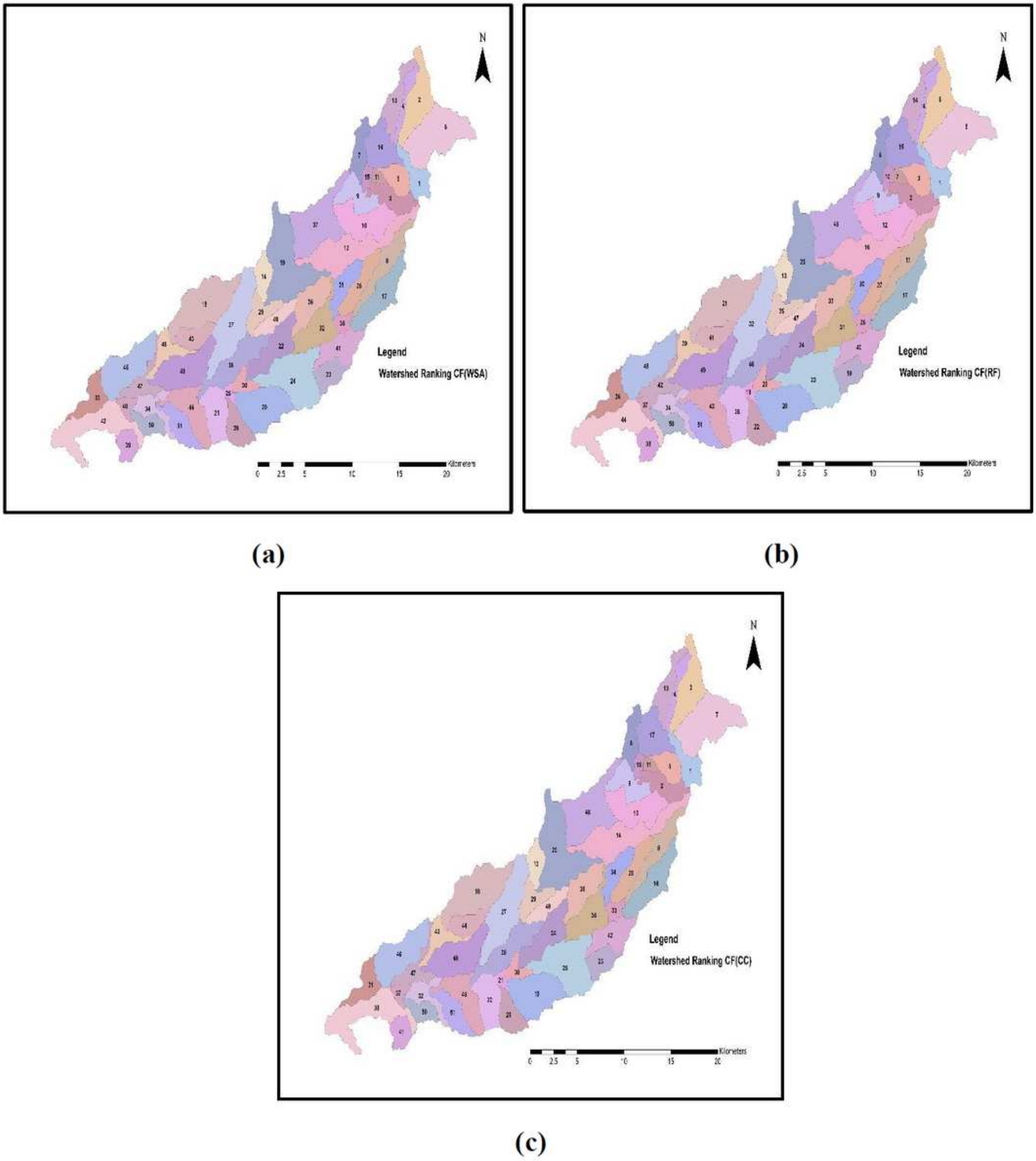


Figure 5

Watershed Prioritized maps using a) CF (WSA) b) CF (RF) and c) CF (CC) Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

