TROPHIC STATE CLASSIFICATION OF KERALA LAKES USING AUTO-TUNED HYBRID AI MODELS

A Project Report

Submitted by

Mr. EBIN JOSEPH

REG NO : TKM20MEAI06

SEMESTER : IV

In partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
IN
Mechanical Engineering (Artificial Intelligence)

Under the guidance of

Dr. Adarsh S

Thangal Kunju Musaliar College of Engineering
Kerala

JULY 2022

DECLARATION

I undersigned hereby declare that the project report "Trophic State Classification of Kerala Lakes Using Auto-Tuned Hybrid AI Models", submitted for partial fulfillment of the requirements for the award of degree of Master of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Dr. Adarsh S. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other university.

Place: Kollam
Date: EBIN JOSEPH

Thangal Kunju Musaliar College of Engineering
Centre for Artificial Intelligence

C E R T I F I C A T E

This is to certify that, this report titled ***TROPHIC STATE CLASSIFICATION OF KERALA LAKES USING AUTO-TUNED HYBRID AI MODELS*** is a bonafide record of the **Project** completed by **EBIN JOSEPH (TKM20MEAI06)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M.Tech in Mechanical Engineering (Artificial Intelligence)** in **APJ Abdul Kalam Technological university** .

Internal Supervisor	Project coordinator	Head of the Department
Dr. Adarsh S Professor Dept of Civil Engg TKMCE	Prof. Sumod Sundar Assistant Professor Centre for Artificial Intelligence TKMCE	Dr. Imthias Ahamad T.P. Professor & HOD Centre for Artificial Intelligence TKMCE

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my project a successful one. I thank him for steering me to the shore of fulfillment under his protective wings

I express my sincere gratitude to **Dr. T A Shahul Hameed**, Principal of TKMCE, and **Dr. Imthias Ahamad T.P.**, Professor and Head of the Department, Centre for Artificial Intelligence, TKMCE, for their constant support and encouragement throughout the project work.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my guide **Dr. Adarsh S**, Professor, Department of Civil Engineering, TKMCE, for his expert guidance, cooperation and immense encouragement. I would like to express my heartfelt thanks to our project coordinator **Prof. Sumod Sundar** Assistant Professor Centre for Artificial Intelligence, TKMCE,for his constant support and encouragement throughout the project work. I also extend my thanks to the entire faculty and staff of the Centre for Artificial Intelligence, TKMCE, who has encouraged me throughout this work.

I also express my thanks to my loving parents, husband, brother, sister and friends, for their support and encouragement in the successful completion of this project work.

EBIN JOSEPH

Abstract

The trophic state is one of the significant environmental impacts that must be monitored and controlled in any aquatic environment. This phenomenon due to nutrient imbalance in water strengthened with global warming, inhibits the natural system to progress. With eutrophication, the mass of algae in the water surface increases and results to lower dissolved oxygen in the water that is essential for fishes. Numerous limnological and physical features affect the trophic state and thus require extensive analysis to asses it. The univariate approach for trophic state classification is based on specified ranges of the cause (Nitrogen, Phosphorus) or response (Chlorophyll-a [Chl-a] and Secchi Depth [SD]) variables or on the variable information expressed in the form of indices. Therefore, in this project, an attempt is made to study about the accurate prediction of TSI using effective combination of different features using Machine Learning(ML), Deep Learning(DL) and hybrid techniques. The framework was applied to a dataset of 11 lakes, 4 reservoirs and 2 ponds in Kerala for the period, 2012-2018. In this study, four different Artificial Intelligence(AI) models were developed for the prediction of Multivariate Trophic State Index(MTSI) and propose the usage of random forest as an effective model. Auto-tuned hybrid models are also proposed for effective Trophic state classification and they show better accuracy than their corresponding stand-alone models.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General

Eutrophication refers to the nutrient enrichment of water bodies particularly with nitrogen and phosphorus compounds, and is considered one of the gravest ecological problems related to lentic water bodies such as ponds, reservoirs, lakes. United Nations Environmental Protection (UNEP) reported that globally 30-40% of lakes and reservoirs show tendency toward varying degrees of eutrophication. It severely deteriorates water quality leading to increased turbidity, cyanobacterial blooms, loss of biodiversity, health hazards, diminishing aquatic growth caused by depletion of oxygen, and foul taste and odour. This, in turn, creates socioeconomic challenges, such as increased water treatment costs, difficulties in fulfilling the criteria for disinfection by-products, and aesthetic damage. Eutrophication management is, hence, the primary step towards conservation of lentic water bodies. Several measures have been conceived to manage eutrophication including re-routing of excess nutrients, alteration of nutrient ratios, and use of herbicides and algaecides. Advanced wastewater treatment (AWT) and diversions are two commonly used techniques for reduction of external nutrients. The external reduction of nutrients often shows little signs of recovery in lentic water bodies; reason is due to internal cycling of phosphorous which can be controlled by intervening with biogeochemical cycles to reduce the phosphorus release from the sediments. However, selection of appropriate treatment method, eutrophication management, and decision-making programme depend largely on the degree of severity of eutrophication. Hence, the fundamental steps for a full conceptual understanding of eutrophication is contained in classification of water bodies into different trophic states combined with quantification of these states. As per Carlson Trophic State Index (TSI), the freshwater bodies have been classified into three possible trophic states: oligotrophic, mesotrophic and eutrophic. Oligotrophic lakes are named so as they host little or no vegetation whereas eutrophic lakes are heavily encroached by scrounging aquatic weeds. To this end, aquatic ecologists and environmental engineers, worldwide, have attempted to accurately determine the level of eutrophication in freshwater bodies.

Both univariate and multivariate statistical approaches have been applied in the past, for classification of trophic states. The univariate approach for trophic state classification is based on specified ranges of the cause (Nitrogen, Phosphorus) or response (Chlorophyll-a [Chl-a] and Secchi Depth [SD]) variables or on the variable information expressed in the

form of indices. Eutrophication is instigated by an interplay of multiple factors (variables). Therefore, individual variables or univariate indices derived from a single variable may not adequately quantify the trophic levels of eutrophication, and a multivariate approach is more appropriate for trophic state assessment. Several multivariate approaches that yield composite indices linking cause and response variables have also been proposed for trophic state assessment. These composite indices are obtained by consolidation of extensive water quality information into a single index, and they enable thorough and continuous assessment of water quality and eutrophication. After determining the TSI by following univariate or multivariate approach, the thresholds are to be estimated to classify the lakes into different trophic states. As the geographic and atmospheric factors significantly influence the eutrophication process, the threshold standards are to be fixed region-specific. Moreover, TSIs are developed on the basis of certain assumptions (limiting nutrient concept), because of which most are site specific, with limited universal applicability. Literature shows that although several multivariate indices have been developed, they vary in the selection of variables and the adopted methodologies. The overlapping of threshold values for each trophic state leads to uncertainty in the classification of lentic water bodies which needs to be addressed. Hence, eutrophication assessment requires a generic framework with discrete threshold values that can be applied to any hydro- climatic region.

Recently, Multivariate Trophic State Index (MTSI) framework was proposed by Kurwattia and Karmakar[2] in 2015, as a robust composite index, to evaluate the trophic state of lentic water bodies and subsequently fix the thresholds for trophic conditions. This multivariate TSI uses multivariate statistical analysis techniques such as Principal Component Analysis (PCA) and Cluster Analysis (CA), incorporating parametric and non-parametric density functions. MTSI is calculated from independent water quality variables, such as: chlorophyll-a (Chl-a) concentration, total phosphorus (TP) concentration and total nitrogen (TN). Measuring the concentration of chlorophyll-a and total phosphorus requires sophisticated techniques that involve chemical experiments, equations, specific analyses of water samples. Consequently, alternative variables or methods for accurately measuring MTSI would represent a significant development in the field. Instead of calculating the index of water quality directly by its original formulas, many studies have predicted many water quality indices based on water parameters other than original parameters of the original formulas. AI based approaches, which have been typically used for this purpose in recent studies, have advantages over the traditional deterministic methods as they reduce the complexity that is associated with a large number of factors and the necessary sophistication of quantifying traditional water parameters. Various AI based approaches have been confirmed to be useful in solving water quality management problems.

Traditional Trophic State Index like Carlson TSI, based on univariate approach was widely used for classification of lakes. Multivariate Trophic State Index based on multivariate approach is a robust alternative for lake classification index, not widely applied. The prediction of TSI using Soft Computing is rare and researchers have hardly attempted the prediction of MTSI using Soft Computing. In this study, different AI models were compared for the prediction of MTSI and proposes the usage of random forest as an effective model. And also hybrid models with effective feature selection are proposed as effective method for Trophic state classification.

1.2 Objectives

The main objectives of the work are listed below:

- To efficiently predict MTSI values using different AI models.
- To find the best combination of input-features using effective feature selection methods.
- To classify the Kerala lakes into different trophic states using hybrid, DL and ensemble models.

1.3 Organisation of Report

The overview of this report is as follows. Chapter 1 deals with the introduction to the topic and mentions about the relevance of MTSI. Chapter 2 discusses about the Literature Survey associated with the project. Chapter 3 gives the methodologies followed in the work. Chapter 4 is based on the Study area and datasets used. Chapter 5 deals with the results obtained in the project. Chapter 6 concludes the discussion on the topic.

Chapter 2

Literature Review

2.1 General

This chapter presents the review of literature describing the various approaches to Trophic State Index (TSI) and its sequential historical development. This chapter also describes the method of prediction of these indices in different hydroclimatic regions using AI models.

2.2 Review of Literature on Trophic State Index

Carlson[1] proposed a tropic state index, which is the most widely used index for classification of lentic water bodies. Carlson Tropic State Index (CTSI) used algal biomass as the basis of this classification. It was a univariate approach. Carlson TSI was attractive because of its theoretical basis and reliance in qualified indicators. Chlorophyll-a pigment, Secchi Depth (SD), Total phosphorus (TP) were the parameters taken in development of index.

Canfield and Hodgson[3] introduced models for the prediction of chlorophyll-a concentrations and Secchi depth using data of Florida lakes. Model yields unbiased estimates of chlorophyll-a concentrations and Secchi depth over a wide range of lake types. They developed a relationship between Secchi depth and nutrient loading and also between chlorophyll and nutrient loading.

Kurwattia and Karmakar[2] proposed Multivariate Tropic State Index (MTSI). It was based on Multivariate Statistical analysis. TP, TN, Chlorophyll-a were selected as the indicator for trophic state classification in this study.
A generic framework was developed for:

- Estimation of MTSI which is based on Principal Component Analysis (PCA)
- Development of threshold of each tropic trophic state using composite parametric and kernel based non-parametric approaches.
- Validation of index was also done using cluster analysis.

Chou et al.[4] developed machine learning models for the prediction of Carlson TSI, using artificial intelligence techniques. Four well known artificial intelligence techniques: Artificial

Neutral Network, Support Vector Machine, Classification and Regression Tree and linear regression were used to predict Carlson TSI. Objective was to develop a versatile water quality modelling approach including single, ensemble and hybrid models to predict the index using surface water temperature,Dissolved Oxygen (DO), Secchi Depth(SD), Chemical Oxygen Demand (COD), and Ammonia(NH3) etc as inputs.

