The Impact of Machine Learning: Harnessing the Power of Developing Water Quality Prediction System

(IJCSNT) International Journal of Communication Systems and Network Technologies
ISSN Number: 2053-6283
Volume 14, Issue No. 02, 71–81

https://journals.mirlabs.in/index.php/ijcsnt
DOI-10.18486/ijcsnt/14.2.007
IJCSNT

R Ramya[1], K. Selvi[2], S. Pradeep[3], V. Samuthira Pandi[4], B. Geetha[5], D Shobana[6]

Abstract

Water color is influenced by sunlight interaction with water constituents, and any change in color often indicates a decline in water quality, making it unsafe for human consumption. Increasing water pollution and natural disasters such as floods and tsunamis have intensified concerns about water safety. Since contaminated water contributes to approximately 35% of global fatalities, rapid water quality assessment is essential. This study presents a Learning-based Quality Prediction System (LQPS) that predicts water quality using machine learning and image processing techniques. The system analyzes water color and key parameters such as pH, turbidity, temperature, and hardness to determine its suitability for drinking and other applications. The proposed model is developed using TensorFlow, Keras, and a Random Forest Classifier (RFC), integrated with IoT for efficient monitoring. Using images captured by mobile devices and Google Earth, the system provides a fast, cost-effective, and reliable preliminary assessment of water quality, supporting timely decision-making and improved water management.

Keywords

Machine Learning, Water Quality Prediction, Machine Learning, LQPS, Random Forest Classifier, RFC

Received: 12 April 2025; **Revised:** 19 July 2025; **Accepted:** 17 August 2025;
Published: 31 August 2025;

[1]Department of Nano Electronics Materials and Sensors, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.
[2]Department of Information Technology, R.M.K Engineering College, Chennai.
[3]Department of Electronics and Communication Engineering, K.S. Rangasamy College of Technology, Tiruchengode.
[4]Centre for Advanced Wireless Integrated Technology, Chennai Institute of Technology, Chennai, Tamil Nadu, India.
[5]Department of Department of Science & Humanities, Prince Shri Venkateshwara Padmavathy Engineering College, Chennai.
[6]Department of Electrical and Electronics Engineering, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India.

Corresponding author:
Email: ramyar.sse@saveetha.com

Introduction

There are a lot of moving parts when it comes to water quality analysis. All of water's many uses are intrinsically related to this idea[1]. Standards must vary according to demands. Predicting water quality is an area that is now seeing a great deal of research. In most cases, the physical and chemical characteristics of water that are most relevant to its use are the ones that decide its quality. Next, we need to decide what values for every variable are considered acceptable and undesirable. Water is deemed suitable for a certain purpose if it satisfies the specified criteria. The water has to be treated prior it may be utilized if it doesn't meet these conditions. There are several physical and chemical characteristics that may be used to evaluate water quality. This means that in reality, it is not feasible to characterize water quality spatially or temporally by observing the behavior of each component separately.

All forms of life depend on water as an essential resource for their continued existence. This includes humans. Ensuring sufficient water quality is of the utmost importance for the well-being of these creatures. Excessive pollution might have a devastating effect on aquatic animals' chances of survival and perhaps put their lives in jeopardy. Rivers, lakes, as well as streams are just a few examples of ambient water sources that may be evaluated according to specific quality requirements. Furthermore, water must not be too salty and must not include pollutants that harm soil and plants in order to preserve the natural equilibrium of ecosystems; this is especially true for irrigation water, which is needed for a variety of purposes. In addition, industrial processes have specific needs for water, which means that this water must have certain properties[2]. Grounds as well as surface water are two examples of naturally existing and affordable freshwater resources, although they may be polluted by human and industrial activities[3]. Aquatic creatures have a tolerance range for pollution levels. Their very survival is at danger when these limits are crossed[4]. Most of the waterways, such as lakes, rivers, and streams, meet certain standards of purity. Drinking water quality is significantly affected by the fast expansion of industry, which leads to a decrease in water quality. A variety of approaches have been proposed for water quality prediction, such as statistical techniques, visual modelling, analytical methods, and predictive algorithms. As the world continues to undergo fast transformation and meet rising needs, the significance of precise water quality predictions continues to grow[11].

