UNIVERSITÀ DEGLI STUDI DI
CASSINO E DEL LAZIO MERIDIONALE

Corso di Dottorato in
Metodi, modelli e tecnologie per l'ingegneria

Curriculum Ingegneria dell'informazione

Ciclo XXXII

A Learning Sensors Platform for Health and Safety Applications

SSD: ING-INF/05

Coordinatore del Corso
Chiar.ma Prof.ssa Wilma Polini

Dottorando
Marco Ferdinandi

Supervisori
Chiar.mo Prof. Luigi Ferrigno
Chiar.mo Prof. Mario Molinara

I would like to dedicate this thesis to my loving parents ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Marco Ferdinandi
February 2020

Acknowledgements

I would like to express my special appreciation and thanks to my supervisors Professor Mario Molinara and Professor Luigi Ferrigno helping and supporting me from both scientific and personal points of view. A special thank is also for Roberto Simmarano, President and CEO of Sensichips s.r.l. for which I was employed as an apprentice for research during my doctorate, who acted as a company supervisor for the conducted research activities. Thank you all, academic and company supervisors, for encouraging my activities and for allowing me to grow as a research scientist. I am very grateful to the entire computer and artificial intelligence as well as the electric and electronic measurements research groups of the University of Cassino and Southern Lazio for letting me to work in a very familiar environment. A great acknowledgment and thankfulness is for ENEA Portici research center and, in particular, Dr. Ettore Massera for the precious scientific support in some of my Ph.D. thesis activities and for giving me access to the laboratory and research facilities. A special thank is also for Gianni Cerro, Ph.D. for his technical and personal support and for the rest of my fellow labmates sharing stimulating discussions, late evenings working on deadlines and for all the fun. Last but not the least, I would like to thank my parents, my sister and my girlfriend for personally supporting me during these 3 years.

Abstract

Nowadays, people and environment health and safety are increasingly threatened by human activities. Industrial processes, or more in general, air and water contamination are worsening the planet life conditions. Increasing people awareness and improving the regulation system play a key role to face and tackle this global phenomenon. To this aim, reliable and low–cost technologies for a pervasive and ubiquitous environmental monitoring are really needed, especially for developing and poorer countries. This Ph.D. thesis has been focused on the development of the SENSIPLUS Embedded System for health and safety applications. It has been designed considering low–cost, miniaturization and low–power as main requirements for the whole developing and experimental phases. More in detail, it is designed according to Internet of Things and Edge Computing paradigms, integrating sensing, elaboration and communication capabilities. Sensing is mainly based on the SENSIPLUS chip, which is a micro analytical tool integrating heterogeneous sensors typologies. As for the elaboration and communication, embedded software based on statistical and artificial intelligence solutions is adopted for data analysis and technologies as Wi–Fi, USB and Bluetooth Low Energy have been integrated to transmit processing results. The embedded software has been tested on low–resources Micro Controller Units as the ESP32, STM32 and CC2541 manufactured by Espressif, STMicroelectronics and Texas Instrument, respectively. Three different applications have been addressed in this thesis: state of health monitoring of activated carbon filters and biofilters; contaminants detection and recognition in air; contaminants detection and recognition in water. Both the hardware and software components have been developed and customized for these three applications and real scenarios experimental activities have been conducted to test and validate the proposed solutions. Positive results have been obtained providing the validation of the developed technology for the addressed applications. The activities carried out for this thesis have different European research projects (Horizon 2020 and European Defence Agency) as background and reference. Furthermore, multiple collaborations with public and private research centers have characterized the design, developing and experimental activities.

Table of contents

List of figures

List of tables

Introduction

Nowadays, people and environment health and safety are continuously threatened by human activities. Industrial processes and daily people habits are worsening health conditions of our planet. Air and water pollution represents a huge hazard for the Earth ecosystems. According to World Health Organization (WHO) 4.2 and 3.8 million deaths are caused by outdoor and indoor contaminated air, respectively, and about 91% of the world's population lives in countries where the air quality indexes exceed the WHO guidelines. As regards water pollution, both freshwater and marine environments are contaminated by human activities. According to WHO estimates, about 2 billion people daily use contaminated drinking water. More than 23% [1] of the global diseases are associated with the environmental pollution, representing a significant weight for the usage of health care services and for their costs. Many studies highlight the association between contaminated air and drinking water with the diffusion of diseases and, more in general, health problems [2][3]. Aiming at improving the health and safety conditions of the Hearth planet, the WHO defines the guidelines, regarding air and water pollution, to help governments with their regulations. Although the reduction of contaminants emission is the key to tackle the environmental pollution, it is worth noting that the development and the adoption of reliable and widely spreadable (miniaturized and inexpensive) monitoring technologies are equally necessary. Indeed, they have positive impact for the phenomenon awareness increase, also in developing and poorer countries, providing a higher spatio-temporal resolution through a pervasive on-line monitoring. This is the basis for deeper analysis, as the detection of the emission sources, helping authorities for more efficient investigations and interventions. In this context, both the academic and industrial research fields are striving to develop suitable environmental pollution monitoring solutions [4] [5] [6] [7]. The main devices which are currently adopted for air and water quality monitoring are cumbersome static or mobile costly stations. In other cases, where specific analysis is needed, more sophisticated laboratory instruments are adopted, requiring the employment of professional technicians. Laboratory based procedures are commonly too slow to be used to fastly respond to accidental contamination events [8]. Furthermore, their high costs are usually the reason for a very sparse diffusion or, in some cases (e.g. poorer and developing countries), non-existent monitoring. The

