Anomaly detection in water quality monitoring: a preliminary solution

L. Ferrigno, L. Gerevini, M. Molinara
Dept. of Electrical and Information Engineering
University of Cassino and Southern Lazio
Cassino, Italy
{ferrigno; luca.gerevini; m.molinara}@unicas.it

G. Cerro
Dept. of Medicine and Health Sciences
University of Molise
Campobasso, Italy
gianni.cerro@unimol.it

C. Bourelly
Sensichips srl
Aprilia, Italy
carmine.bourelly@sensichips.com

***Abstract*—In smart city framework, the water monitoring through an efficient, low–cost, low–power and IoT–oriented sensor technology is a crucial aspect to allow, with limited resources, the analysis of contaminants eventually affecting wastewater. In this sense, common interfering substances, as detergents, cannot be classified as dangerous contaminants and should be neglected in the classification. By adopting classical machine learning approaches having a finite set of possible responses, each alteration of the sensor baseline is always classified as one out of the predetermined substances. Consequently, we developed an anomaly detection system based on one-class classifiers, able to discriminate between a recognized set of substances and an interfering source. In this way, the proposed detection system is able to provide detailed information about the water status and distinguish between harmless detergents and dangerous contaminants.**

***Index Terms*—water monitoring, sensor network, anomaly detection, machine learning**

I. Introduction

Sensing technologies for water pollution monitoring have been widely studied in the scientific literature, e.g. see the extensive reviews in [1]. Measurements from sensors are usually based on Electrochemical Impedance Spectroscopy (EIS) [2], a frequency domain technique that evaluates the response of an electrochemical cell to a low amplitude sinusoidal perturbation. After measurement acquisition, data analysis for contaminants classification is principally based on machine learning methods. For example, in [3], Artificial Neural Network and Principal Component Regression are used to estimate nitrate concentration in ground water.

To ensure a widespread diffusion of pollutants detection systems in wastewater, low cost, compactness and low consumption (which would make them suitable even for low maintenance) become fundamental. An ideal monitoring system should be able to detect and/or distinguish between thousands of different substances, even in very complex conditions of high water turbidity, sensor degradation, which can quickly become important, and so on. Furthermore, such a system should be able to distinguish polluting substances also on the basis of their dangerousness level and treatment processes. This work is part of a project where the ultimate goal is the creation of an end-to-end identification system (from sensing to classification) capable of detecting a predefined set of substances commonly considered as dangerous and indicative of an anomalous use of water. The proposed measurement system is based on a proprietary embedded IoT–ready Micro-Analytical Sensing Platform (MASP henceforth) of size 1.5×1.5 mm and 1.5 mW power absorption [4].

Due to time and cost issues of collecting measures for any substance that could be found in wastewater, the faced machine learning problem should be considered as an "open-set recognition" problem, i.e., a classification problem where the training data only gives a partial view of the application domain. A first step to solve such a problem is the detection of unknown substances before the classification of the known dangerous contaminants. Thus, in this paper, stemming from our past experience in sensors and data processing [4], we present an anomaly detection system based on one class classifiers able to discriminate between a known set of substances and an interfering source so as to reduce or avoid wrong classifications or false alarms. Experiments have been conducted on a dataset realized in our laboratories to show that the proposed system is able to provide detailed information about the water status and distinguish between harmless detergents and dangerous contaminants.

II. The SENSIPLUS ecosystem

The overall system, namely SENSIPLUS, is reported in Fig. 1 and is composed of a sensor layer (also known as Smart Cable Water (SCW)) and a processing sub-system, namely Sensiplus Deep Machine (SDM).

SENSIPLUS is a proprietary technology of Sensichips s.r.l. developed in collaboration with the University of Pisa. Specifically, to work with wastewater, the system has been customized by hosting it on a printed circuit board endowed with both measuring and sensing capabilities.

The physical principle adopted to achieve the goal is the electrical impedance. By the analysis of the single components, it is possible to see: the SCW, endowed with 6 Interdigitated Electrodes (IDEs) metalized through different materials; the processing object (SDM), composed of: (i) an ESP8266 Micro Controller Unit running a customized software for data conveying; and (ii) a host controller .

TABLE I
TPR RESULTS USING BEST HYPERPARAMETERS.

Classifier	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10	Mean	SD
OCCSVM	85.32	100.0	99.82	98.06	100.0	87.99	100.0	64.61	82.54	70.39	88.87	12.49

TABLE II
TNR RESULTS ON TEST SET.

