

Deep learning-based production assists water quality warning system for reverse osmosis plants

K. Udayakumar* and N. P. Subiramaniam

Department of Electronics and Communication System, Nehru Arts and Science College, Coimbatore, India

E-mail: elect.professor@yahoo.in

Abstract

Classifying water quality irregularities in reverse osmosis (RO) production plants requires suitable systems to provide intelligent warnings to the operators or supervisors who are engaged in executing corrective procedures applicable to production. The suggested deep learning methods are of utmost importance to identify at once variations in water quality irregularities in plants through competent classification methods, thereby enabling a reduction of burden for operators. In this paper, two types of LSTM-CNN based classification techniques are suggested to classify water quality features temporally into grades based on corrective actions that classify irregularity conditions of water quality on the basis of corrections. Distinct control methods are used for experiments to find water quality irregularities from variables, namely, pH, TDS, ORP, and EC, which aim to assist the production line. This proposed method enables automatic diagnosis and warning systems about water quality in RO plants. For classification, LSTM-CNN was trained with data recorded from eight plants of west and north parts of Chennai region. This research is meant to demonstrate particularly the top-level classification job for quality alerts. The features obtained from 4,096 time series array data using LSTM-CNNs achieved sensitivity to 97% and accuracy to 98%.

Key words: convolutional neural networks (CNN), electrical conductivity (EC), long short-term memory (LSTM), oxidation-reduction potential (ORP), pH, total dissolved solids (TDS)

Highlights

- Technology for nature is constantly moved forward to improve data collection of regular use.
- The best environmental observing systems that use the most advanced technologies.
- Near the beginning of a particular stage to point out and frequent checking are demanding tasks for water classification.
- In this paper, we examine classification of drinking water by applying machine learning classifier.

INTRODUCTION

Controlling the water quality continuously in reverse osmosis (RO) production plants and the necessary timely corrective action to improve or maintain the water quality during processes is a major real-time problem. Many production plants quite commonly use measuring instruments to monitor water quality physicochemical parameters and warn about adverse variations so as to perform instantly corrective actions in real time. Without a proper warning system, the remedial actions may get delayed. This

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

may lead to operators carrying out inappropriate actions. From just a display of the water quality parameters by instruments, some new or inexperienced operators may find it hard to interpret their values instantly and to perform suggested corrective measures quickly. Thus, water quality monitoring in real time requires more interpretation instead of just the values of physicochemical parameters like pH, electrical conductivity (EC), oxidation-reduction potential (ORP), and total dissolved solids (TDS).

Presently, many advanced ideas are proposed in designing advance automatic control methods of water quality (Zemouri 2003). In this regard, deep learning techniques are used as a useful mechanism to assist production operators in their decision-making to derive more intelligible solutions from crude values of the parameters displayed by the instruments. The fundamental idea behind this is that the measured values of water quality, which are directly simple, are to be processed into judicious and realistic information to carry out necessary corrective actions in real time. This too permits one to use the inexperienced or new operators in production. Customarily, for the water quality that is associated with numerical values, now, a meaningful qualitative interpretation system is suggested. This qualitative supported interpretation enables easier corrective actions on water quality in production using artificial intelligence algorithms.

The main challenge in automated assessment of water quality in a RO production plant is the complexity in interpretation of the values of physicochemical and microbiological parameters in real time and the use of unskilled labor involves quality risk. However real-time continuous monitoring which provide data for any variations that occur in the production process reflects the status of water quality. The main aim of the present investigation is to propose a classification-based intelligible representation of water quality based on real-time basic physicochemical parameters, specifically, pH, TDS, and ORP so as to perform meaningful corrective actions if any variations occur in quality in a RO production plant. The proposed method, which utilizes three basic physicochemical measurements, can possibly predict sensible details for the operators or supervisors to reach decisions on corrective measures in order to improve or maintain the water quality. RO plants across Chennai and other regions of North Tamil Nadu generally use underground water of bore-wells in treating the water for quality like pH, EC, TDS, and ORP parameters. The normal values of various parameters (Central Pollution Control Board 2007; Loganathan *et al.* 2011; Anoop Raj & Muruganandam 2014; Nandhakumar *et al.* 2015) of the underground water in specified areas are that pH value is between 6.4 and 7.6, TDS is between 600 and 6,500 mg/l, ORP is between 37 and 143 mV, and EC is between 489 and 1,561 μ S/cm. The adjoining regions of specified areas showed many inconsistencies in values of the chosen parameters (Central Pollution Control Board 2007). The WHO has recommended values of the parameters of drinking water (WHO 2011): pH must be in the range of 6.4 and 7.6, TDS must be in the range of 300–500 mg/L, ORP must be in the range of 300 and 500 mV, and EC must be less than 400 μ S/cm.

Most previous research works have just aimed to monitor the water quality of natural and man-made sources and reservoirs, and also the water quality of RO plants for drinking. Among these works, most were based on the Internet of Things (IoT) ad hoc monitoring of water quality. The novelty here is intended to include an artificial intelligence-based quality warning system to a RO plant for the production of drinking water. Further, the suggested system alerts the operator about the quality status of water during production and allows inexperienced operators to carry out appropriate remedial actions.

RELATED WORKS

The current research topic in water quality control is detection of abnormal physicochemical properties of drinking water with greater accuracy, thus online controlling and correcting water quality become a tougher task. A real-time system of water quality monitoring and abnormality detection

based on multiclass support vector machines (MSVM) is proposed that suits RO production plants. The suggested system is realistic to furnish user-friendly information for production controlling operators to make decisions and perform operative corrections.

Muharemi *et al.* (2019) described many methods for detecting variations or abnormalities taking place in time series data of water quality. Their work too discussed and proposed solutions to overcome certain problems when handling time series data. They used logistic regression, linear discriminate analysis (LDA), support vector machines (SVM), deep neural network (DNN), artificial neural network (ANN), long short-term memory (LSTM), and recurrent neural network (RNN) to apply over time series data and performed evaluation using F score.

Zhang *et al.* (2017) proposed a new abnormality detection technique on water quality data applying two time-moving windows, which can detect abnormalities by patterns in real-time data history. The technique was on the basis of a statistical model, combination of autoregressive linear model. They verified the proposed idea utilizing three months' data of water quality such as pH obtained from a real-time water quality monitoring station of a river system.

Ahmed *et al.* (2020) used IoT systems for water quality monitoring in real-time and machine learning methods like ANN and SVM to predict WQI, and detect abnormal events, namely, purposeful pollution to allow real-time pollution detection and control operation. Their proposed method also applied regression, correlation analysis, hierarchical clustering, etc., to study multiple directions to include learning technique to support decision-making.

