INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 2)
Available online at: www.ijariit.com

Intelligent model for predicting water quality

Ashwini K.
ashwinikrishnan06@gmail.com
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

J. Janice Vedha
vedhajanice@gmail.com
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

D. Diviya
diviyadurai@gmail.com
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

M. Deva Priya
m.devapriya@skct.edu.in
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

ABSTRACT

Over the decades, water pollution has been a real threat to the living species. The real-time monitoring of drinking water is nothing less than a challenging task. This paper aims to design and develop a low-cost system for the real-time monitoring of water quality using the Internet of Things (IoT) and Machine Learning (ML). The physical and chemical parameters of water such as temperature, level, moisture, humidity and visibility are measured using respective sensors. ESP8266, the core controller is employed to process the measured values from the sensors. The data acquired from Sensors are sent to the Django server. Random Forest (RF) and K-Nearest Neighbours (KNN) algorithm are used in the analysis and prediction of water quality.

Keywords— *Water quality, Internet of Things (IoT), Machine Learning (ML), K-Nearest Neighbours (KNN), Random forest*

1. INTRODUCTION

Water has become a vital resource due to the increase in population and scarcity. Though about 75% of earth's surface is covered with water, the level of freshwater is getting reduced due to various reasons (Koditala 2018) like an increase in global industrial wastes and over-utilization of land and sea resources. The quality of water available to people has decreased greatly. Nitrogen and Phosphorus are the major water pollutants.

Water quality monitoring techniques can be divided into physical and chemical analysis methods and biological monitoring methods.

- **Physical analysis methods:** It is used for finding transparency, colour, temperature, turbidity and odour
- **Chemical analysis methods:** It is used for analysing Electrical Conductivity (EC), pH, Total Solids (TS), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Fluorides, Total Dissolved Solids (TDS), Total Suspended Solids (TSS), Total Hardness, Calcium Hardness, Magnesium Hardness, Nitrates, Phosphates, Sulphates, Chlorides, Dissolved Oxygen (DO), Free Carbon-di-oxide, Potassium and Sodium
- **Biological Monitoring Methods**: It involves the qualitative analysis of planktons (Zooplankton and Phytoplankton)

Presently, water quality monitoring has become a difficult task because of global warming and an increase in population (Daigavane and Gaikwad 2017). To deal with this issue, it is necessary to develop a better system that monitors water quality parameters. Turbidity is the measure of water purity obtained by analysing the number of particles in the water. If the turbidity is high, there are chances for the increased presence of impurities that may cause bacterial diseases like cholera and typhoid. The temperature of water increases with an increase in turbidity. The water quality is predicted based on turbidity, level, moisture content and flow of water.

2. RELATED WORK

Fault diagnosis using Wireless Sensor Networks (WSNs) is one of the current research areas in the field of water quality monitoring (Liu et al. 2018). Achieving high fault diagnostic accuracy of water quality monitoring and controlling is a challenging task. A hybrid water quality monitoring device fault diagnosis model based on Multiclass Support Vector Machines (MSVM) is proposed. It is seen that the prediction method based on RBDT-MSVM is effective and feasible.

Yue &Ying (2011) have designed a water quality monitoring system using WSN powered by solar panel. The prototype system is designed and implemented using nodes powered by solar panels. Data from various sensors including pH, turbidity and oxygen density are sent to the Base Station (BS). The system has advantages such as low carbon emission and power consumption and flexible deployment.

Verma (2012) have proposed a water surveillance system to control the level of contamination. In India, manual water

quality surveillance methods dramatically exacerbate water quality deterioration. This paper gives the requirements and suitability of WSN in water quality management. As WSNs support real-time, continuous and dynamic measurement, they can act as an early warning system and trigger an appropriate alarm in hazardous situations.

O' Flyrm et al (2007) have proposed a 'Smart Coast' multi-sensor system for water quality monitoring. It aims to provide a platform capable of meeting the monitoring parameters of the Water Framework Directive (WFD). The main parameters include temperature, pH, turbidity, phosphate, dissolved oxygen, conductivity and water level. The developed WSN platform offers various capabilities.