Saghi et al.[5] investigated the prediction of TSI by Artificial Neural Network. TSI process, suggested by Florida Department of Environmental Protection, was analysed using artificial neutral networks. Feed forward neutral network with one input layer, one hidden layer and one output layer was applied using MATLAB neutral network toolbox. Input data are nitrogen cycle parameters (Total Nitrogen(TN), Nitrate(NO3), Ammonium(NH4)), phosphorus cycle parameters (Total Phosphorous (TP), Phosphate(PO4)), and parameters that will be changed by eutrophication: Chl-a, SD, DO and output data is TSI. Predicted output data showed an R^2= 0.8377 with real output data.

Sabino et al.[6] proposed using a hybrid CT-ANN model (classification tree artificial neural network) to evaluate the trophic state based on the chosen important features. Eight initial features were removed by using the classification tree as a multidimensional reduction strategy for feature selection. Chlorophyll-a, phosphorus, and Secchi depth make up the remaining predictors with significant effects. The 20 artificial neurons in a two-layer ANN were built to evaluate the trophic status of input characteristics. The crucial variables of learning time, cross-entropy, and regression coefficient were used to model the neural network. The accuracy of the ANN model, which was utilised to evaluate the trophic state using 11 predictors, was 81.3 %. The hybrid classification tree-ANN model, which used three predictors, produced an accuracy rate of 88.8% and a cross-entropy performance of 0.096495. The modelled hybrid classification tree-ANN offers improved accuracy in determining the trophic condition of the aquaponic system, according to the results.

Hu et al.[7] seek to design an algorithm to estimate the trophic condition of inland waters Water bodies are split into two categories: algae-dominated water (Type I) and turbid water after the turbid water index was used to determine the optical water types on each pixel (Type II). The trophic status index (TSI), introduced by Carlson, was then derived using the algal biomass index (ABI), which was constructed based on water categorization (1977). The results in Type I water (R^2= 0.62, N = 282) and Type II water (R^2 = 0.57, N = 132) demonstrated a high degree of precision. A machine learning technique and a number of band-ratio techniques were defeated by the ABI-derived TSI. Using surface reflectance data from Landsat-8, this model was used to calculate the trophic status index for 146 lakes in eastern China from 2013 to 2020. Compared to the middle sections of the Yangtze River and Huai River basin, the annual mean TSI for lakes in the lower portions of the Yangtze River basin was higher. The retrieval algorithm demonstrated the applicability to other sensors with a total accuracy of 83.27 percent for the moderate-resolution imaging spectroradiometer (MODIS) and 82.92 percent for the Sentinel-3 OLCI sensor, highlighting the potential for high-frequency observation and large-scale simulation capability.

Zhu and Mao[8] used key environmental elements (water temperature and wind field) were taken into account throughout the modelling procedure in order to increase the preci-

sion of remotely sensed estimates of the trophic status index (TSI) of inland urban water bodies. These environmental variables are simple to assess and exhibit a significant link with TSI. Then, using remote sensing and environmental parameters, a backpropagation neural network (BP-NN) was used to create the TSI estimation model. We chose the best set of input variables based on the performance of the BP-NN after the model was trained and validated using the TSI defined by five water trophic indicators acquired for the period between 2018 and 2019. Our findings show that the water temperature and single-band reflection measurements from Sentinel-2 satellite imagery can be combined as input variables to achieve the best performance. Generally speaking, the predicted maps based on our suggested model demonstrate considerable seasonal fluctuations and spatial peculiarities in the water trophic status, showing the potential to carry out cost-effective, RS-based TSI estimation research on complicated urban water bodies abroad.

2.3 Summary

Traditional Trophic State Index like Carlson TSI, based on univariate approach was widely used for classification of lakes. Multivariate Trophic State Index based on multivariate approach is a robust alternative for lake classification index, not widely applied. The prediction of TSI using Soft Computing is rare and researchers have hardly attempted the prediction of MTSI using AI methods.

Chapter 3

Methodology

This chapter describes the methodologies adopted in the study. Theory of lake classification and methodologies for determination of TSI using statistical techniques has been summarized in section 3.1 and 3.2. It also outlines the theoretical details of AI techniques which is used for the prediction of MTSI.

3.1 Theory of Lake Classification

Trophic state is referred to as the entire mass of the biomass in a body of water at a certain moment and place. The biological reaction to nutrient additions to aquatic bodies is called the trophic state. The trophic state of the lake is determined by the amount of dissolved nitrogen, phosphorus, and other biologically valuable nutrients in the water. To gauge its biological state, the water trophic is employed. Three categories of trophic status are used to categorise lakes: oligotrophic, mesotrophic, and eutrophic lakes. When a lake's trophic index is higher, it may be categorised as hypereutrophic.

Oligotrophic - A lake that is oligotrophic has a low level of production as a result of its low nutrient concentration. Due to the restricted growth of algae in the lake, the waters of these lakes are typically fairly clear. These lakes have excellent drinking water quality. Such lakes are home to aquatic animals like lake trout that demand cold, well-oxygenated waters. Due to the chilly lake waters, oligotrophic lakes are typically located in colder parts of the planet where nutrient mixing is uncommon and slow.

Mesotrophic - Mesotrophic lakes are those with a medium level of productivity. These lakes often feature clear water and medium nutrient levels, along with aquatic plants.

Eutrophic - Because of the availability of nutrients in this lake, particularly nitrogen and phosphorus, eutrophic lakes have high levels of biological production. Due to the high quantities of oxygen that a lot of plants growing in the lake give, eutrophic lakes initially speed up the multiplication and expansion of Lake Fauna. However, when things go too far, the lake becomes overpopulated with plants or algal blooms, which causes the lake fauna to suffer from the high levels of respiration caused by the live vegetation. Both natural eutrophication and anthropogenic environmental impact are possible causes.

Hypereutrophic - These lakes experience issues brought on by excessive plant and algae growth as a result of an abundant availability of growth nutrients. As a result of the excessive proliferation of algae or aquatic plants, these lakes exhibit little clarity. Lower than three feet is the lowest point of visibility in these lakes. Additionally, hypereutrophic lakes contain more than 40 micrograms/liter of total chlorophyll and more than 100 micrograms/liter of phosphorus. Dead zones may form beneath the water's surface as a result of the development of algae, which frequently suffocates the aquatic life.

3.2 Determination of Trophic State Index

3.2.1 Carlson TSI

Carlson[1] proposed a tropic state index, which is the most widely used index for classification of lentic water bodies. Carlson Tropic State Index (CTSI) used algal biomass as the basis of this classification. It was a univariate approach. Carlson TSI was attractive because of its theoretical basis and reliance in qualified indicators. Chlorophyll-a pigment, Secchi Depth (SD), Total phosphorus (TP) were the parameters taken in development of index.

Separate indices are derived for each parameter:

$$TSI(SD) = 60–14.41ln(SD) \tag{3.1}$$

$$TSI(Chl) = 9.81ln(Chl) + 30.6 \tag{3.2}$$

$$TSI(TP) = 14.42ln(TP) + 4.15 \tag{3.3}$$

3.2.2 Development of MTSI

The development of Multivariate Trophic State Index (MTSI) and subsequent identification of the threshold values involves the use of multivariate outlier removal and Principal Component Analysis(PCA) to make the approach a generic one. The methodology involves two phases- the estimation of MTSI and estimation of threshold, for which two frameworks are used. The first framework is based on PCA, estimates the MTSI using the available cause and response variables. The second framework is based on amalgam of parametric and non-parametric approaches, is used to develop thresholds for each trophic state.

PCA is widely applied for transformation of original variables into new, uncorrelated variables known as principal components (PC), which are linear combinations of the original variables. PCA hence helps to reduce the dimensionality of the dataset; it is useful when the causal variables are highly correlated and PCA can be advantageously used to the combined dataset of all trophic states once multivariate outliers have been removed. The use of PCA as a weighting technique, which was used to develop the composite index, offers the possibility of adding countless positively correlated causal variables in the development of index. Further, the weights are assigned on the basis of the data rather than by the analyst; hence, the approach is objective.

In the PCA based approach, the linear combination of original variables, that explains inherent variation to the greatest possible extent are used in developing the composite index, based on comparison between the maximum variance explained by the first PC to those explained by the other PCs. If the Eigenvalue exceeds 1 only for the first PC, the coefficients (weights) of variables in the first PC can be used to construct a composite index (MTSI).

The coefficients derived from the first PC are standardised such that their sum equals 1. The MTSI may then be expressed as follows:

$$MTSI = x_{\mathrm{i}}(var_{\mathrm{i}}) + x_{\mathrm{i+1}}, ... + x_{\mathrm{n}}(var_{\mathrm{n}}) \tag{3.4}$$

Eq. (3.4) is used for calculating MTSI for each trophic state dataset. As the proposed methodology uses both parametric and nonparametric approaches for threshold calculation, the MTSI values are initially analysed for the parametric approach. Outliers are detected and removed, as the Gaussian parametric approach is sensitive to outliers.

3.3 Proposed Framework

The proposed system is used to classify the MTSI data into its respective trophic states using different ensemble and Deep Learning models. Figure 3.1 shows the entire architecture of the proposed model. The entire system can be divided into six: Data pre-processing,MTSI prediction, Assessment of prediction models, Trophic state classification, training and testing data hybrid and simple models and finally performance analysis and comparison of these classifiation models.

