2022 IEEE International Conference on Smart Computing (SMARTCOMP) | 978-1-6654-8152-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/SMARTCOMP55677.2022.00070

A new Dataset for Detection of Illegal or Suspicious Spilling in Wastewater through Low-cost Real-time Sensors

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

Andrea Ria
Italian National Council of Reasearch (CNR)
Pisa, Italy
andrea.ria@ieiit.cnr.it

F.Magliocca, L.Ruscitti, R. Simmarano,
Sensichips srl
Aprilia, Italy
{francesco.magliocca; luca.ruscitti; roberto.simmarano}@sensichips.com

A.Trynda, P.Olejnik
Central Forensic Laboratory of the Police
Poland
{piotr.olejnik, anna.trynda}@policja.gov.pl

***Abstract*—The spilling of suspicious or illegal substances in wastewater poses a serious global threat to human health. Low-cost sensor technologies enabling wastewater continuous monitoring are an important tool that can help face this problem. In this paper, electrical impedance measurements on different sensors are proposed in order to have a dataset of raw data to be used to perform the classification of possible contaminants in a water environment. In detail, the sensor technology is based on a proprietary multi-sensing platform called SENSIPLUS, which is arranged in a suitable set-up able to carry out measurements in water for prolonged times. Sensors metalized with different materials are jointly used to exploit sensitivity diversity to different contaminants. An ad-hoc measurement procedure has been designed, including data acquisition during the warm-up period, contaminant injection, and steady-state conditions. The dataset proposed in this paper has been acquired in different European laboratories (Italy and Poland) and is made publicly available for testing new data analysis or machine learning techniques for the detection and classification of ten wastewater dangerous "contaminants" (https://aida.unicas.it/icprchallenge2022/).**

***Index Terms*—Wastewater, Spilling Detection, Machine Learning, Pollution Monitoring, Smart City, IoT**

I. Introduction

Human settlements produce a huge amount of waste, a considerable part of them are discarded through the sewage network which became a collector of domestic, industrial, and agricultural pollutants. In addition, there is the fraudulent spilling phenomena that contribute to increase quickly the pollutant concentration, with devastating consequences on the environment, and with a weak probability of identifying the criminals. As a result, the wastewater contamination, by different types of organic and inorganic substances, represents a hazard to public health and the environment. In the Smart Cities framework, a lot of efforts are directed to solve these problems by both reducing the amount of waste (for example by adopting better behaviors) and by effetely supervising the sewage network in order to map the wastewater collection system and detect sources of pollution within a metropolitan area, in order to optimize water treatment systems or identify anomalies in wastewater, which could also indicate fraudulent spills. Sensing technologies for water pollution monitoring have been widely studied in the scientific literature, e.g. see the extensive reviews in [1]–[5] for detect changes in wastewater composition. This kind of sensors doesn't directly measure the pollutant concentration because, in general, the precise identification requires laboratory instrumentations that are not compatible with monitoring specifications in terms of cost and simplicity. Data collected with a different type of measurements (like conductivity, impedance, or turbidity) are more suitable for the monitoring application. The main disadvantage of this approach is that the data are not easy to read due to the number of sensors or the not perfect correlation between the sensor value and the pollutant. The adopted methods mainly focus on the application of electrodes of different metals [6] or covered by sensing films [7], and optical sensors [8]. Measurements from sensors are usually based on Electrochemical Impedance Spectroscopy (EIS) [9]–[11], a frequency domain technique that evaluates the response of an electrochemical cell to a low amplitude sinusoidal perturbation. The joining of a large number of sensors, not easy understandable data, not perfect correlation between sensor value and pollutant could make difficult extract valuable knowledge from measured data. Machine Learning (ML) or Deep Learning (DL) approaches could help in this way. To extract the information from such amount of data particularly useful is the Machine Learning (ML) or Deep Learning (DL) approach. In literature ML applications are reported in [12], Artificial Neural Network and Principal Component Regres-