Liu *et al.* (2019) focused on a prediction model of water quality which needed data of higher quality. They constructed and operated a smart system of water quality monitoring on the basis of the IoT, which generated too much big data at a higher speed and thus made water quality data complex. They designed and established a drinking-water quality time-series prediction model to predict water quality big data by means of state of the art deep learning (DL) method using LSTM. They obtained drinking-water quality data from the automatic water quality monitoring station of Guazhou Water Source of the Yangtze River in Yangzhou.

Karthick *et al.* (2018) focused on designing a system to measure parameters, namely, dissolved oxygen, chloride, ammonia, nitrate, hardness, and turbidity to determine the water quality in well-water using IoT-based machine learning for water quality monitoring system. They accessed space-time data that promote easy detection of real-time instances to study the impact of pollution on the quality of water in regular water bodies.

Yafra & See (2016) addressed the impacts of water pollution effectively when data are studied and the quality of water is predicted in advance. They discussed several early research works and emphasized the work needed to be performed with respect to accuracy, reliability, efficiency, together with usability and of the present water quality control methods. They aimed to construct a prediction model of water quality by means of its parameters using time-series analysis and ANN.

Sakizadeh (2016) used three algorithms of ANN type comprising early stopping ANN, ensemble ANN, and Bayesian regularized ANN. The author used 16 ground water variables from 47 wells and springs to predict water quality index.

Ashwini *et al.* (2019) aimed to design and implement an inexpensive system for the IoT and machine learning-based real-time water quality monitoring. The physicochemical parameters of water, namely, temperature, humidity, moisture, and visibility were recorded using corresponding sensors. They used ESP8266 as a main controller for processing the recorded values from the sensors. They applied random forest (RF) and K-nearest neighbor (KNN) techniques in analyzing and predicting the quality of water.

Muhammad *et al.* (2015) proposed an appropriate classification method for water quality on the basis of machine learning techniques. They studied and contrasted the performances of different classification methods for the purpose of identification of important characteristics that assisted in classification of the water quality of Kinta River, Perak, Malaysia. They tested five algorithms and

compared their effectiveness. They found that the Lazy model applying K Star algorithm was the best among the five models.

Vijay & Kamaraj (2019) collected samples from the bore wells of Vellore district, Tamil Nadu, India that were largely utilized for drinking. They measured the following water quality parameters, namely, pH, TDS, EC, chloride, sulfate, nitrate, carbonate, bicarbonate, metal ions, trace elements for the purpose of analysis. They focused on prediction of water quality by applying machine learning classifiers like RF, Naïve Bayes, and C5.0 to predict with higher efficiency and accuracy.

SIGNIFICANCE OF VALUES OF WATER QUALITY PARAMETERS

The parameters of drinking water according to WHO standards are that the pH value must be in the range of 6.4–7.6, the TDS from 300 to 500 mg/L, the ORP from 300 to 500 mV and EC less than 400 μ S/cm WHO (2011). Low or high pH value indicates the pollution level due to heavy metal and chemicals. Low or high pH value makes water taste and smell unpleasant while drinking and corrodes metal pipes and appliances that carry the water. Low or high pH value results in adverse medical conditions such as acid reflux, high blood pressure, diabetes, high cholesterol, etc. Negative ORP exhibits anti-oxidant property which prevents aging and improves cellular health, whereas positive ORP value exhibits oxidative property which cannot supply electrons to neutralize free radicals. The presence of free radicals affects the absorption of protein and lipid, and the structure of DNA. EC is the measure of conductivity of water due to the presence of inorganic salts, while TDS is a measure of both organic substances and inorganic salts. Low TDS or EC can have ill-health effects in the long term like mineral removal from body tissue, whereas high TDS or EC although not a significant effect for ill health, contributes to the esthetic property of water like taste, smell, cleanliness, etc. However, the presence of substances like pesticides and fertilizers has a huge effect on health conditions such as cancer, tumor growth, kidney disease, liver and pulmonary functions, etc. Elevated TDS level also raises the water temperature and pollution. The increased temperature of water in water bodies has related variations on climate (Valipour *et al.* 2020).

PROPOSED SYSTEM

Several remarkable efforts have been made to design advanced automatic water quality monitoring and control methods (Meghwani & Dewangan 2017). The deep learning techniques are commonly used as a necessary basic mechanism to assist decisions and derive comprehensive solutions from raw data values. Out of various deep learning techniques, LSTM-based convolutional neural networks (LSTM-CNN) is considered better on the basis of its potentiality during training, especially huge data dimensions. LSTM-CNN remarkably and extensively can be applied for recognizing pattern, computing density and regression (Hochreiter & Schmidhuber 1997). Its speedy convergence is well suited to derive a complete optimal model for classification with less complication. LSTM-CNN technique is incorporated as a proposed system to monitor the water quality in RO production plants and classify the quality of water from an operator's perspective in performing remedial actions in production online. The common problem perceived in classification of water quality is classifying it into multiple classes – highly alkaline, highly acidic, high TDS, high ORP, and low ORP as quality irregularities. Multiple sensor architecture-based classification system for water quality monitoring is shown in Figure 1.

Architecture and components of monitoring system

Figure 2 depicts the proposed architecture of real-time water quality monitoring for an RO plant. The suggested system contains some basic water quality sensors, namely, pH, TDS, ORP, and EC and is

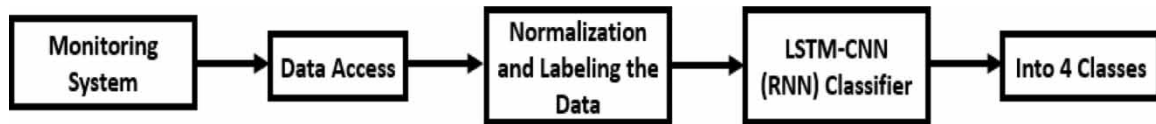


Figure 1 | Functional block diagram of proposed monitoring and classification system.

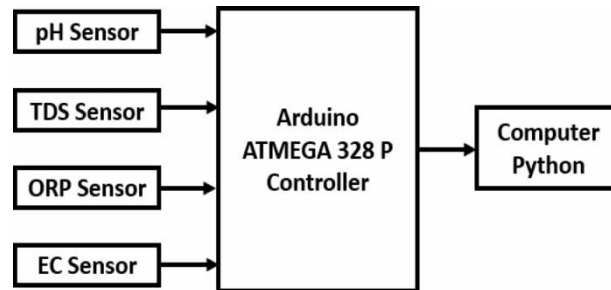


Figure 2 | Monitoring system architecture.

connected with Arduino as the central controller. This main controller accepts the sensor signals and passes them to the computer for analysis. The controller that is used here is Arduino ATMEGA 328p.