Menon et al (2012) have developed a quality monitoring system for river water using WSN. The proposed system mainly comprises of a signal conditioning, processing, wireless communication and the power module. The pH value sensed is transmitted to the BS through ZigBee communication using signal conditioning and processing techniques.

Yuan et al (2018) have used computer image processing technology and computer vision to analyse the fish behaviour in real-time and predict the water quality. The sensors monitor the movement velocity, rotation angle, spatial standard deviation and body colour which characterize the behaviour changes of the fish. Long Short-Term Memory (LSTM) neural network is used to classify the parameters of the fish and predict the pollution of water**.**

Wu et al (2017) have developed an underwater robotic dolphin which has in-built sensors to monitor water quality. The dolphin's pressure shell is made of Polyformaldehyde (POM) and double dynamic seals of Glyd rings. It uses a Central Pattern Generator (CPG) based controller for seamless propulsion of dolphin. Based on both laboratory and field experiments, self-propelled and self-contained robotic dolphins are well-suited for aquatic mobile sensing.

Jingmeng et al (2011) have developed a system that uses fuzzy mathematics comprehensive evaluation model and spatial cluster analysis to predict water quality based on the following evaluation parameters: Dissolved Oxygen, Total Nitrogen (TN), Total Carbon (TC), Chemical Oxygen Demand (COD), Ammonia Nitrogen, Nitrate Nitrogen and Total Phosphorus (TP). The data collected through various sensors are pre-processed and analysed using cluster analysis method.

Vijayakumar et al (2016) have proposed a real-time monitoring system for predicting water quality. It aims to provide a sensor-based system that introduces cloud computing architecture into IoT to make the sensor data obtainable worldwide.

Kedia (2015) has proposed a water quality monitoring model for rural areas. It is a sensor cloud-based economical project. It includes embedded sensor systems and deals with the challenges and economic activity of the system involving Government and mobile network operators. The system contacts the Government directly to take actions based on the seriousness of the quality issue.

Kumar et al (2014) have proposed a Solar-based advanced water quality monitoring system. It aims to provide energy to sensors in a WSN.

3. DECISION SUPPORT SYSTEM (DSS)

Currently, random samples are collected at various locations every week/ month and analysed. This method is inefficient, as water samples cannot be concurrently collected from all areas. Human intervention is required to monitor the quality of water. This method is not suitable for highly populated countries like India and China as they are costly (Koditala and Pandey 2018)**.**

The conventional role of current WSN systems is often limited to data acquisition and transmission. However, the increasing computational capabilities of wireless embedded technologies create exciting opportunities to close the gap between sensing and processing.

The system predicts the time to water the plants and the amount of needed (Viani et al 2017). Further, Fuzzy Logic (FL) which is used involves longer run time for system and real-time responses offer restricted usage for input variables and demands more fuzzy grades to provide accuracy.

The existing systems either collect or analyse the data, but do not include a mechanism to simultaneously collect and analyse data in real-time.

4. INTELLIGENT PREDICTION MODEL (IPM)

An Intelligent Prediction Model (IPM) is proposed, wherein sensors are deployed to measure the parameters of water. As Arduino UNO and Raspberry Pi are heavy computing devices, NodeMCU is used in the proposed model as it is lightweight and also cost-effective. The data from sensors are passed to the NodeMCU microcontroller which has an inbuilt Wi-Fi module.

The sensor data are sent to the Django server. Machine learning algorithms are involved to analyse the water's parameters and classify into different categories based on their purpose of use - drinking, industrial, irrigation, domestic and dirty (not usable). This system is completely automated and does not demand any human intervention (Koditala & Pandey 2018)**.**

The quality and the purpose of water taken as a sample are predicted using Machine Learning (ML) algorithms namely Random Forest (RF) and K-Nearest Neighbour (KNN).

