OBSERVATION:

Water quality indexing

Water quality depends on several physiochemical parameters and complex to understand for a common person. It is difficult to judge the water quality based on the variation of multiple water quality parameters. Water quality index (WQI) is an important and unique rating to depict the overall quality of water in a single term, and also it is easy to understand for everyone to select the right quality of water or to select appropriate treatment technique to improve the quality of water Tyagi et al. Water quality measurements are often processed using various mathematical models to find the score for water quality status. Several models for water quality indexing have been explored using different methods such as weighted average method, weighted geometric mean method, minimum operator and hybrid methods. Most of these methods process different water quality parameters using basic mathematical operations with standard weights and calculated sub-index.

Few of the algorithmic approaches such as partial least mean square regression-based Simental. and fuzzy based Abietol. have been tried to make water quality prediction process more intelligent and less time-consuming. As it is difficult to correlate single water quality parameter with water quality index, researchers have used several water quality parameters to predict water quality index such as pH, TDS, TSS, DO, BOD, COD and many more. Selection of water quality parameters to predict water quality index is a very important step and should be selected on the location and application basis. Selection of a large number of input parameters makes WQI prediction model complex and time-consuming. Five different water quality parameters such as pH, TDS, salinity, conductivity and ORP have been selected according to Rajasthan water quality scenario and further used for water quality indexing. Target water quality index values have been calculated using weighted arithmetic index method Sengupta and Dalwani for collected water quality samples data from various villages of Rajasthan. Standard and ideal values have been taken from Bureau of Indian Standards (BIS) for drinking water. Various statistical methods have been implemented to predict water quality index and further compared based on their accuracy and complexity.

Further, artificial neural network-based learning scheme with three different training techniques also have been explored to predict water quality index and results have been compared with results obtained with statistical modeling methods.

Statistical modeling

Various statistical methods have been implemented to predict water quality index, and results have been compared based on their prediction accuracy. Principle component regression (PCR) Jolliffe partial least square regression (PLSR) Sim et al. and multiple linear regression (MLR) Grégoire have been implemented on acquired water quality parameters data to predict water quality index. Selection of latent input variables to generate reliable prediction models is a crucial step. As a smaller number of input latent variables can loss the information, while a large number of input variables can lead toward complexity. Variance analysis has been performed to select the optimal number of PLS components. Overall five input latent variables have been selected for the water quality index prediction model based on the variance analysis.

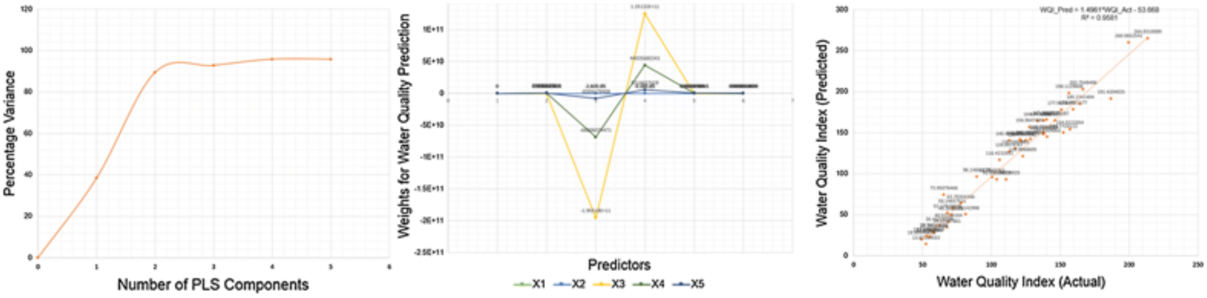

Fig. 13

Water quality indexing prediction using partial least mean square regression (PLSR)

MLR, PCR and PLSR statistical modeling techniques have been used for water quality index prediction model, and it has been observed that PLSR method predicts water quality index less accurate ($R^2 = 95.81\%$) compared to other two methods. PCR and MLR predict better results ($R^2 = 99.99\%$) compared to PLSR.