Among the most valuable resources, water is essential to the survival of all living things. The health of aquatic organisms and, therefore, people who rely on them is jeopardized when water quality declines due to pollution. Keeping an eye on water quality and making sure marine life can survive is therefore of the utmost importance. In order to reduce and control water pollution, it is essential to understand water quality challenges and challenges[12]. Environmental water management programmes have been developed by various governments throughout the globe in an effort to understand the state of the marine ecosystem. Polluted water, together with inadequate sanitation and hygiene, is responsible for the annual deaths of two million people, and about one billion people do not have accessibility to safe drinking water. Therefore, it is essential to maintain the quality of the freshwater. For a diversion plan to be sustainable in the long run, water quality is paramount. Since funds have to be reallocated to fix water distribution infrastructure whenever a problem arises, water of low quality could be expensive as well[5]. In order to achieve these goals and provide affordable, safe drinking water, there has been an increase in the need for better water management as well as water quality control. In order to resolve these concerns, it is essential to conduct thorough evaluations of freshwater sources, disposal systems, and organizational monitoring practices.

Related Study

Seventy percent of Earth's surface contains water, making it a crucial resource for all forms of life[6]. Water quality has been declining at a frightening rate due to rapid industrialization and urbanization, which has resulted in terrible illnesses. The modern idea of real-time monitoring is rendered irrelevant by the standard methods of water quality estimation, which include costly and laborious lab and statistical analysis. The dire effects of subpar water quality call for a less time-consuming and cost-effective solution. This research delves into a range of supervised machine learning approaches for estimating the water quality index and the water quality class, two unique categories established by the WQI. The water quality index is a solitary metric that describes the overall water quality. Four parameters—temperature, turbidity, pH, and total dissolved solids—are utilized by the suggested technique. With an absolute mean error of 1.9642 as well as 2.7273, respectively, gradient boosting and degree 2 polynomial regressions were the most effective algorithms in predicting the WQI out of all those that were used. Alternatively, the WQC is best classified using multi-layer perception having a setup of (3, 7), which achieves an accuracy of 0.8507. With only a few parameters, the suggested technique seems to be reasonably accurate, which bodes well for its potential application in real-time water quality detecting systems[6].

There have been a number of contaminants that have been threatening water quality in recent years. As a result, the control of water pollution has made modelling and prediction of water quality extremely essential[7]. In order to forecast water quality and categorize it, this study develops sophisticated artificial intelligence algorithms. Two types of artificial neural networks—a non-linear auto-regressive neural network as well as a deep learning algorithm with long short-term memory—have been created for the purpose of WQI prediction. Furthermore, WQC forecasting has made use of three machine learning algorithms: Naive Bayes, K-nearest neighbors, and support vector machine (SVM). The created models were assessed using a number of statistical factors, and the dataset in question contains seven important parameters. The outcomes demonstrated that the suggested models are capable of producing reliable WQI predictions and water quality classifications based on increased resilience. When it came to predicting WQI values, the NARNET model outperformed the LSTM by a hair, while the SVM method had the best WQC prediction accuracy at 97.01%. In addition, during testing, both the NARNET and LSTM models were quite accurate, with the regression coefficient differing slightly between the two (RNARNET=96.17% and RLSTM=94.21%). Research of this promising nature has the potential to make a substantial impact on water management[7].

Water quality metrics modeled are crucial to any study of aquatic systems. Conventional modelling approaches typically include laborious procedures and rely on datasets with massive amounts of input data that is either unknown or poorly described[8]. The use of AI results in a mathematical framework that is both flexible and able to detect complicated and non-linear correlations among input as well as output data. Various human and developmental activities have severely degraded the Johor River Basin. Thus, it is crucial to establish a water quality forecasting system as a potent tool for improved water resource management. Multi-Layer Perceptron Neural Networks, Adaptive Neuro-Fuzzy Inference Systems, and Radial Basis Function Neural Networks are among the utilized modelling methodologies. The data collected from experiments and monitoring stations might be contaminated by noise signals due to both random and systematic mistakes. An accurate forecast is challenging to achieve because of data noise. As a result, a method that relies on past data of the water quality indicator and is enhanced with wavelets has been suggested, and it is based on a Neuro-Fuzzy Inference System. Ammoniacal nitrogen, suspended solids, and pH are the key water quality indicators in the study area. Three assessment procedures or evaluation methods have been employed to gauge the effects on the model. Partitioning the weights of the connections in the neural network determines the relevance of each input parameter in the network, which is the first step in the assessment process. To contrast, the first two evaluation steps determine the best input for building models with a single parameter, while the last two steps provide the best input for building models with a combination of parameters. Scenarios 1 as well as Scenario 2 were presented throughout these proceedings. One scenario builds a model to forecast water quality variables at each station, whereas the other builds a model using an estimate of the same variable at the station upstream. The twelve input factors determine both possibilities. We validated WDT-ANFIS using field data from 2009 to 2010. Compared to all of the suggested models, the WDT-ANFIS model significantly improved the accuracy of water quality parameter predictions. Additionally, all water quality measures at all stations showed significant improvements between 0.5 and 5 percent under Scenario 2, making it the superior performance compared to Scenario 1. After the suggested model was validated, it was discovered that all water quality metrics could be accurately predicted by the model (R2 values more than 0.9)[8].