2693-8340/22/$31.00 ©2022 IEEE
DOI 10.1109/SMARTCOMP55677.2022.00070

sion are used to estimate nitrate concentration in groundwater, whereas in [6] partial least square discriminant analysis is applied to detect explosive precursors in wastewater. Also, DL solutions have been proposed, e.g. in [13] Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN) were adopted for detection and classification of chemicals in seawater. As well known, training and testing ML or DL algorithms require a considerable volume of data that are not easy to collect due to the need to design and realize a sensor network able to collect data in situ or in a controlled environment with reliable and labeled data. The proposed dataset, joined with others in literature [14] [15], helps the growth of prevention or intervention algorithms useful for limiting the harmful impact on the environment. With this aim, a Pollutant Wastewater dataset (PoWds) has been acquired and released to support computational sustainability researchers to test their models, systems, or algorithms. PoWds include a selection of impedance component measurements of 5 sensors with eleven toxic substances dissolved in wastewater medium. Each pollutant is measured ten times in order to evaluate also the measurement repeatability and allows the training and test of the algorithms.

Similar datasets collected with the same sensor platform has been used in different papers [16]–[24].

In section II the measuring setup is described, the acquired dataset with a focus on experiment structure is reported in section III. In Sections IV and V, there are the guidelines for correctly dividing the dataset into training-dataset and test-dataset, and for correctly evaluating the algorithms performance. In section VI some conclusions are drawn.

II. Setup

The system architecture of the acquisition system is composed of two layers:

(i) a sensing layer based on the SENSIPLUS chip (customized in a commercial solution called Smart Cable Water – SCW) with different interdigitated electrodes metalized through different materials. SENSIPLUS is a proprietary technology of Sensichips developed in collaboration with Prof. Paolo Bruschi team of the University of Pisa.

(ii) a data collection and communication layer with both hardware and software components.

A. Smart Cable Water

The SCW is based on a proprietary multi-sensor microsystem able to monitor the presence of toxic chemicals (TICs), pollutants, hydrocarbons, and organics in water. At the core of SCW there is SENSIPLUS (a chip of size 1.5 × 1.5 mm and 1.5 mW), the Sensichips micro-sensor platform (see fig. 1) that can interrogate on-chip and off-chip sensors with its versatile and accurate Electrical Impedance Spectrometer (EIS) and Potentiostat.

Analytics performed with EIS allows exploiting RedOx dynamics of catalytic noble metals to aid chemical discrimination plus measurement of conductivity and permittivity spectra.

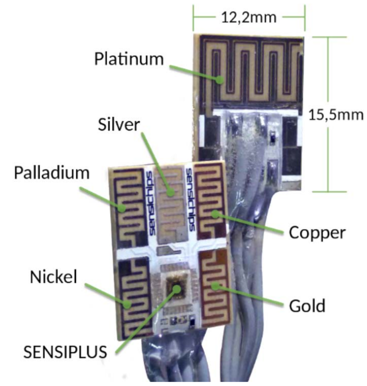


Fig. 1: Smart Cable Water.

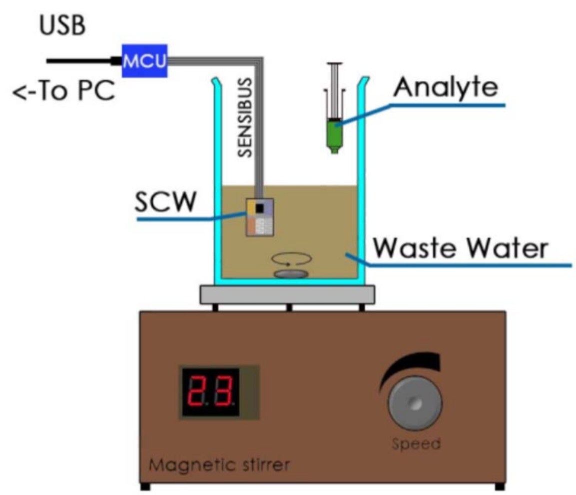


Fig. 2: Measurement setup

The SCW can be flooded in water since the measuring part is protected by a transparent acrylic resin. The SCW board includes several sensors, on the front face there are five small (estimated surface $9.95\ mm^2$ each) Interdigitate Electrodes (IDEs) made of Copper and metalized with Gold, Copper (oxide), Nickel, Silver, Palladium. Back board has only one big (estimate surface $19.61\ mm^2$) IDE still copper-based and metalized with metal platinum.