Fig. 1 shows the modules in the proposed Intelligent Prediction Model (IPM).

4.1 Pre-processing

The water samples are taken as input and the data collected from the sensors are passed to the Django server. The data is frequently collected and stored in the server. Django server simultaneously picks the data and pre-processes them using Label Encoding (LE) and Normalization. In LE, data is classified into labels for prediction of output. In Normalization, the values are normalized in a certain pattern. From the collected data, a model is prepared. The model is trained based on a certainly trained dataset using supervised learning.

4.2 Classification

Both structured as well as unstructured data can be classified. Classification is a technique where data is categorized into classes by predicting the class labels/categories of the data. In this paper, classification is done using RF and KNN to predict the purpose of the water sample.

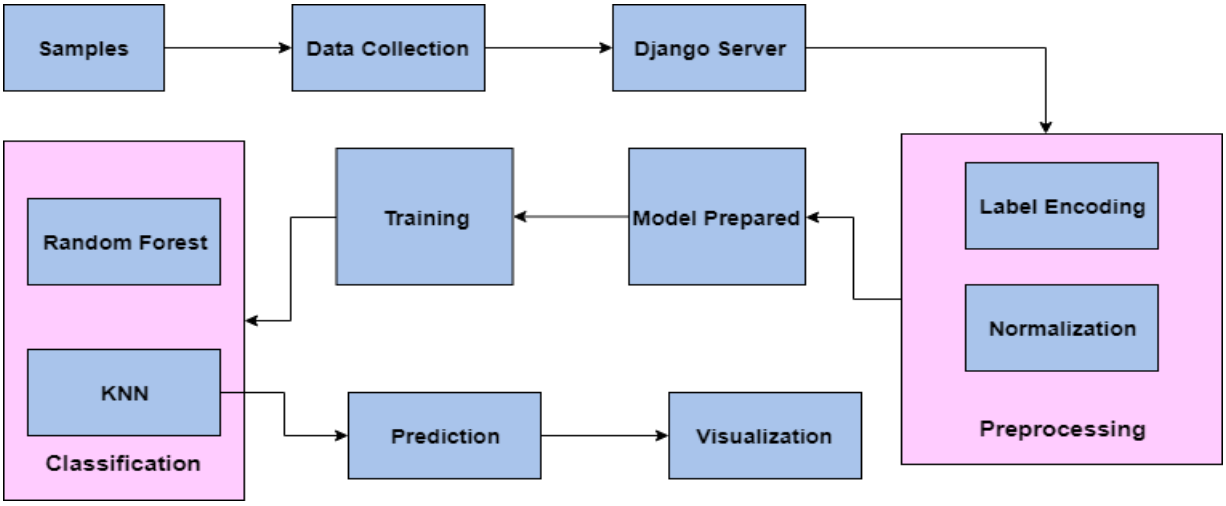

Fig. 1: Intelligent Prediction Model (IPM)

4.3 Random Forest (RF)

Random Forest (RF) algorithm consists of an ensemble of simple tree predictors, each capable of producing a response when presented with a set of predictor values. The response may be a class membership for classification problems, which associates or classifies a set of independent predictor values with one of the categories present in the variable.

It consists of an arbitrary number of simple trees which determine the final outcome. The ensemble of simple trees votes for the most popular class in case of classification problems. The use of tree ensembles leads to significant improvement in the prediction accuracy.

Each tree response depends on a set of predictor values chosen independently (with replacement) and the distribution of trees in the forest, which is a subset of the predictor values of the original data set. The subset optimal size for predictor variables is given by $\log_2 M+1$, where 'M' is the number of inputs.

The predictions of the RF are taken as the average of predictions of the trees.

Random Forest Prediction:

$$p = \frac{1}{k}\sum_{k=1}^{k} k^{th} \text{ tree response} \quad (1)$$

Where 'k' runs over the individual trees in the forest.