widespread of the Internet of Things (IoT) paradigm [9] in a continuously increasing number of applications has furthered its adoption also in environmental monitoring [10]. The concept of integration of multiple technologies as sensing and communication in cheap platforms, which is the basis of the IoT paradigm, can be considered a breakthrough in the environmental monitoring field. Indeed, IoT enable the development of solutions for continuous and real time data collection and ubiquitous deployment of sensor nodes. A further technological advance is represented by edge computing [11][12]. It consists in performing most of data elaboration directly on-field devices positioned near the sensors at the network edges. This paradigm represents a meaningful improvement in terms of costs, power consumption and data security with respect to centralized approaches [13]. Bandwidth costs, power usage, latency and privacy can be optimized performing proper data processing algorithms directly on the data collecting device. As for this emerging technologies for environment monitoring, based on requirements as reduced sizes and costs, reliability of measurements and their analysis have not been fully reached yet. Acquired data quality still represents an open issue [5][8] for a very large deployment. Aspiring to a such diffusion, these technologies must be characterized by few false positives and false negatives, low costs and easy integration, usage and maintenance. As underlined in [14], low-costs sensors are usually affected by the interference with different substances (low selectivity) and by the dependence on the environmental conditions. Many sensor platforms are already commercially available and others are under development in many research projects, both for air and water monitoring [15] [8]. Although they are still far from replacing the current bulky devices, as laboratory instruments or monitoring stations, they represent an important alternative to increase the awareness of citizens, utilities companies and authorities.

This Ph.D. thesis main aim is represented by the development of the SENSIPLUS Embedded System, an IoT and Edge Computing solution endowed with sensing, processing and communication capabilities for health and safety applications. It is built around the SENSIPLUS (SP) chip which is developed by Sensichips s.r.l. [16] and the Department of Information Engineering of the University of Pisa. The SP chip is a micro–analytical platform integrating heterogeneous sensors and representing the core business for the Italian company Sensichips s.r.l. for which I have been employed as apprentice for research activities during the Ph.D. course. As partner of different European research projects (Horizon 2020 and European Defence Agency), Sensichips has been engaged in the development of key-technologies for the following listed applications:

- State of health monitoring of Activated Carbon Filters (ACF) and Biofilters;
- Detection and recognition of indoor air contaminants;

- Detection and recognition of contaminants in industrial waste water;

These applications have represented the testing scenarios of the developed SENSIPLUS Embedded System. Furthermore, as partner of the same research projects or for pure scientific interest, collaborations with different public and private research centers have been pursued (e.g. ENEA Portici, Arescosmo, Centro Diagnostico Baronia).

While the final aim is the development of prototypes customized for each specific application, from a general point of view, the main common activities can be summarized as follows:

- Design and development of Printed Circuit Boards (PCBs) for the integration of the SP chip;
- Software development for measurement acquisition, processing, displaying and storage for mobile Android devices, desktop Windows or Linux PC and embedded systems;
- Metrological characterization of the SP chip as generic measuring instrument and of the sensors adopted in each specific application;
- Experimental campaigns for measurements acquisition;
- Development of processing solutions based on statistical methods and Artificial Intelligence (AI) techniques as Machine Learning (ML) and Deep Learning (DL) architectures for raw sensors data analysis.

This thesis is organized as follows: Chapter 1 provides details regarding the SENSIPLUS Embedded System, highlighting both hardware and software components. General architecture and hardware and software solution developed and exploited for all applications are described. Each of the following chapters is fully dedicated to the addressed applications. Specif customization details, state of the art analysis, experimental activity and obtained results are provided in each of them. More in detail, Chapter 2 describes the activity regarding the state of health monitoring of ACF and Biofilters while chapter 3 and 4 reports details regarding contaminant detection and recognition in indoor air and water, respectively.

