Surface Water Pollution Detection using Internet of Things

Uferah Shafi[1], Rafia Mumtaz[2], Hirra Anwar[3], Ali Mustafa Qamar[4], Hamza Khurshid[5]
[1,2,3,4,5]School Of Electrical Engineering and Computer Science
National University of Science and Technology
[1]ushafi.dphd18seecs@seecs.edu.pk, [2]rafia.mumtaz@seecs.edu.pk, [3]hirra.anwar@seecs.edu.pk,
[4]mustafa.qamar@seecs.edu.pk, [5]hkhurshid.bese16seecs@seecs.edu.pk

Abstract: **Water is one of the primary requisites and crucial for sustaining the quality of life. In Pakistan its significance is more than ordinary due to the agrarian nature of the economy. Owing to increasing trend in urbanization and industrialization, the quality of water is continuously declining. For this purpose, we propose an Internet of Things (IoT) based water quality system capable of measuring the quality of water in near real time. The proposed solution is based on World Health Organization (WHO) defined water quality metrics. For this purpose, a real time embedded prototype has been developed to record the water quality parameters from the water samples collected from various sources across the study area. The hardware solution sends data to cloud for real time storage and processing. The processed data can be remotely monitored and water flow can be controlled using our developed software solution comprising of mobile app and a dashboard. In addition to water quality monitoring and control system, the predictive analysis of the collected data has been performed. For training purposes a dataset has been obtained from Pakistan Council of Research in Water Resources (PCRWR). Machine learning algorithms have been applied for classification of water quality and the experimental results indicate that deep neural network outperforms all other algorithms with an accuracy of 93%. The preliminary results have shown a high potential of scaling up this concept to an advanced level.**

Keywords— Water quality monitoring; Internet of Things; embedded systems, Deep Learning

I. INTRODUCTION

Water quality degradation is one of the most critical problems currently faced by Pakistan. Pakistan is ranked as 17th country across the globe, which is facing acute water crisis. Owing to the lack of water quality monitoring systems in place, population in rural and urban regions residing closer to industrial sector is compelled to use contaminated water for domestic purposes, which causes various acute diseases leading to increase in mortality rate [1-2]. Despite of this alarming situation, this problem remains unaddressed by the government. A number of research efforts have been made regarding the improvement of water quality however; these efforts have not been realized as practical systems at the national level. Today, when world's trend has been moved towards smart cities where every system is automated, Pakistan still relies on traditional systems due to challenges such as lack of literacy, institutional capability to deliver technology, and economic constraints**.** Owing to the dearth of real time water quality assessment and decision support systems in Pakistan, water quality assessment is currently carried out only in research laboratories where data is processed in non-real time. Mostly, the currently available supportive tools for water quality are manual based on human intervention rather than using technology, which is a promising alternative to the traditional complex and ineffective approaches. To address this challenge, we propose an Internet of Things (IoT) based system that is capable of analyzing the water quality with competitive accuracy. To ensure accurate and reliable analysis in monitoring water quality we need a large number of samples, where, IoT resolves such issues of data collection, analysis and communication.

The parameters required for monitoring the quality of water, include turbidity, temperature, dissolved oxygen level, pH level, dissolved ammonium, potassium, nitrate and conductivity. World Health Organization (WHO) has defined safe ranges for each of the water quality parameters as shown in Table-1 [3].

Table 1 Water parameters with safe ranges

Sr#	Parameter	Safe Range
1	pH	6.5 to 8.5
2	Turbidity	0 to 5 NTU
3	Hardness as CaCo3	500 mg/l
4	Conductance	2000 µS/cm
5	Alkalinity	500mg/l
6	Dissolved Solids	1000mg/l
7	Nitrate as NO2	<1mg/l
8	Fecal Coliform	Nil Colonies/ 100ml
9	Calcium	200mg/l

The temperature and dissolved oxygen level indicates the level of contamination, pH level shows the concentration of hydrogen ions in water [4].

Other dissolved solid particles indicate the inorganic and organic content present in water [5].

The proposed system records real time data through various sensors which is further analyzed to rank the quality of water based on the safe ranges set by WHO. For remote monitoring and remote water flow control a mobile app has been developed. For data analysis a number of machine learning algorithms have been applied to classify the quality of water as good or poor.