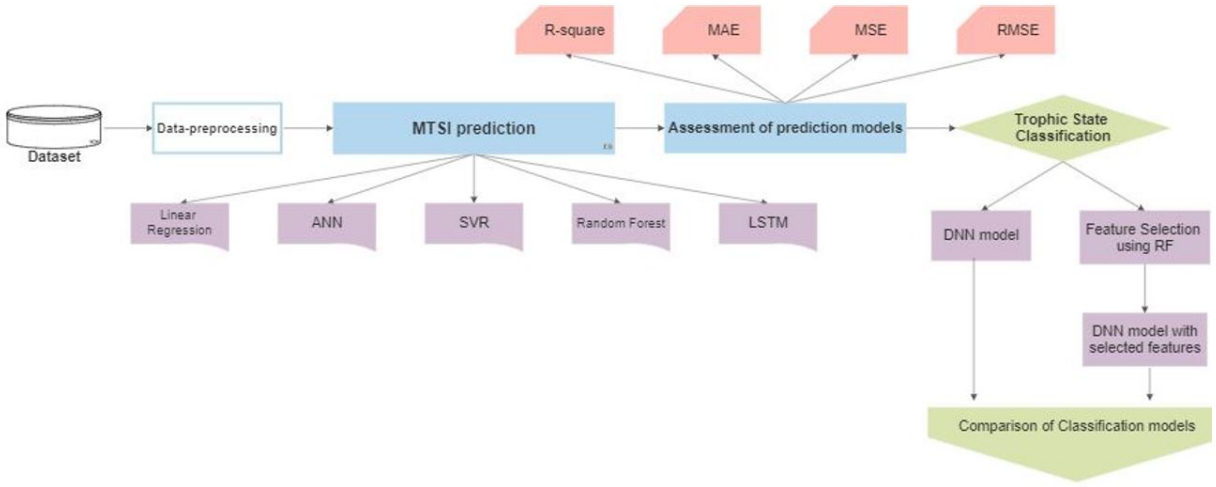


Figure 3.1: Proposed Framework

The effective dataset for the work is compiled by aggregating the individual datasets containing the features and MTSI values of different water-bodies across Kerala. Outliers from data are provided with the highest MTSI values and also the data is processed adequately for classification purpose. Then the data is trained and tested for MTSI prediction using different ML models: Linear regression, ANN, SVR and Random Forest. And the performance indices of these models are assessed and compared for evaluating the better model. Mean Absolute Error, Root Mean Square Error, R-square values are the performance indices used for prediction evaluation. In the second phase of the work, the dataset is classified into different trophic states: oligotrophic, mesotrophic, eutrophic and hyper-eutrophic. Trophic State Classification is done using different hybrid models which include

combination of Random Forest(for feature selection) with different ensemble and DL models. This include 1d-CNN, DNN, J48 tree and XGBoost algorithms with all the parameters autotuned using the keras-tuner function. Also a comparison is done between the classification performances of DNN and the hybrid RF+DNN model, thereby showing the effectiveness of efficient feature selection as shown in the Fig 3.1 .

3.4 Performance measures

The equations below shows the different performance indices for regression models. Model with minimum errors and maximum correlation coefficient is considered as the best model. Here, n=number of samples, y_i= actual MTSI values, f_i = predicted MTSI values and y = mean of actual MTSI values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(y_i - f_i)| \tag{3.5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2 \tag{3.6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2} \tag{3.7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - y)^2} \tag{3.8}$$

Equations below shows the different performance indices for classification models. Model with maximum precision, recall and accuracy is considered as the best model. Here, TN represents the number of True Negative values whose predicted value is negative and its negative , FP represents the number of False Positive values whose predicted value is positive but its negative(Type 1 error), FN represents the number of False Negative values whose predicted value is negative but its positive(Type 2 error) and TP represents the number of True Positive values whose predicted value is positive and its positive.

$$Precision = \frac{TP}{TP + FP} \tag{3.9}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.10}$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.11}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.12}$$

3.5 AI methods for prediction of MTSI

Traditionally, empirical trophic state indices of lentic water bodies often defined based on changes in concentration of effective factors (nutrients) and its consequences (increase in chlorophyll a), have been used as an efficient tool. But these parameters are very difficult to measure. The main aim of this study is the determination of MTSI from easily measurable water quality parameters such as dissolved oxygen (DO), temperature, conductivity, chemical oxygen demand (COD), turbidity, potassium, pH etc. AI techniques (soft computing tools) can be used for developing models for prediction of MTSI. Theoretical background of following AI techniques used in this study are described here:

- Artificial Neural Network (ANN)
- Support Vector Regressor (SVR)
- Linear Regression (LR)
- Random Forest (RF)

3.5.1 Artificial Neural Network (ANN)

Similar to the extensive network of neurons in the brain, an artificial neural network is made up of interconnected groups of nodes. A network of neurons joined by synapses makes up an ANN. ANN acquire weight and threshold through examples. Training determines these weights. The average of all mistakes is the error in the training epoch. Three layers make up its fundamental structure: the input layer, the hidden layer, and the output layer as shown in Fig 3.2. Artificial neurons are interconnected. An artificial neuron is represented by each circular node, and a connection between the output and input of one artificial neuron is represented by an arrow. It is a method of adaptive learning. Here, an impulse is created from the input data.

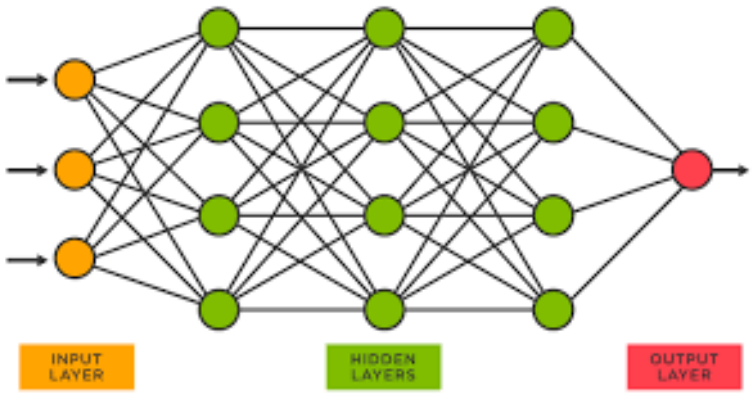


Figure 3.2: ANN Structure

3.5.2 Support Vector Regression (SVR)

Popular supervised machine learning techniques for classification and regression include support vector machines. It is a formal definition of a separating hyperplane discriminative

classifier. Building a hyperplane that may minimise the sum of the distances from the data points to the hyperplane is the basic objective of the support vector machine for regression (SVR). It is applied to build an input-output model that addresses nonlinear regression issues. In other words, the algorithm generates an ideal hyperplane that classifies fresh samples given labelled training data. In SVR, a fixed mapping process is used to first map the input onto n-dimensional feature space. Then, a linear model is built in this feature space as shown in Figure 3.3. During training, SVR kernel functions (linear, radial basis, polynomial, or sigmoid) are used to identify support vectors along the function surface.

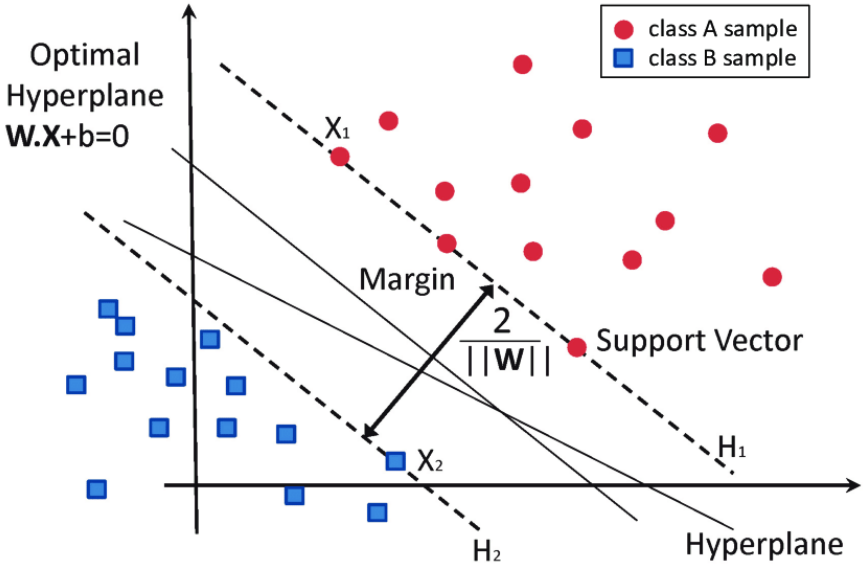


Figure 3.3: SVR Feature Space

3.5.3 Linear Regression (LR)

A machine learning algorithm based on supervised learning is linear regression. It executes a regression operation. Regression uses independent variables to model a goal prediction value. It is mostly used to determine how variables and forecasting relate to one another. Regression models vary according to the amount of independent variables they use, the type of relationship they take into account between the dependent and independent variables, and other factors.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y (output). Hence, the name is Linear Regression. In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model. Hypothesis function for Linear Regression:

$y = \theta 1 + \theta 2 * x$

While training the model we are given:

x: input training data (univariate – one input variable (parameter))

y: labels to data (supervised learning).

When training the model – it fits the best line to predict the value of y for a given value

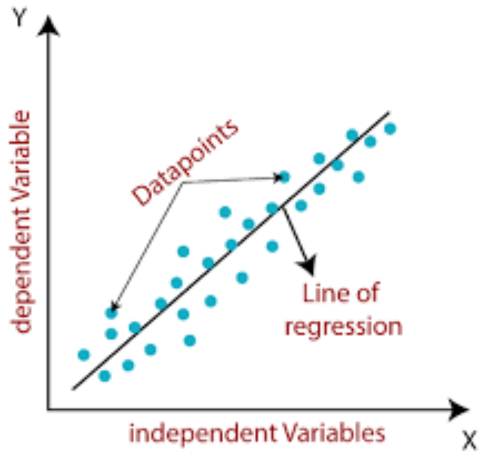


Figure 3.4: Illustration of linear regression

of x as shown in Fig 3.4. The model gets the best regression fit line by finding the best $\theta 1$ *and* $\theta 2$ *values.*

$\theta 1 : intercept$

$\theta 2 : coefficient\ of\ x$

3.5.4 Random Forest (RF)

An ensemble learning method is the RF approach. It has been utilised to solve a number of prediction issues successfully. It is a machine-learning approach that integrates a sizable collection of decision trees to boost the Classification And Regression Trees (CART) method's predictive performance. The end output of RF is the average of all the decision trees, and each tree is created using a randomly chosen bootstrap sample from the original data set. The RF requires a fairly small number of parameters to be defined in comparison to regression approaches. The number of variables employed in each tree-building procedure and the total number of trees constructed in the forest are the only two requirements. The number of trees built in the forest has significant influence on the result of RF. The insufficient number of trees would result in poor forecasting performance, while the excessive number of trees may lead to complicated predictors.