RFs can flexibly incorporate missing data in the predictor variables. If at a particular point in the sequence of trees, a predictor variable is selected at the root node for which some cases have no valid data, then the prediction for those cases is simply based on the overall mean at the root node. Hence, it is neither necessary to eliminate cases with predictors have missing data, nor compute split statistics.

4.4 K-Nearest Neighbours

To classify objects based on closest training examples in the feature space, the KNN algorithm is used. KNN algorithm is a type of lazy learning where the computation is ranked until classification and the function is locally approximated. It is the fundamental classification technique used when there is no knowledge of data.

It considers the training set as a whole during learning and assigns a class to each query represented by the majority label of its KNNs in the training set. The Nearest Neighbor (NN) rule is the straight forward form of KNN when K=1.

The samples are classified based on the surrounding samples. Thus, if the classification of a sample is not known, it could be predicted by using the classification of the NN samples. The performance of a KNN classifier is determined by the choice of 'K' and by the distance metric applied.

4.5 K-Nearest Neighbour Predictions

After gathering the values of 'K', predictions are made based on the KNN examples. In regression, KNN prediction is taken as the average of its outcome:

$$y = \frac{1}{k}\sum_{i=1}^{k} y_i \quad (2)$$

Where,
y_i: i^{th} case of the examples sample
y: Prediction (outcome) of the query point

In classification problems, KNN predictions are based on a voting model where the winner is used to label the query.

Thus, KNN analysis is discussed without giving any attention to the distance between 'K' nearest examples to the query point. In other words, K-neighbours have a uniform influence on predictions irrespective of the relative distance from the query point. An alternative approach is to use arbitrarily large values of 'K' with more importance given to cases closest to the query point. This would be achieved by using 'distance weighting'.

4.6 Distance Weighting

KNN indications are based on instinctive speculation that objects closer in distance are similar. It aids in discriminating between the KNNs in making predictions. This is achieved by proposing a set of weights 'W', one for each NN defined by the relative closeness of each neighbour.

$$W(x, p_i) = \frac{\exp(-D(x, p_i))}{\sum_{i=1}^{k} \exp(-D(x, p_{i}))} \quad (3)$$

Where,
D(x, p): Distance between the query point 'x' and the i^{th} case 'p_i' of the example sample

The weights defined above will satisfy:

$$\sum_{i=1}^{k} W(X_0, X_i) = 1 \quad (4)$$

Thus, for regression problems,

$$y = \sum_{i=1}^{k} W(X_0, X_i)\, y_i \qquad (5)$$

For classification problems, the maximum value of the equation is considered for each class variable.

Thus, when K >1, the Standard Deviation (SD) for prediction is given by,

$$\text{errbar} = \mp\sqrt{\frac{1}{K-1}\sum_{i=1}^{k}(y - y_i)^2} \qquad (6)$$

KNN algorithm has several advantages which include simplicity, effectiveness, intuitiveness and competitive classification performance. It is robust to noisy training.

Further, the analysis is done using Scikit-learn, an ML library in Python. Finally, the result is visualized to depict the nature of the water.

5. IMPLEMENTATION

The core controller collects the sensor values, processes them and transfers the data through the internet. Node MCU is used as a core controller. The sensor data is transferred using the Wi-Fi system. The block diagram shows the connection between the temperature sensor, level sensor, moisture sensor and the LDR sensor with the core controller.

Figure 2 shows the hardware setup. Figure 3 shows the connection diagram.