The details of literature survey of the existing research in this domain is discussed in Section II. Section-III describes our proposed water quality monitoring system. Experiments & results are discussed in Section-IV, and Section-V specifies the conclusion and future work.

978-1-5386-8354-5/18/$31.00 ©2018 IEEE

II. LITERATURE RIVIEW

This section reviews the background knowledge on various practical approaches and sensors available for analyzing the water quality. A case study to monitor the water quality of southern Lahore, Pakistan is presented in [6] using turbidity, pH and dissolved solid levels. Similarly, another case study is presented in [7] to assess the water quality in Southern Sindh, Pakistan using pH level, turbidity and other dissolved matters in water. Considering the water quality parameters, another water quality analysis is performed in Rawalpindi [8]. In each case of [6-8], multiple samples of water are collected and analyzed and the results obtained are very unsatisfactory. Because the calculated values of water quality parameters didn't meet the WHO standard values. A comprehensive review of water quality in Pakistan is presented in [9] in which consequences of low quality water on human health have been discussed in detail.

As trend has been moved towards IoT based solution in every field of life so during last decade, tremendous research has been noticed in IoT based systems for water quality monitoring. In [10], a real time IoT based water quality monitoring system is developed using various water quality sensors connected to the Raspberry Pi model B+ microprocessor. In [11], an effective monitoring solution is presented for the surface water quality by using a low cost sea water probe that are capable of measuring the water conductivity, temperature, chlorophyll-a and turbidity. Another IoT based system is presented in [12] using drone technology along with water sensors for monitoring quality. The real time parameters captured by the sensors are sent to the master drone through the RF modem and the IoT modem uploads it to the server in order to analyze the water pollution levels. An IoT based system for water quality monitoring is proposed in [13] in which pH sensors and temperature sensors have been used to calculate pH level and temperature level of water later to measure water quality. In [14], a cost effective water quality monitoring system is proposed in which three water quality parameters are used i.e. temperature, turbidity and pH using different sensors. Water quality analysis of waste water using wireless sensors network technology is presented in [15], in which conductivity and dissolved oxygen are also considered along with pH level and temperature to assess the quality of water, 3G communication module was used for communication between sensors and database.

Currently, machine learning models are being used worldwide to make systems more intelligent by performing predictive analysis. A predictive analysis on drinking water is presented in [16], in which water quality is ranked using pH, turbidity and dissolved solids. Linear regression model is applied on measured parameters values to calculate the correlation among calculated parameter. Similarly in [17], Fuzzy Neural network is applied on dataset which is collected for three years to measure quality of water using water quality parameters.

There are some systems working worldwide on monitoring of water contamination. EWDS [18] is an event detection system which detects anomalies in water and provides early warning of contamination in the case when deviations are observed in one or more water quality parameters from safe range. A surveillance and response dashboard [19] is another system which provides online services i.e. water monitoring and public health surveillance, system alerts if water contamination event has occurred. Canary [20] is one more framework that provides multiple services including water quality monitoring, public health surveillance and security monitoring.

As literature highlights the current situation in Pakistan where water quality is monitored only in laboratories. No proper system is available that can be deployed at domestic level to assess the quality of water and to control the flow of water. In context of Pakistan, we propose a cost effective real time system which not only monitor the water quality but also control the flow of water remotely. The next Section presents comprehensive explanation of proposed methodology.

III. PROPOSED METHODOLOGY

We propose an IoT based water quality monitoring system that comprises of hardware and software solution. The proposed hardware solution consists of micro controller (ATMega328), water quality measuring sensors such as pH, turbidity and temperature sensors, solenoid valve for flow control, and WiFi shield for sending data to cloud.