	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10	Mean	SD
OCCSVM	99.67	99.67	99.67	99.67	99.67	99.69	99.67	99.69	99.68	99.68	99.68	0.008

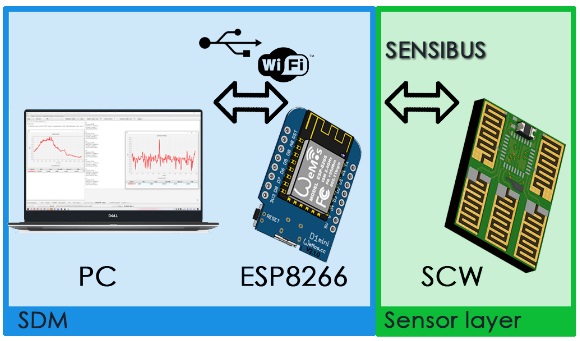

Fig. 1. The adopted acquisition and processing system

A. Classification

Our final goal is to build a multiclass classifier system with the capability of distinguish among a known set of substances flowing through the wastewater. Given the complexity of the real scenario where every day there are plenty of spills of different nature (e.g. organic materials, soaps, industrial products etc.), we have to face a problem known as ”open-set recognition”. In a real context, we must avoid to input samples of unknown classes to our final multiclass classifier system, otherwise we will end up by giving the wrong results, which would make our system useless. Accordingly, here we propose the use of One-Class classifiers for the Anomaly Detection, used as a kind of false positive reduction system.

In this paper we propose a Novelty Detection system based on One-Class Classifier SVM [5] where the training set contains only positive samples and we are interested to find out if a new sample is an outlier (aka novelty) [6];

B. Experiments

The dataset consists of 10 data acquisitions for each of the 9 substances of interest (8 pollutants plus water), totaling 90 data acquisitions carried out as in [4].

To evaluate the ability of our classifiers to correctly predict the class of a new unseen sample we used 10–fold cross-validation. In each of the 10 iterations, 9 acquisitions (one for each substance) were used as validation set, and the remaining 81 acquisitions (9 for each substance) as training set.

The performance of each classifier was evaluated on a separate test set composed of 10,000 samples of 3 interfering substances, for a total of 30,000 test samples.

As performance evaluation metrics, we computed the True Positive Rate for the validation set I and the True Negative Rate for the test set II.

III. CONCLUSIONS

In this paper an anomaly detection in a water monitoring system has been presented. For the experimental phase three substances have been considered as an outlier: washing machine detergent, dishwasher detergent and sodium chloride. The TNR obtained on test set (99.68 % with OCCSVM) gives the evidence that the proposed approach is promising.

IV. ACKNOWLEDGMENTS

SENSIPLUS is a proprietary technology of Sensichips developed in collaboration with Prof. Paolo Bruschi team of University of Pisa. The research leading to these results has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement SYSTEM No. 787128. The authors are solely responsible for it and that it does not represent the opinion of the Community and that the Community is not responsible for any use that might be made of information contained therein. This work was also supported by MIUR (Minister for Education, University and Research, Law 232/216, Department of Excellence).

REFERENCES

[1] S. N. Zulkifli, H. A. Rahim, and W.-J. Lau, “Detection of contaminants in water supply: A review on state-of-the-art monitoring technologies and their applications,” *Sensors and Actuators B: Chemical*, vol. 255, pp. 2657 – 2689, 2018.

[2] A. M. Syaifudin, K. Jayasundera, and S. Mukhopadhyay, “A low cost novel sensing system for detection of dangerous marine biotoxins in seafood,” *Sensors and Actuators B: Chemical*, vol. 137, no. 1, pp. 67–75, 2009.

[3] G. Charulatha, S. Srinivasalu, O. Uma Maheswari, T. Venugopal, and L. Giridharan, “Evaluation of ground water quality contaminants using linear regression and artificial neural network models,” *Arabian Journal of Geosciences*, vol. 10, no. 6, p. 128, Mar 2017.

[4] A. Bria, G. Cerro, M. Ferdinandi, C. Marrocco, and M. Molinara, “An iot-ready solution for automated recognition of water contaminants,” *Pattern Recognition Letters*, vol. 135, pp. 188–195, 2020.

[5] M. Amer, M. Goldstein, and S. Abdennadher, “Enhancing one-class support vector machines for unsupervised anomaly detection,” in *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, ser. ODD ’13. New York, NY, USA: Association for Computing Machinery, 2013, p. 8–15. [Online]. Available: https://doi.org/10.1145/2500853.2500857

[6] B. Lamrini, A. Gjini, S. Daudin, P. Pratmarty, F. Armando, and L. Travé-Massuyès, “Anomaly detection using similarity-based one-class svm for network traffic characterization,” in *DX@Safeprocess*, 2018.