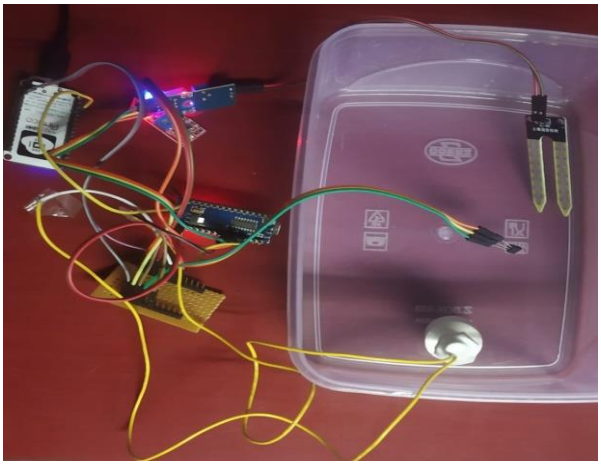

Fig. 2: Hardware Setup

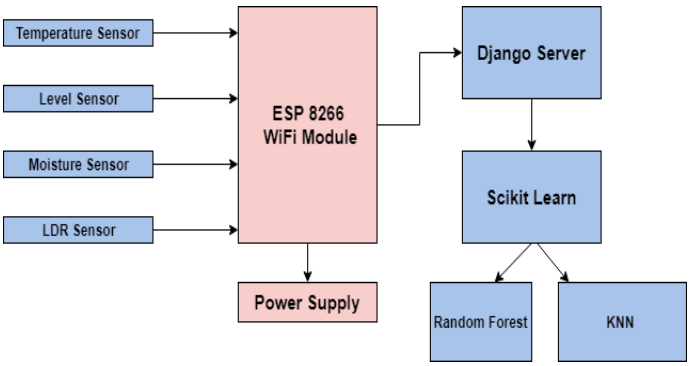

Fig. 3: Connection Diagram

5.1 LM35

To sense the external temperature, the LM35 module is used. It involves a temperature range of 55℃ to 150℃ degrees centigrade with an accuracy of 0.5℃. It operates between 3V to 30V voltage range. It gives the temperature of the given water sample.

5.2 Moisture sensor

A soil moisture sensor is used to measure the amount of water contained in a material, such as soil on a volumetric basis. For accurate measurement, a soil temperature sensor is used for calibration. It is used for detecting salt content in the water sample.

5.3 LDR sensor

A light-dependent resistor sensor is a light-controlled variable resistor. It is used to analyse the presence of impurities in the water by measuring the light intensity. The water quality is decided by the number of impurities present in the water. LDRs are made from semiconductor materials. It gives the visibility of the water sample, which if clear, means that it is not contaminated.

5.4 Level sensor

To measure the level of liquids in reservoirs or deep tanks, hydrostatic pressure level sensors are used. By measuring the level of water, their usage can be predicted. It is used to give the level of the sample.

5.5 NodeMCU

NodeMCU is a microcontroller with an inbuilt Wi-Fi module. It requires 3.3V current. It has 17 GPIO pins and 1 analog pin. It supports 802.11 b/g/n. It is economical and efficient when compared to other microcontrollers. It is an open source IoT platform. It has a firmware which runs on the ESP8266 Wi-Fi SoC from Espressif Systems, and hardware which is based on the ESP-12 module.

The data collected is transferred to Django which is a high-level Python Web framework that encourages rapid development and clean pragmatic design. It is imported for further analysis.

5.6 Scikit-learn

Scikit-learn is a free software ML library which is used for implementing ML algorithms including KNN and RF. The dataset and CSV files are trained to give appropriate prediction results for the given water sample.

6. RESULT AND DISCUSSION

In this project, the data collected by the sensors are used to determine the contamination of water and the results are analysed. The implementation cost of the system is minimal. Moreover, efficient algorithms are also being used to coordinate the information collected by the sensors and hence there is no data loss.

The two phases of classification are training and testing.

- **Training:** It is the process of learning to the label from the examples. The training process can be either supervised or unsupervised. Here, the supervised mode is used for training.
- **Testing:** It is the process of checking how well the classifier has learnt to label the unseen samples.

The classification efficiencies of the algorithms are shown in Table 1.

Table 1: Diagnostic Accuracy

Classifier	Diagnostic Accuracy (%)
Random Forest (RF)	96
K-Nearest Neighbours (KNN)	97

The graphs (figure 4 and figure 5) show the number of samples categorised based on the purpose of water as drinking water, water for industrial, agriculture and home usage or dirty water.