Fig 3.5 shows the ensemble RF learning method for regression is used in the supervised learning model. To produce a more precise forecast, it combines several model tree techniques. On a variety of issues, including those involving non-linear relationships, it typically delivers excellent results. There is no interpretability, overfitting is easily possible, and we must decide how many trees to include in the model, among other drawbacks.

class sklearn.ensemble.RandomForestRegressor(n-estimators=100, criterion='squared-error', max-depth=None, min-samples-split=2, min-samples-leaf=1, min-weight-fraction-leaf=0.0, max-features=1.0, max-leaf-nodes=None, bootstrap=True, oob-score=False, n-jobs=None, random-state=None,verbose=0, warm-start=False, ccp-alpha=0.0, max-samples=None)

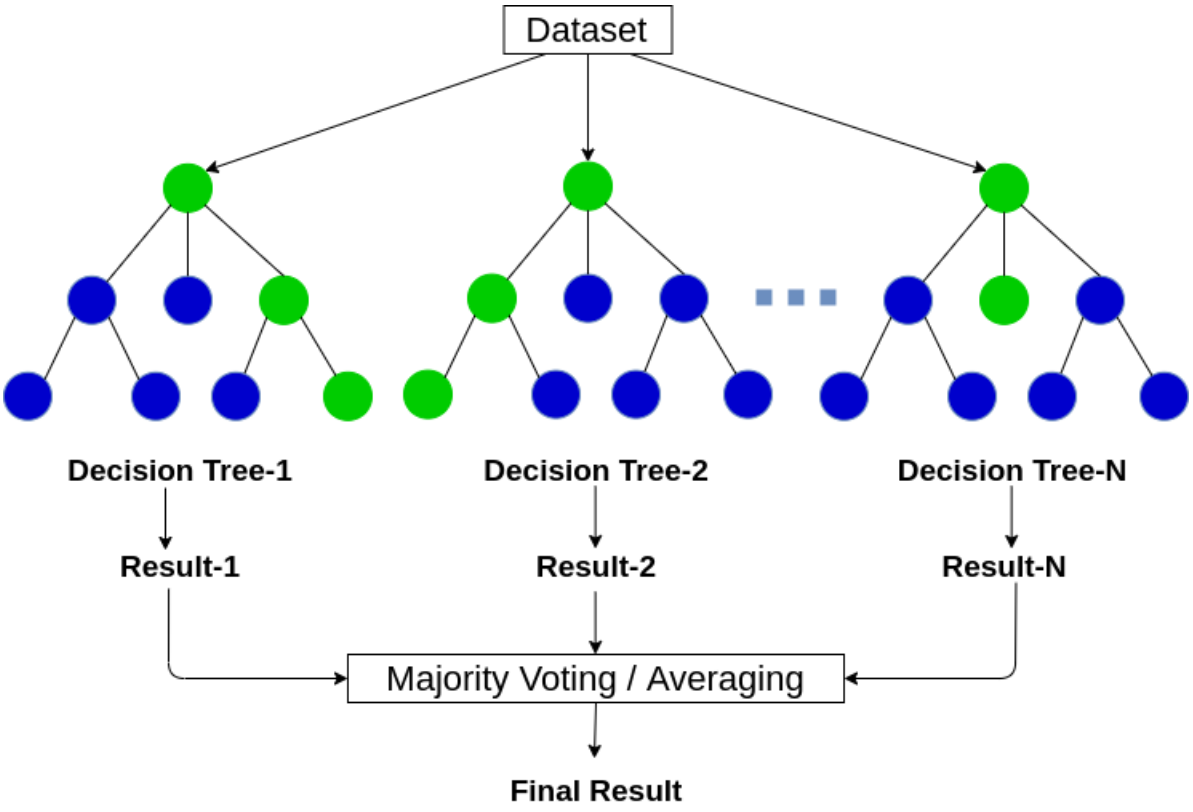


Figure 3.5: Working of Random Forest

3.6 Trophic State Classification Models

Different Hybrid models were developed by coupling Random Forest algorithm with various ensemble and Deep learning models for effective Trophic State Classification. Here Random Forest model was used as an efficient tool for feature selection. In this work, different models used for classification are DNN, 1D CNN, J48 and XGBoost classifiers [16]. The parameters associated with each models are auto-tuned using the keras_tuner function.

3.6.1 XGBoost Classifier

XGBoost or eXtreme Gradient Boost is a gradient boosting algorithm that comes under ensemble learning shown in the Fig 3.6. This is a flexible machine learning algorithm used for regression, classification, and feature scoring tasks. One of the main characteristics of the XGBoost algorithm is that it uses a gradient descent algorithm to minimize the loss function. Since this is an ensemble learning approach, the algorithm uses a decision tree as the predictor. Predictions are made sequentially by minimizing the error of the previous tree. Weights for the new tree are assigned based on their performance in the previous tree. The final output is the sum of outputs from each predictor. Apart from regression and classification, XGBoost also returns the feature importance score based on the number of times each feature is utilized in a tree. This feature score is mapped into some threshold values, which will return the number of features and accuracy when trained. So, XGBoost will help find the optimum number of features with maximum accuracy [11].

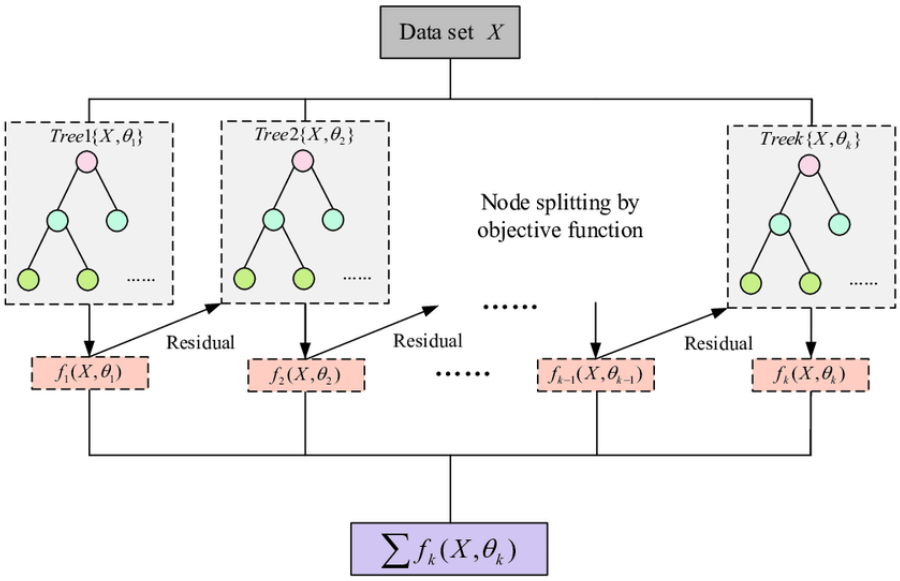


Figure 3.6: Working of XGBoost Classifier

3.6.2 1D CNN

Fig 3.7 represents a general framework of 1d CNN model.

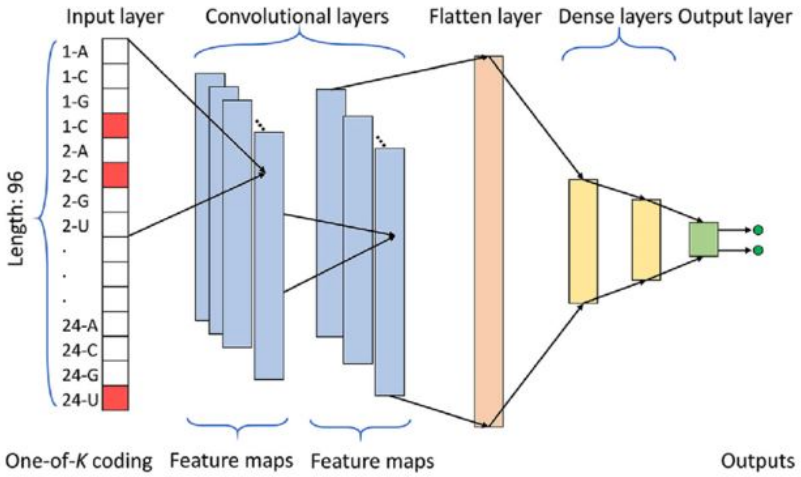


Figure 3.7: Layers in 1D CNN

3.6.3 J48 Algorithm

The J48 algorithm selects the property of the data from each node of the tree that divides its set of samples into subsets that are enriched in one class or the other. The normalised information gain, which is determined by the difference in entropy, serves as the splitting criterion. The attribute used to determine the decision is the one with the largest normalised information gain. The J48 algorithm then employs a divide-and-conquer strategy to

recursively work on the partitioned sub lists and builds a decision tree based on the greedy algorithm. [13].

3.7 Parameter Auto-tuning

First, the keras-tuner packages are installed and the RandomSearch function in it is used for getting the best combination of the parameters.

```
import keras
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import Flatten
from keras.layers import Dropout
from keras.layers.convolutional import Conv1D
from keras.layers.convolutional import MaxPooling1D

def build_model(hp):

    filter_num = hp.Int('filters', min_value=50, max_value=150, step=1)
    act_func = hp.Choice('Activation Func', ['relu', 'elu', 'sigmoid', 'tanh'])

    model = Sequential()
    model.add(Conv1D(filters=filter_num, kernel_size=3, activation=act_fn, input_shape=(X_train.shape[1],X_train.shape[2])
    model.add(MaxPooling1D(pool_size=1))
    model.add(Flatten())
    model.add(Dense(100, activation='relu'))
    model.add(Dropout(0.2))

    model.add(Dense(y_train.shape[1], activation='softmax'))

    model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model
```

```
tuner = RandomSearch(
    build_model,
    objective='accuracy',
    seed=1,
    max_trials=16,
    executions_per_trial=2,
    directory='dir',
    project_name='set6')
```

Figure 3.8: Auto-tuning in 1D CNN

A Random search for the best combination of parameters prescribed by the user in the function is done effectively. And also the criterion for the parameter-tuning has to be set in advance. Here in the Figure 3.8, selected parameters for tuning include the number of nodes or units in the convolution layer and the activation function to be used. And these parameters are tuned based on the objective to maximize the accuracy [13].

In case of integer search, the step size and the maximum and minimum values for search can be set. And in the case of Activation function, all the activation function to be checked can be inserted as an array. Also the seed, maximum trails and executions per trails can all be set prior to the search. As a result, the best combination of prescribed features are obtained. This combination of tuned parameters can be used as input to the program for getting maximum performance.

Chapter 4

Study Area and Datasets used

4.1 Study Area

The Kerala State Pollution Control Board (KSPCB) is a body of the Department of Health and Family Welfare, Government of the State of Kerala, India. Inorder to control pollution,

