aditya.pdf
May 20, 2021
6305 words / 31792 characters

aditya c

aditya.pdf

Sources Overview

19%

OVERALL SIMILARITY

1	Rajesh Kumar Yadav, Adarsh Jha, Aditya Choudhary. "IoT based prediction of water quality index for farm irrigation", 2021 International ... CROSSREF	6%
2	Kookmin University on 2020-06-02 SUBMITTED WORKS	3%
3	Indian Institute of Science, Bangalore on 2019-05-26 SUBMITTED WORKS	2%
4	Delhi Technological University on 2020-05-21 SUBMITTED WORKS	2%
5	Delhi Technological University on 2018-12-02 SUBMITTED WORKS	1%
6	Delhi Technological University on 2020-05-24 SUBMITTED WORKS	<1%
7	Mihai Bojescu. "Treatment of 27 Different Types of Cancer", Research Square, 2021 CROSSREF POSTED CONTENT	<1%
8	Indian School of Mines on 2019-05-09 SUBMITTED WORKS	<1%
9	Kookmin University on 2020-06-02 SUBMITTED WORKS	<1%
10	text-id.123dok.com INTERNET	<1%
11	mafiadoc.com INTERNET	<1%
12	mgutheses.in INTERNET	<1%
13	Liverpool John Moores University on 2021-02-28 SUBMITTED WORKS	<1%
14	spectrum.library.concordia.ca INTERNET	<1%
15	Can Bülent Karakuş. "Assessment of relationship between land use/cover and surface water quality trends within the riparian zone: a ... CROSSREF	<1%
16	Taylor's Education Group on 2013-01-15 SUBMITTED WORKS	<1%

17	Universiti Tenaga Nasional on 2021-01-31 SUBMITTED WORKS	<1%
18	Zouheira Ibn Ali, Asma Gharbi, Moncef Zairi. "Evaluation of groundwater quality in intensive irrigated zone of Northeastern Tunisia", Gr... CROSSREF	<1%
19	aclweb.org INTERNET	<1%
20	Ajman University of Science and Technology on 2021-05-12 SUBMITTED WORKS	<1%
21	City University of Hong Kong on 2007-04-24 SUBMITTED WORKS	<1%
22	Higher Education Commission Pakistan on 2012-10-15 SUBMITTED WORKS	<1%
23	The University of Manchester on 2021-04-23 SUBMITTED WORKS	<1%
24	University of Technology, Sydney on 2018-10-12 SUBMITTED WORKS	<1%
25	Creusen, I. M., L. Hazelhoff, P. H. N. de With, Onur G. Guleryuz, and Robert L. Stevenson. "<title>A semi-automatic traffic sign detection... CROSSREF	<1%
26	Mohamad Sakizadeh. "Artificial intelligence for the prediction of water quality index in groundwater systems", Modeling Earth System... CROSSREF	<1%
27	University of Lancaster on 2020-12-16 SUBMITTED WORKS	<1%
28	Liverpool John Moores University on 2021-03-15 SUBMITTED WORKS	<1%
29	University of Bedfordshire on 2019-11-07 SUBMITTED WORKS	<1%

Excluded search repositories:

- None

Excluded from Similarity Report:

- Bibliography
- Quotes

Excluded sources:

- None

IoT based prediction of water quality index for farm irrigation

B.Tech Major Project Project report

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF
BACHELOR OF TECHNOLOGY
IN
COMPUTER ENGINEERING
by

Adarsh Jha
2K17/CO/022
Aditya Choudhary
2K17/CO/024

Under the supervision of
Dr. R K Yadav
Assistant Professor
COE Department

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of engineering)
Bawana Road, Delhi-110042

May, 2021

DECLARATION

We/I hereby certify that the work which is presented in the Major Project-II entitled **IoT based prediction of water quality index for farm irrigation** in fulfilment of the requirement for the award of the Degree of Bachelor of Technology in **Computer Engineering** and submitted to the Department of Information Technology, Delhi Technological University, Delhi is an authentic record of my/our own, carried out during a period from January to May 2021, under the supervision of **Dr. Rajesh Kumar Yadav**. The matter presented in this report has not been submitted by us/me for the award of any other degree of this or any other Institute/University. The work has been published/accepted/communicated in SCI/ SCI expanded/SSCI/Scopus indexed journal OR peer reviewed Scopus indexed conference with the following details:

Title of the Paper: **IoT based prediction of water quality index for farm irrigation**
Author names (in sequence as per research paper): **Dr Rajesh Kumar Yadav, Adarsh Jha, Aditya Choudhary**
Name of Conference/Journal: **2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)**
Conference Dates with venue (if applicable): **25-27 March**
Have you registered for the conference (Yes/No)?: **Yes**
Status of paper (Accepted/Published/Communicated): **Published**
Date of paper communication: **20 February 2021**
Date of paper acceptance: **24 February 2021**
Date of paper publication: **12 April 2021**

Adarsh Jha **Aditya Choudhary**

2K17/CO/022 **2K17/CO/024**

Student(s) Roll No., Name and Signature

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I further certify that the publication and indexing information given by the students is correct.

Place: ________________ **Supervisor Name and Signature**

Date: ________________

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

We Adarsh Jha, Aditya Choudhary, Roll no – 2K17/CO/022, 2K17/CO/024 , hereby declare that the project Dissertation titled "**IoT based prediction of water quality index for farm irrigation**" which is submitted by us to the Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology (Computer Engineering), is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi

Date: 21 May 2021

Adarsh Jha

Aditya Choudhary

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

We hereby certify that the Project Dissertation titled "**IoT based prediction of water quality index for farm irrigation**" which is submitted by Adarsh Jha, Aditya Choudhary, Roll Nos. – 2K17/CO/022, 2K17/CO/024 Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date: 21 May 2021

Dr. R K Yadav

SUPERVISOR

ACKNOWLEDGEMENT

We would like to express great appreciation to our supervisor Dr R. K. Yadav for his valuable and constructive suggestions during the planning and development of this project. His willingness to give his time so generously has been very much appreciated. We owe a special debt to our parents and friends for showing their generous love and care throughout the entire period of this time.