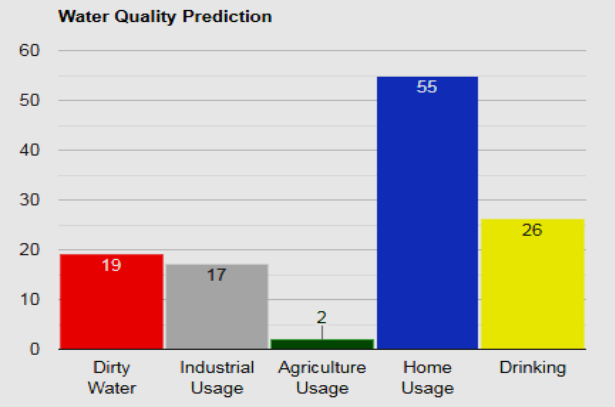

Fig. 4: Analysis using RF

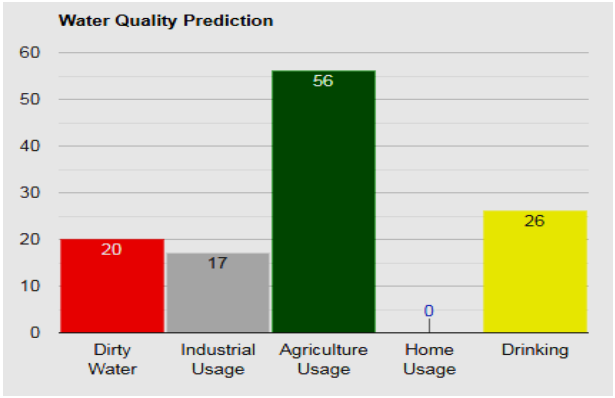

Fig. 5: Analysis by KNN

Figure 6 to figure 9 show the Visibility, Solidity, levels and Temperature of the proposed scheme.