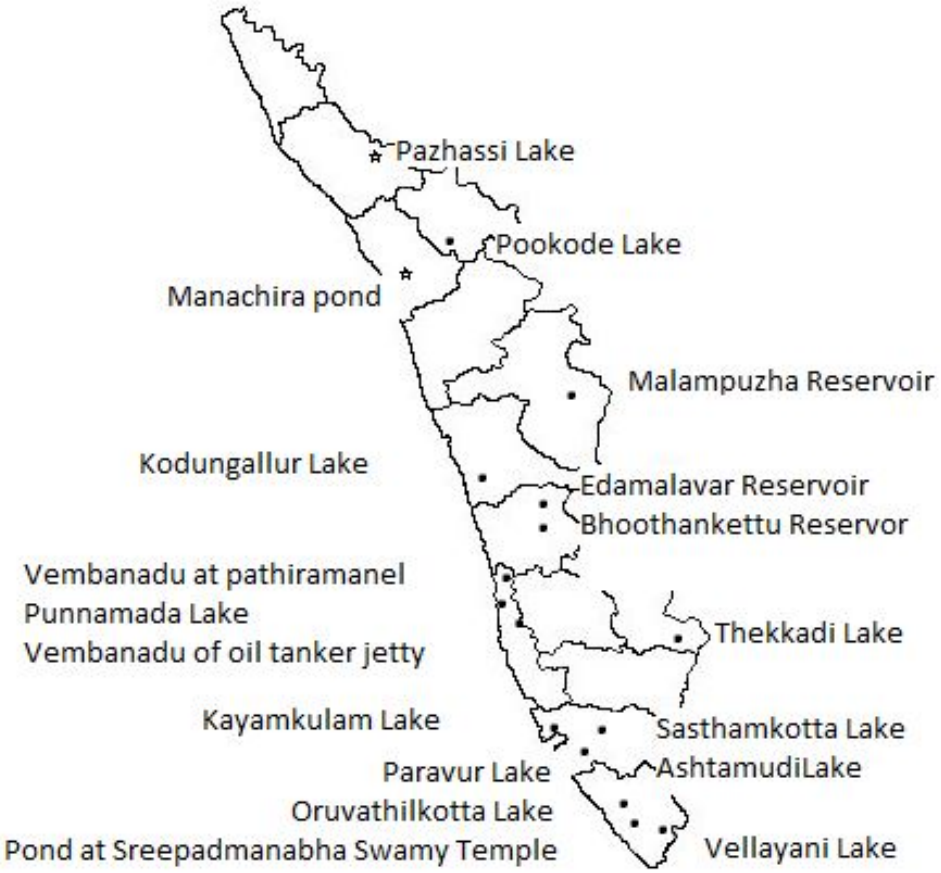


Figure 4.1: Study area

the Board has conducted numerous examinations of subsurface water, solids, and air. Anuja et al.[10] collected the water quality parameters from KSPCB as part of their final year project thesis. They also carried out several data preprocessing in the raw data available and calibrated MTSI values using the methodology proposed in Kurwattia and Karmakar[2] work. KSPCB collect water quality data on monthly basis for 18 sampling station of lentic water bodies across Kerala. This includes 11 lakes, 4 reservoirs and 2 ponds. The lakes are Oruvathilkotta, Sasthamkotta, Ashtamudi at Kollam, Paravur, Vembanadu oil tanker jetty, Thekkadi, Kodungalloor, Kayamkulam, Punnamada, Pookote, Vembanad at Pathiramanal, Vellayani. The reservoirs include Malampuzha, Bhoothathankettu, Edamalayar, Pazhassi and ponds include Pond at Sree Padmanabha Swamy Temple, Mananchira pond at Kozhikode as shown in Fig 4.1.

4.2 Dataset

The dataset used for the study includes the concentration of different factors associated with eutrophication of these lentic water bodies. The water quality parameters available with KSPCB are water temperature, dissolved oxygen (DO), pH, conductivity, biochemical oxygen demand (BOD), nitrate, turbidity, phenolphthalein alkalinity, total alkalinity, chlorides, chemical oxygen demand (COD), total kjeldahl nitrogen, ammonia nitrogen, hardness, calcium, magnesium, sulphate, sodium, total dissolved solids, total suspended solids, total fixed solids, phosphate, boron, potassium, fluoride as shown in Table 4.1.

Table 4.1: Input features and ranges

Input Features	**Range**
Nitrate	(0.009, 7.5)
Potassium	(0.03, 820)
PH	(1.2, 8.7)
Chloride	(1, 32900)
COD	(1.6, 3152)
DO	(1.3, 66)
Conductivity	(19, 437000)
Temperature	(5.8, 37.5)
Turbidity	(0.01, 46.6)
BOD	(0.1, 8.4)
TN	(0.01, 8.36)
Phosphate	(0.0101, 2.023)

It is a Multivariate Statistical analysis. TP, TN, Chlorophyll-a were selected as the indicator for trophic state classification in this study, considering the lentic waterbodies

of Kerala. A generic framework was developed for estimation of MTSI which is based on Principal Component Analysis (PCA) and development of threshold of each tropic trophic state using composite parametric and kernel based non-parametric approaches. Validation of index was also done using cluster analysis. The corresponding MTSI values were calibrated for each data using this proposed model for prediction.

Chapter 5

Results and Discussions

In the first phase of the work, Multivariate Trophic State Index (MTSI) values for the data was predicted using different Machine Learning and ensemble models: Linear Regression(LR), Artificial Neural Network(ANN), Support Vector Regression(SVR) and Random Forest(RF). And their performance indices were compared for finding the better model.

5.1 MTSI Prediction

The actual vs predicted value plots of results obtained from various prediction models experimented on the dataset are shown from Fig 5.1 to 5.5.

5.1.1 Linear Regression

The experiment conducted on the dataset with Linear Regression Model shows lesser accuracy. The performance indices obtained are shown in the Figure 5.1.

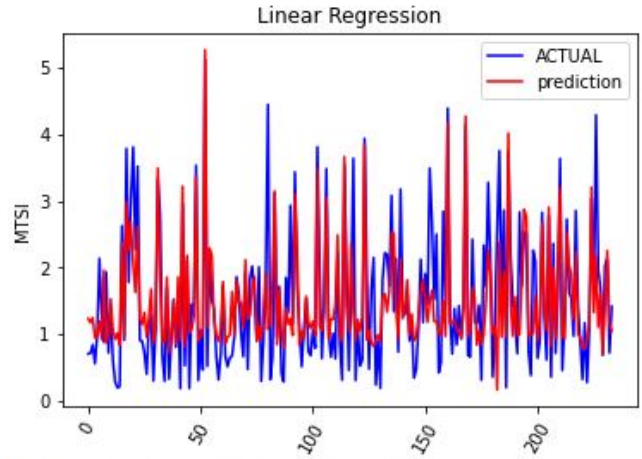


Figure 5.1: Actual vs Predicted MTSI values for Linear Regression Model

5.1.2 Artificial Neural Network

The experiment conducted on the dataset with ANN Model shows less accuracy. The performance indices obtained are shown in the Figure 5.2.

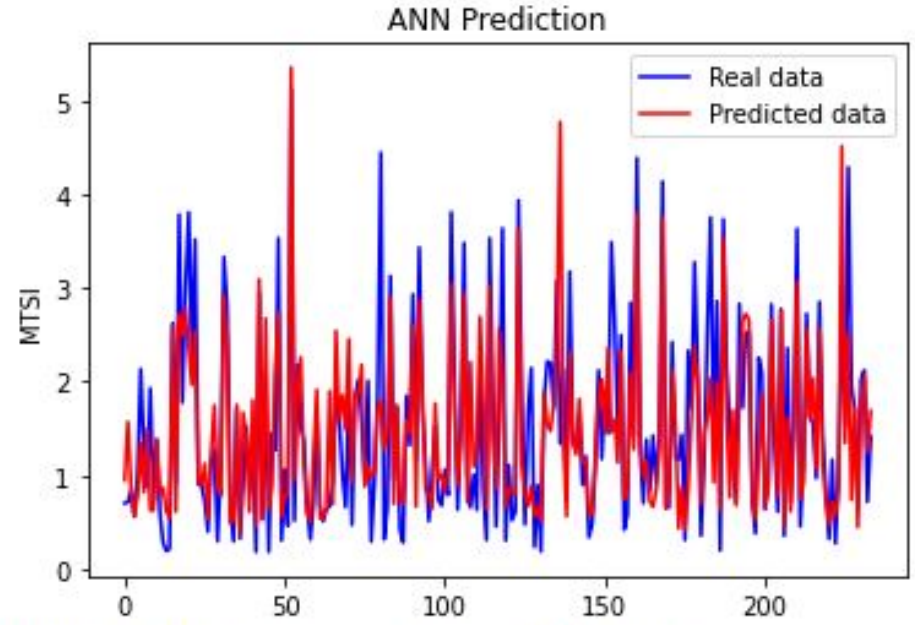


Figure 5.2: Actual vs Predicted MTSI values for ANN Model

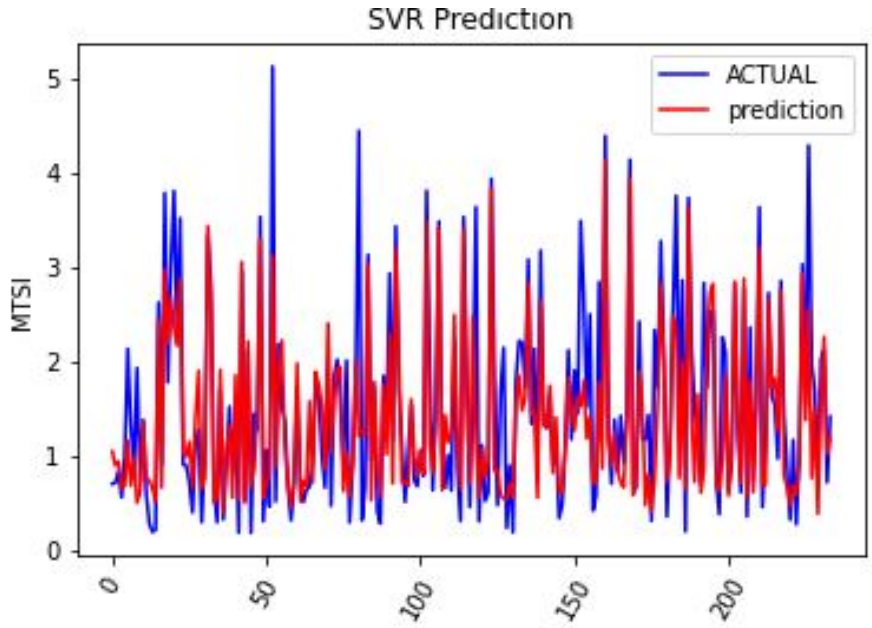
Parameters used in the ANN regressor model are as follow:
Optimizer='adam', Loss='mean squared error' and Activation function='relu'.

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_4 (Dense)             (None, 32)                320

 dense_5 (Dense)             (None, 1)                 33

=================================================================
Total params: 353
Trainable params: 353
Non-trainable params: 0
_________________________________________________________________
```

Figure 5.3: Layers of ANN

5.1.3 Support Vector Regression

The experiment conducted on the dataset with SVR Model shows better performance as shown in Fig 5.4.
Parameters used in the SVR prediction model are as follow:
kernel='rbf', gamma='auto', uses 1 / n-features. C=1.0, epsilon = 0.1.