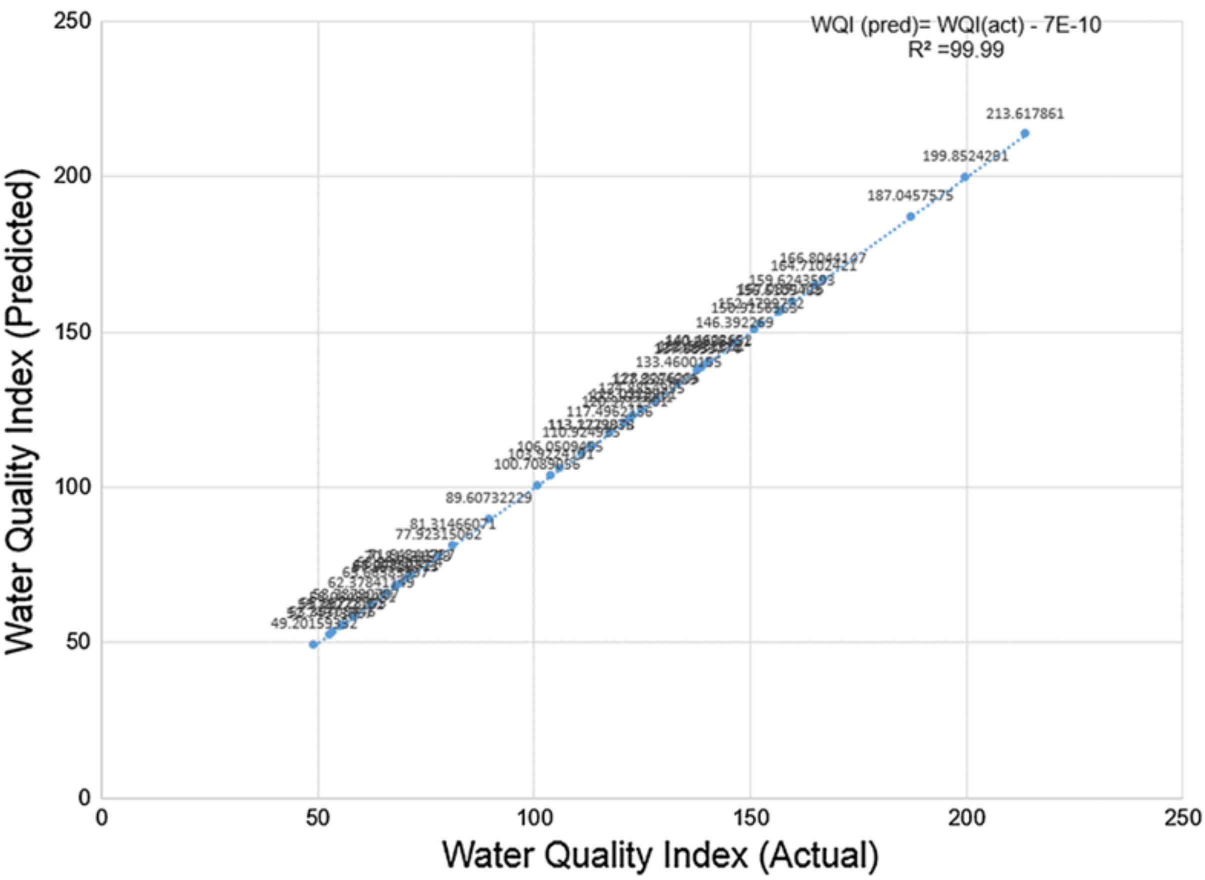

Fig. 14

Predicted water quality index using multiple linear regression (MLR) method

The reason behind the higher accuracy could be the method used for target water quality index prediction. As weighted arithmetic index uses the linear combination of water quality parameters and further calculates water quality index.

Artificial neural network (ANN)-based modeling

ANN-based modeling technique also has been explored to predict water quality index. Multi-input and single output-based neural network architecture has been used to predict water quality index. Five water quality parameters have been used as inputs in the input layer and single output water quality index in the output layer. The optimal number of neurons in the hidden layer has been achieved using trial and error method (4 trials (neuron count = 2, 4, 6 and 8) for each training algorithm). Three different training techniques such as Levenberg Marquardt, Bayesian regulation and conjugate gradient have been explored and implemented in MATLAB 2015a to achieve best results. Further, results have been compared on the basis of mean square error, number of neurons and accuracy. Hidden layer neurons have been varied from 2 to 8, and obtained results have been compared in the context of complexity, computation time and accuracy. Total water samples dataset (50 samples) have been divided into three different datasets such as training, testing and validation. Training dataset consists of 70% of the dataset, while testing and validation consist of 15% each. Out of all three training techniques, conjugate gradient performs worst and predict results with highest mean square error (749–0.5439 for 2–8 neurons in hidden layer) and least accuracy (R^2= 86.38–99.98 for 2–8 neurons in hidden layer), while Bayesian regulation performed best with least mean square error (0.005×10^{-5}) and with highest accuracy (R^2= 100%) at very less number of neurons in the hidden layer (4 neurons). Performance of all three training techniques has been compared based on the variation of mean square error with the variation of the number of neurons in the hidden layer (Table 1).