Place: Delhi

Date: 21 May 2021

Adarsh Jha

Aditya Choudhary

ABSTRACT

Each step of agriculture, be it preparation of soil, adding fertilizer, irrigation, sowing and harvesting requires proper analysis of soil, water, sunlight, weather etc. for better crop yield which is time consuming and costly. Water quality Index is used as a measure of quality of water and a model consisting of various linear and non-linear regression models is used for prediction of water quality index based on parameters measured using sensor data. In this work, we develop a model for prediction of the Water Quality Index for Irrigation purposes (IWQI) mainly based on Sodicity and Salinity. Developing this index can save cost and time for chemical tests. Five out of thirteen parameters mainly indicating salinity which are Cl^{-}, Na^{+}, HCO^{3-}, SAR and EC were used to calculate the IWQI Index. We reduce the five parameters using correlation analysis to three parameters. Finally, we develop a classification algorithm to predict the IWQI classes using seven multiple algorithms. Best performance was by Random Forest Algorithm and overall results for other algorithms were satisfactory. The developed algorithms can be incorporated in precision farming to prevent salt based damage on the crops.

Keywords: Irrigation Water, Analytical models, Salinity (geophysical), Sociology, Water quality, Predictive models, Tools

Contents

List of figures

List of tables

LIST OF ABBREVIATIONS

IOT - Internet of Things

ML - Machine Learning

WSN - Wireless Sensor Networks

SVR - Support Vector Regressor

MLP - Multilayer Perceptron

WQI - Water quality Index

SVM - Support Vector Machine

ANN - Artificial Neural Network

MAE - Mean Absolute Error

MSE - Mean Squared Error

BOD - Biological Oxygen Demand

DO - Dissolved Oxygen

PH - Potential Hydrogen

1 Introduction

According to the economic survey of 2018[1] in India, around half of the workforce is involved in the agriculture sector. The contribution of the sector to the GDP is only 16% which has reduced significantly from 50% in 1950. The decline is not limited to India and is observed in the rest of the world. This low productivity depends on many factors, one of the major one is wastage of farming resources, money and time. Also, the majority of people involved in farming are from rural areas who are poor and have insufficient knowledge regarding farming practices. Each step in farming, be it preparation of soil, adding Fertilizers, irrigation or harvesting, requires proper analysis of soil nutrients, water quality, weather, sunlight etc for improving productivity.

With improvement in technology, various farming equipment have been developed aiming to increase productivity and proper utilization of resources. The use of the Internet of Things, which consists of various devices connected through a network interacting with computers to transfer data, is becoming feasible with the internet and computers getting faster and cheaper.

Various sensors can be placed in a farming environment to obtain large, real-time datasets such as irrigation water(pH, TDS, various chemical concentrations), soil(moisture, pH, nutrients), weather(temperature, humidity etc). These datasets can then be analyzed and used to train machine learning models that will help inadequate utilization of resources, determining which crop could be suitable for a specific environment and increasing the efficiency of crop production.

Before irrigation, it is necessary to test and evaluate the quality of water according to a set of standards so as to prevent damaging the soil and crops. There needs to be a measure of water quality based on data gathered from various sensors. The Water Quality Index, defined by the world health organization, is used in this work as the quality measure for irrigation water. A model is developed consisting of classification techniques to predict the water quality Index class based on parameters measured by the IoT sensors.

In this work, a classification model for prediction of Irrigation Water Quality Index(IWQI) is developed that is based on salinity and salt tolerance of the sample. The following lists our contribution:

- The five parameters measured using underwater sensors are converted to quality measurement indications and further used to calculate water quality index. The indices are further used to calculate quality classes.
- Using various dimensional reduction techniques, the five parameters are reduced to three parameters.

- The resultant three parameters are used to develop a model for classifying the water samples based on classes of IWQI. Seven algorithms i.e. Support Vector Machine Classifier, MLP based classifier, Ensemble based Gradient Boosting, Bagging, Bayesian classifier, Decision Tree Classifier and Random Forest.

- The best performing classifier is chosen after a rigorous evaluation. This is further selected for future use in farming systems.

The rest of the report is structured as described: Section 2 describes similar works related to this project, section 3 describes the preliminary concepts leading to IWQI model, Section 4 discusses the dataset for water quality and the classification algorithm, Section 5 describes the various experimental settings and results of applying the classification algorithm on the dataset and finally Section 6 gives the concludes the project and gives some future possibilities.

2 Related Works

Many equipment have been developed for helping in agriculture that includes modern farming tools such as mower, sprayer, seed drill etc. Also, with recent developments in IOT and cloud computing, many new and cheap sensors have come out that can help in agriculture. Also, many methods have been developed which make use of these equipment for efficient farming. Precision agriculture[2] is the process developed which involves gathering various time series, geolocation and individual data, processing and analyzing them, and making use of them for improving productivity in agriculture. Any step of agriculture where data can be gathered can come under precision agriculture. For example, in irrigation systems data can be gathered and analyzed to find the quality of water, determining and classifying soil and fertilizing it accordingly which will help in better fertilizing.

Vij. et al[3] proposes the use of wireless sensor networks for measuring various parameters such as temperature, moisture, water level, weather etc. , passes it through Support vector regression(SVR), Random Forest Regression algorithms to classify soil type and predict the amount of water required for irrigation. Janani and Jebakumar[4] measure soil, plant and water data and pass it through a MLP to estimate the irrigation amount.