```
Mean Absolute Error: 0.3858784988648869
Mean Squared Error: 0.33678084855913953
Root Mean Squared Error: 0.5803282248513677
R-square value  0.6874156053397076
```

Figure 5.4: Actual vs Predicted MTSI values for SVR Model

5.1.4 Random Forest

The experiment conducted on the dataset with Random Forest Model shows best accuracy. The performance indices obtained are shown in the Figure below:

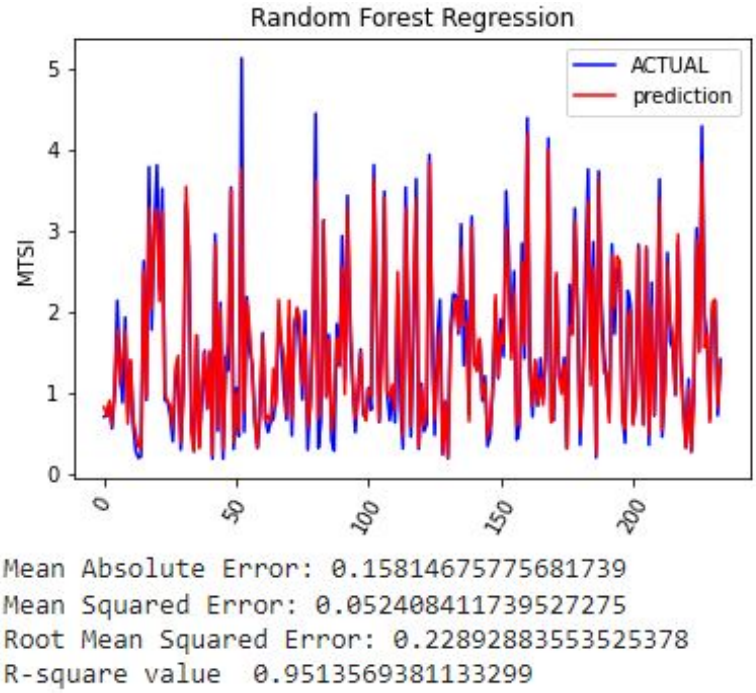


Figure 5.5: Actual vs Predicted MTSI values for Random Forest Model

5.1.5 Comparison Of Prediction Models

Comparison of performance indices for the four prediction models are shown in the Table 5.1. Its clearly evident from the table that Random Forest shows the least MAE, MSE and RMSE values and the highest R-square value for both training and testing results. Hence, Random Forest is selected as the better one among the prediction models.

Table 5.1: Performance Indices of MTSI Prediction Models for Training and Testing

	TRAINING				TESTING			
Performance Measures	LR	ANN	SVR	RF	LR	ANN	SVR	RF
MAE	0.507	0.391	0.414	**0.152**	0.483	0.437	0.385	**0.158**
MSE	0.471	0.281	0.388	**0.047**	0.436	0.390	0.336	**0.052**
RMSE	0.686	0.530	0.623	**0.216**	0.660	0.624	0.580	**0.228**
R^2	0.474	0.685	0.566	**0.947**	0.595	0.637	0.687	**0.951**

5.1.6 Violin and Box Plot Comparison

Figure 5.6 and 5.7 shows the Violin and Box Plot for the Actual and Predicted MTSI values for different prediction models. Violin plot represents the dispersion of Mtsi values and Box plot gives the range, median, interquartile range etc.

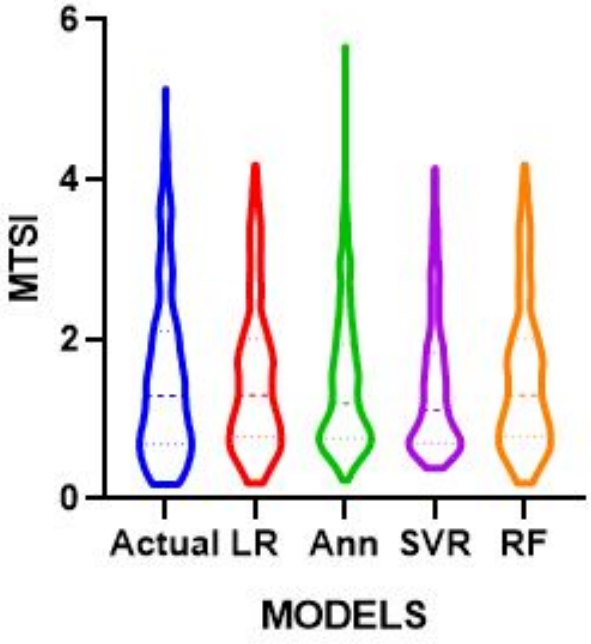


Figure 5.6: Violin Plot Comparison

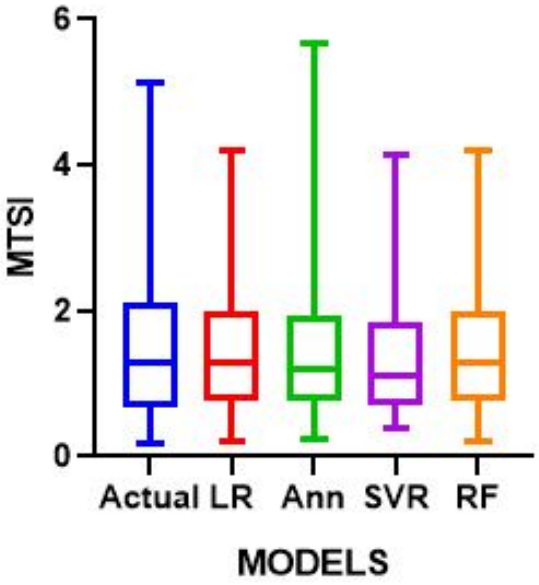


Figure 5.7: Box Plot Comparison

5.1.7 Radar Plot Comparison

Figure 5.8 shows the radar plot for different prediction models. It clearly shows RF as the better model as it has maximum R2 score and minimum errors

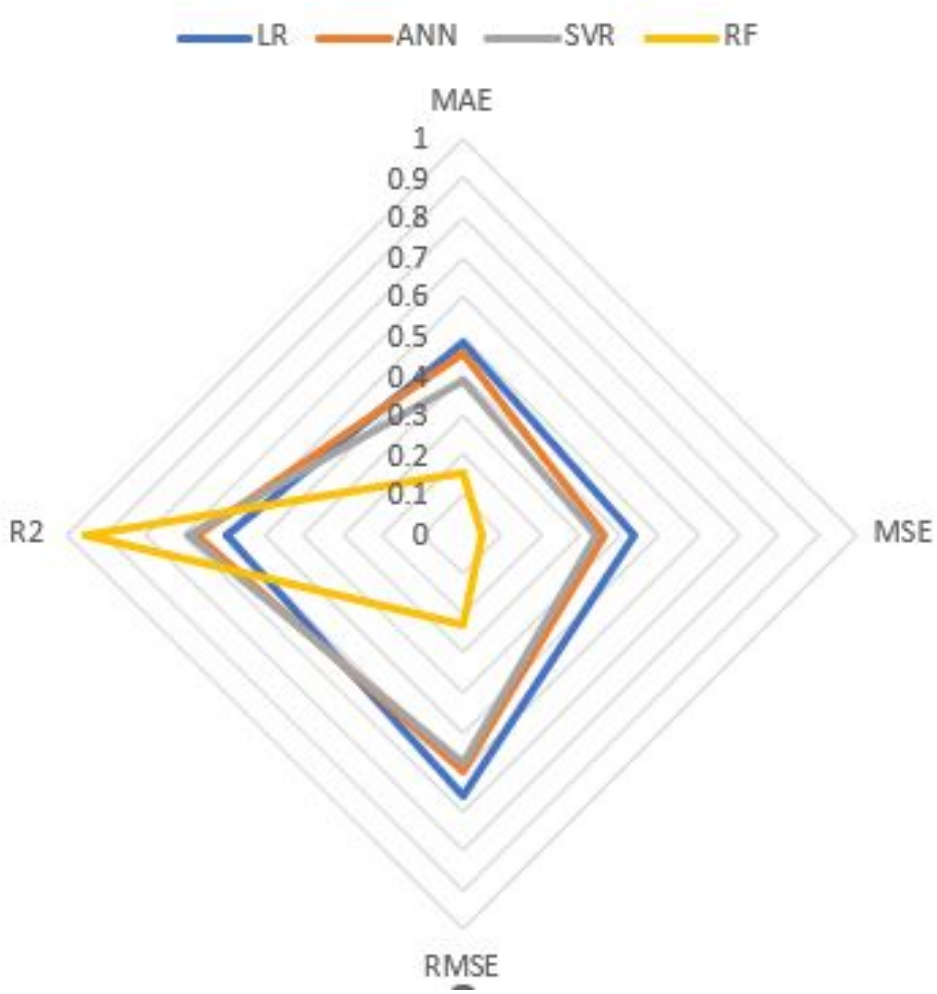


Figure 5.8: Radar Plot Comparison

5.2 Trophic State Classification

In the next phase of the work, the KSPCB dataset was transformed adequately for Trophic State Classification. The dataset was divided into different Trophic States according to the ranges of MTSI values. The MTSI thresholds for initial classification is shown in Table 5.2.

Table 5.2: MTSI Thresholds

Trophic State	Lower Limit	Upper Limit
Oligotrophic	0.2905	0.87
Mesotrophic	0.792	1.36
Eutrophic	1.365	2.34
Hypereutrophic	2.35	-

Different hybrid models were developed by coupling Random Forest algorithm with various ensemble and Deep learning models for effective Trophic State Classification. Here Random Forest model was used as an efficient tool for feature selection. In this work, different models used for classification are DNN, 1D CNN, J48 and XGBoost classifiers. The parameters associated with each models are auto-tuned using the keras_tuner function.

5.3 Feature Selection Using Random Forest

RF was used as a multidimensional reduction technique for feature selection, which eliminates insignificant features [14].

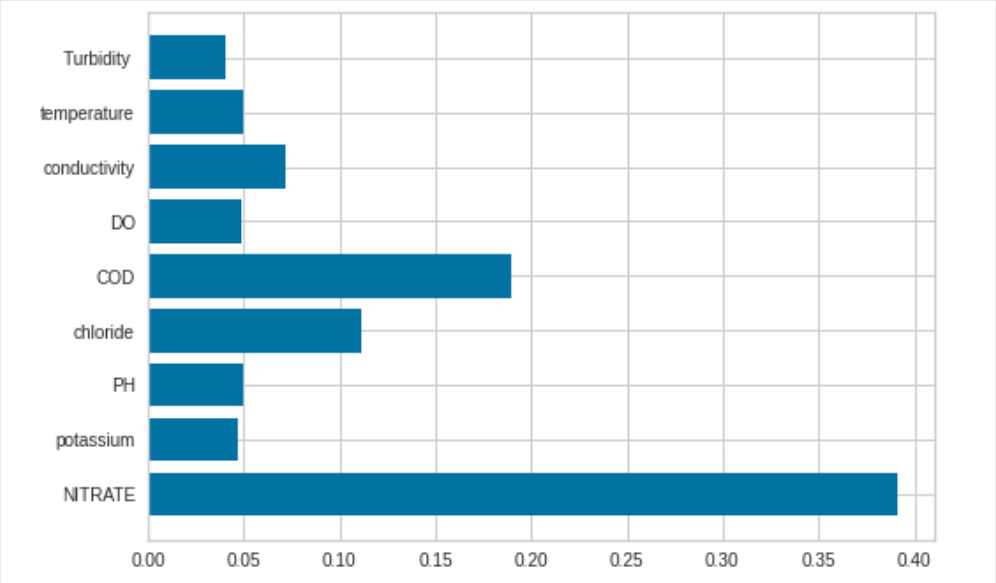


Figure 5.9: RF_Feature Importance

It collects the feature importance values which can be accessed via feature_importances attribute after fitting the Random Forest model shown in Fig 5.9.

Features selected from RF model based on the feature importance value include:

- Nitrate
- PH
- Chloride
- COD
- Conductivity

These 5 features are contributing the 80% of the complete TSI prediction. Better results are obtained when considering these features (than 9 features as input) as over fitting is reduced in this case.

5.4 Hybrid Classification Models

5.4.1 1D CNN

In the initial step, the important features are selected making use of the Feature_Importance attribute in Random Forest. Then the trophic states are classified based on the selected features using 1d-CNN. Fig 5.10 shows the representation of layers used in the model.

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d_4 (Conv1D)           (None, 3, 128)            512

 conv1d_5 (Conv1D)           (None, 1, 128)            49280

 dropout_2 (Dropout)         (None, 1, 128)            0

 max_pooling1d_2 (MaxPooling  (None, 1, 128)           0
 1D)

 flatten_2 (Flatten)         (None, 128)               0

 dense_4 (Dense)             (None, 96)                12384

 dense_5 (Dense)             (None, 4)                 388

=================================================================
```

Figure 5.10: Layers used in 1D CNN

Fig 5.11 and 5.12 shows the confusion matrix and loss curve for trophic state classification using 1d CNN. Different performance measures for classification are formulated in Table 5.3.