Agriculture and industry cannot function without freshwater[9]. As a first step in managing freshwater resources, water quality testing is essential. Many people, particularly pregnant women and children, are becoming ill or even dying as a result of the lack of access to clean drinking water, according to the annual report from the World Health Organization. Whether you're going to use the water for drinking, chemical spraying (Pesticides, etc.), or watering animals, you must always check the water's quality beforehand. One way to ensure that the water you consume is safe is to get it tested. In light of this, proper water monitoring is fundamental to ensuring that water is safe, clean, and sterile. In order to determine if water is safe to drink, to detect disease outbreaks, to validate methods and preventative measures, and to examine the legal operation of water sources, water testing is essential. One measure of water's suitability for a given use is its quality, which takes into account the physical, chemical, as well as biological properties of the water[9].

In the modern world, water quality monitoring is a must-have strategy for survival[10]. Many modern problems with water quality assessments are the product of human activity, including urbanization, industry, farming methods, and other forms of agriculture. Processing water quality analyses isn't easy; it involves going to the water collection station by hand, collecting water samples, analyzing them in a lab, entering the results into a database, etc. To overcome these obstacles, we will employ technologies that are based on artificial learning models. There is a wide range of ways to developing and implementing machine learning algorithms for water quality analysis. This survey includes all available water sources, including groundwater, ponds, and lakes, along with rivers. Future research will be based on the survey results[10].

Methodology

An integral part of studying water pollution in order to learn about water ecological preservation is water quality forecasting, which is a crucial approach to water planning, management, and monitoring. Consequently, improving a practical and realistic approach to water quality prediction is of the utmost importance. In order to avoid unexpected changes in water quality as well as offering remedies, it is crucial to predict water quality in the future. Therefore, accurate predictions of future water quality updates could potentially more than only guarantee the safety of people's drinking water; they may also direct fishing efforts and protect biodiversity. The following figure Fig.1 shows the flow diagram of the proposed system.

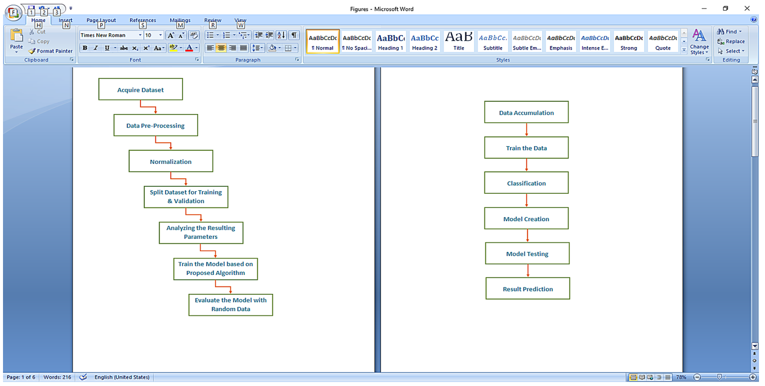

Figure 1. Flow Diagram

Predicting future water quality is essential for protecting existing water supplies and finding long-term solutions to water scarcity. To safeguard water resources in a particular region from contamination, it is necessary to have a scientific basis, and an accurate depiction of the quality of water can represent the current and future trends of water pollution. When compared to the current technique Random Forest Classifier (RFC), the suggested system Learning based Quality Prediction System (LQPS) can reliably predict the water quality. The following figure Fig.2 describes the proposed system design.