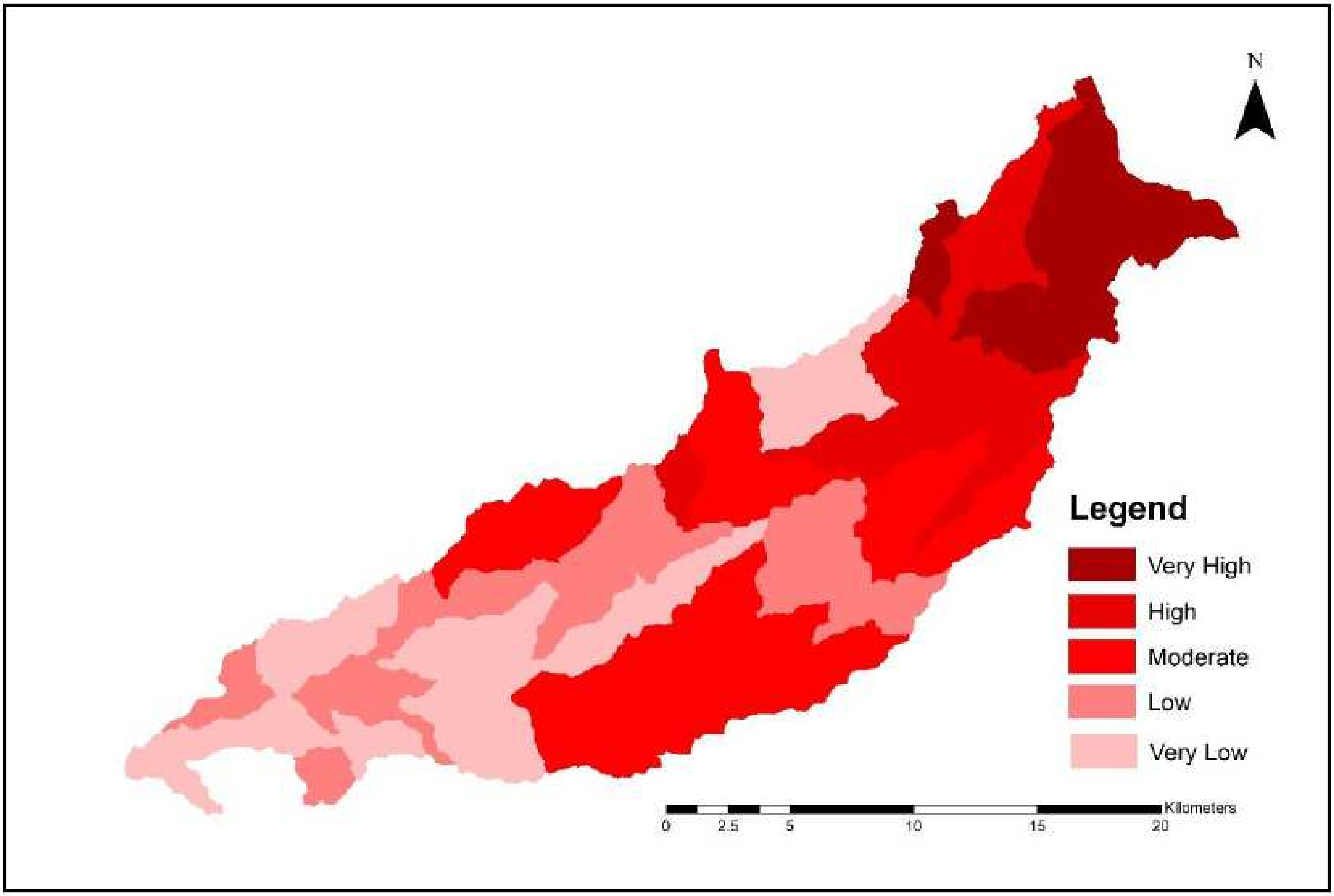
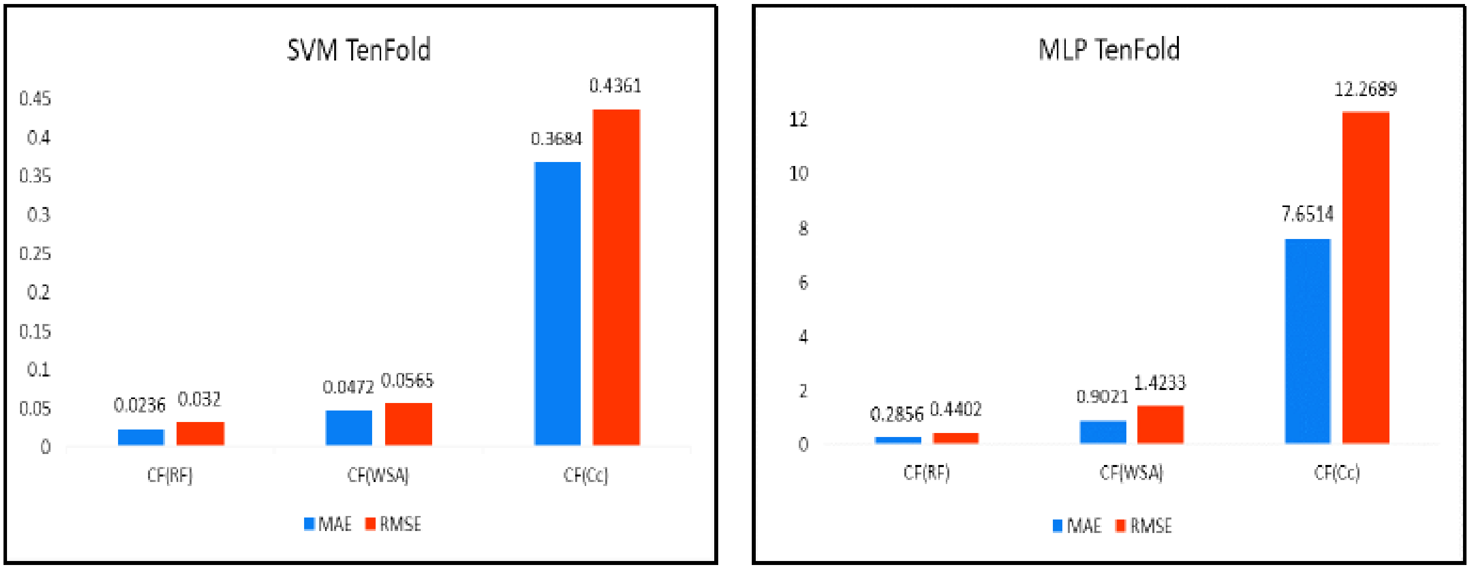