Irrigation Water quality Indices(IWQIs) are used for determining the quality of water and acts as a benchmark for whether the given water source can be used for irrigation or not. Meireles et al[5] developed a WQI for irrigation water which reduces thirteen parameters to 5 parameters for water samples using factor and principal component analysis. A model using Neural Networks for prediction of IWQI was developed by Sayiter et al using IWQI[6]. An indexed based approach was used by Singh et al for performing classification using multiple criteria decision analysis (MCDA) approach which indicated the importance of IWQ.It used 12 parameters and was able to classify water to five categories[7].Various dimensional reduction techniques like PCA and FA were used for reducing 22 parameters to 7 parameters, their weights being derived from PCA by Hossam et a[9].

In the proposed work, we use the IWQI developed by Meireles et al and develop a classification model to predict the IWQI class.

3 Preliminaries

3.1 Irrigation Water Quality based on Sodicity and Salinity

To study the effect of water on the health of crops and soil, it is recommended to calculate the water quality index(WQI). It represents the aggregate of physical and chemical properties of water and is calculated by multiplying weights and values of the parameters. For this project, we calculate and use the WQI for irrigation(IWQI) formulated by Meireles that is based on the effect of salt toxicity of water on crops. The given subparts describes the five parameters measured for water quality, the relative weights and IWQI formula calculation.

3.1.1 Parameters used for IWQI calculation

The Water Quality Index formulated in this project is a modified version of that given by Meireles et al. They reduced the thirteen parameters used to represent water quality to five parameters using various dimensional reduction techniques such as PCA. The reduced parameters are: Chloride ion, Sodium ion, Specific Electrical Conductivity, and SAR(also known as Sodium absorption ratio that indicates sodicity)[10]. Conductivity is calculated by passing electric current through electrodes dipped in solutions and SAR is given using equation 1 that affects the infiltration rate of water as well as soil permeability[11]. Sodicity being in excess or deficient degrades the crops and increases soil toxicity. Relative weights of the five parameters are given in Table 3.1.

$$SAR = \frac{Na^{+}}{\sqrt{\frac{Ca^{2+}+Mg^{2+}}{2}}} \tag{3.1}$$

Here,

Ca^{2+}:-Calcium ion Conc.

Na^{+}:-Sodium ion Conc.

Mg^{2+}:-Magnesium ion Conc.

3.1.2 Quality Measurement Values and IWQI

The water quality parameters’ value should not be in excess or deficient or else they may lead to salt toxicity to plants that may inhibit their growth. Also, the parameters are in different units which makes it difficult to aggregate them together. Therefore, they are converted to a normalized value between (0-100) also known as quality measurement scores(q_i). They are calculated using the formula provided in equation 2 and the limiting parameter values given in table 3.3.

Table 3.1 Relative Weights of IWQI parameters

Parameter for Quality Index	**Rel. Weights**
Specific Electrical Conductivity (EC)	0.2142
Sodium Conc. (Na^{+})	0.2003
Bicarbonate Conc. (HCO^{3-})	0.2012
Chloride Conc. (Cl^{-})	0.1954
Sodium Absorption Ratio(SAR)	0.1829

These are then multiplied with the weights provided in table 3.1 and aggregated using eqn. 3 to obtain the Water Quality Index For Irrigation(IWQI). The value of IWQI lies in the range 0-100 and divided into salinity based classes which is provided in Table 3.2.

$$q_i = q_{imax} - \frac{(x_{ij} - x_{inf}) * q_{iamp}}{x_{amp}} \tag{3.2}$$

Here,

q_i:- values for quality measurement,

q_{imax}:- maximum value for a given WQI class,

q_{iamp}:- amplitude for q_i,

x_{ij}:- value of x,

x_{inf}:- min. value of x in WQI class,

x_{amp}:- Amplitude range of class

$$IWQI = \sum_{i=1}^{n} q_i w_i \tag{3.3}$$

Here: w_i :-relative weights given in table 1

Table 3.2 Classification ranges for water based on IWQI values

IWQI range	**Soil Type**	**Plant Type**
85 – 100	Recommended for any kind of soil	No effect on most crops
70 – 85	Recommended for soil having moderate permeability	Avoid for plants having low salt tolerance capacity
55 – 70	Recommended for soil having moderate or high permeability	Avoid plants having low to meds. salt tolerance
40 – 55	Highly Permeable Soil samples	Plants having high tolerance of salinity
0 – 40	Avoid use for soils	Prevent use in plants

Table 3.3 quality measurement limiting values (qi)

q_i	**E.C**	**S.A.R**	Na^+	Cl^-	HCO^{3-}
85 - 100	0.22	2.001	2.04	1.003	1.23
60 - 85	0.715	3.02	3.08	4.01	1.48
35 - 60	1.530	6.01	6.010	7.12	4.53
0 - 35	$E.C < 0.220$ or E.C>1.530	$S.A.R < 2.001$ or S.A.R	$Na^+ < 2.04$ or Na^+	$Cl^- < 1.003$ or Cl^-	$HCO^{3-} < 1.23$ or $HCO^{3-} > 4.53$

3.2 Classification Algorithms

ANN based Classifier

Artificial Neural Networks are derived from the structure of our brain where dendrites receive the message which is passed through axon [12]. In ANN, neurons are responsible for receiving the input and producing the output after applying the activation function. Each neuron has a weight which increases or decreases as the learning proceeds. Typically, neurons are aggregated into layers. These layers transform the given input, finally producing and output which gets tested with the actual output and the error gets propagated back.

SVM Classifier

The Support Vector Machine classifier presents one of the most robust prediction methods. Its objective is to find a hyperplane with maximum margin separation which distinctly classifies data points [13].

Gradient Boost Classifier

Gradient Boosting Algorithm uses weak learners and makes changes in it to construct a strong learner [14]. Decision trees are used by this algorithm as its weak learners although it's prone to overfitting.