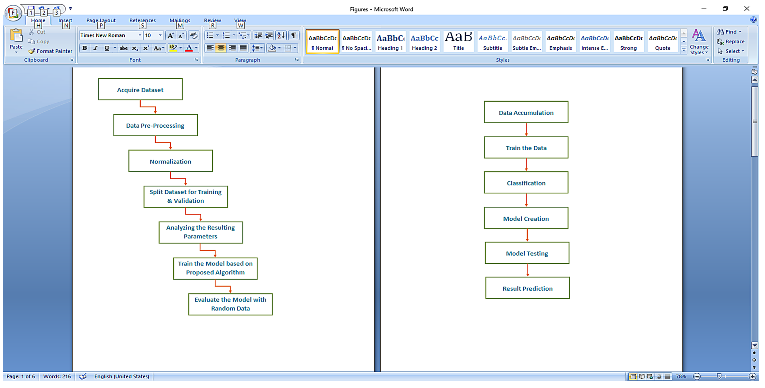

Figure 2. System Design

(i) Issue Identification: The current objective is to identify the issue statement. Making accurate water quality predictions using a learning-based quality prediction system is the current obstacle.
(ii) Data Extraction: Gathering information from one or more sources in order to store it, analyze it, or process it elsewhere is what data extraction is all about.

(iii) Pre-processing the Data: Data processing is a key component in enhancing the data analysis quality. What we call "data processing" actually involves collecting and transforming different parts of data into something more valuable and applicable.
(iv) Fixing Missing Values: There are a number of ways to fill in blanks in data. The most common approach to dealing with numeric column missing values is using means.
(v) Water Quality Index: This comprehensive metric takes a number of elements into account when determining water quality.
(vi) Data Visualization: It is presenting data visually so that trends, patterns, and correlations may be more easily seen.

There are a lot of competing clients with increasing demands, so the executives' productive water is a limited asset. Understanding the interest in water accessibility from several sources along with water quality is necessary for adjusting between the water market's interests. Water quality has been deteriorating at an alarming rate due to rapid development. One of the primary factors contributing to the alarming rise in disease rates is the more helpless water quality consequences.

Results and Discussions

Eleven water quality factors were used to determine the general capacity of the water's quality for each dataset. Aromatic carbon dioxide, electrical conductivity, pH, chlorine compounds, sulfuric acid, chemicals and roughness were some of the metrics that were tested. The selection of parameters was influenced by the fact that they are all important metrics with clearly defined and often observed water quality standards. The comprehensive prediction models presented in the research, on the other hand, are parameter number independent. The following figure, Fig-3 represents the accuracy evaluations of the proposed scheme called Learning based Quality Prediction System (LQPS), in which it is cross-validated with the conventional learning model called Random Forest Classifier (RFC) to evaluate the efficiency of the proposed scheme in clear manner. The same is depicted over the following table, Table-1 in descriptive manner.

Table 1. Accuracy Analysis

Epochs	RFC (%)	LQPS (%)
100	91.26	97.24
125	91.34	97.31
150	91.28	97.76
175	91.37	97.84
200	91.38	97.47
225	91.41	97.82
250	91.43	97.92
275	91.46	98.02
300	91.49	98.12

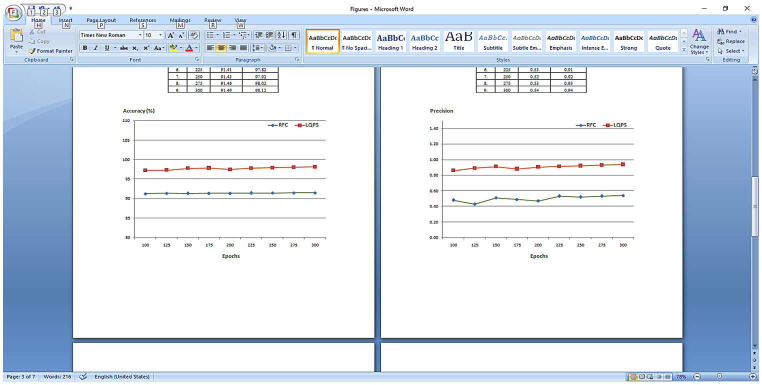

Figure 3. Accuracy Analysis

The following figure, Fig-4 represents the precision ratio of the proposed scheme called LQPS, in which it is cross-validated with the conventional learning model called RFC to evaluate the efficiency of the proposed scheme in clear manner. The same is depicted over the following table, Table-2 in descriptive manner.