Hardware components of constructed monitoring system

The hardware used in building the monitoring system are pH sensor of SKU: SEN0161, shown in [Figure 3\(a\)](#), TDS/EC sensor of SEN0244, shown in [Figure 3\(b\)](#), ORP sensor of Atlas EZO, shown in [Figure 3\(c\)](#), temperature sensor DS18B20, shown in [Figure 3\(d\)](#) and Arduino microcontroller with LCD, shown in [Figure 3\(e\)](#) and [3\(f\)](#), respectively.

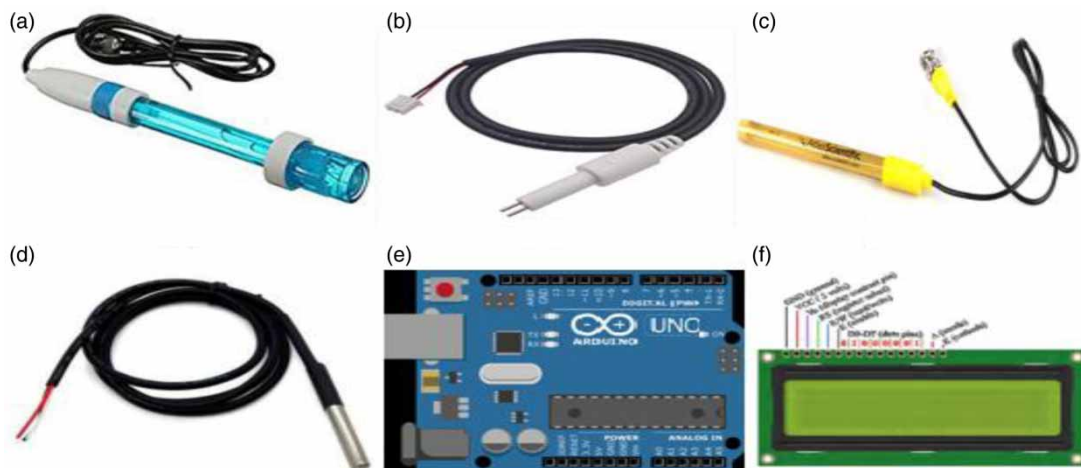


Figure 3 | Components that comprise the water quality monitoring system: (a) pH sensor, (b) EC/TDS sensor, (c) ORP sensor, (d) temperature sensor, (e) Arduino, (f) LCD display.

pH sensor

A pH sensor is an electronic apparatus which determines the pH value of aqueous solution. The pH measure demonstrates the level of the alkaline or acidic property of the solution. It can be derived

numerically by applying a logarithm and it takes a value between 0 and 14 with 7 for a neutral solution. Values less than 7 indicate the solution is acidic while a value above 7 is basic. The sensor operates at the voltage of 5 V and it is fairly consistent with Arduino.

The sensor includes three distinct electrodes: (a) glass, (b) reference, and (c) gel compound. The pH value is featured as a negative logarithmic measure that demonstrates hydrogen ion concentration of a solution:

$$pH = -\log[H^+]$$

TDS/EC sensor

A TDS sensor indicates the total dissolved solids that present in a aqueous solution; in other words, it is a total concentration of dissolved solid particles. Dissolved ionized solids, such as salts and minerals, that present will raise the EC in a solution. As it is a volumetric dependent estimation of ionized dissolved solids, EC can also be treated to determine TDS. EC is unlikely to change significantly in the presence of dissolved organic solids, such as sugar particles and suspended organic colloids.

The TDS and specific conductance of water is generally stated by the equation

$$TDS = k_e EC$$

where TDS is denoted in mg/L and EC is denoted in mS/cm at a temperature of 25 °C. The k_e is a correlation value and it ranges between 0.55 and 0.8 (Atekwana *et al.* 2004).

Conductivity is the estimation of an ability of the solution to permit the flow of an electric current. Conductivity is obtained by applying Ohm's law, $V = IR$; in which voltage (V) is obtained by multiplying the values of current (I) and resistance (R). Thus, the resistance is calculated by obtaining the ratio between voltage and current. When a voltage is supplied to the conductor, the current starts to flow, which is subject to the conductor resistance. Then, the conductivity is easily described by the reciprocal value of the resistance of a solution present between electrodes. The more ions that present in the solution means the conductivity will likely increase.

ORP sensor

The ORP determines the ability of the solution to act like a reduction or oxidizing agent (Botlagunta *et al.* 2015). This is utilized to determine the ability of oxidization by the chlorine present in water or to measure whether the neutral equilibrium point has been attained in the reaction of oxidization/reduction. Water treatment using chlorine will mainly act like disinfectant in order to remove microbes, namely, protozoans, bacteria, and viruses that commonly thrive in water bodies, reservoirs, and mains. Chlorine ranges up to 4 mg/L in water are recognized to be safe for drinking. Exposure to or drinking continuously water with excess chlorine may elevate the risk of being affected by respiratory problems such as asthma, especially in children, tumor growth in intestine and kidney, and tissue and cell lesions (Botlagunta *et al.* 2015).

Temperature sensor

The temperature of water will vary according to the conditions of day-night and climate. The temperature of water bodies is subject to the environmental season and geographical location. The temperature of regular drinking water may be roughly in the range of 7–44 °C. If unusual chemicals are present in drinking water, the temperature of the water rises, which surpasses the

normal values. The range of temperature that the sensor measures will be between approximately -55 and $+125$ °C.

Arduino microcontroller

Arduino is an openly available platform normally used for developing applications in electronics. Arduino includes elements of both programmable hardware which acts like a microcontroller and a software IDE (Integrated Development Environment) which executes on a computer and uses script with C and C# codes to upload it onto Arduino hardware. Various types of Arduino are available in stores. Here, in this project, AtMEGA 328p UNO is chosen as it serves sufficient digital-analog IO pins for connecting numerous sensors at one time. AtMEGA 328p requires a 5 V power supply and works at the clock frequency of 16 MHZ. Arduino can be hooked to a PC through universal serial bus (USB) and can save data of 2 KB. It comprises 14 digital I/O pins and 6 analog pins of input. The above-mentioned four sensors are connected to four digital-analog IO pins of Arduino (Hari Sudhan *et al.* 2015; Škultéty *et al.* 2018).

Liquid crystal display (LCD)

A LCD of 16×2 is connected to Arduino to display the measured sensor data on screen. This has 2 rows with 16 characters in a row. Thus, it can show a total of 32 characters and each character is displayed in 5×8 dots of pixel.