Random Forest Classifier

The random forest classifier uses multiple decision trees as an ensemble [15]. Each tree gives an output which contributes to the final result. The classifier is based on the idea that a group of classifiers outperforms a single classifier.

Decision Tree Classifier

A decision tree is defined as a flowchart like structure where nodes represent an if-else condition, each branch represents the outcome whereas leaf nodes represent the actual class label [16]. Here, the path from root to leaf gives us the classification rule for that class. Commonly used algorithms for splitting are: Gini impurity, Chi-Square and Information Gain.

Naive Bayes Classifier

In this classifier, we use a Bayes theorem based model but there is no dependence of relationships between the features[17].

Bagging Classifier

It uses randomized decision trees, Bagging is quite similar to the random forests but differ in how the decision trees are constructed [18]. Here each tree has a random sample of features which lead to de-correlated trees.

4 Proposed Work

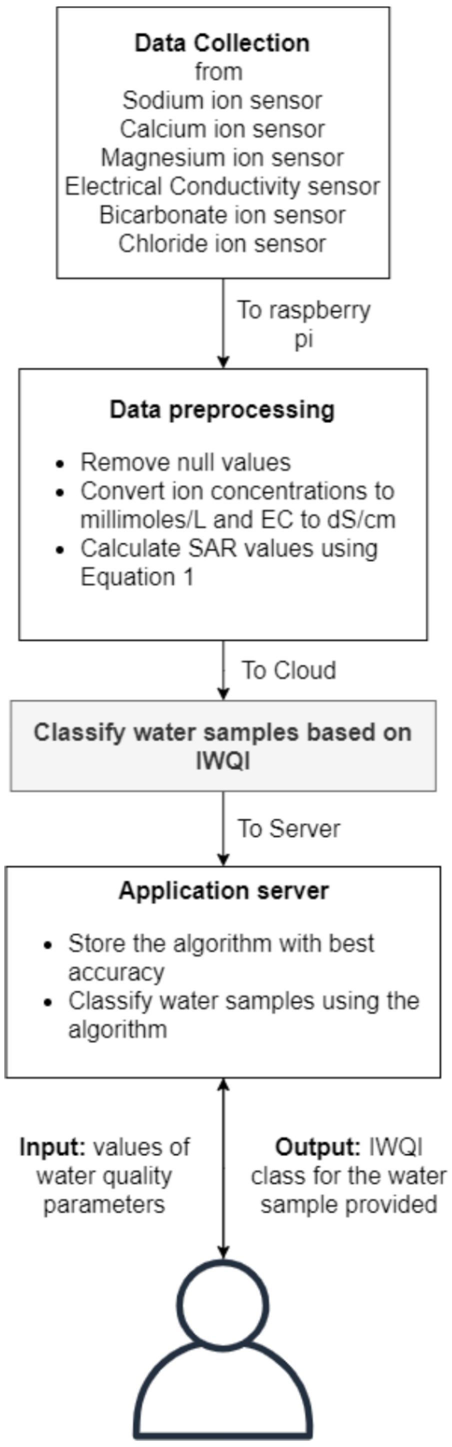


Fig 4.1 Overall architecture of the System

A classification model is proposed in this section that was described in section 3.1.2 for predicting IWQI class. For irrigation purposes, this model is capable of being used by agricultural IoT systems to quantitatively and qualitatively analyze water samples. Water quality parameters like Chloride, Sodium, Magnesium, Bicarbonate, Calcium ion concentrations with electrical conductivity can be measured from a sensor network. To the cloud after processing in a local machine like Raspberry pi which gets data from sensor network, final output is sent. Relationship of Water quality parameters, IWQI classes are analysed by the cloud using the data it obtained. Users can interact through web or mobile using the application server which has results of classification saved. Section 4.2 focusses on data analysis portion involving creation of a classification model with dataset description in section 4.1 for this work.

4.1 Dataset-Major Ions Brackwish Dataset

U.S.A Geological Survey provided us with the dataset which was utilized in the given work. Cl^-, Mg^{2+}, Ca^{2+}, EC and HCO^{3-} were only used for IWQI out of nutrient, metals and dissolved solids for IWQI from the dataset consisting of 66 thousand rows. Chloride, Bicarbonate, Sodium, Magnesium and Calcium ion concentration represents Cl^-, HCO^{3-}, Na^+, Mg^{2+} and Ca^{2+} respectively[19]. millimoles/L is obtained from mg/L in which the concentrations were measured. To reach dS/cm EC was divided by 100. To calculate SAR, Mg^{2+}, Ca^{2+} and $Na^+$are used in Equation 1.

4.2 Irrigation Water Quality Index based classification

A classification model was constructed and applied to major ions dataset for this work. The model will be constructed using the datatset which has all essential features required for IWQI described in Section 3.1.1 using three steps: Obtaining quality measurement values from given values, IWQI which is calculated by calculated values with relative weights for each parameter, according to range assigning to a class given in Table 2 and choosing the best of them out of different classification techniques

4.2.1 Quality Measurement Values

Different units with multiple parameters can be obtained from the given dataset. The quality of water becomes low if the values of parameters are significantly large or small. Therefore according to given limits for parameters the values need to be normalized. Using Equation 2 and Table 3, quality measurement values (q_i) can be obtained by converting the values. Other parameters can be combined with quality measurement value (q_i) which represents a dimensionless quantity. For the dataset used, values present in specific ranges of q_i are given by Table 4. From given below table we can see that either excess or deficit value of Na^+, Cl^- and SAR with most of the samples having optimal value of E.C.