Table 2. Precision Analysis

Epochs	RFC	LQPS
100	0.48	0.86
125	0.43	0.89
150	0.51	0.91
175	0.49	0.88
200	0.47	0.91
225	0.53	0.91
250	0.52	0.92
275	0.53	0.93
300	0.54	0.94

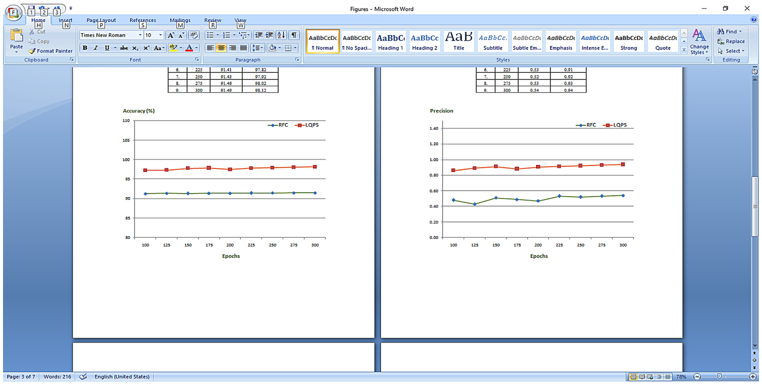

Figure 4. Precision Analysis

The following figure, Fig-5 represents the recall ratio of the proposed scheme called LQPS, in which it is cross-validated with the conventional learning model called RFC to evaluate the efficiency of the proposed scheme in clear manner. The same is depicted over the following table, Table-3 in descriptive manner.

Table 3. Recall

Epochs	RFC	LQPS
100	0.56	0.74
125	0.51	0.78
150	0.53	0.79
175	0.59	0.76
200	0.61	0.71
225	0.58	0.75
250	0.61	0.73
275	0.62	0.73
300	0.64	0.72

The following figure, Fig-6 represents the F1-Score of the proposed scheme called LQPS, in which it is cross-validated with the conventional learning model called RFC to evaluate the efficiency of the proposed scheme in clear manner. The same is depicted over the following table, Table-4 in descriptive manner.

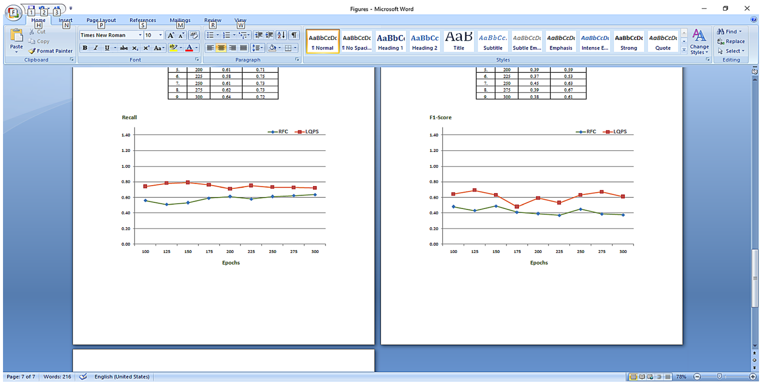

Figure 5. Recall

Table 4. F1-Score

Epochs	RFC	LQPS
100	0.48	0.64
125	0.43	0.69
150	0.49	0.63
175	0.41	0.48
200	0.39	0.59
225	0.37	0.53
250	0.45	0.63
275	0.39	0.67
300	0.38	0.61