PROPOSED DEEP LEARNING MODEL

The aim of this study is the advancement of a classification model to classify water quality by applying and capitalizing on state-of-the-art deep learning methods. Convolutional layer is featured by its potential to derive beneficial information and learn the inner characterization of time-series data, whereas LSTM networks are functional for recognizing long- and short-term associations. The main aim of the suggested model is to effectively unite the merits of both deep learning methods. Ultimately, the suggested model of CNN-based LSTM, contains two major blocks. The first constituent contains the layers of convolution and pooling, where complex mathematical processes are carried out to generate features for the input data, whereas the second constituent uses the developed features obtained from LSTM and dense layers. In the latter part, a short explanation about the layers of convolution, pooling, and LSTM is provided as they establish the main part of the suggested model.

The layers of convolution and pooling (Rawat & Wang 2017) are particularly framed as data pre-processing layers which have the job to process the input data and obtain beneficial knowledge which will be applied as an input, normally to a fully connected network (FCN) layer. In particular, the convolutional layers perform convolution operations between original input data and convolution kernels generating new values for features. The input data must be structured to a matrix shape because this method is primarily aimed to extract features from the dataset of images (Krizhevsky *et al.* 2012). The convolutional kernel can be regarded as a small window, in comparison with the size of the input matrix, which comprises the values of coefficients into a matrix shape. This small window performs sliding throughout the size of the input matrix using convolutional operation on every sub-region in such a way that this designated window covers the whole matrix. The outcome of the entire operation will be a convolved matrix which provides the value of features indicated by the values of coefficients and the dimensional size of the implemented kernel. By the use of various convolution kernels on data input, numerous convolved features can be created, which may be

generally quite helpful comparing first primary features of the data input, hence improving the performance of the proposed model.

The convolutional layers are generally continued with the layer of a rectified linear unit, i.e., non-linear activation function, and subsequently with a pooling layer. A pooling layer performs a sub-sampling operation which obtains specific values out of convolved features and generates a smaller dimensional matrix. Alternatively, in the same way of conducting operations on the convolution layer, the pooling layer uses a tiny moving window which accepts the values of every sub-region of the convolved features as inputs and one fresh value as an output, which is denoted by a process that the pooling layer is determined to perform. For instance, average pooling and maximum pooling compute the average and maximum values of every sub-region. Consequently, the pooling layer generates the new matrices which are regarded as abridged forms of the convolved features generated by the convolution layer. Operation of the pooling layer can support the model by being more stable for any small variations on the input so that the output values due to pooling operations do not vary.

The neural network of LSTM (Ilakkiya & Sharmila 2017) is a specific form of recurrent neural network (RNN) which has the capability to learn long-term relations by using feedbacks. A standard RNN seeks to address the issue of feed-forward neural network, termed 'lack of memory,' which is accountable in exhibition of underperformance on time-series and sequence issues. This model uses cyclic connections in the hidden layers for the purpose of acquiring short-term memory and has the opportunity to acquire details from the data of time-series and sequences. However, RNN experiences the popular vanishing gradient issue which limits the model in learning long-term relations. Hence, LSTM is capable of addressing this issue by keeping beneficial details on storage bins and fading redundant details, thereby attaining, overall, higher returns in comparison to standard RNN.

Every LSTM component is formed using memory bin and three principal gateways: the input, the output, and ignore. Due to this scheme, the LSTM handles development of a controlled flow of data by determining which data must be ignored and which have to be retained, hence controlling learning of long-term relations. More precisely, the input gateway as well as a second gateway c_t^* manages the new data which is retained in the memory state c_t at a time t . The ignore gateway f_t manages the earlier data which need to disappear or need to be retained in the memory bin at a time $(t - 1)$, whereas the output gateway o_t manages which data might be used for the memory bin output. To recap, Equations (2)–(5) concisely explain the operations carried out by an LSTM component.

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(U_g x_t + W_g h_{t-1} + b_g) \quad (2)$$

$$c_t^* = \tanh(U_c x_t + W_c h_{t-1} + b_c) \quad (3)$$

$$c_t = g_t \odot c_{t-1} + i_t \odot c_t^* \quad (4)$$

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o) \quad (5)$$

in which x_t indicates the input, W_* and U_* are the weights of the matrices, b_* are bias vectors, σ is a sigmoid function, and the operator \odot indicates element-wise multiplication. Lastly, the hidden state h_t which comprises the memory bin and its output is evaluated as follows:

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

Realizing that by chance numerous LSTM layers are stacked, both the memory and hidden states c_t and h_t , respectively, of every LSTM layer are dispatched as inputs to the next layer of LSTM.

To realize the proposed model, some remarkable characteristics of existing models are used. The CNN-LSTM comprises two convolution layers with filters of 32 and 64 and of size (2,)

correspondingly, succeeded by the layer of pooling, a layer of LSTM and a layer of one neuron as an output. A summary of the suggested model of CNN-LSTM architecture is shown in Figure 4 and its layer specifications are mentioned in Table 1.

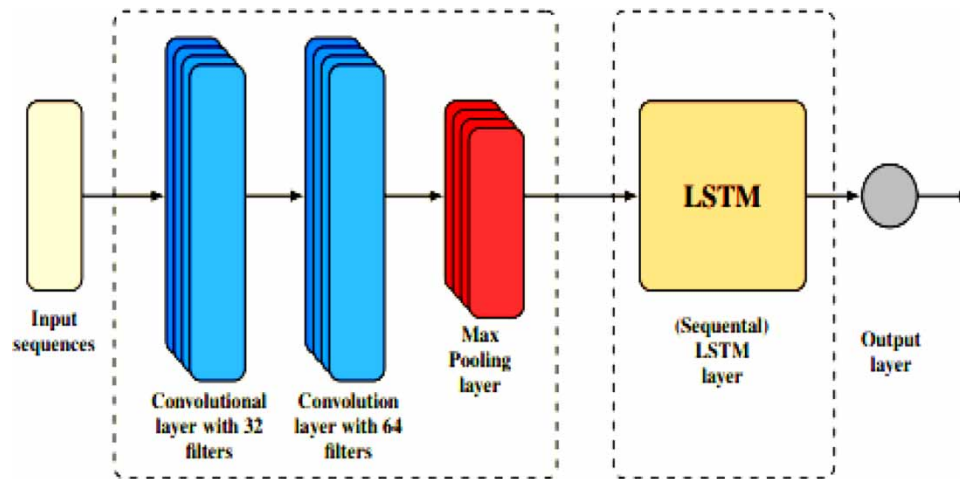


Figure 4 | The architecture of the proposed model of CNN-LSTM with two convolution layers, a pooling layer, a LSTM layer, and an output layer.