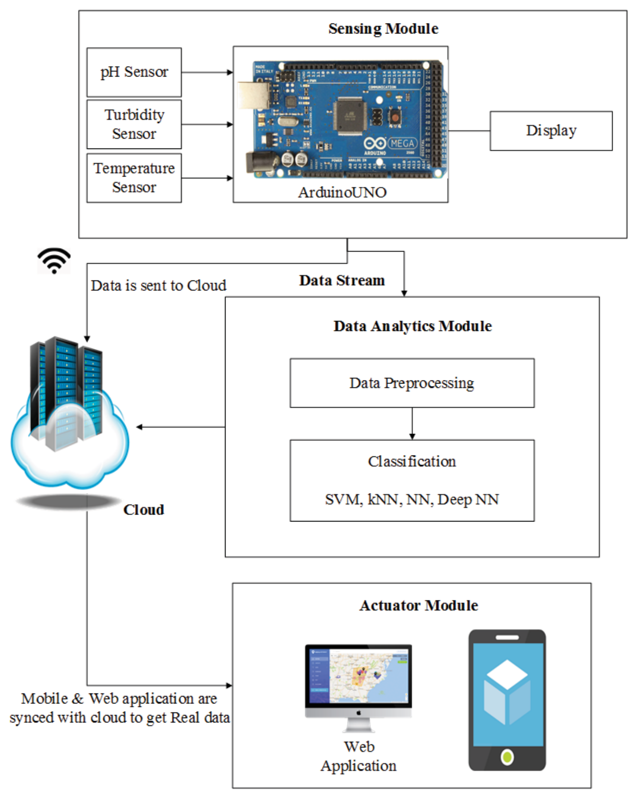

Figure 1- System Architecture

The software solution consists of a mobile app and dashboard to remotely monitor water quality and its flow control. Proposed system architecture is shown in Figure 1.

The data collected from sensors is sent to cloud for real time storage and analysis. IEEE 802.11(WiFi) protocol is used for communication between sensors and cloud. The processed data is compared against WHO defined standard safe ranges for water quality (as shown in Table 1). The comparison specifies whether the collected sample comply with WHO standard or not. The end user can visualize the assessed water quality parameters on a mobile app.

Subsequently, a data analysis is performed on dataset consisting of 667 samples gathered from 11 different water sources of Pakistan using machine learning algorithms. The proposed system is based on modular design (as shown the in Figure-1); details of each module are discussed below:

1. Sensing Module:

This module consists of three most influencing water quality measuring sensors i.e pH sensor, turbidity sensor and temperature sensor. These sensors are connected to Arduino board, for collecting data which is further sent to cloud with an interval of 30 seconds. Arduino WiFi shield has been used for the communication purposes. The connection diagram of the hardware solution is shown in Figure 2:

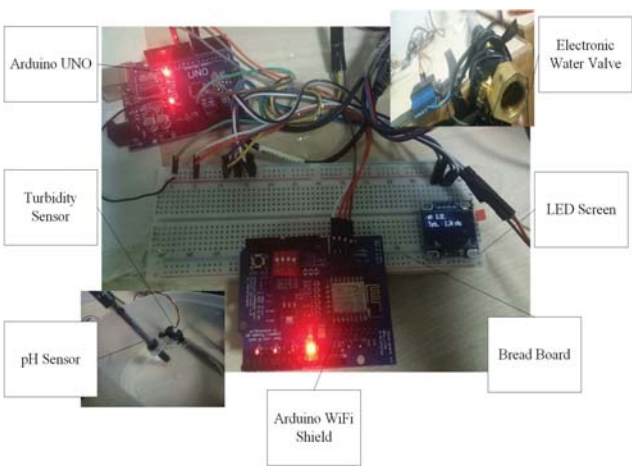

Figure 2- System Hardware Components

The pH sensor is used to measure the acidic or alkaline nature of water. Referring to Table 1, if pH > 8.5, the water is considered as alkaline. If pH < 6.5, water is considered as acidic. Another key parameter to determine the water quality is turbidity, which measures the transparency of the water and level of dissolved solids. Turbidity level ranges between 0 to 10, whereas according to WHO standards the turbidity level of drinking water should lie between 0 to 5NTU (as shown in Table 1). Ideally drinking water should have turbidity less than 2NTU, but turbidity level up to 5NTU is considerable as normal for drinking perspective. Temperature is another important parameter which directly influenced the water quality. Therefore, temperature is also monitored while calculating pH and turbidity of water. The complete flow chart of the proposed solution is shown in Figure-3:

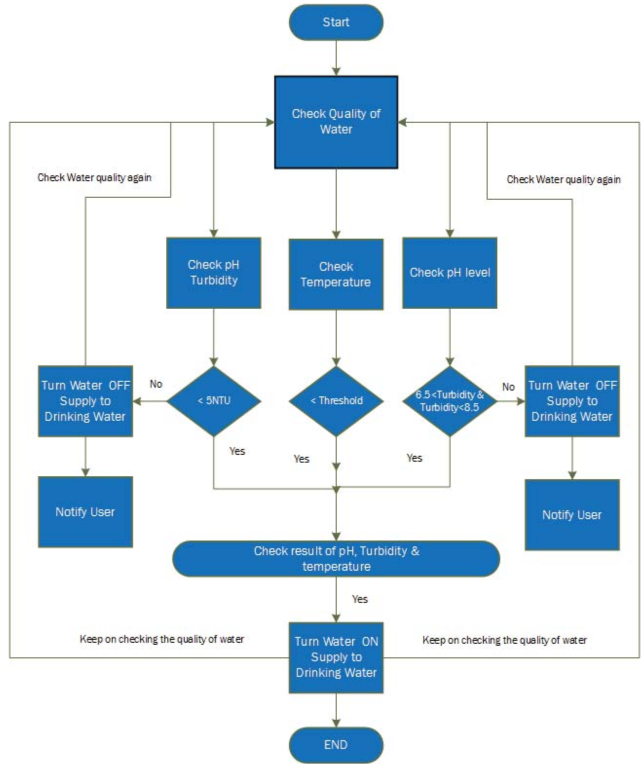

Figure 3 - Flow Diagram

2. Data Analytics Module:

The second module is data analytics module in which predictive analysis is performed on the dataset acquired from PCRWR. The dataset consists of 667 samples gathered from 11 different water sources of Pakistan with multiple water quality parameters. The water samples are used for training purpose in machine learning algorithms such as SVM, NN and kNN [21]. While using SVM, rbf kernel is applied based on the nature of our dataset. With K nearest neighbor (kNN) Euclidean distance has been applied as a similarity metric and k is set to 3. With Neural Network, single layer NN is applied with stochastic gradient descent of batch size of 28 with 15000 epochs of back propagation. To improve accuracy, deep NN [22] has been applied with 3 hidden layers excluding input and output layers with same batch size and epochs as defined for NN. First hidden layer consists of 100 nodes, second layer consists of 60 nodes, and third layer consists of 30 nodes. The 2/3 of the dataset has been used for training purposes while 1/3 has been used for testing purposes. Once the algorithm has been trained and tested, the real time data collected from the sensing module is used to test the quality of water.

3. Actuator Module:

The third module is actuator module which provides remote water flow control system along with remote monitoring. LCD screen connected to the sensing module displays the real time values of various parameters of water quality. A solenoid valve is connected to the water container to control the water flow based on water quality.

A mobile app has been developed so that the end user can remotely monitor the water quality parameters as the updated readings are continuously being transmitted to cloud after every 30 seconds. The interface of mobile application is shown in Figure-4

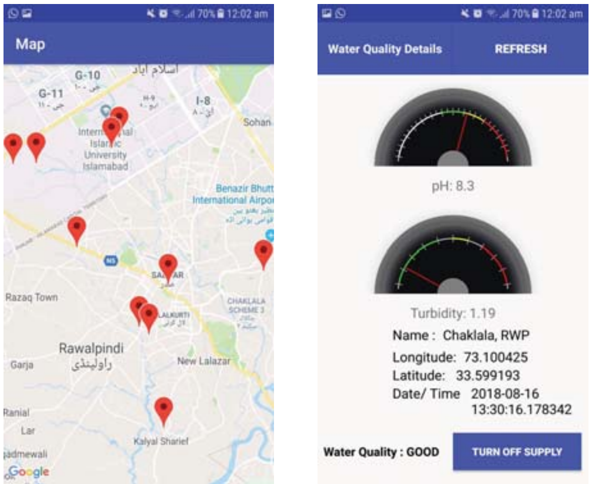

Figure 4 - Mobile Application Interface

IV. Results and Discussion

In order to train the selected machine learning algorithms, the data collected from 11 different places consists of 667 samples with 10 features as mentioned in Table-1. For classification purposes, class label is defined as binary i.e. good, poor. The dataset is labeled according to the WHO standards. For performance evaluation of these machine learning algorithms, accuracy, precision and recall have been calculated using equations 1,2,3 and presented in Table-2.

$$\text{Accuracy [23]} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$\text{Precision [23]} = TP/(TP+FP) \quad (2)$$

$$\text{Recall [23]} = TP/(TP+FP) \quad (3)$$

Where TP indicates the number of positive samples correctly classified as positive and TN shows the number of negative samples correctly classified as negative. However, FP indicates the number of negative samples incorrectly classified as positive; and FN shows the number of positive samples incorrectly classified as negative.