Measures are performed with five of the six SCW's IDEs: Gold, Platinum, Copper, Nickel, Silver. For the first two sensing elements, two frequencies are evaluated: 78125 Hz, and 200 Hz. For the other IDEs, only one frequency (200 Hz) is measured. Measure two different frequencies allow the evaluation of diverse effects, because the interaction between the pollutant and the sensors depends on the stimulus frequency.

B. Measurement setup

The measurement setup, illustrated in 2, is composed of the measurement chain (SCW, Micro Control Unit, PC), a magnetic stirrer, 300 ml beaker in which is injected the analyte.

A magnetic stirrer equipped with a 25 mm anchor simulates the water movement: all measurements are performed under the same water stirring conditions. Anchor rotation is set to 50 rpm. At this speed, 100 ml of water does not incorporate air bubbles, which causes noisy measurement, but the water movement still emulates the sewage working conditions. In order to have any measure independent from the others, between each measure, cleaning beaker anchor and sensors is fundamental. All equipments have been cleaned with soap and freshwater, sometimes the beaker requires a hard cleaning with concentrated 95% sulphuric acid to remove the wastewater residues.

C. Adopted sensors

Each acquired sample contains an array of 12 values proportional to resistance and capacitance of the IDEs, evaluated at two distinct stimulus frequencies (200 Hz and 78 kHz) so as to exploit different sensor frequencies responses.

At the lower frequency, we measured resistance and capacitance with five IDEs (Platinum, Gold, Silver, Nickel, and Copper), whereas at the higher frequency we measured only resistance with Gold and Platinum IDEs. These frequencies have been chosen by considering the Randles equivalent circuit for the IDE electrodes immersed in wastewater.

III. Dataset

All acquisitions were made between 2019 and 2021 with different sensors and in two different laboratories (Polish, Italian) and are available in [25].

The dataset contains measures obtained on 10 substances plus the background (the wastewater – WW):

The measurement protocol has been divided into two steps:

- the first 600 samples are acquired in warm-up mode, in this period of time sensors are exposed to WW only;
- the next 1000 samples are acquired after analyte injection.

In this way, each acquisition contains 1,600 samples where the first 600 samples are measured in WW and the remaining 1000 samples are measured with the analyte mixed to WW. We can observe that obviously in the case of the 10 acquisitions of WW no substance has been injected during the entire acquisition and so all the 1600 samples are measured in only WW.

The acquisition of each sample requires about 1.6 seconds and consequently, each cycle is about 40 minutes long. The dataset consists of 10 data acquisitions for each of the 11 substances (10 pollutants plus WW) carried out with the measurement protocol previously described.

In the Figure 3 is reported an example of the signals measured for different substances (ACETONE, HYDROCHLORICACID – HCL, etc.), for different frequencies (200Hz, 78kHz) in the case of PLATINUM IDE. The step in the middle of each acquisition corresponds to the injection of the substance. Let's make some observations about the temporal trend of the measurements:

- often, the value of the baseline (the starting point of the curve) changes between different acquisitions: this phenomenon is partially related to poisoning, aging, and to the differences in sensors;
- the shape of the curves after the injection time could be very different among different experiments;
- the curve should not be interpreted as a "time series", because the slope after the injection and in general its shape, is strongly related to the speed of the injection of the substance that in the real scenario could have huge variability.