Table 4.1 Count of Quality measurement values in each class of IWQI

Range of IWQI	$Q_{HCO^{3-}}$	Q_{EC}	Q_{Cl^-}	Q_{Na^+}	$Q_{S.A.R}$
0 - 35	5563	6356	32451	31412	32635
35 - 60	9952	19	7	5	767
60 - 85	22335	3577	6	2795	1935

85 - 100	2346	26751	5284	3832	2654

4.2.2 Calculating IWQI and developing a classification model

From Equation 4 we will obtain IWQI by the values of quality measurement obtained from the section 3.1.2 and multiplying them to their relevant relative weights which is given in Table 1. Quality ranges given in Table 2 are used to classify samples into quality classes. Pearson's Correlation from Table 5 is used to obtain a Correlation matrix with quality parameters.

Following inferences were obtained:

1. Na^{+} and EC have a very high correlation with Cl^{-} .
2. All parameters are highly correlated with Na^{+}.
3. All parameters are highly correlated with EC except SAR.
4. EC has a high correlation with HCO^{3-}.
5. Na^{+}, Cl^{-} and EC has a high correlation with IWQI.

$$r = \frac{\sum(U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum(U_i - \bar{U})^2(V_i - \bar{V})^2}} \tag{4.1}$$

Here,

r:- coefficient of correlation

U_i:- u-variable's value for the given sample

$\overline{U}$:- u-variable's mean for the given value

V_i:- v-variable's value for the given sample

$\overline{V}$:-v-variable's mean value for our sample

All combinations of parameters were tried for predicting IWQI using the classification model but Cl^{-}, Na^{+} and EC were found out to be the best. The results obtained are similar either using 3 or 5 parameters.

This implies that other reduction techniques aren't required and correlation analysis gives us excellent parameters. This will in turn help us save costs for the parameters which now aren't needed to be calculated.

Table 4.2 correlation matrix for the five parameters along with IWQI

	Cl^-	Na^+	E.C	HCO^{3-}	S.A.R	I.W.Q.I
Cl^-	1	0.467	0.506	0.159	0.083	0.44
Na^+	0.467	1	0.482	0.297	0.56	0.607
E.C	0.506	0.482	1	0.835	0.018	0.423
HCO^{3-}	0.159	0.297	0.835	1	0.018	0.293
S.A.R	0.083	0.56	0.018	0.018	1	0.302
I.W.Q.I	0.44	0.607	0.423	0.293	0.302	1

Method having best accuracy with best F2 score are to be used for predicting water quality classes from the classification algorithms given in Section 3.2. For the given dataset, to predict the best classification algorithm we use various ensemble, traditional and deep learning techniques and compare their performance with each other.

The overall algorithm is mentioned in algorithm 1 and flowchart in Fig. 2.

Input: EC, Na^+, HCO^{3-}, SAR and Cl^- are the 5 parameters from water quality dataset which are utilized.

Output: Classification algorithm with best accuracy and F2 score

1: For each parameter, quality measurement values(q_{param}) are computed.

2: From Table II classes are to be assigned and IWQI is to be computed using the Equation 3

3: Cl^-,Na^+,EC are to be chosen as features from the dataset

4: Labels of dataset are to be chosen as IWQI

5: Train test split is to be performed on the given dataset

6: 7 classifiers from Section III-B are to be applied on the training set , evaluation to be done on the testing set using the metrics accuracy and F1 score.

7: return classification algorithm with best accuracy and F2 score

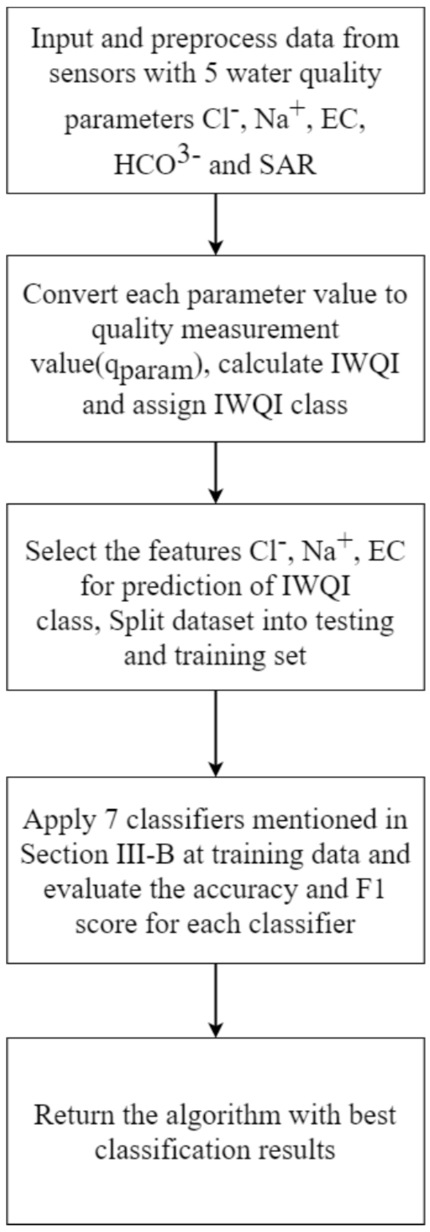


Fig 4.2 Flowchart for classification of water sample based on IWQI

5 Experiments and Results

This section describes the setup required to execute the models in section 5.1, evaluation metrics for classification techniques and results(5.3) and evaluation results(5.4) of various classification algorithms contained in our models.

5.1 Experimental Setup

The proposed models were implemented in Python3 and trained using Google Collaboratory.
Following are the details of experimental setup for the project:
Operating System: Any Unix or Windows NT based
Programming Language: Python3
Python Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

The libraries are loaded to the colab notebook using pip, for training the model GPU runtime is used for fast training and we load the dataset from the website link itself in the colaboratory so that anyone can easily run and test the code.