Chapter 1

Development of the SENSIPLUS Embedded System

1.1 Overview

The SENSIPLUS Embedded System is a novel solution for people and environment health and safety applications, developed in this Ph.D. thesis work, where sensing, elaboration and communication capabilities have been developed and integrated according to IoT and edge computing paradigms [11][12][17]. A schematic representation of the SENSIPLUS Embedded system is provided in Figure 1.1. It is mainly composed of:

- SP chip: a versatile and heterogeneous sensing platform, further described in section 1.2, capable to interact with the external physical world through its sensing elements and measuring capabilities;
- SENSIPLUS Deep Machine (SDM): a hardware and software module, whose complete description is provided in section 1.3, dealing with data acquisition, pre-processing and analysis.

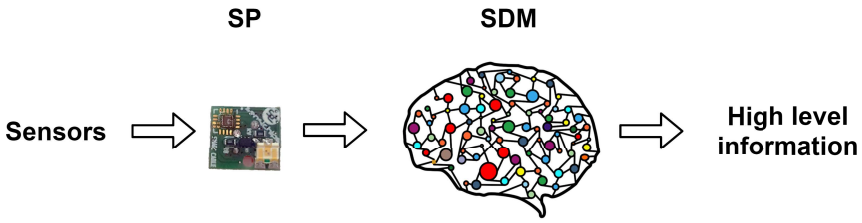


Fig. 1.1 SENSIPLUS Embedded System

The embedded system general architecture and the main hardware and software components adopted in the different addressed application scenarios are presented in this chapter. Details regarding the system customization carried out for each application, as the specific adopted sensors and the involved pre-processing and analysis techniques, are reported in relative chapters.

1.2 SENSIPLUS chip

The SP chip is a very low power (1.5 mW) micro–analytical sensing platform developed by the Italian company Sensichips s.r.l. and the Department of Information Engineering of the University of Pisa. The SP chip, whose logical architecture is depicted in Figure 1.2, is characterized by a tiny silicon CMOS integrated circuit of 3mm x 3mm in size endowed with the capability to perform continuous (DC) and alternate (AC) electrical measurements on a versatile and configurable Analog Front End (AFE). The latter can be connected up to 4 external and 11 internal analog ports and can be used to read a wide range of sensors typologies (resistive and capacitive, as chemiresistors, temperature, humidity and pressure sensors) requiring high resolution measurements. Both the two-wire and four-wire electrical measurements method can be applied on the external analog ports. The Electrochemical Impedance Spectroscopy (EIS) technique is implemented through the usage of the Lock-In Amplifier (LIA) method. The electrical measuring capability and the external analog ports are the basis to use the SP chip as a real measuring instrument. A large functions set can be selected configuring the digital SP registers through specific commands. To this aim, 3 distinct serial communication protocols are available: SPI, I2C and SENSIBUS (one-wire proprietary protocol with 6 bytes for addressing). Sensichips s.r.l. and the Department of Information Engineering of the University of Pisa develop a SP release every year (*RUNx*, where *x* identifies the release number). The boards developed for the activities reported in this thesis work integrate the fourth (RUN4) and fifth (RUN5) SP releases.

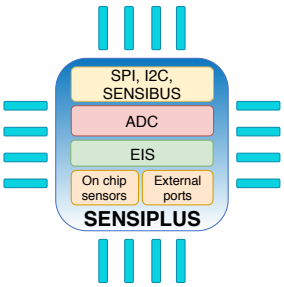


Fig. 1.2 SENSIPLUS logical architecture

1.3 SENSIPLUS Deep Machine

The SDM is a module based on a double layer (hardware and software) architecture dealing with data acquisition, pre-processing and analysis as main operative functions. The two layers and relative components are reported in the following:

- Hardware Layer: it is composed of a Micro-Controller Unit (MCUs) and/or a Host device, which is involved depending on the specific operation to be performed. Different Printed Circuits Boards (PCBs) (reported in section 1.4) have been designed and developed as part of this layer for the experimental testing campaigns.
- Software Layer: it is based on a stacked multi-level architecture characterized by a hierarchical logical model. As further detailed in section 1.5, it involves the Application Programming Interface (API), namely SENSIPLUS API, a pre–processing software which computes data transformation and, finally, a data analysis tool which is based on elaboration algorithms (analytical or ML models) for high level information extraction. Furthermore, it is organized in an IoT architecture as reported in section 1.3.1.

The SDM block diagram, reported in Figure 1.3, highlights the three operative phases: data acquisition, pre–processing and analysis. The involved software (SW) and hardware (HW) components are indicated on top and bottom sides of each block, respectively. As regards the HW components, the host device is optional in all three blocks since the reported operation can be directly performed by the MCU. More in detail, the acquisition phase is performed through the interaction with the SP chip to set the needed parameters, to perform measurements and to gather values. Such operations are executed by the SENSIPLUS API, whose details are provided in section 1.5.1.