IV. Training and Test

For each substance, 9 acquisitions have been added to the training set and 1 has been added to the test set. The following table report the number of samples for each substance in the training and the test set:

In particular, for wastewater, in the training set there are:

- 600 samples coming from warm-up phase, for each of the 10 pollutants and for each of the 9 acquisitions (for a total of 54000 samples);
- 1600 samples coming from 9 acquisitions made only in wastewater (for a total of 14400 samples);

while in the test set there are:

- 600 samples coming from warm-up for each of the 10 substances and from 1 acquisition (for a total of 6000 samples);
- 1600 samples coming from 1 acquisition made only in wastewater.

Each acquisition is contained in a .csv file named as follow: 1_Experiment_19-11-2019_16-15_GLOBAL_L1_7N6_SWW_ACETICACID_1_ADC.csv

In particular the name contains:

- The number of the experiment: 1_Experiment
- Date and hour of the acquisition: 19-11-2019_16-15
- Id of the adopted sensor: L1_7N6
- Substance: ACETICACID (this name will not be present in the test set!)
- Data type: ADC $\longrightarrow$ Analog to Digital Converter values

while the data column follows the structure showed in the "Feature name" columns reported in the Table I. For what concern the data organization, training data are stored in a folder that contains a sub-folder for each substance. Each sub-folder contains 9 files, one for each acquisition (see Figure 4 for more details).

The test data, as well as training data, are stored into a folder that contains a sub-folder for each substance, while each sub-folder contains only 1 file (a single acquisition) (see Figure 5 for more details).

As shown in Figure 6 each row of the .csv file, represents a sample containing a value for all the features (as reported in Table 2). Each column represents the trend over time of the single feature. Please remember that the first 600 rows are measured in only wastewater, while the remaining 1000 samples are measured with the current substance (ACEDICACID in this example, as reported in the filename). The total number of rows for each file is equal to 1600 and the substance has been injected at the sample 600.

TABLE I: Selected features (sensors)

	IDE	Frequency of Acqusition	Physical values	Feature name
1	PLATINUM	78kHz	RESISTANCE	OFFCHIP_PLATINUM_78kHz_RESISTANCE
2	GOLD	78kHz	RESISTANCE	OFFCHIP_GOLD_78kHz_RESISTANCE
3	PLATINUM	200Hz	RESISTANCE	OFFCHIP_PLATINUM_200Hz_RESISTANCE
4	PLATINUM	200Hz	CAPACITANCE	OFFCHIP_PLATINUM_200Hz_CAPACITANCE
5	GOLD	200Hz	RESISTANCE	OFFCHIP_GOLD_200Hz_RESISTANCE
6	GOLD	200Hz	CAPACITANCE	OFFCHIP_GOLD_200Hz_CAPACITANCE
7	COPPER	200Hz	RESISTANCE	OFFCHIP_COPPER_200Hz_RESISTANCE
8	COPPER	200Hz	CAPACITANCE	OFFCHIP_COPPER_200Hz_CAPACITANCE
9	SILVER	200Hz	RESISTANCE	OFFCHIP_SILVER_200Hz_RESISTANCE
10	SILVER	200Hz	CAPACITANCE	OFFCHIP_SILVER_200Hz_CAPACITANCE
11	NICKEL	200Hz	RESISTANCE	OFFCHIP_NICKEL_200Hz_RESISTANCE
12	NICKEL	200Hz	CAPACITANCE	OFFCHIP_NICKEL_200Hz_CAPACITANCE

TABLE II: Details of chemical composition of substances

	Substances	Brief description
1	WASTEWATER	Real domestic sewage $pH = 7.4$, conductivity = 1.341 mS
2	ACETICACID	CH_3COOH
3	ACETONE	C_3H_6O
4	AMMONIA	NH_3
5	ETHANOL	C_2H_5OH
6	FORMICACID	CH_2O_2
7	HYDROCHLORICACID	HCl
8	HYDROGENPEROXIDE	H_2O_2
9	PHOSPHORICACID	H_3PO_4
10	SODIUM_HYPOCHLORITE	$NaClO$
11	SULPHURICACID	H_2SO_4