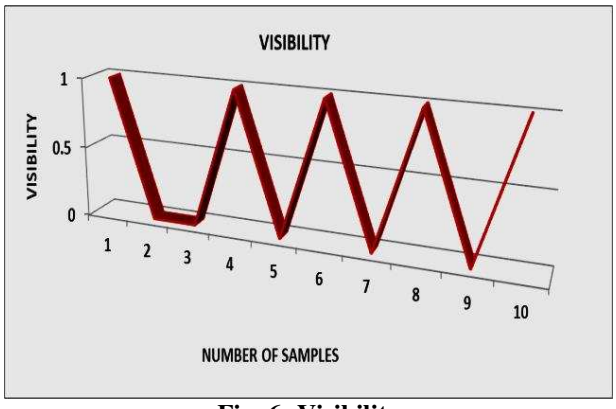

Fig. 6: Visibility

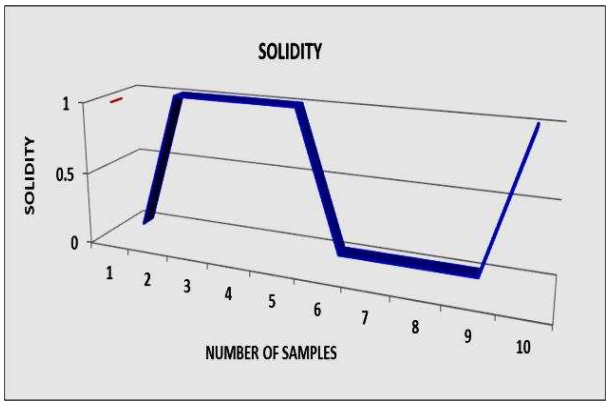

Fig. 7: Solidity

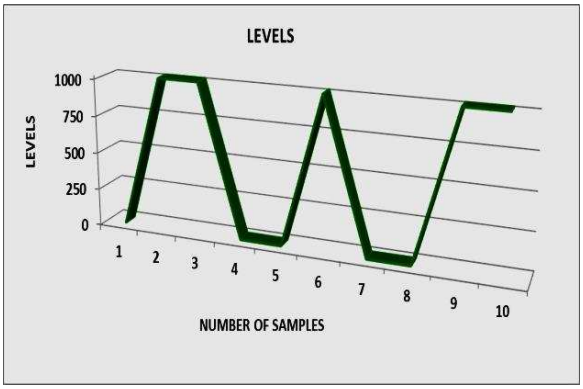

Fig. 8: Levels

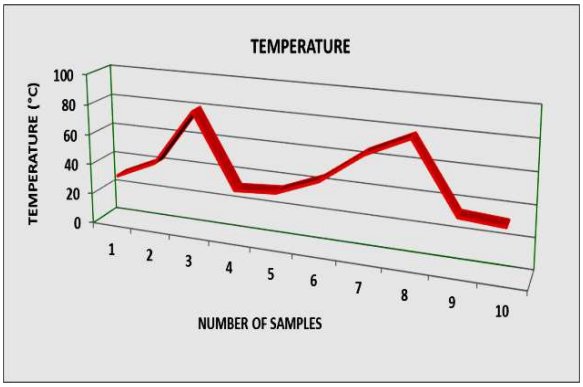

Fig. 9: Temperature

7. CONCLUSION

The proposed Intelligent Prediction Model (IPM) delivers an economical and practical solution to monitor the quality of water without any human intervention. To solve the water quality issues, this system uses various technologies such as the Internet of Things (IoT) and Machine Learning (ML). The problems of human survival are dealt with a certain extent. The existing system could be improved by merging the Computer Image Processing (CIP) technologies to improve accuracy, applicability and reliability of the system. In addition, water quality parameter management and decision support systems can be used in sewage treatment plants and in units of water quality prediction and management.

8. REFERENCES

[1] Koditala, N. K., & Pandey, P. S., Water Quality Monitoring System using IoT and Machine Learning, in Proceedings of the IEEE International Conference on Research in Intelligent and Computing in Engineering, pp. 1-5, 2018.

[2] Daigavane V, & Gaikwad, M. A., Water Quality Monitoring System Based on IoT, Advances in Wireless and Mobile Communications, 2017.

[3] Liu, S., Xu, L., Li, Q., Zhao, X., & Li, D., Fault Diagnosis of Water Quality Monitoring Devices Based on Multiclass Support Vector Machines and Rule-Based Decision Trees, IEEE Access, Vol. 6, pp. 22184-22195, 2018.

[4] Yue, R., & Ying, T., A water quality monitoring system based on wireless sensor network & solar power supply, in Proceedings of the IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems, pp. 126-129, 2011.

[5] Verma, S., Wireless Sensor Network application for water quality monitoring in India, in Proceedings of the IEEE National Conference on Computing and Communication Systems, pp. 1-5, 2012.

[6] O' Flynn, B., Martinez-Catala, R., Harte, S., O' Mathuna, C., Cleary, J., Slater, C., & Murphy, H., Smart Coast: a

Wireless Sensor Network for Water Quality Monitoring, in Proceedings of the 32nd IEEE Conference on Local Computer Networks, pp. 815-816, 2017.

[7] Menon, K. U., Divya, P., & Ramesh, M. V., Wireless Sensor Network for River Water Quality Monitoring in India, in Proceedings of the 3rd IEEE International Conference on Computing Communication & Networking Technologies, pp. 1-7, 2012.

[8] Yuan, F., Huang, Y., Chen, X., & Cheng, E, A., Biological Sensor System using Computer Vision for Water Quality Monitoring, IEEE Access, Vol. 6, pp. 61535-61546, 2018.

[9] Wu, Z., Liu, J., Yu, J., & Fang, H., Development of a Novel Robotic Dolphin and its Application to Water Quality Monitoring, IEEE/ASME Transactions on Mechatronics, Vol. 22, No. 5, pp. 2130-2140, 2017.