The pre–processing phase deals with data manipulation operations as noise filtering, normalization, measurements adjustment and transposition through statistical approaches. Finally, the data analysis block is related to the application of analytical or ML models for the extraction of high level information, as events classification or physical quantities regression. As regards

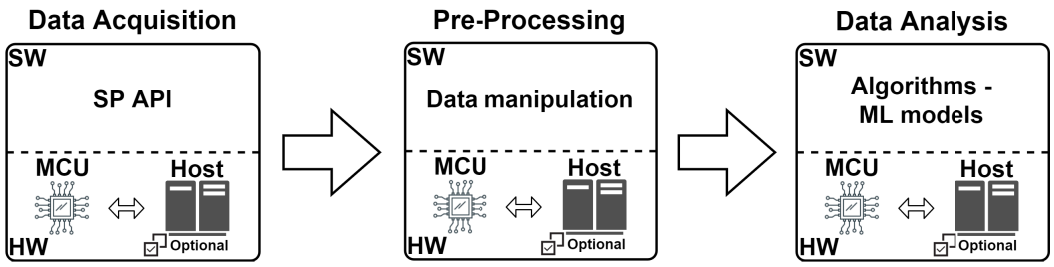


Fig. 1.3 SDM logical diagram

the pre-processing and data analysis block, the specific adopted features strongly depend on the application and are consequently described in relative chapters.

1.3.1 SDM architecture

The designed SDM architecture is characterized by multiple levels, arranged as shown in Figure 1.4.

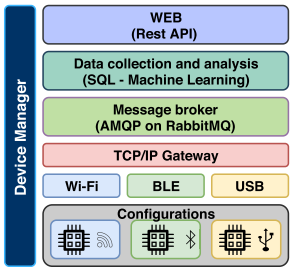


Fig. 1.4 SDM software architecture

The main items of SDM software architecture are reported in the following:

- WEB: a web dashboard front-end, based on REST API, allows to monitor and control the rest of the system;
- Data collection and analysis: a Java EE server based on Model View Control (MVC) as architectural pattern, Struts2 as framework and MySQL as Relational DataBase Management System (RDBMS) MySQL.
- Message broker: the open source RabbitMQ based on the Advanced Message Queuing Protocol (AMQP) has been involved as for data exchange purposes;
- TCP/IP: this level implements the TCP/IP functionalities allowing to transmit data acquired through lower levels to the message broker;
- Configurations: this layer is composed of the hardware elements arranged according different possible configurations and communication technologies (Wi–Fi, USB, Bluetooth Low Energy);
- Device Manager: a cross-cutting module dealing with the communication between server and on–field devices. It is designed to compute operations as software management (e.g. firmware upgrading), diagnostics through proper tests of the configurations and security. As for the latter, different privileges are assigned to users according to the account typology (as administrator and end–user). While the administrator has access

to the whole set of functionalities and data, the end–user can only monitor data coming from the configuration nodes associated with the account.

1.3.2 Configurations

A highly configurable SDM has been designed and developed around the SENSIPLUS chip aiming to address the different application scenarios needs. A general architecture of the SDM configurations, showing the different available solutions, is depicted in Figure 1.5. It is mainly characterized by multiple SP chips, a MCU and different communication protocols and technologies. As regards the interaction with the external world, the SDM configurations are endowed with communication modules based on USB, Wi–Fi or Bluetooth. On the other hand, three serial protocols are available for MCU and SP chips communication: SPI (Serial Peripheral Interface), I2C (Inter Integrated Circuit) and SENSIBUS (a proprietary protocol). The latter is characterized by 6-bytes for addressing, allowing to connect a very large number of SP chips. Furthermore, different communication modes, as unicast (one to one), multicast (one to many) and broadcast (one to all), can be adopted for the communication with SP chips through SENSIBUS protocol. Depending on the serial protocol, each configuration is characterized by a maximum number of chips (noc): 1 noc for SPI, 255 nocs for I2C and 2^{48} nocs for a single SENSIBUS configuration. As shown in Figure 1.5, the multicast communication mode of SENSIBUS protocol allows to create multiple clusters of SP chip on the same wire.