TABLE III: Number of samples per substances

	Substances	Samples in training set	Samples in test set
1	WASTEWATER	**68400** $(600 * 10 * 9 + 1600 * 9)$	**7600** $(600 * 10 * 1 + 1600 * 1)$
2	ACETICACID	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
3	ACETONE	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
4	AMMONIA	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
5	ETHANOL	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
6	FORMICACID	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
7	HYDROCHLORICACID	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
8	HYDROGENPEROXIDE	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
9	PHOSPHORICACID	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
10	SODIUM_HYPOCHLORITE	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
11	SULPHURICACID	**9000** $(1000 * 9)$	**1000** $(1000 * 1)$
	Total samples	**158400**	**17600**

V. Evaluation

The proposed dataset represents a multi-class classification problem and challenges the machine learning researcher community. The suggested performance measure of the obtained results is the Matthews Correlation Coefficient MCC for multi-classification (also called Rk statistic) that should be of course evaluated on the test-set.

Starting from these definitions:

- **True positive (TP) –** The number of correctly identified samples. The number of samples measured in presence of one of the 10 substances of interest.
- **True negative (TN) –** The number of correctly identified negative samples, i.e., samples measured in wastewater.
- **False positive (FP) –** The number of wrongly identified samples, i.e., a commonly called a "false alarm". The number of samples classified as one of the 10 substances of interest but measured in wastewater.
- **False negative (FN) –** The number of wrongly identified negative samples. The number of sample classified as wastewater but measured in presence of some of the 10 substances of interest.

The MCC is:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

VI. Conclusions

This paper presents a new dataset for detecting illegal or suspicious spilling in wastewater through low-cost real-time sensors. The dataset, collected in two different laboratories in Europe, is available for download at the following URL [25] and available for free for research. It is important to notice that water quality usually is monitored by laboratory analyses, performed by experienced professionals who use complex and expensive tools. With this kind of approache, due to the costs