5.2 Hyperparameters Settings

The hyperparameter settings for the classification algorithms are provided in Table 5.1 and they parameters are described below:

Table 5.1 Hyperparameter settings for Classification Model

Classifier	**Parameter Settings**
Bagging Classifier	base_estimator:SVC, n_estimators:10
DecisionTree	criterion: gini, splitter: best, min_samples_split: 2
Naive Bayes	alpha: 1.0, class_prior: None, class_count: [7059. 15826. 5057. 2055. 402.]
Gradient Boosting	loss: deviance, learning rate: 0.1, n_estimators: 100
Random Forest	n_estimators: 100, criterion: gini, base_estimator_: DecisionTreeClassifier
SVM	C: 1.0, decision_function_shape: 'ovr', tol:0.001
MLP	hidden_layer_sizes:128, activation='relu', learning_rate=0.001

5.3 Evaluation Metrics for Classification Algorithms

The evaluation metrics used to test the algorithms used in our model are described below:

Accuracy

It can be defined as the ratio of correct predictions to the total predictions. It helps us to find out how much data is classified correctly by the given model. It's given by equation 5.1.

Confusion Matrix

It is a square matrix in which many values are classified to which class is depicted. This gives us the overall performance of the classification model and helps us to find what is wrong with the given model. It can be used to calculate various other metrics such as precision, recall, f-measure and accuracy.

Coefficient of determination

This metric is used to measure the relationship between two variables. It is also known as R squared and sometimes referred to as a goodness of fit. In simple terms, it is the ratio of explained variance to total variance. Its value is between 0 and 1 where 1 depicts perfect fit and 0 depicts that model fails to predict correct output at all.

Mean square error

It's defined as the average of the squared errors of absolute difference between predicted and true outputs. It is used to measure the quality of the estimator and the closer its value to zero, the better the predictions of the model. Though it indicates the data or not.

Mean absolute error

It is defined as the average difference of absolute errors. Like the mean square error, it indicates the accuracy of the estimator and the closer its value is to zero, the accurate the predictions of the model.

Precision

It can be defined as the ratio of accurately classified values to the total positively classified values of that class. High precision indicates that a higher proportion of positively classified data was correct. Its given by equation 5.2

Recall

It can be defined as the ratio of accurately classified values to the actual values belonging to that class. High precision indicates that a higher proportion of actually positive classes were classified as positive. It's given by equation 5.3.

F1 Score

It is a metric which captures both the relevance of precision and recall, captures features of both and puts it into a single score. It can be used to rate whether the given classifier performs better or not. It is high only when both precision and recall are high. It's given by equation 5.4.

$$Accuracy = \frac{TPos + TNeg}{TPos + TNeg + FPos + FNeg} \quad (5.1)$$

$$Precision(P) = \frac{TPos}{TPos + FPos} \quad (5.2)$$

$$Recall(R) = \frac{TPos}{TPos + FNeg} \quad (5.3)$$

$$F1Score = 2 * \frac{P * R}{P + R} \quad (5.4)$$

where,

TPos :- The observations that are correctly classified for an affirmative class

TNeg :- Observations that are correctly classified for a given -ve class

FPos :- Incorrect classifications for an affirmative class

FNeg :- Incorrect classifications for a given -ve class

5.4 Results of Model Based on IWQI Classes

With Cl^-, Na^+ and EC as features and labels as IWQI class, these things are used in algorithm 1 to obtain the given metrics in Section 5.3 whose details are given in Section 4.2.2.

Table 5.2 Result of Classification for our model

Methods	**Accuracy**	**Precision**	**Recall**	**F1**
DecisionTree	0.832	0.728	0.740	0.733
Naive Bayes	0.536	0.105	0.200	0.138
Gradient Boosting	0.858	0.762	0.757	0.760
Random Forest	**0.879**	**0.790**	**0.765**	**0.776**
SVM	0.845	0.762	0.711	0.726
Bagging	0.813	0.750	0.664	0.690
MLP	0.846	0.734	0.743	0.738

We have very less data for IWQI range 85-100 which only accounts for 1.3% of the total given samples due to which our model isn’t able to predict it correctly for the given dataset. Our model is not adversely affected though by this phenomenon. It was seen that random forest has best accuracy of 87.9% with gradient boosting having second best accuracy 85.8%. The worst accuracy of 53.6% was obtained by Naive bayes. The results are represented in Table 5.2. For five-fold cross validation the results are in table 5.3. From the results we observe that every method is capable of classifying water samples correctly except Naive Bayes.

Table 5.3 Cross Validation Scores for the model

Methods	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**
Bagging Classifier	0.829	0.823	0.814	0.830	0.813
Decision Tree	0.833	0.836	0.828	0.823	0.839
Naive Bayes	0.518	0.518	0.518	0.518	0.519
Gradient Boosting	0.862	0.866	0.869	0.859	0.862
Random Forest	**0.866**	**0.870**	**0.870**	**0.862**	**0.871**
SVM	0.853	0.849	0.846	0.841	0.848
MLP	0.847	0.853	0.857	0.847	0.849