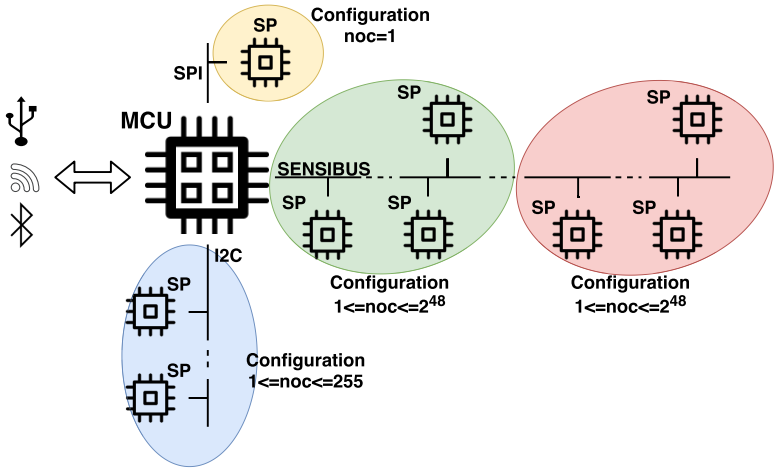


Fig. 1.5 SDM node architecture

The availability of multiple possibilities (Wi–Fi, Bluetooth Low Energy and USB) to be reached by a remote host device allows to build different network topologies. Each configuration can be used directly as a set of smart sensors controlled by a a host (e.g. PC, tablet, smartphone

etc..) via Universal Serial Bus (USB)/Bluetooth Low Energy (BLE)/Wi-Fi or integrated within a general architecture via TCP/IP.

1.4 SDM Hardware Layer

The different application scenarios addressed in this thesis have required the development of PCBs endowed with specific features. According to the application requirements, a particular Micro Controller Unit (MCU), overall power consumption, communication technologies and computational performance have been the mainly considered factors. Three PCBs endowed with different wireless and wired communication technologies and integrating the SP chip have been developed. The open source and cross platform Electronics Design Automation (EDA) software, namely KiCad®, has been used for the boards design. This software has allowed to design both the electrical circuits schematics and the boards layout. The common main requirements which have been considered for the design and development process of the different, following presented, SDM prototypes have been the miniaturization (i.e. small sizes) and low power consumption.

1.4.1 Bluetooth Low Energy board based on CC2541

A Bluetooth Low Energy (BLE) board has been developed in order to have a low power consumption solution for applications requiring battery operation and long lifetime. To this aim, the CC2541 Texas Instrument (TI) MCU endowed with the BLE module, has been chosen and integrated in the board. A picture of the board, whose sizes are 4.5 cm x 2.8 cm x 1.7 cm, is reported in Figure 1.6.

The main elements composing the PCB are listed in the following:

- CC2541 (Texas Instrument) MCU;
- SP chip (RUN4);

Fig. 1.6 BLE board prototype

- Coin cell battery socket;
- USB connector for battery charging purposes;
- 7-segments led digit;
- Programming connector.

1.4.2 Wi-Fi and USB Board based on ESP8266 and STM32L0 MCUs

The Wi-Fi and USB board has been developed in order to address applications requiring very small sizes and ultra-low power consumption. To this aim, a modular device based on two stackable boards, whose picture is provided on the lefts side of Figure 1.7, has been developed. It is composed of a board integrating the SP chip and a STM32L0 MCU (central part of Figure 1.7) and a board integrating the ESP8266 MCU in the ESP-01 package (right side of Figure 1.7). The STM32L0 MCU is produced by STMicroelectronics and is an ultra-low power consumption MCU, while the ESP8266 is produced by Espressif and designed for low-power IoT applications. The STM32L0 based board is designed to handle measurements acquisition and elaboration while the other one is for communication purposes only. For this reason, the overall power consumption can be minimized by using only the STM32L0 board or turning off the ESP8266 when not necessary (in stacked mode). The sizes of the device in stacked mode are 3 cm x 1.4 cm x 0.8 cm and can be further reduced by using only the board integrating the SP chip when no wireless communication is needed.

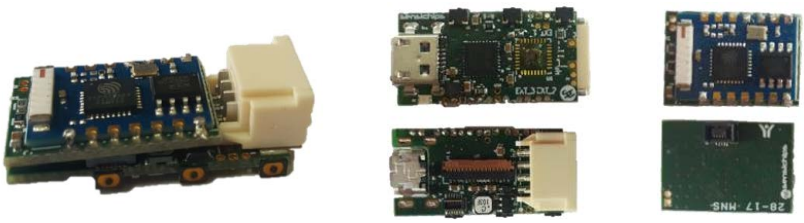


Fig. 1.7 USB and Wi-Fi board prototype

1.4.3 Smart-Cable Board based on SP RUN5

A very tiny (1 cm x 0.9 cm x 0.5 cm in sizes) board integrating only the SP chip, a led and few other electrical components (resistors and capacitors) has been developed for wired usage.