Table 1

Performance comparison of three different training techniques used of water quality index prediction

Training algorithm	Number of neurons	Testing R^2	MSE (testing)
Bayesian regulation	2	99.98	0.03×10^{-5}
Bayesian regulation	4	99.99	0.005×10^{-5}
Bayesian regulation	6	99.99	6.138×10^{-5}
Bayesian regulation	8	99.99	0.005×10^{-5}
Scaled conjugate gradient	2	86.35	749
Scaled conjugate gradient	4	94.44	673.4
Scaled conjugate gradient	6	96.9	99
Scaled conjugate gradient	8	99.98	0.543
Levenberg–Marquardt	2	99.98	3166.3×10^{-5}
Levenberg–Marquardt	4	99.99	2.1826×10^{-5}
Levenberg–Marquardt	6	99.99	0.663×10^{-5}

Training algorithm	Number of neurons	Testing R^2	MSE (testing)
Levenberg–Marquardt	8	99.99	0.0001212×10^{-5}

MLR and PCR also have shown the similar level of accuracy as ANN-based approach has shown. MLR and PCR are just linear models and take less computation time (less additions and multiplications) compared to neural network approach. MLR-based prediction model for water quality index prediction has used to train smartphone-based water quality monitoring system. Further, water quality index has been predicted for water samples collected from nearby villages of Rajasthan

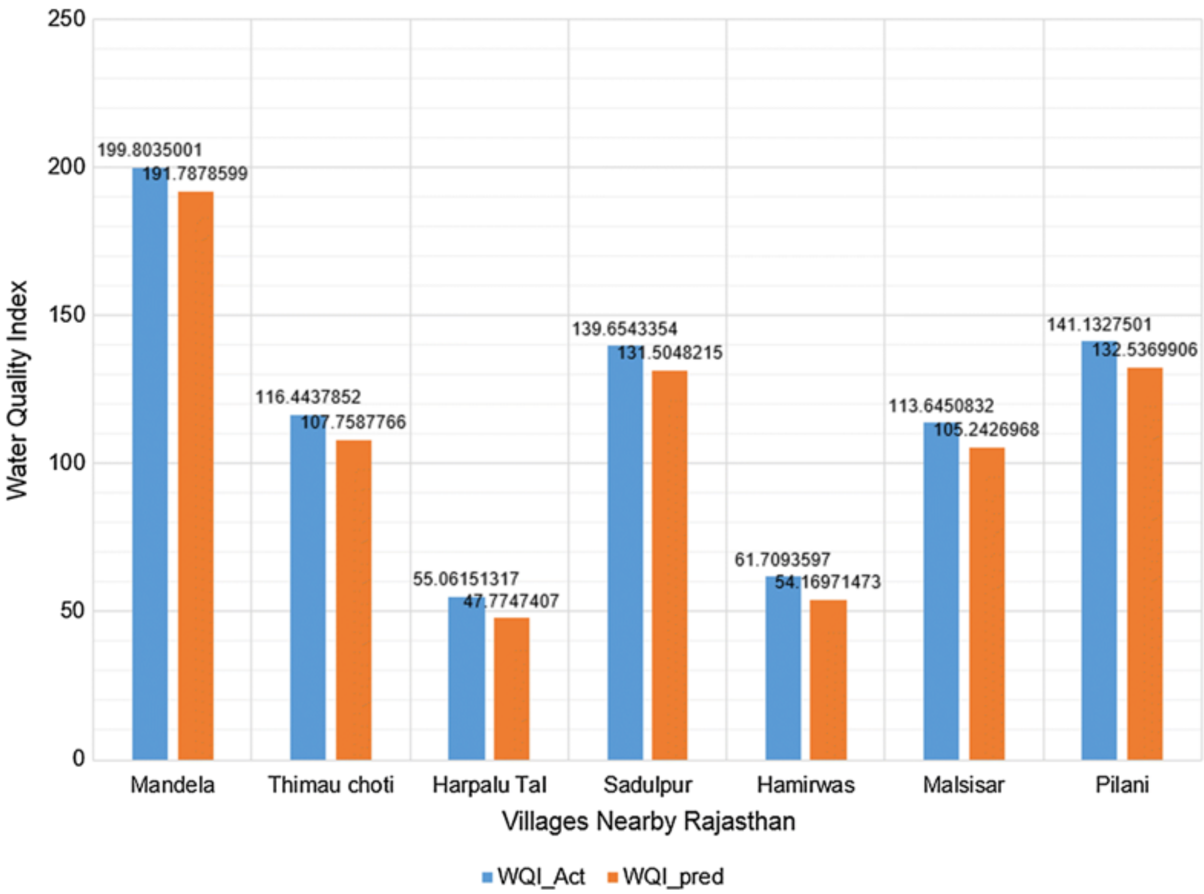

Fig. 15

Water quality indexing of nearby villages of Rajasthan

Conclusion

This paper reported a handy smartphone-based battery-operated water quality measurement system to measure water quality for different applications. A smartphone-based application interface with various advanced features such as location-based data collection, storage on the cloud, integration with Google map for quick judgment, sensor calibration and many more has been introduced. System has been tested for various water quality measurement applications. Different statistical and ANN-based algorithms have been explored to predict water quality index. Presented work consists of various innovations such as location-based water quality measurement and storage, water quality indexing prediction, hassle-free operation, on-site calibration and many more. Presented work has been compared with past literature and it has been found that system could be a low-cost tool for water quality measurement at various rural and urban places. The overall system is cost-effective, portable and easy to use for villagers or any unskilled person.