Variable importance analysis can increase the accuracies of the models

The variable importance measure must be weighted sums of the absolute regression coefficients.

The proposed prediction system will iteratively test the model with training and testing datasets

Cross-validation can used to evaluate method for reducing scales of overfitting and increasing accuracy of the model

Use a minimal number of parameters with cheap sensors to predict water quality

Accurate model can be selected based on the outcome in the model evaluation

Stratified sampling strategy is used to mitigate the uneven distribution of training and testing dataset

The data distribution in the testing data should not affect the training data set.

Data modeling to use the past dataset to inform the future effort

Each data needs to be in different measures to analyze the quality

The size of training datasets should not be less than the number of training parameters required in the model.

Feature selection helps to simplify the procedure and reduce computational cost of analysis

Some of the variables can be eliminated due to the meaningless analysis

Parameters like temperature, turbidity, pH and dissolved solids can be used

Network structure selection method is proposed to identify the corelated input parameters

Evaluating the effect of substantial nutrient loads on overall water quality

Keep the data design

The timeline of the measurements must be recorded

A method like neuro-fuzzy interference system can be implemented which is capable of integrating linear and non-linear relationships in dataset.

The data mining techniques will be used for applying the classification method for water quality application

Massive dataset and strong correlation between parameters will make the best prediction.

Prediction can also be taken from the historical dataset

Using supervised learning algorithm, water quality class can be predicted

Various techniques can be included to predict the quality within the application.