Table 1 | Parameter specification of the proposed model

Model	Description
CNN-LSTM	Convolutional layer with 32 filters of size (2,) Convolutional layer with 64 filters of size (2,) Max pooling layer with size (2,) LSTM layer with 100 units

DATA COLLECTION

Now, pH, TDS, ORP, and temperature data were sampled by random method: thus, 20 data randomly at 12 various instants in a particular day randomly selected from a span of 8 weeks with the span of 6 months from the eight RO plants situated in western and northern suburbs of extended parts of Chennai with the frequency of eight times in a week for every plant. The collected data were then grouped into 64 arrays each of length 240 for each parameter and for each plant. These data measured from the eight RO plants are shown as sample graphs in Figure 5 for pH data, Figure 6 for ORP data, Figure 7 for TDS data, Figure 8 for EC data, and Figure 9 for temperature data.

Inferences are deduced from Table 2 as follows:

- pH data measured at various plant sites have demonstrated that the groundwater utilized for treatment is alkali rather than acidic.
- ORP data measured at various plant sites have demonstrated that the groundwater utilized for treatment is in the range from 89 to 107 mV which is less than the WHO recommendation values of 300–500 mV.
- TDS data measured at various plant sites have demonstrated that the groundwater utilized for treatment is in the range of 999–1,320 (mg/L) which is far greater than the WHO recommendation values of 300–600 mg/L.

- EC data measured at various plant sites have demonstrated that the groundwater utilized for treatment is in the range of 1,100–1,500 (μS/cm) which is far greater than the WHO recommendation values of 400 μS/cm.
- The proposed model is designed to classify the groundwater quality status to carry out remedial action in RO plants based on major quality irregularities, namely, high pH, high TDS, and low ORP.
- Accessing data and grouping the water quality irregularities in relation to remedial action likely to be carried out in production plants are key elements for experiments.

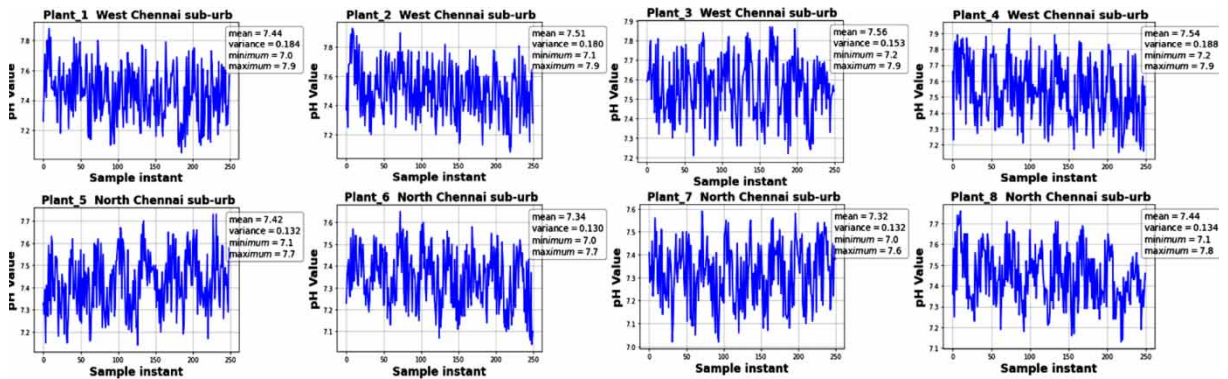


Figure 5 | pH data recorded from RO plant at eight locations, Plant_1 to Plant_4 from West Chennai sub-urban and Plant_5 to Plant_8 from North Chennai sub-urban.

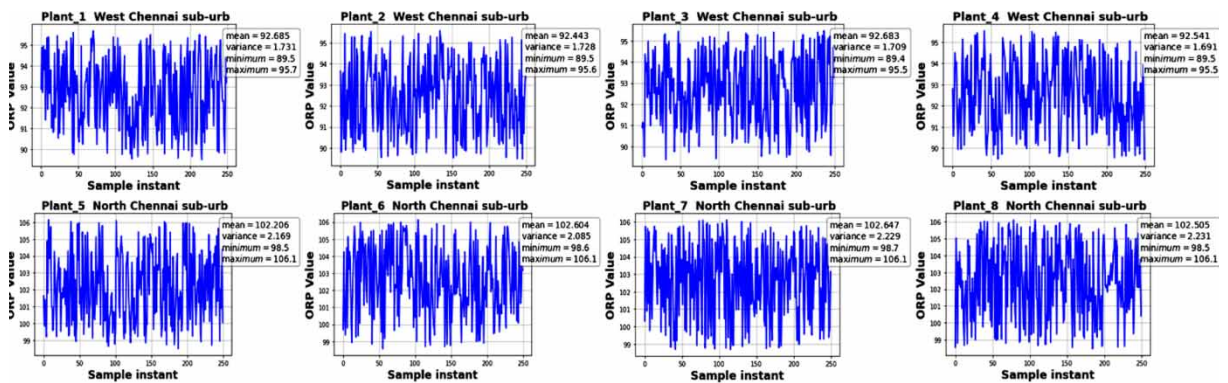


Figure 6 | ORP data recorded from RO plant at eight locations, Plant_1 to Plant_4 from West Chennai sub-urban and Plant_5 to Plant_8 from North Chennai sub-urban.

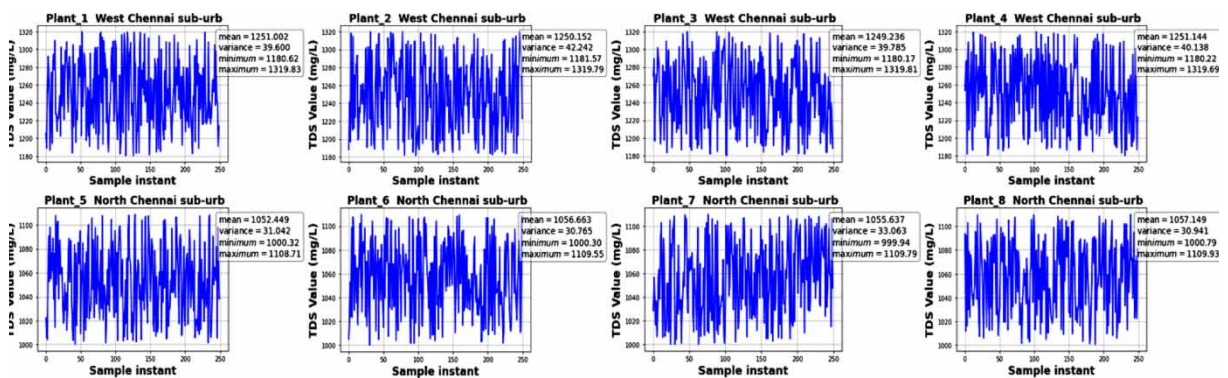


Figure 7 | TDS data recorded from RO plant at eight locations, Plant_1 to Plant_4 from West Chennai sub-urban and Plant_5 to Plant_8 from North Chennai sub-urban.