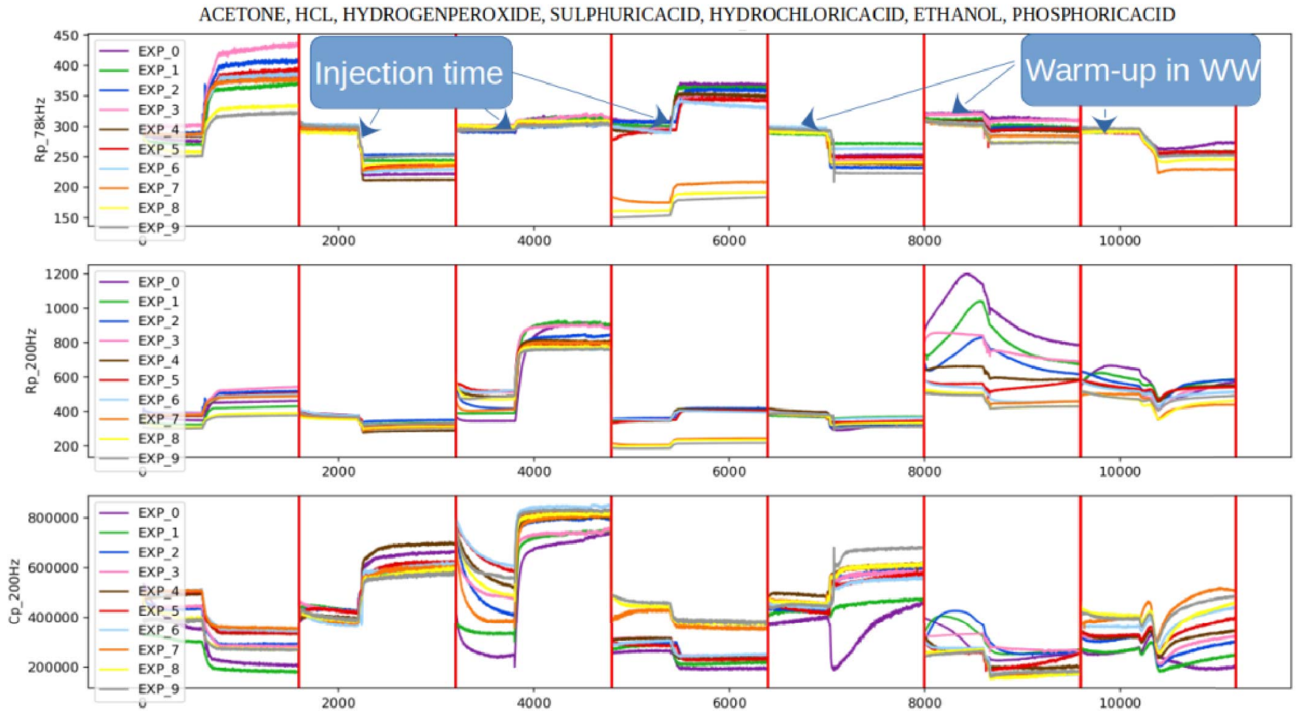


Fig. 3: Example of acquired signals.

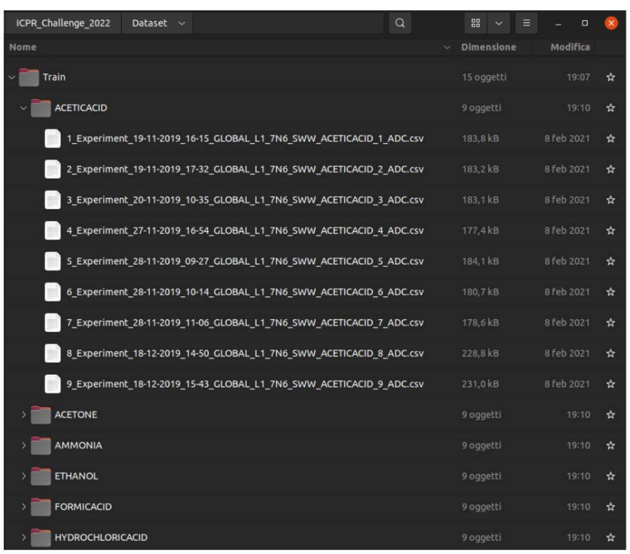


Fig. 4: File System for training.

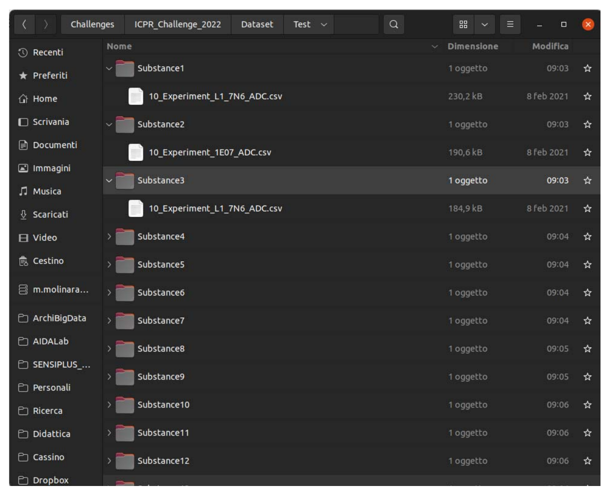
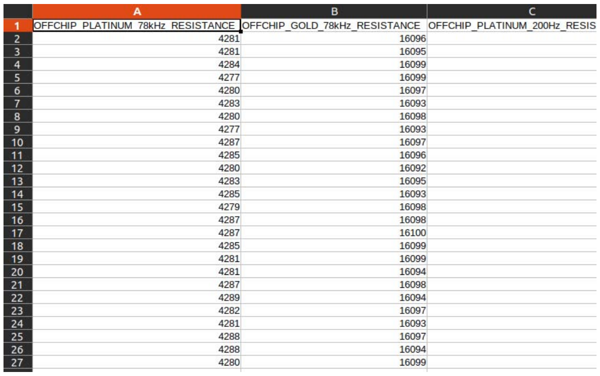


Fig. 5: File System for test.

and the time necessary for the water quality analysis, it is pretty difficult to prevent environmental disasters.