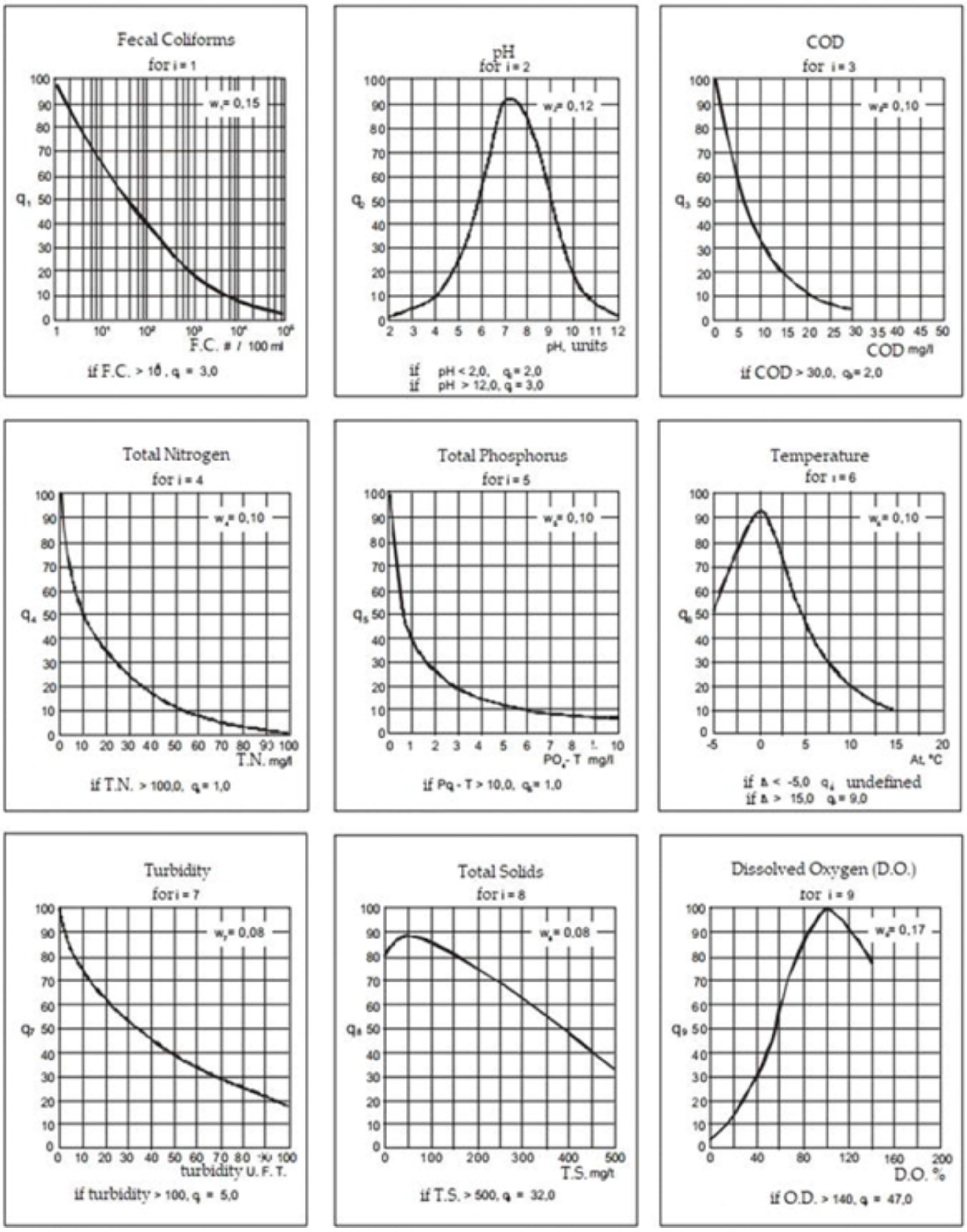
6 Conclusion and Future Work

The purpose of this project was to develop smart solutions for agriculture for reducing time, cost and resource utilization. This work mainly focuses on proper utilization of water for irrigation purposes. Data from various chemical sensors are gathered and Water Quality For Irrigation(IWQI) is obtained. A model is developed for classification based on WQI class using fewer parameters that can be obtained using cheap sensors. The 5 parameters used to calculate IWQI were reduced to Three parameters which are Sodium, Chloride and Electrical Conductivity of Water sample and are used as features for training the classification algorithm. Best performance was given by Random Forest Classification and then by Gradient Boosting Classifier.

The index developed here can be modified to accommodate other types of effects like soil acidity, granularity, temperature, precipitation etc. Here, we have only used groundwater data, which can be extended to other sources like rivers, lakes, canals etc. In future, we plan to work on other phases of agriculture such as soil preparation, crop selection, fertilizing and harvesting and developing models to analyze and assist in them. We plan to develop a mobile/web application that will provide an interface for using functionalities developed in this project. We plan to explore other possible domains of IOT such as networking, developing software for hardware like raspberry PI and Arduino and automation in Farming.

APPENDIX A

The Q-values are calculated according to "Fig. A.1"



[Figure A.1] Q values for various parameters

The following table can be used to obtain Q-values of Fecal coliform, BOD and Dissolved Oxygen

Table A.1 Q-val Conversion chart for FC, BOD and DO

Fecal Coliforms (per 100 ml)	**Q-FC**	**BOD (mg/L)**	**Q-BOD**	**Dissolved Oxygen (saturation)**	**Q-DO**
1	99	0	100	0	2
2	91	1	95	5	5
3	86	2	80	10	7
4	82	3	67	15	10
5	80	4	61	20	12
6	78	5	56	25	15
7	76	6	51	30	19
8	74	7	46	35	23

9	73	8	42	40	30
10	72	9	38	45	37
20	63	10	34	50	44
30	58	11	30	55	51
40	55	12	28	60	57
50	52	13	25	65	66
60	50	14	23	70	75
70	48	15	20	75	81
80	47	16	18	80	87
90	45	17	16	85	91
100	44	18	14	90	95
200	37	19	13	95	98

The following table can be used to obtain Q-values of pH, Temperature and Nitrate.