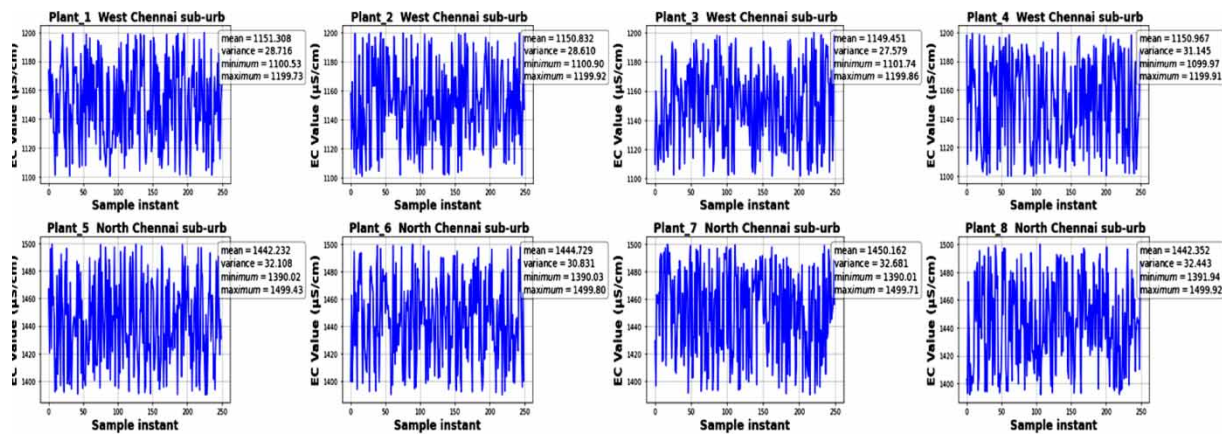


Figure 8 | EC data recorded from RO plant at eight locations, Plant_1 to Plant_4 from West Chennai sub-urban and Plant_5 to Plant_8 from North Chennai sub-urban.

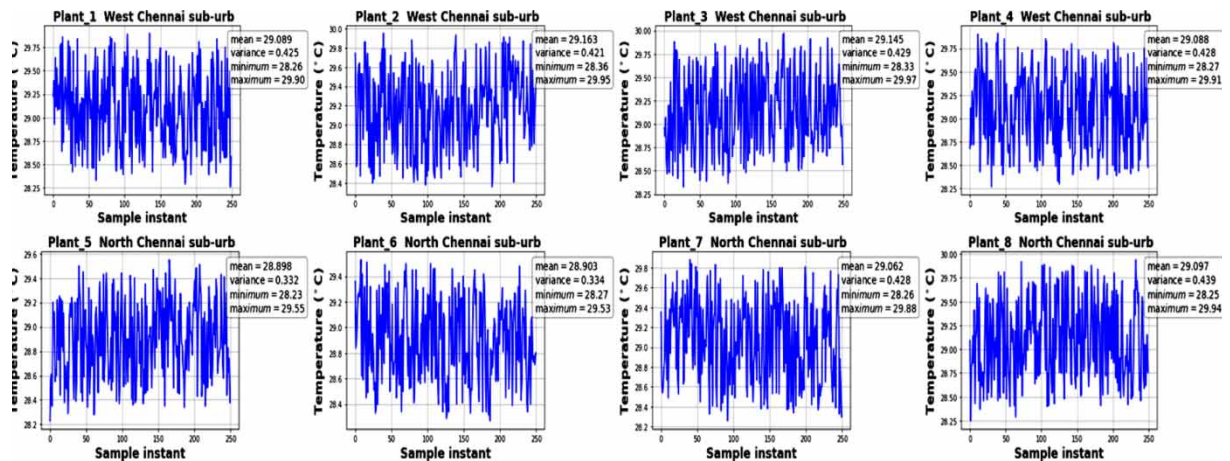


Figure 9 | Temperature data recorded from RO plant at eight locations, Plant_1 to Plant_4 from West Chennai sub-urban and Plant_5 to Plant_8 from North Chennai sub-urban.

For training the LSTM-CNN and to classify the irregularities denoted in terms of grades, 1,920 data samples were collected from eight RO plants around north and west Chennai sub-urbans. 240 data samples at various occasions in a day were collected from every RO plant. For the purpose of deciding remedial action which is supposed to be performed by the plant operator to adhere to water quality standards, the class targets are tabulated on the basis of water quality irregularities, as shown in Table 3.

ANALYSIS AND RESULTS

This section describes the experiments performed and the results obtained from the classifier of LSTM – CNN and by using preferred classes as target, as described in Table 3.

Table 4 depicts the statistics of confusion matrices of the classifiers of LSTM – CNN and Figure 10 shows the confusion matrices of the classifiers when classifying the water quality irregularity.

A confusion matrix is a commonly used technique to denote the classifier performance on a group of sample data where the true values are available. From the confusion matrix depicted in Figure 10, there exists four likely predicted levels. The classifier has been supplied with 4,096 sample data

Table 2 | The minimum, maximum, mean, standard deviation (SD) and mean/SD values of recorded pH, ORP, TDS, EC, and temperature data

S. no.	Parameter	Plant Id	Max	Min	Mean	Std. dev.	SD/Mean in %
1	Ph	P1	7.9	7.0	7.44	0.184	2.47
2		P2	7.9	7.1	7.51	0.180	2.4
3		P3	7.9	7.2	7.56	0.153	2.02
4		P4	7.9	7.2	7.54	0.188	2.49
5		P5	7.7	7.1	7.42	0.132	1.78
6		P6	7.7	7.0	7.34	0.130	1.77
7		P7	7.6	7.0	7.32	0.132	1.8
8		P8	7.8	7.1	7.44	0.134	1.8
9	ORP (mV)	P1	95.7	89.5	92.68	1.73	1.87
10		P2	95.6	89.5	92.44	1.72	1.86
11		P3	95.5	89.4	92.68	1.71	1.85
12		P4	95.5	89.5	92.54	1.69	1.83
13		P5	106.1	98.5	102.21	2.17	2.12
14		P6	106.1	98.6	102.61	2.09	2.04
15		P7	106.1	98.7	102.65	2.22	2.16
16		P8	106.1	98.5	102.51	2.23	2.18
17	TDS (mg/L)	P1	1,319.8	1,180.6	1,251.02	39.60	3.17
18		P2	1,319.8	1,181.6	1,250.15	42.42	3.39
19		P3	1,319.8	1,180.1	1,249.24	39.78	3.18
20		P4	1,319.7	1,180.2	1,251.14	40.14	3.21
21		P5	1,108.7	1,000.3	1,052.45	31.04	2.95
22		P6	1,109.5	1,000.3	1,056.66	30.76	2.91
23		P7	1,109.8	999.4	1,055.64	33.06	3.13
24		P8	1,109.9	1,000.8	1,057.15	30.94	2.93
25	EC (µS/cm)	P1	1,199.3	1,100.5	1,151.3	28.72	2.49
26		P2	1,199.9	1,100.9	1,150.8	28.61	2.49
27		P3	1,199.8	1,101.7	1,149.5	27.58	2.4
28		P4	1,199.9	1,099.9	1,150.9	31.15	2.71
29		P5	1,499.4	1,390.0	1,442.2	32.11	2.23
30		P6	1,499.8	1,390.0	1,444.7	30.83	2.13
31		P7	1,499.7	1,390.0	1,450.2	32.68	2.25
32		P8	1,499.9	1,391.9	1,442.4	32.44	2.25
33	Temperature (°C)	P1	29.90	28.26	29.09	0.425	1.46
34		P2	29.95	28.36	29.16	0.421	1.44
35		P3	29.97	28.33	29.14	0.429	1.47
36		P4	29.91	28.27	29.09	0.428	1.47
37		P5	29.55	28.23	28.89	0.332	1.15
38		P6	29.53	28.27	28.90	0.334	1.16
39		P7	29.88	28.26	29.06	0.428	1.47
40		P8	29.94	28.25	29.09	0.439	1.51