Hereafter referred as Smart-Cable board, whose picture is reported in Figure 1.8, it directly makes available the SP digital interfaces (SPI, I2C and SENSIBUS), allowing to be reached by

Fig. 1.8 Smart Cable board prototype

external devices through wired connection. Multiple Smart-Cable boards can be simultaneously placed on the same wire configuring a wired network. The very small sizes of both boards and cable are designed to enable the integration of such technology in various infrastructure typologies.

1.5 SDM Software Layer

The SDM software layer is endowed with a generic Application Programming Interface (API), namely SENSIPLUS API, and modules for data transmission, pre–processing and analysis specifically designed and developed for each application scenario.

The SENSIPLUS API architecture, as further detailed in section 1.5.1, is designed to model the SP chip as a parametric measuring device providing high level functionalities.

Initially implemented only in Java programming language, it has been used as common module for Android, Windows and Linux applications (1.5.2), allowing to acquire, elaborate, display and store measurements carried out through the SP chip.

In order to meet the requirements of low power consumption and small sizes of the addressed applications, the SP API has been implemented also in C programming language for the integration on resources limited MCUs. Furthermore, lightweight embedded software has been built on the SP C API, for a stand-alone usage of the developed hardware boards in experimental testing and demonstration activities.

1.5.1 SENSIPLUS API

The SP API represents the way a software application can deal with the set of functionalities needed for the interaction with SP chip. It is based on a stacked multi-layer software architecture

characterized by a hierarchical logical model. Each layer provides services and functionalities to the upper level and exploits the ones provided by the lower level. As shown on the right column of Figure 1.9, the SP API architecture is composed of 3 levels. Each level abstracts a hardware component to which is logically linked as virtual interaction (horizontal arrow in Figure 1.9).

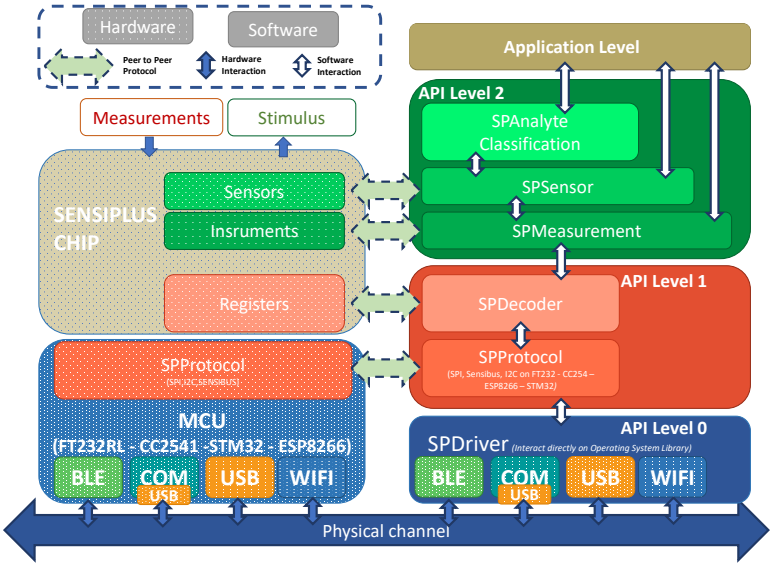


Fig. 1.9 SENSIPLUS API Architecture

More in detail, each API level and relative functionalities are listed in the following:

- API level 0: it represents the communication driver level. Various communication technologies can be adopted to connect the host (Android or Windows/Linux device) and the MCU which is connected to the SENSIPLUS chip, as BLE, WiFi and USB.
- API level 1: this level deals with decoding and encapsulating in data packets high level instructions. More in particular, the SPDecoder translates instructions strings (written in a human readable format) in byte sequences according to specific rules depending on the SP chip, while the SPProtocol creates specific data packets according to the serial communication protocol used between the MCU and the SP chip.
- API level 2: representing the higher software library API level, it is the unique level reachable from the application level. It is endowed with three distinct entities: SPAnalyte-Classification, SPSensor and SPMeasurement. The SPAnalyteClassification is composed of a series of methods and algorithms which are the core of any high level data analysis performed on raw acquired data. This entity can optionally use a ML models depending on the specific task to be performed and according to the application specifications. The

SPSensor entity deals with the SP sensors. It is endowed with a series of methods where all the measuring settings for each specific sensor are contained and used. Finally, the SPMeasurement module allows a deeper and more expert usage of the SP chip as generic measuring instrument. It is possible to set all the measuring settings as the stimulus parameters (e.g. sinusoidal frequency and amplitude etc..) or signal conditioning variables (e.g. amplifier gain).