Table A.2 Q-Val Conversion chart for pH, Temperature and Nitrate

pH	Q	Temp. change (deg C)	Q	Nitrate (mg/L)	Q
2.4	3	-7	66	3	90
2.6	3	-6	70	4	70
2.8	4	-5	74	5	65
3	4	-4	78	6	60
3.2	5	-3	82	7	58
3.4	6	-2	85	8	56
3.6	7	-1	89	9	53
3.8	8	0	93	10	51
4	9	1	89	12	48
5.2	33	7	61	24	33
5.4	38	8	56	26	31

5.6	44	9	50	28	29
5.8	49	10	45	30	27
6	55	11	40	32	25
6.2	60	12	36	34	23
6.4	68	13	34	36	21
6.6	75	14	33	38	19
6.8	83	15	31	40	18
7	88	16	29	42	16
7.2	92	17	27	44	15
7.4	92	18	26	46	13
7.6	92	19	24	48	12
7.8	90	20	22	50	10

References

[1] R. S. Ayers, D. W. Westcot and others, Water quality for agriculture, vol. 29, Food and Agriculture Organization of the United Nations Rome, 1985.

[2] L. Breiman, "Bagging predictors," Machine learning, vol. 24, p. 123–140, 1996.

[3] L. Breiman, "Random forests," Machine learning, vol. 45, p. 5–32, 2001.

[4] V. Chinnusamy, A. Jagendorf and J.-K. Zhu, "Understanding and improving salt tolerance in plants," Crop science, vol. 45, p. 437–448, 2005.

[5] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Annals of statistics, p. 1189–1232, 2001.

[6] R. Gentili, R. Ambrosini, C. Montagnani, S. Caronni and S. Citterio, "Effect of soil pH on the growth, reproductive investment and pollen allergenicity of Ambrosia artemisiifolia L.," Frontiers in plant science, vol. 9, p. 1335, 2018.

[7] R. Gentili, R. Ambrosini, C. Montagnani, S. Caronni and S. Citterio, "Effect of Soil pH on the Growth, Reproductive Investment and Pollen Allergenicity of Ambrosia artemisiifolia L.," Frontiers in Plant Science, vol. 9, p. 1335, 2018.

[8] M. E. Grismer and K. M. Bali, "Drought Tip: Use of Saline Drain Water for Crop Production," 2015.

[9] B. Gupta and B. Huang, "Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization," International journal of genomics, vol. 2014, 2014.

[10] M. H. Hassoun and others, Fundamentals of artificial neural networks, MIT press, 1995.

[11] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," International Journal of Data Mining & Knowledge Management Process, vol. 5, p. 1, 2015.

[12] H. S. Jahin, A. S. Abuzaid and A. D. Abdellatif, "Using multivariate analysis to develop irrigation water quality index for surface water in Kafr El-Sheikh Governorate, Egypt," Environmental Technology & Innovation, vol. 17, p. 100532, 2020.

[13] D. La Mora-Orozco, H. Flores-Lopez, H. Rubio-Arias, A. Chavez-Duran, J. Ochoa-Rivero and others, "Developing a water quality index (WQI) for an irrigation dam," International journal of environmental research and public health, vol. 14, p. 439, 2017.

[14] J. L. Lerios and M. V. Villarica, "Pattern Extraction of Water Quality Prediction Using Machine Learning Algorithms of Water Reservoir," International Journal of Mechanical Engineering and Robotics Research, vol. 8, 2019.

[15] W. McKinney and others, "pandas: a foundational Python library for data analysis and statistics," Python for High Performance and Scientific Computing, vol. 14, p. 1–9, 2011.

[16] A. C. M. Meireles, E. M. d. Andrade, L. C. G. Chaves, H. Frischkorn and L. A. Crisostomo, "A new proposal of the classification of irrigation water," Revista Ciencia Agronomica, vol. 41, p. 349–357, 2010.

[17] K. P. Murphy and others, "Naive bayes classifiers," University of British Columbia, vol. 18, p. 60, 2006.

[18] W. S. Noble, "What is a support vector machine?," Nature biotechnology, vol. 24, p. 1565–1567, 2006.

[19] A. Orgiazzi, C. Ballabio, P. Panagos, A. Jones and O. Fernández-Ugalde, "LUCAS Soil, the largest expandable soil dataset for Europe: a review," European Journal of Soil Science, vol. 69, p. 140–153, 2018.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, p. 2825–2830, 2011.

[21] S. L. Qi and A. C. Harris, "Geochemical Database for the Brackish Groundwater Assessment of the United States," US Geological Survey data release, 2017.

[22] P. Shrivastava and R. Kumar, "Soil salinity: a serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation," Saudi journal of biological sciences, vol. 22, p. 123–131, 2015.

[23] S. Singh, N. C. Ghosh, S. Gurjar, G. Krishan, S. Kumar and P. Berwal, "Index-based assessment of suitability of water quality for irrigation purpose under Indian conditions," Environmental monitoring and assessment, vol. 190, p. 29, 2018.

[24] Y.-Y. Song and L. U. Ying, "Decision tree methods: applications for classification and prediction," Shanghai archives of psychiatry, vol. 27, p. 130, 2015.

[25] S. Sunder, "India economic survey 2018: Farmers gain as agriculture mechanisation speeds up, but more R&D needed," The Financial Express, 2018.

[26] S. Van Der Walt, S. C. Colbert and G. Varoquaux, "The NumPy array: a structure for efficient numerical computation," Computing in science & engineering, vol. 13, p. 22–30, 2011.

[27] A. Vij, S. Vijendra, A. Jain, S. Bajaj, A. Bassi and A. Sharma, "IoT and Machine Learning Approaches for Automation of Farm Irrigation System," Procedia Computer Science, vol. 167, p. 1250–1257, 2020.

[28] R. K. Yadav, R. Dev, T. Mann and T. Verma, "IoT Based Irrigation System," International Journal of Engineering & Technology, vol. 7, p. 130–133, 2018.

[29] S. Yıldız and C. B. Karakuş, "Estimation of irrigation water quality index with development of an optimum model: a case study," Environment, Development and Sustainability, p. 1–16, 2019.

[30] N. Zhang, M. Wang and N. Wang, "Precision agriculture—a worldwide overview," Computers and electronics in agriculture, vol. 36, p. 113–132, 2002.