Accuracy is the fraction between number of correctly classified instances and number of incorrectly classified instances. Precision is the ratio between correctly classified positive instances and total positive classified instances, while recall is the ratio between correctly classified positive instances and true positive and false positive.

From Table-2 it is evident that the accuracy of deep NN is higher as compared to other classifier. This higher accuracy is achieved mainly due to additional hidden layers in deep NN which makes it more efficient to learn complex relationships in the dataset.

Table 2 - Classification Report

Parameter	SVM	NN	Deep NN	kNN
TP	160	129	145	104
FN	2	31	15	56
FP	23	10	5	15
TN	123	138	15	133
Accuracy	0.91	0.86	0.93	0.76
Precision	0.93	0.87	0.94	0.79
Recall	0.90	0.85	0.93	0.76

The algorithm which has shown an accuracy closer to deep NN is SVM. This is mainly due to the robust nature of SVM with all kinds of dataset and its key property of maintaining large margin for classification. The accuracy of single layer NN is 86% while kNN shows lowest accuracy of 76%. The other performance measures such as TP, FP, TN and FN are also presented in Table-2.

To provide the proof of concept of the proposed system, 10 different samples of water are collected from 10 different locations in Rawalpindi & Islamabad as shown in Table-3. All readings recorded from sensors are sent to server to assess the water quality, which is subsequently sent to mobile app for remote monitoring. The water is considered to be of safe quality if pH level is between 6.5 to 8.5 and turbidity level is between 0 to 5 NTU; otherwise the quality of the water is regarded as poor. The turbidity, pH and temperature sensors keep on checking the quality of water by taking readings and send them to mobile app via cloud in real time. The real time water quality results generated by our proposed system are shown in Table-3 where Long & Lat are the longitude and latitude of the water source:

Table 3- Statistics of Collected Samples

Sr #	Long	Lat	Place	pH	Turbidity	Water Quality
S1	33.577	73.039	Harley Street, RWP	7.2	2.30	Good
S2	33.642	72.979	Mohalla Riazbad, RWP	7.1	2.40	Good
S3	33.573	73.044	Tahli Mohri, RWP	7.0	2.30	Good
S4	33.594	73.054	Saddar Metro, RWP	7.9	2.32	Good
S5	33.609	73.009	Chor chok, RWP	8.0	2.31	Good
S6	33.536	73.052	Gulshanabd, RWP	7.2	1.18	Good
S7	33.599	73.100	Chaklala, RWP	8.3	1.19	Good
S8	33.653	73.030	Ravigroup I/10-3 ISB	8.1	1.22	Good
S9	33.643	73.988	SEECS, NUST, ISB	7.6	2.24	Good
S10	33.649	73.026	I/10-3 ISB	7.0	2.32	Good

In the case the quality of the water declines and breaches the defined safe range, the system will automatically turn off the water supply. In addition to this, the mobile app also provides a control to the end user to turn OFF the water supply by clicking the 'Turn ON/OFF Supply' button. The interface of the mobile application is shown in Figure-4, where left screen shows the location of the water sources and right screen shows the interface of the water quality statistics.

V. Conclusion & Future work

We proposed an IoT based solution to monitor the water quality in real time. The proposed system provides remote monitoring of water quality assessment along with water flow control via a mobile app. Four machine learning algorithms including Support Vector Machine (SVM), k Nearest Neighbor (kNN), single layer neural network and deep neural network have been applied for classification of water quality and experimental results revealed that deep neural network outperforms all other algorithms with an accuracy of 93%.

To exploit the full potential of the proposed system, it will be extended to a commercial product that can be practically deployed as a decision support system in the industrial sector, water quality monitoring stations and for domestic use. The business value of the product lies in measuring the quality of water and making a decision to control the water flow.in real time. This system has the potential to effectively utilize to overcome the challenges of water quality in agriculture sector and various industries.

Acknowledgement

This project has been funded by School of Electrical Engineering and Computer Science, National University of Science and Technology, Islamabad Pakistan and dataset is provided by PCRWR.