1.5.2 Developed Applications

Two distinct software applications have been developed to interact with the SDM hardware and software components and the SP chip. Personal Hazards Monitor (PHM) and Sensiplus Winux are the applications name for Android and Windows/Linux operating systems, respectively. These applications have been exploited to test the different use cases where the SP technology has been involved. They both allow a direct interaction with available SDM configurations performing electrical measurements and gathering acquired values. Measured data are plotted, as shown in Figures 1.10 and 1.11, during the acquisition and stored in files for post–processing. As regards PHM application, the main views implemented to deal with electrical impedance and sensors measurements are shown. More in detail, from the left to the right, the reported views correspond to: the GUI shown at the application starting, the electrical impedance parameters setting, the list of available sensors and the graphical representation of acquired data. Furthermore, Sensiplus Winux application is endowed with the capability to perform measurements on multiple sensors. An array of sensing elements can be instantiated in a proper file (XML format) which is loaded and parsed by the software. Once data structures have been prepared the sequential sweep of measurements on each sensor is computed in cycle and repeated a parametric number of repetitions. Multiple electrical impedance quantities,

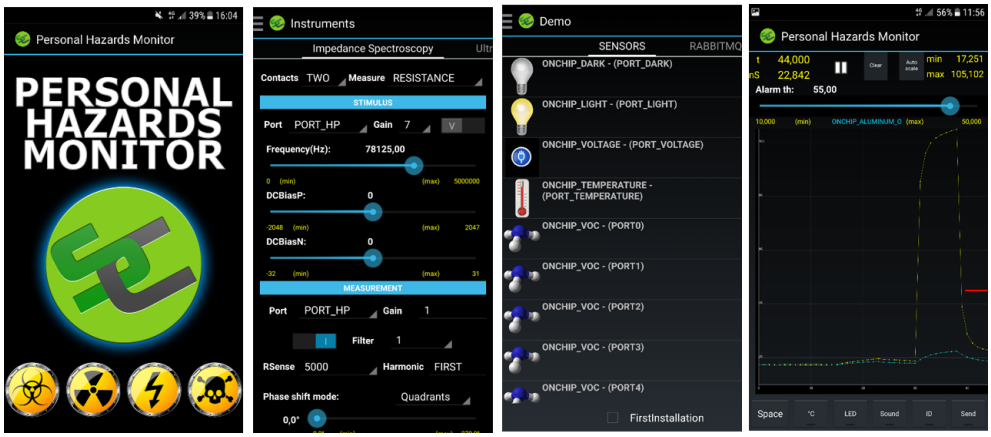


Fig. 1.10 Personal Hazards Monitor (PHM) GUI

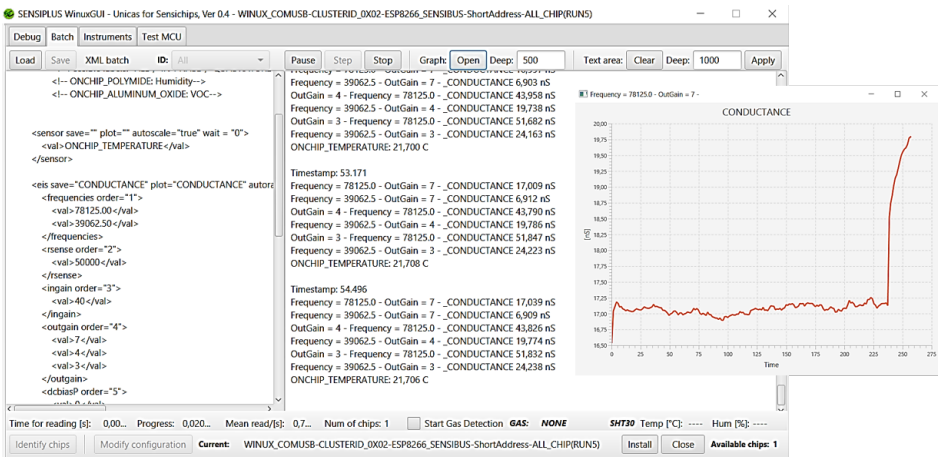


Fig. 1.11 Sensiplus Winux GUI

as modulus and phase but also resistance and capacitance values calculated according to the selected equivalent circuit, can be plotted and stored for each sensor. As shown in Figure 1.11, acquired measurements from sensors are reported both on a plot window and in the text view of the main GUI. As regards the measurement storage, distinct files (CSV format), reporting all measuring parameters (e.g. stimulus frequency and amplification gain), are saved for each sensor and a further global file, where only the plotted quantities are reported, is stored. While the former files allows to analyze the sensors behavior from the measuring settings point of view, the latter has been designed for an easier usage in post–processing, having all needed data in a unique container.

Chapter 2

State of Health Monitoring of Activated Carbon Filters and Biofilters

2.1 Introduction

Technological advances carried out in electronics, sensors and data science fields are furthering the deployment of novel paradigms and solutions, as sensor networks and IoT [18], in many different scenarios. Industrial [19], homeland security [20] and predictive maintenance [21] [22] applications represent fields where continuous and near real time monitoring are required for people and environmental health and safety guarantee.