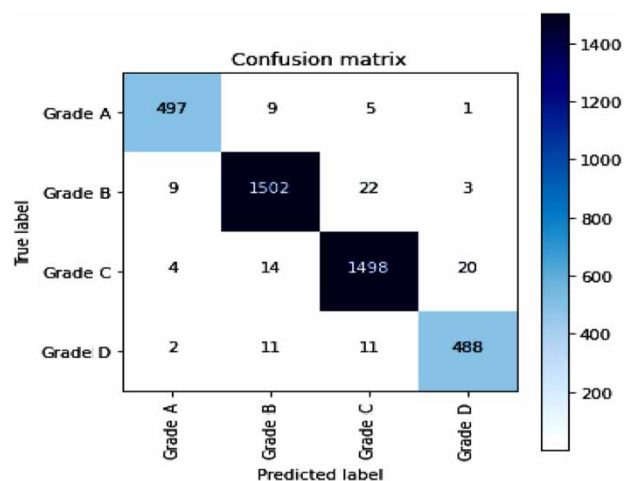
belonging to various classes: 512 samples in Grade A, 1,536 samples in Grade B, 1,536 samples in Grade C, and 512 samples in Grade D. Grade classes are described in Table 3. While predicting for the presence of anomaly in the 512 supplied sample data of Grade A class, the classifier correctly

Table 3 | Grading of water quality to monitor in RO production plant

Water quality	Ph	ORP	TDS	Corrective action
Class 1	High	Low	High	Required
Class 2	High	Low	Normal	Required
Class 3	Normal	Normal	High	Required
Class 4	Normal	Normal	Normal	Normal

Table 4 | The statistics of the confusion matrix statistics of the classifiers

Class/Model/Parameters	Grade A <i>LSTM – CNN</i>	Grade B <i>LSTM – CNN</i>	Grade C <i>LSTM – CNN</i>	Grade D <i>LSTM – CNN</i>
Accuracy	99.26	98.34	99.12	99.41
Precision	97.07	97.78	97.52	95.31
Sensitivity	97.57	97.86	97.12	95.43
Specificity	99.56	98.67	98.52	99.33

**Figure 10** | Confusion matrix for *LSTM – CNN* classifier on grading the water quality irregularities.

predicted 497 Grade A data as Grade A, and wrongly predicted 9 data of Grade A into Grade B, 5 data of Grade A into Grade C, and 1 data of Grade A into Grade D. Similarly, for 1,536 supplied sample data of Grade B class, the classifier correctly predicted 1,502 Grade B data as Grade B, and wrongly predicted 9 data of Grade B into Grade A, 22 data of Grade B into Grade C, and 3 data of Grade B into Grade D. For 1,536 supplied sample data of Grade C class, the classifier correctly predicted 1,498 Grade C data as Grade C, and wrongly predicted 4 data of Grade C into Grade A, 14 data of Grade C into Grade B, and 20 data of Grade C into Grade D. For 512 supplied sample data of Grade D class, the classifier correctly predicted 500 Grade D data as Grade D, and wrongly predicted 2 data of Grade D into Grade A, 11 data of Grade D into Grade B, and 11 data of Grade D into Grade C.

The commonly used basic terms are described as follows: true positives (TP) are the cases predicted correctly in the correct classes, true negatives (TN) predicted wrongly in the wrong classes, false positives (FP) (termed as ‘Type I error’) predicted correctly in the wrong classes, and false negatives (FN) (termed as ‘Type II error’) predicted wrongly in the correct classes.

The following metrics that are regularly calculated from a confusion matrix for the classifier are: accuracy which describes how well the classifier predicts correctly the events (both positive and negative); precision which describes how well the classifier is correct when predicting positive events; sensitivity which describes how well the classifier predicts the positive events as positive event; and specificity which describes how well the classifier predicts the negative events as negative event.

DISCUSSION

Table 4 shows the results derived from applying the LSTM-CNN model. As evidenced, subject to applying the described parameters, LSTM-CNN showed improved performance with an accuracy of 99.03%.

In general, the outcomes showed that LSTM-CNN could be considered an appropriate choice in order to classify water quality as a quality alert system for RO plant drinking water production. One of the foremost characteristics of the LSTM-CNN model is that the necessary network size is directly realized in the network's structural process. Hence, the LSTM-CNN model has smaller network size and provides greater classification accuracy.

Although the above-discussed method could find online water quality classification over the intervals if sampling is done simply by very vital parameters, this research is distinguished by the fact it always encountered restrictions, namely, shortage of on-site data. The constraint on water quality data is more serious compared to quantitatively. Further, the period of time series water quality samples and the number of recorded variables might be considered as limitations. Despite the good amount of water quality data applied in the current investigation, the water quality data sample size is moderate for precise modeling. Hence, the aforementioned method is proposed and verified by using a reasonable amount of on-site plant data.

CONCLUSION

The aim of the current research is to classify the water quality irregularities into grades. Reverse osmosis plant operators will be able to perform remedial action using this grading which will also be quite intelligible. The classification considered here will be a four-target class challenge as a whole in detection of water quality irregularities in RO plants using LSTM-CNN classifiers. The proposed classifiers have shown some encouraging signs of providing intelligible warning about water quality irregularities which require to be rectified with precise treatment in the necessary circumstances. Early researches that were performed aimed to estimate and classify the water quality of water reservoirs or bodies. However, the proposed method is focused on investigating drinking water quality produced by RO plants that is straightaway supplied to the community. This approach that sets out effective control methods for production robustly improves the water quality throughput.