References

[1] Asia, Draft South. "Water vision 2025, Country Report, Pakistan." (2000).

[2] Kahlown, M. A., et al. "Water quality status, national water quality monitoring programme." Fourth Technical ReportPCRWR 5 (2006).

[3] Ali, Maqbool, and Ali Mustafa Qamar. "Data analysis, quality indexing and prediction of water quality for the management of rawal watershed in Pakistan." Digital Information Management (ICDIM), 2013 Eighth International Conference on. IEEE, 20

[4] Tyagi, Shweta, et al. "Water quality assessment in terms of water quality index." American Journal of Water Resources1.3 (2013): 34-38.

[5] World Health Organization. Guidelines for drinking-water quality: second addendum. Vol. 1, Recommendations. World Health Organization, 2008..

[6] Haydar, S., M. Arshad, and J. A. Aziz. "Evaluation of drinking water quality in urban areas of Pakistan: A case study of Southern Lahore." Pakistan Journal of Engineering and Applied Sciences (2016).

[7] Memon, Mehrunisa, et al. "Drinking water quality assessment in Southern Sindh (Pakistan)." Environmental monitoring and assessment 177.1-4 (2011): 39-50.

[8] Hashmi, Imran, Shaukat Farooq, and Sara Qaiser. "Chlorination and water quality monitoring within a public drinking water supply in Rawalpindi Cantt (Westridge and Tench) area, Pakistan." Environmental monitoring and assessment 158.1-4 (2009): 393.

[9] Azizullah, Azizullah, et al. "Water pollution in Pakistan and its impact on public health—a review." Environment International37.2 (2011): 479-497.

[10] Vijayakumar, N., & Ramya, R. (2015, March). The real time monitoring of water quality in IoT environment. In Circuit, Power and Computing Technologies (ICCPCT), 2015 International Conference on (pp. 1-4). IEEE.

[11] Adamo, F., Attivissimo, F., Carducci, C. G. C., & Lanzolla, A. M. L. (2015). A smart sensor network for sea water quality monitoring. IEEE Sensors Journal, 15(5), 2514-2522.

[12] Agarwal, A., Shukla, V., Singh, R., Gehlot, A., & Garg, V. (2018). Design and Development of Air and Water Pollution Quality Monitoring Using IoT and Quadcopter. In Intelligent Communication, Control and Devices (pp. 485-492). Springer, Singapore.

[13] Rose, K., S. Eldridge, and L. Chapin. "The Internet of Things: An Overview–Understanding the Issues and Challenges of a More Connected World. The Internet Society (ISOC)." (2015).

[14] Faruq, Md Omar, et al. "Design and implementation of cost effective water quality evaluation system." Humanitarian Technology Conference (R10-HTC), 2017 IEEE Region 10. IEEE, 2017.

[15] Siregar, Baihaqi, et al. "Monitoring quality standard of waste water using wireless sensor network technology for smart environment." ICT For Smart Society (ICISS), 2017 International Conference on. IEEE, 2017.

[16] Krishnan, K. Sri Dhivya, and P. T. V. Bhuvaneswari. "Multiple linear regression based water quality parameter modeling to detect hexavalent chromium in drinking water." Wireless Communications, Signal Processing and Networking (WiSPNET), 2017 International Conference on. IEEE, 2017.

[17] Sun, Hongxiang, and Yuanhua He. "Research and Application of Water Quality Evaluation of a Certain Section of Yangtze River Based on Fuzzy Neural Network." Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), 2017 International Conference on. IEEE, 2017.

[18] Ostfeld, Avi, and Elad Salomons. "An early warning detection system (EWDS) for drinking water distribution systems security." World Water & Environmental Resources Congress 2003. 2003.

[19] Hasit, Yakir, and John Vogtman. "Benefits Of A Surveillance And Response System (SRS) Dashboard-The Philadelphia Water Department Experience." (2014).

[20] Hart, David, et al. "CANARY: a water quality event detection algorithm development tool." World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat. 2007.

[21] Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." Emerging artificial intelligence applications in computer engineering 160 (2007): 3-24.

[22] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." Neural networks 61 (2015): 85-117.

[23] Sokolova, Marina, and Guy Lapalme. "A systematic analysis of performance measures for classification tasks." Information Processing & Management 45.4 (2009): 427-437.