Figure 5.11: Confusion Matrix of 1D CNN

Table 5.3: Performance Indices of 1D CNN

	Precision	**Recall**	**F1-score**
Oligotrophic	0.61	0.83	0.71
Mesotrophic	0.44	0.29	0.35
Eutrophic	0.58	0.54	0.56
Hypereutrophic	0.80	0.73	0.76
Avg	0.61	0.60	0.59
Accuracy			**0.61**

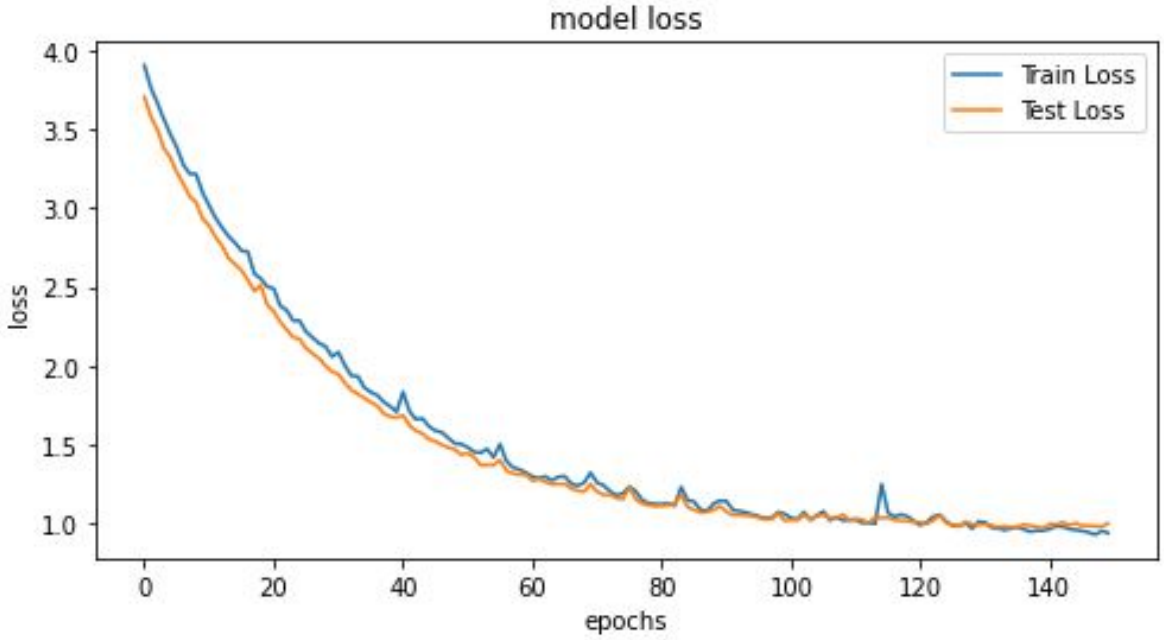


Figure 5.12: Loss Curve of 1D CNN

5.4.2 J48 algorithm

In the initial step, the important features are selected making use of the Feature_Importance attribute in Random Forest. Then the trophic states are classified based on the selected features using J48 algorithm in Waikato Environment for Knowledge Analysis (WEKA) software. Different performance measures for classification are formulated in Table 5.4.

Table 5.4: Performance Indices of Hybrid DNN Model

	Precision	**Recall**	**F1-score**
Oligotrophic	0.71	0.74	0.72
Mesotrophic	0.51	0.29	0.37
Eutrophic	0.37	0.46	0.41
Hypereutrophic	0.46	0.61	0.52
Avg	0.54	0.53	0.53
Accuracy			**0.54**

5.4.3 XGBoost Classifier

In the initial step, the important features are selected making use of the Feature_Importance attribute in Random Forest. Then the trophic states are classified based on the selected features using XGBoost Classifier. Fig 5.13 and Table 5.5 shows the confusion matrix and performance indices respectively for trophic state classification.

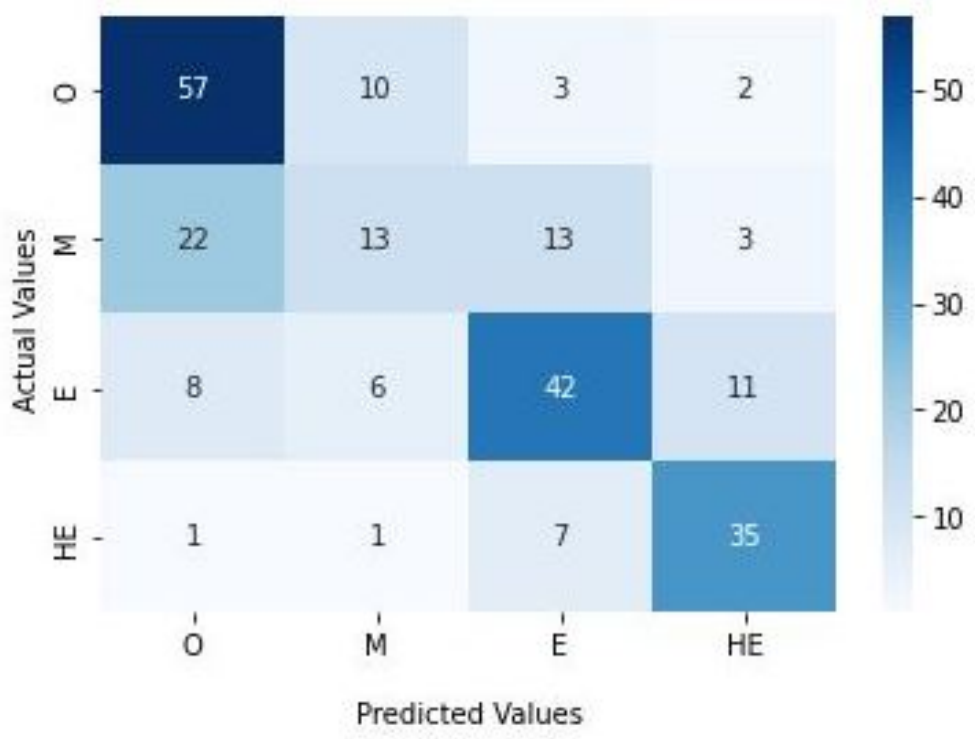


Figure 5.13: Confusion Matrix of XGBoost

Table 5.5: Performance Indices of XGBoost

	Precision	**Recall**	**F1-score**
Oligotrophic	0.58	0.85	0.69
Mesotrophic	0.45	0.25	0.33
Eutrophic	0.65	0.51	0.57
Hypereutrophic	0.77	0.82	0.79
Avg	0.61	0.61	0.59
Accuracy			**0.62**

5.5 Comparison of Simple DNN vs Hybrid DNN Model

In the final phase of the work, the performance comparison is done simple DNN (which utilises all the 9 features for trophic state classification) and hybrid RF+DNN model (which utilises the advantage of efficient feature selection). In hybrid model, first the important features are selected making use of the Feature_Importance attribute in Random Forest. Then the trophic states are classified based on the selected features using DNN [17].

5.5.1 Performance Indices Comparison

Its clearly evident from the Tables below that the classification accuracy of Hybrid model(62%) is better than that of Simple DNN model(56%). Also the averages of precision, recall and f1-score are better for the hybrid model. This improvement in performance shows the importance of effective feature selection for Trophic State Classification.

Table 5.6: Performance Indices of Simple DNN Model

	Precision	**Recall**	**F1-score**
Oligotrophic	0.62	0.76	0.69
Mesotrophic	0.47	0.43	0.45
Eutrophic	0.55	0.43	0.48
Hypereutrophic	0.57	0.59	0.58
Avg	0.55	0.55	0.55
Accuracy			**0.56**

Table 5.7: Performance Indices of Hybrid DNN Model

	Precision	**Recall**	**F1-score**
Oligotrophic	0.65	0.79	0.71
Mesotrophic	0.43	0.25	0.32
Eutrophic	0.65	0.63	0.64
Hypereutrophic	0.69	0.80	0.74
Avg	0.61	0.63	0.61
Accuracy			**0.63**

5.5.2 Analysis of Confusion Matrices Obtained

Figure 5.14 and 5.15 shows the confusion matrices for simple and Hybrid DNN model respectively.

Figure 5.14: Confusion Matrix of Simple DNN Model

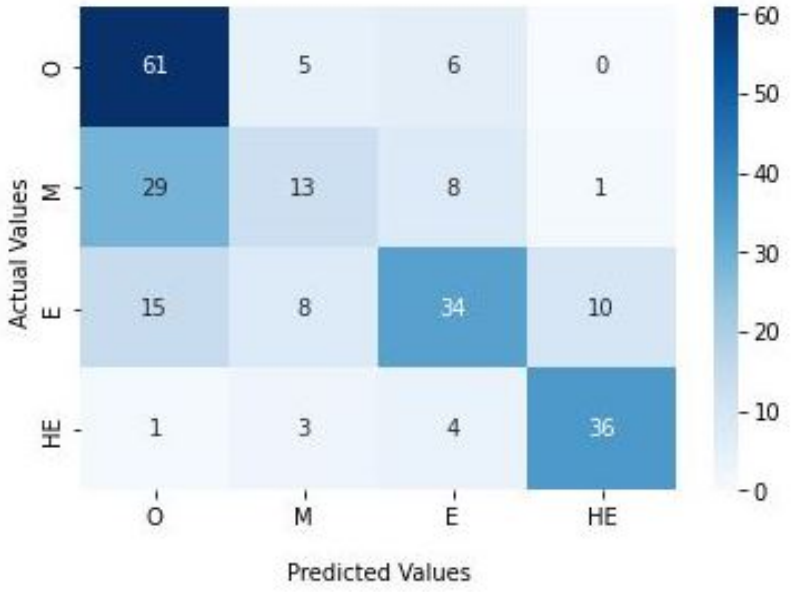


Figure 5.15: Confusion Matrix of Hybrid DNN Model

5.5.3 Loss Curve Comparison

In Figure 5.16, the loss curve seems to cease decreasing and remain almost a constant for rest of the epochs. But the loss curve for the hybrid model is decreasing at a higher rate than the simple DNN model. So it can be concluded that the better performance is obtained for the hybrid model.