Figure 6

Categorised Soil erosion vulnerability map Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



(a)

(b)

Figure 7

MAE and RMSE values for CF(RF), CF(WSA) and CF (CC) using SVM and MLP algorithm

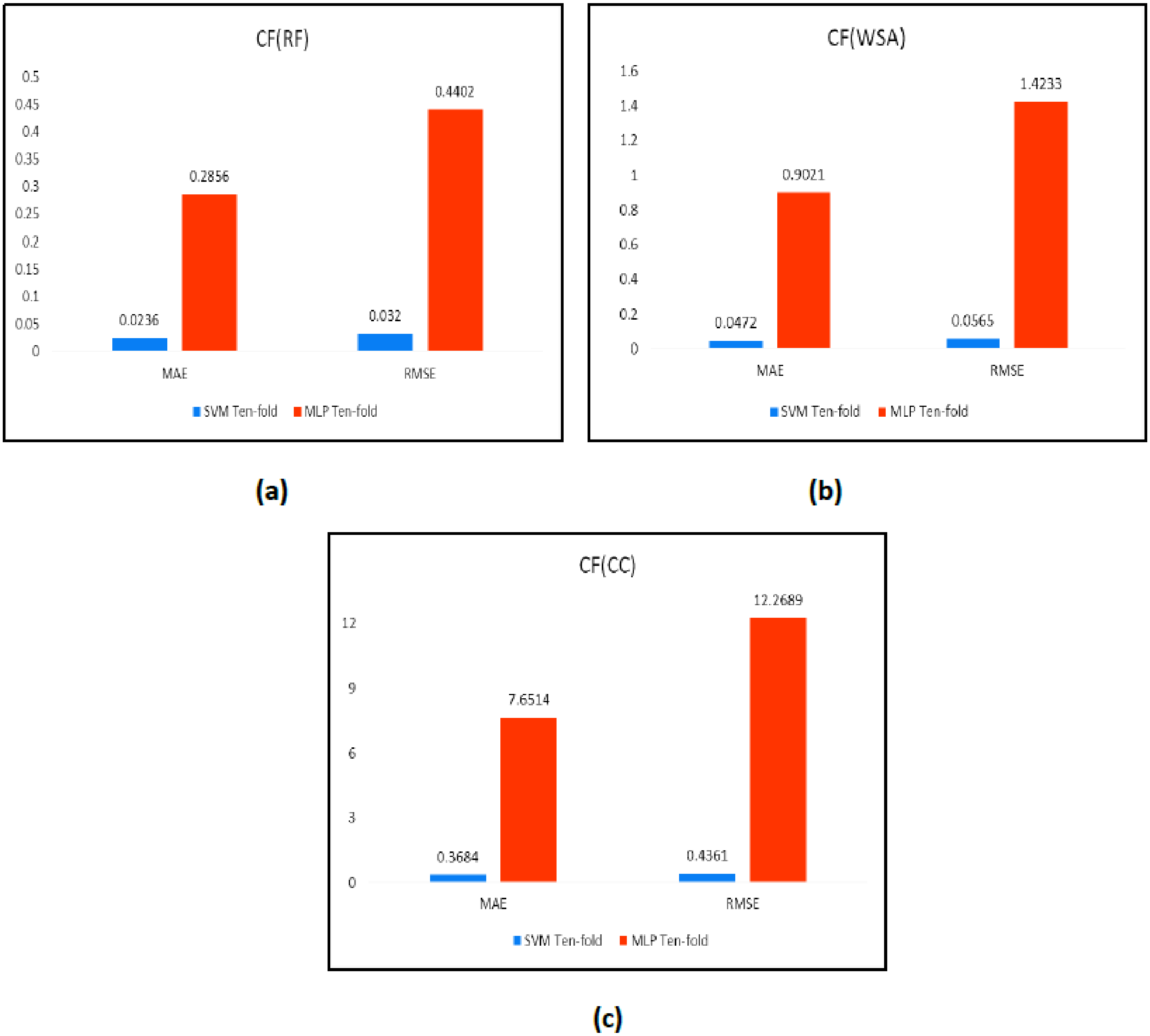


Figure 8

a) MAE and RMSE variation for CF(RF) b) MAE and RMSE variation for CF(WSA) c) MAE and RMSE variation for CF(CC) using SVM and MLP Ten-fold algorithm