[10] Jingmeng, W., Xiaoyu, G., Wenji, Z., & Xiangang, M., Research on Water Environmental Quality Evaluation and Characteristics Analysis of TongHui River, in Proceedings of the IEEE International Symposium on Water Resource and Environmental Protection, Vol. 2, pp. 1066-1069, 2011.

[11] Vijayakumar, N., & Ramya, R., The Real-time Monitoring of Water Quality in IoT Environment, in Proceedings of the IEEE International Conference on Innovations in Information, Embedded and Communication Systems, pp. 1-5, 2015.

[12] Kedia, N., Water Quality Monitoring for Rural Areas-a Sensor Cloud-based Economical Project, in Proceedings of 1st IEEE International Conference on Next Generation Computing Technologies, pp. 50-54, 2015.

[13] Kumar, R. K., Mohan, M. C., Vengateshapandiyan, S., Kumar, M. M., & Eswaran, R., Solar based Advanced Water Quality Monitoring System using Wireless Sensor Network, in Proceedings of the International Journal of Science, Engineering and Technology Research, Vol. 3, No. 3, pp. 385-389, 2014.

[14] Viani, F., Bertolli, M., Salucci, M., & Polo, A., Low-cost Wireless Monitoring and Decision Support for Water Saving in Agriculture in Proceeding of IEEE Sensors Journal, Vol. 17, No. 13, pp. 4299-4309, 2017.

[15] Prasad, A. N., Mamun, K. A., Islam, F. R., & Haqva, H., Smart Water Quality Monitoring System, in Proceedings of the 2nd IEEE Asia-Pacific World Congress on Computer Science and Engineering, pp. 1-6, 2015.

[16] Krishnan, K. S. D., & Bhuvaneswari, P. T. V., Multiple Linear Regression based Water Quality Parameter Modelling to detect Hexavalent Chromium in Drinking Water, in Proceedings of the IEEE International Conference on Wireless Communications, Signal Processing and Networking, pp. 2434-2439, 2017.

[17] Dinniy, M. F., Barakhbah, A. R., & Kusumaningtyas, E. M., Quality Measurement Classification for Water Treatment using Neural Network with Reinforcement Programming for Weighting Optimization, in Proceedings of the IEEE International Conference on Knowledge Creation and Intelligent Computing, pp. 126-133, 2016.

[18] Arun, D., & Pais, A. R., Optimal Placement of Sensor Nodes for Water Quality Measurement, in Proceedings of the 12th IEEE International Conference on Wireless and Optical Communications Networks, pp. 1-9, 2015.

[19] Ramesh, M. V., Nibi, K. V., Kurup, A., Mohan, R., Aiswarya, A., Arsha, A., & Sarang, P. R., Water Quality Monitoring and Waste Management using IoT, in Proceedings of the IEEE Global Humanitarian Technology Conference, pp. 1-7, 2017.

[20] Alberto, W. D., del Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A., & de los Ángeles, B. M., Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality, A Case Study: Suquía River Basin, Water research, Vol. 35, No. 12, pp. 2881-2894, 2001.

[21] Huiting, W., Jingyu, S., & Hongmin, W., Application of Aquatic Organisms in Water Quality Monitoring, Environment and Development, Vol. 4, pp. 115, 2018.

[22] Zhenhua, L., Supervision and Management Information System for Rural Drinking Water Project Construction, in Proceedings of the 3rd IEEE International Conference on Intelligent System Design and Engineering Applications, pp. 1372-1375, 2013.

[23] Imen, S., Chang, N. B., Yang, Y. J., & Golchubian, A., Developing a Model-based Drinking Water Decision Support System Featuring Remote Sensing and Fast Learning Techniques, in Proceedings of the IEEE Systems Journal, Vol. 12, No. 2, pp. 1358-1368, 2018.

[24] Cloete, N. A., Malekian, R., & Nair, L., Design of Smart Sensors for Real-time Water Quality Monitoring, IEEE Access, Vol. 4, pp. 3975-3990, 2016.