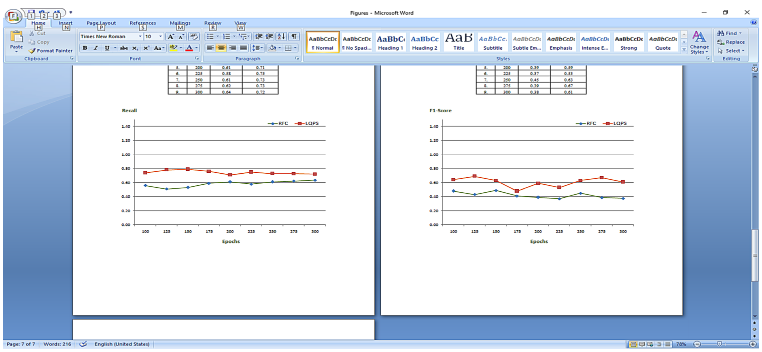

Figure 6. F1-Score

Conclusion

An essential resource for life, portability assesses the standard of water. Lab analyses were once the only way to determine water quality, but they were also very time-consuming and costly. This research investigated a different machine learning approach to water quality prediction utilizing just a handful of basic water quality variables. A collection of typical supervised machine learning methods was employed for the estimation. It would alert the proper authorities if it detected water of poor quality prior to its release for consumption. Hopefully, fewer people will drink water that isn't up to par, which would cut down on cases of typhoid and diarrhoea. Management and decision-makers in the future would benefit from future capabilities that are the product of a predictive evaluation based on anticipated values.

References

[1] Md. Saikat Islam Khan, Nazrul Islam, et al., "Water quality prediction and classification based on principal component regression and gradient boosting classifier approach", Journal of King Saud University - Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2021.06.003, 2022.

[2] Ahmed A. S. et al "Estimation of Groundwater Quality in Arid Region or Semiarid Region by Using Statistical Methods and Geographical Information System Technique", International Journal of Photoenergy, https://doi.org/10.1155/2022/7902301

[3] Xijuan Wu, Qiang Zhang, et al., "A Water Quality Prediction Model Based on Multi-Task Deep Learning: A Case Study of the Yellow River, China", Water, https://doi.org/10.3390/w14213408, 2022.

[4] Mohamed SaadAhmed et al "Investigation of Groundwater Hydro-Geochemistry, Excellence, and Anthropoid Wellbeing Hazard in Dry Zones Using the Chemometric Method" Advances in Materials Science and Engineering, Volume 2022, https://doi.org/10.1155/2022/4903323

[5] Jitha P Nair and M S Vijaya, "River Water Quality Prediction and index classification using Machine Learning", Journal of Physics: Conference Series, DOI:10.1088/1742-6596/2325/1/012011, 2022.

[6] Umair Ahmed, Rafia Mumtaz, et al., "Efficient Water Quality Prediction Using Supervised Machine Learning", Water, https://doi.org/10.3390/w11112210, 2019.

[7] Theyazn H.H Aldhyani, Mohammed Al-Yaari, et al., "Water Quality Prediction Using Artificial Intelligence Algorithms", Applied Bionics and Biomechanics, DOI:10.1155/2020/6659314, 2020.

[8] Boobalan, S., Das, S., Pandi, V.S., Swain, K.P., Palai, G, "Generation of multiple signals using single photonic structure at visible regime: a proposal to realize the harmonic generation", Optical and Quantum Electronics, 2021, 53(8), 463.

[9] Samuthira Pandi.V et al., (2021), "A spectral-effcient 1 Tbps terrestrial free- space optics link based on super-channel transmission", Optical and Quantum Electronics, Springer, ISSN - 03068919, 1572817X, Vol.53, Issue 252, Impact factor 1.842

[10] K. Kalaivanan and J. Vellingiri, “Survival Study on Different Water Quality Prediction Methods Using Machine ”, An International Quarterly Scientific Journal, iginal Research Paper https://doi.org/10.46488/NEPT.2022.v21i03.032, 2022.

[11] Mahmoud Y. Shams, Ahmed M. Elshewey, et al., “Water quality prediction using machine learning models based on grid search method”, Multimed Tools Appl, https://doi.org/10.1007/s11042-023-16737-4, 2023.

[12] Sangamesh Sirsgi, Anoopkumar Elia, NandKishore Rao, Sanjay Patil, “Design and Fabrication of Battery Operated Paddy Transplanting Machine”, International Journal of Communication Systems and Network Technologies, Vol.10, No.2, pp.103-123, 2021. DOI- 10.18486/ijcsnt/10.2.133

[13] K. K. Arora, G. S. Tomar and S. S. Agrawal, “Studying the Role of Data Quality on Statistical and Neural Machine Translation,” *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, Bhopal, India, 2021, pp. 199-204, doi: 10.1109/CSNT51715.2021.9509604.

[14] Yogalakshmi S and Mahalakshmi A, “Efficient Water Quality Prediction for Indian Rivers Using Machine Learning”, Asian Journal of Applied Science and Technology, DOI:10.38177/ajast.2021.5111, 2021.