In order to improve the efficiency of the prevention activity it's needed an effective and reactive water analysis through a large number of distributed measurement systems. In this context, low-cost microsensors capable to combine a good measurement accuracy as well as good reliability, represent one of the best solution. From the major benefit of this kind of systems, there is the possibility to use the paradigm of the Internet of Things (IoT) and consequently the application of the Artificial Intelligence (AI) techniques to analyze and exploit the information contained in the generated data.

In this sense the system presented in this work, perfectly meets the requirements that a low-cost microsensor system should have in order to build an effective and reactive distributed measurement system. Finally one of the main aim



	A	B	C
1	OFFCHIP_PLATINUM_78kHz_RESISTANCE	OFFCHIP_GOLD_78kHz_RESISTANCE	OFFCHIP_PLATINUM_200Hz_RESIS
2	4281	16096	
3	4281	16095	
4	4284	16099	
5	4277	16099	
6	4280	16097	
7	4283	16093	
8	4280	16098	
9	4277	16093	
10	4287	16097	
11	4285	16096	
12	4280	16092	
13	4283	16095	
14	4285	16093	
15	4279	16098	
16	4287	16098	
17	4287	16100	
18	4285	16099	
19	4281	16099	
20	4281	16094	
21	4287	16098	
22	4289	16094	
23	4282	16097	
24	4281	16093	
25	4288	16097	
26	4288	16094	
27	4280	16099	

Fig. 6: An example of CSV file for training.

of the presented dataset is to improve the benefit, in terms of accuracy and reliability, of the application of the Artificial Intelligence techniques in the field of water quality analysis.

VII. Acknowledgment

The research leading to these results has received funding from the European Unions Horizon 2020 research and innovation program 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 the 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. Zhuiykov, "Solid-state sensors monitoring parameters of water quality for the next generation of wireless sensor networks," *Sensors and Actuators B: Chemical*, vol. 161, no. 1, pp. 1 – 20, 2012.

[2] M. H. Gholizadeh, A. M. Melesse, and L. Reddi, "A comprehensive review on water quality parameters estimation using remote sensing techniques," *Sensors*, vol. 16, no. 8, p. 1298, 2016.

[3] 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.

[4] X. Qin, F. Gao, and G. Chen, "Wastewater quality monitoring system using sensor fusion and machine learning techniques," *Water research*, vol. 46, no. 4, pp. 1133–1144, 2012.

[5] W. Bourgeois and R. Stuetz, "Use of a chemical sensor array for detecting pollutants in domestic wastewater," *Water research*, vol. 36, no. 18, pp. 4505–4512, 2002.

[6] C. Desmet, A. Degiuli, C. Ferrari, F. S. Romolo, L. Blum, and C. Marquette, "Electrochemical sensor for explosives precursors' detection in water," *Challenges*, vol. 8, no. 1, 2017.

[7] J. K. Atkinson, M. Glanc, M. Prakorbjanya, M. Sophocleous, R. P. Sion, and E. Garcia-Breijo, "Thick film screen printed environmental and chemical sensor array reference electrodes suitable for subterranean and subaqueous deployments," *Microelectronics International*, Apr. 2013.

[8] T. P. Lambrou, C. C. Anastasiou, C. G. Panayiotou, and M. M. Polycarpou, "A low-cost sensor network for real-time monitoring and contamination detection in drinking water distribution systems," *IEEE Sensors Journal*, vol. 14, no. 8, pp. 2765–2772, Aug 2014.