The notable merit of using directly measured water quality parameters like pH, ORP, TDS, EC, and temperature in training the LSTM-CNN classifiers will allow the classifiers to directly learn water quality features and enable intelligible warning information for production operators or supervisors. For better performance of classification, the data that utilized were recorded and designated tags based on the amount of irregularities.

In practice, the data fed to the classifiers of LSTM-CNN types in grading the water based on quality irregularities may not predict exactly the grades until they are appropriately organized into classes. The LSTM-CNN classifiers have no issue on learning to detect the water quality features for drinking because the data types are health based. For the purpose of classification of worst case water quality irregularities, the time of learning involved in training will be significantly less. The challenge of the

proposed model is its integration into wireless sensor networks in detection of subtle variations on quality irregularities.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Ahmed, U., Mumtaz, R., Anwar, H., Mumtaz, S. & Qamar, A. M. 2020 **Water quality monitoring: from conventional to emerging technologies**. *Water Supply* **20** (1), 28–45.
- Anoop Raj, J. R. & Muruganandam, L. 2014 Occurrence of perchlorate in drinking, surface, ground and effluent water from various parts of South India. *International Journal of Innovations in Engineering and Technology (IJJET)* **4** (4), 1–7.
- Ashwini, K., Vedha, J. J., Diviya, D. & Deva Priya, M. 2019 Intelligent model for predicting water quality. *International Journal of Advance Research, Ideas and Innovations in Technology* **5** (2), 70–75.
- Atekwana, E. A., Atekwana, E. R., Rowe, R. S., Werkema, D. D. & Legall, F. D. 2004 **The relationship of total dissolved solids measurements to bulk electrical conductivity in an aquifer contaminated with hydrocarbon**. *Journal of Applied Geophysics* **56** (4), 281–294.
- Botlagunta, M., Bondili, J. S. & Mathi, P. 2015 Water chlorination and its relevance to human health. *Asian Journal of Pharmaceutical and Clinical Research* **8** (1), 20–24.
- Central Pollution Control Board 2007 **Status of Groundwater Quality in India – Part I**. Central Pollution Control Board, Ministry of Environment and Forests, Government of India.
- Hari Sudhan, R., Ganesh Kumar, M., Udhaya Prakash, A., Devi, A. & Sathiya, S. 2015 Arduino AtMEGA-328 microcontroller. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering* **3** (4), 27–29.
- Hochreiter, S. & Schmidhuber, J. 1997 **Long short-term memory**. *Neural Computing* **9** (8), 1735–1780.
- Ilakkiya, J. & Sharmila, P. 2017 **Monitoring water contamination using ORP sensor**. *International Journal for Research in Applied Science & Engineering Technology* **5**. doi:10.22214/ijraset.2017.11383.
- Karthick, T., Gayatri, D., Kohli, T. S. & Snigdha, P. 2018 Prediction of water quality and smart water quality monitoring system in IoT environment. *International Journal of Pure and Applied Mathematics* **118** (20), 3969.
- Krizhevsky, A., Sutskever, I. & Hinton, G. 2012 Imagenet: classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* **25** (2), 1097–1105.
- Liu, P., Wang, J., Sangaiah, A. K., Xie, Y. & Yin, X. 2019 **Analysis and prediction of water quality using LSTM deep neural networks in IoT environment**. *Sustainability* **11**, 2058.
- Loganathan, D., Kamatchiammal, S., Ramanibai, R., JayakarSanthosh, D., Saroja, V. & Indumathi, S. 2011 **Status of groundwater at Chennai city, India**. *Indian Journal of Science and Technology* **4** (5), 566–572.
- Meghwani, P. & Dewangan, K. 2017 **Real time water quality monitoring and control system**. *International Journal for Research in Applied Science & Engineering Technology* **5** (IX), 1–3.
- Muhammad, S. Y., Makhtar, M., Rozaimie, A., Abdul Aziz, A. & Jamal, A. 2015 **Classification model for water quality using machine learning techniques**. *International Journal of Software Engineering and Its Applications* **9** (6), 45–52.
- Muharemi, F., Logofătu, D. & Leon, F. 2019 **Machine learning approaches for anomaly detection of water quality on a real-world data set**. *Journal of Information and Telecommunication* **3** (3), 294–307.
- Nandhakumar, S., Varun, K. & Sathyanarayanan, N. 2015 Interpretation of groundwater quality around Ambattur Lake, Chennai, Tamil Nadu. *Journal of Chemical and Pharmaceutical Research* **7** (4), 1626–1633.
- Rawat, W. & Wang, Z. 2017 **Deep convolutional neural networks for image classification: a comprehensive review**. *Neural Computing* **29** (9), 2352–2449.
- Sakizadeh, M. 2016 **Artificial intelligence for the prediction of water quality index in groundwater systems**. *Modeling Earth Systems and Environment* **2**, 8.
- Škultéty, E., Pivarčiová, E. & Karrach, L. 2018 **The comparing of the selected temperature sensors compatible with the Arduino platform**. *Management Systems in Production Engineering* **26** (3), 168–171.
- Valipour, M., Bateni, S. M., Gholami Sefidkouhi, M. A., Raeini-Sarjaz, M. & Singh, V. P. 2020 **Complexity of forces driving trend of reference evapotranspiration and signals of climate change**. *Atmosphere* **11** (10), 1081.
- Vijay, S. & Kamaraj, K. 2019 Ground water quality prediction using machine learning algorithms in R. *International Journal of Research and Analytical Reviews* **6** (1), 743–749.
- WHO 2011 **Guidelines for Drinking Water Quality Recommendations**, 4th edn, Vol. 1. World Health Organization, Geneva, Switzerland.
- Yafra, K. & See, C. S. 2016 **Predicting and analyzing water quality using machine learning: a comprehensive model**. In *IEEE Long Island Systems, Applications and Technology Conference (LISAT)*.

- Zemouri, M. R. 2003 *Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamique. Application à la e-maintenance (Contribution to the monitoring of production systems using dynamic neural networks: Application to e-maintenance)*. PhD Thesis, Université de Franche-Comté, France.
- Zhang, J., Zhu, X., Yue, Y. & Wong, P. W. 2017 Seventh International Conference on Innovative Computing Technology (INTECH). In *A Real-Time Anomaly Detection Algorithm/or Water Quality Data Using Dual Time-Moving Windows*. IEEE, pp. 36–41.

First received 6 October 2020; accepted in revised form 16 November 2020. Available online 21 December 2020