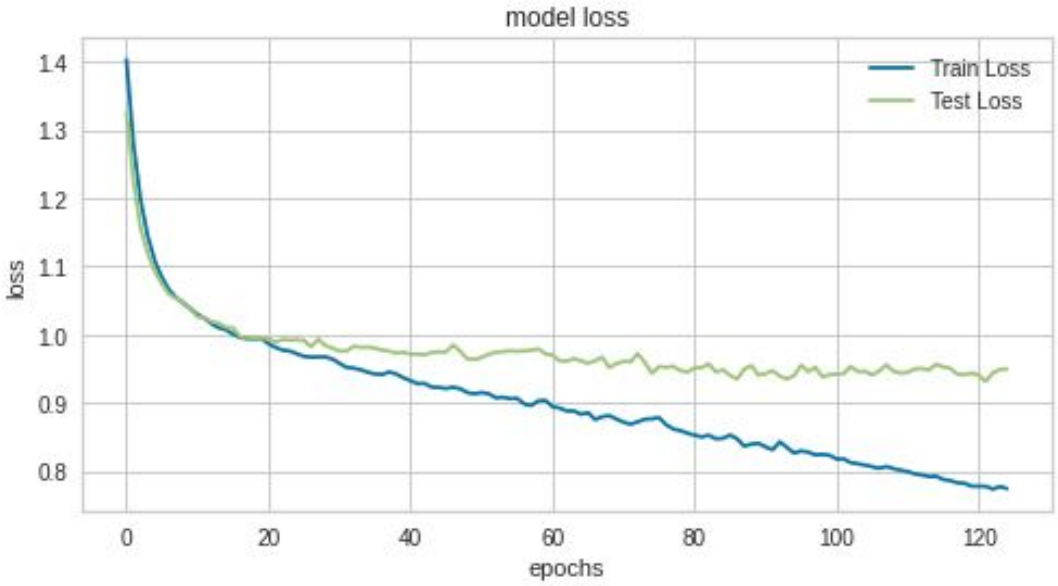


Figure 5.16: Loss Curve of Simple DNN Model

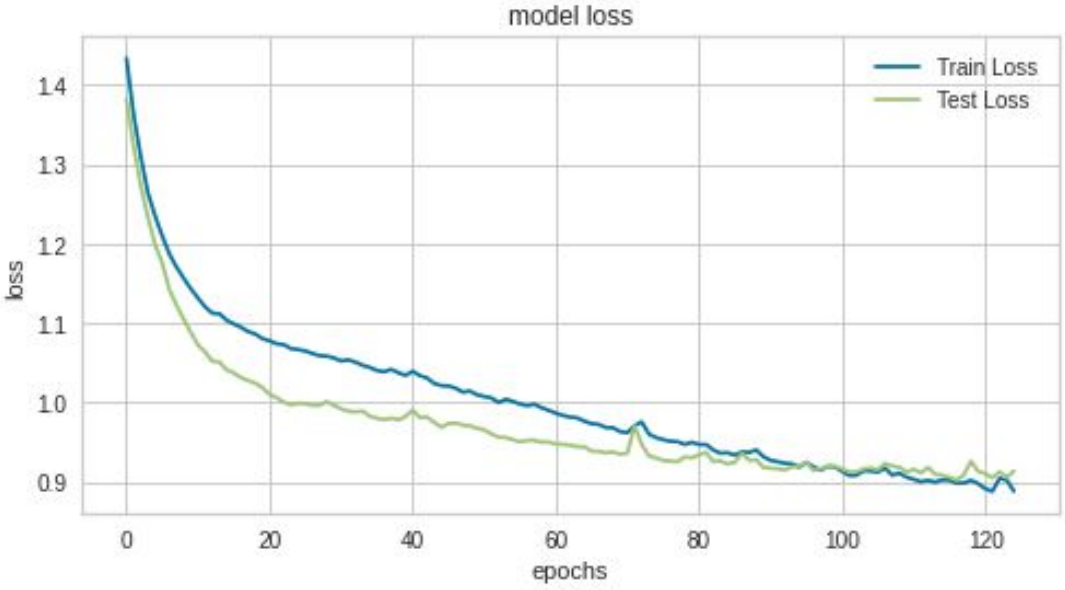


Figure 5.17: Loss Curve of Hybrid DNN Model

5.6 Comparative Study of different Classification models

Figure below shows the radar plot for comparison of different performance measures (Precision, Recall,F1 score and Accuracy) for different classification model used.There is not much difference between the performance indices, whereas hybrid (RF+DNN) shows a slight edge over the others.

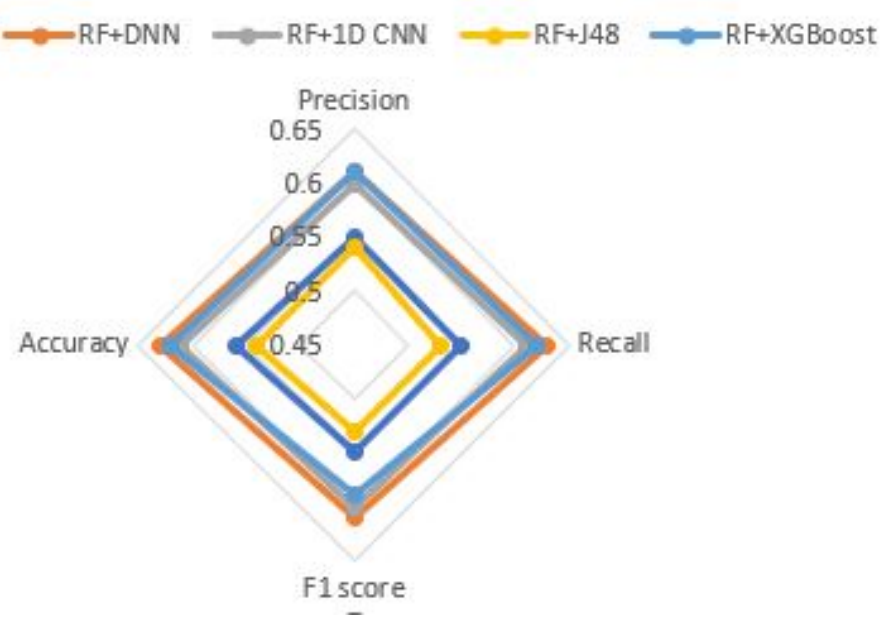


Figure 5.18: Radar Plot for Classification models performances

Chapter 6

Conclusion

Conservation and management of lentic water bodies require an effective method of evaluating its trophic states. However the development of TSI by statistical methods require parameters like TN, TP and BOD, which are difficult to obtain. Traditional Trophic State Index like Carlson TSI, based on univariate approach was widely used for classification of lakes. Multivariate Trophic State Index based on multivariate approach is a robust alternative for lake classification index, not widely applied. The prediction of TSI using Soft Computing is rare and researchers have hardly attempted the prediction of MTSI using Soft Computing.

Hence, AI models can be adopted to predict the MTSI using easily measurable water quality parameters. Easily available and efficient features can be used for prediction using AI algorithms. The framework was applied to the KSPCB dataset of 11 lakes, 4 reservoirs and 2 ponds in Kerala for the period, 2012-2018. In this study, MTSI prediction was done using four regression models : LR, ANN, SVR and RF and autotuned hybrid models were proposed for effective Trophic state classification.

- RF obtained better results($R^2 > 0.9$) among the regression models for MTSI prediction.
- Combination of Nitrate, PH, Chloride, COD and Conductivity were selected as input-features using RF model based on feature importance values.
- Trophic State Classification was performed with selected features using different DL and ensemble models : DNN, 1D CNN, J48 and XGBoost algorithms.
- Comparison of different performance indices shows that Hybrid (RF+DNN) model performs better than stand-alone DNN model.

References

[1] Carlson RE. A trophic state index for lakes: Limnology and Oceanography. March. 1977.

[2] Kurwattia TB, Karmakar S. Development of Modified Trophic State Index for Lakes. International Journal of Global Technology Initiatives. 2012 Mar 29;1(1):A71-8.

[3] Canfield DE, Hodgson LM. Prediction of Secchi disc depths in Florida lakes: impact of algal biomass and organic color. Hydrobiologia. 1983 Feb;99(1):51-60.

[4] Chou JS, Ho CC, Hoang HS. Determining quality of water in reservoir using machine learning. Ecological informatics. 2018 Mar 1;44:57-75.

[5] Saghi H, Karimi L, Javid AH. Investigation on trophic state index by artificial neural networks (case study: Dez Dam of Iran). Applied Water Science. 2015 Jun;5(2):127-36.

[6] Concepcion II RS, Loresco PJ, Bedruz RA, Dadios EP, Lauguico SC, Sybingco E. Trophic state assessment using hybrid classification tree-artificial neural network. International Journal of Advances in Intelligent Informatics. 2020 Mar 1;6(1):46-59.

[7] Hu M, Ma R, Cao Z, Xiong J, Xue K. Remote Estimation of Trophic State Index for Inland Waters Using Landsat-8 OLI Imagery. Remote Sensing. 2021 Jan;13(10):1988.

[8] Zhu S, Mao J. A Machine Learning Approach for Estimating the Trophic State of Urban Waters Based on Remote Sensing and Environmental Factors. Remote Sensing. 2021 Jan;13(13):2498.

[9] Ellina G, Papaschinopoulos G, Papadopoulos BK. Variables' classification via equivalence relations for the trophic state of a Mediterranean ecosystem. Water Environment Research. 2021 Apr 2.

[10] Anuja P K, Aggie S, Ammu B, Anargha A. Prediction of Multivariate Trophic State Index for Lentic Water Bodies Using Soft Computing Techniques. Undergraduate Final Year Project, TKMCE. 2019

[11] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 785-794).

[12] Li D, Zhang J, Zhang Q, Wei X. Classification of ECG signals based on 1D convolution neural network. In2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom) 2017 Oct 12 (pp. 1-6). IEEE.

[13] Huang C, Li Y, Yao X. A survey of automatic parameter tuning methods for metaheuristics. IEEE transactions on evolutionary computation. 2019 Jun 7;24(2):201-16.

[14] Li X, Chen W, Zhang Q, Wu L. Building auto-encoder intrusion detection system based on random forest feature selection. Computers Security. 2020 Aug 1;95:101851.

[15] Yadav AK, Chandel SS. Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model. Renewable Energy. 2015 Mar 1;75:675-93.

[16] Miikkulainen R, Liang J, Meyerson E, Rawal A, Fink D, Francon O, Raju B, Shahrzad H, Navruzyan A, Duffy N, Hodjat B. Evolving deep neural networks. InArtificial intelligence in the age of neural networks and brain computing 2019 Jan 1 (pp. 293-312).

[17] Sze V, Chen YH, Yang TJ, Emer JS. Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE. 2017 Nov 20;105(12):2295-329.

[18] Adamovich BV, Zhukova TV, Mikheeva TM, Kovalevskaya RZ, Luk'yanova EV. Long-term variations of the trophic state index in the Narochanskie Lakes and its relation with the major hydroecological parameters. Water resources. 2016 Sep;43(5):809-17.