[9] 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.

[10] X. Li, K. Toyoda, and I. Ihara, "Coagulation process of soymilk characterized by electrical impedance spectroscopy," *Journal of Food Engineering*, vol. 105, no. 3, pp. 563–568, 2011.

[11] P. Geng, X. Zhang, W. Meng, Q. Wang, W. Zhang, L. Jin, Z. Feng, and Z. Wu, "Self-assembled monolayers-based immunosensor for detection of escherichia coli using electrochemical impedance spectroscopy," *Electrochimica Acta*, vol. 53, no. 14, pp. 4663 – 4668, 2008.

[12] 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.

[13] S. N. Dean, L. C. Shriver-Lake, D. A. Stenger, J. S. Erickson, J. P. Golden, and S. A. Trammell, "Machine learning techniques for chemical identification using cyclic square wave voltammetry," *Sensors*, vol. 19, no. 10, 2019.

[14] T. Mohanakavitha, R. Divahar, T. Meenambal, K. Shankar, V. S. Rawat, T. D. Haile, and C. Gadafa, "Dataset on the assessment of water quality of surface water in kalingarayan canal for heavy metal pollution, tamil nadu," *Data in brief*, vol. 22, pp. 878–884, 2019.

[15] B. Srivastava, S. S. Sandha, V. Raychoudhury, S. Randhawa, V. Kapoor, and A. Agrawal, "Building an open, multi-sensor, dataset of water pollution of ganga basin and application to assess impact of large religious gatherings," in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2020, pp. 1–6.

[16] G. Betta, G. Cerro, M. Ferdinandi, L. Ferrigno, and M. Molinara, "Contaminants detection and classification through a customized iot-based platform: A case study," *IEEE Instrumentation Measurement Magazine*, vol. 22, no. 6, pp. 35–44, Dec 2019.

[17] M. Ferdinandi, M. Molinara, G. Cerro, L. Ferrigno, C. Marrocco, A. Bria, P. Di Meo, C. Bourelly, and R. Simmarano, "A novel smart system for contaminants detection and recognition in water," in *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, June 2019, pp. 186–191.

[18] G. Cerro, M. Ferdinandi, L. Ferrigno, and M. Molinara, "Preliminary realization of a monitoring system of activated carbon filter rli based on the sensiplus® microsensor platform," in *2017 IEEE International Workshop on Measurement and Networking (M N)*, Sep. 2017, pp. 1–5.

[19] G. Cerro, M. Ferdinandi, L. Ferrigno, M. Laracca, and M. Molinara, "Metrological characterization of a novel microsensor platform for activated carbon filters monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 10, pp. 2504–2515, Oct 2018.

[20] P. Bruschi, G. Cerro, L. Colace, A. De Iacovo, S. Del Cesta, M. Ferdinandi, L. Ferrigno, M. Molinara, A. Ria, R. Simmarano, F. Tortorella, and C. Venettacci, "A novel integrated smart system for indoor air monitoring and gas recognition," in *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, June 2018, pp. 470–475.

[21] 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.

[22] M. Molinara, M. Ferdinandi, G. Cerro, L. Ferrigno, and E. Massera, "An end to end indoor air monitoring system based on machine learning and sensiplus platform," *IEEE Access*, vol. 8, pp. 72 204–72 215, 2020.

[23] B. A., F. L., G. L., M. C., M. M., B. P., C. M., M. G., R. A., C. G., S. R., T. G., and V. M., "A false positive reduction system for continuous water quality monitoring," 2021, p. 311 – 316.

[24] B. C., B. A., F. L., G. L., M. C., M. M., C. G., C. M., and R. A., "A preliminary solution for anomaly detection in water quality monitoring," 2020, p. 410 – 415.

[25] [Online]. Available: https://aida.unicas.it/icprchallenge2022/