Air contamination is a widely concerning phenomenon for which many technologies are already adopted and continuously improved. The usage of technologies capable to remove contaminants, as air filters, can surely be considered an important solution for people and environment health protection.

Activated Carbon Filters (ACF) [23] are currently employed in different scenarios, as military and industrial fields to protect soldiers form toxic substances and to prevent polluting emissions. ACs capability to remove contaminants from the air is based on their pore structure, responsible for trapping and adsorbing undesired substances molecules. Nowadays, ACF maintenance is commonly based on a time-based schedule. More in detail, they are replaced according to an estimated expiration time [24] and often used in a clogged and exhaust state. ACF state of health and exhaustion time depend on different factors as environmental conditions (e.g. temperature and humidity) and the specific typology of contaminants present in the working scenario. For these reasons, a monitoring device capable to measure and analyze, through specific algorithms, the state of health and the residual life of an air ACF is required to assure a safer usage. In such a way, predictive maintenance can be implemented in industrial

air purification system reducing failures and consequent polluting emissions. Furthermore, a safer usage of the respiratory mask's cartridges can be reached by integrating a such technology. An accurate notification, based on continuous and real time measurements and elaborations, is provided when the filter is reaching the saturation level.

A further technology, often adopted for purification purposes both for air and water regards the biofiltration [25]. It is based on the decomposition and metabolization of the undesired compounds. Pollutants are captured and biologically degraded through the living film material adopted for the biofilter [26]. The emissions are passed through an organic porous medium, that is a special bed filled with natural material such as wood chips or barks. Temperature and relative humidity are usually controlled to handle and guarantee the microorganisms (e.g. bacteria) to flourish. Colonies of aerobic microorganisms living on the filter biofilm are able to degrade contaminants present in the gaseous flow. As for the ACF, a reliable solution capable to estimate the state of health of biofilters provide the possibility to perform safer usages. Monitoring their the state of health allows, for instance, to implement a proper control of temperature and humidity condition of the material, on which strongly rely microorganism life, and to provide early warning notification when specific dangerous levels are reached.

The work carried out in the Ph.D. thesis aims at the development of monitoring device prototype both for ACs and biological filters. The same approach based on the SENSIPLUS embedded system has been adopted and electrical impedance has been exploited as measuring technique. After a detailed analysis of the solution adopted and proposed by the scientific community for ACF and biofilters state of health monitoring, the developed prototypes and the experimental activities are reported. The research activity described in this chapter and relative to ACF has brought to the works proposed in [27][28][29]. The research activities for the development of a state of health monitoring system for ACF and biofilters have been conducted with the collaboration with Arescosmo and Centro Diagnostico Baronia (CDB), respectively, which have provided the needed materials for real scenarios experiments.

2.2 State of the Art

Different approaches have been adopted in the scientific research field for the development of both a measurement technique and a Residual Life Indicator (RLI) estimation algorithm for ACF, which still represent on open issue. The evaluation of the state of health of an air filter based on activated carbons is commonly made with destructive or off–line approaches. Filters are opened and different techniques are adopted for the analysis of the residual life of carbons. Obviously, this approach cannot be adopted for online monitoring for which other techniques are required. Furthermore, ACF state of health cannot be analyzed with the level

of impediment to airflow as for several other filtration systems. Infact, ACF does not provide any measurable reduction of airflow even when the activated carbon pores are saturated by unwanted vapors. This has made the development of a generic End of Service Life Indicator (ESLI) or RLI sensor for ACF filters quite challenging. Company 3M has commercialized an Organic Vapor Service Life Indicator which can only operate in specific environments. The 3M ESLI provides a visual indication, exploiting the capability of a polymer to change the color during the adsorption of target compounds. Chemical tracers are usually used for the evaluation of the residual adsorption capability of the carbons. A study of the relationship between the residual life of an ACF and the resonance frequency of a resonator cavity is provided in [30]. This work is based on the consideration that the adsorption of air contaminants causes a variation of the carbons dielectric constant of the filter. Electrical Impedance variation relation with the AC adsorption of SO2 is highlighted by Rubel et al. [31], while Electro-Mechanical Impedance Spectroscopy (EMIS) is applied on ACs in [32] to evaluate their state of health through the observation of both electrical and mechanical carbons properties. EMIS allows to analyze structural deformation of the carbons, while the EIS provides information related to the carbons' dielectric properties. Some specific examples of the dependence of activated carbons' electrical properties on contaminants adsorption are provided in [32] [33]. Depending on the material used for the electrodes connected to activated carbons, DC or AC electrical measurements can be used. More in detail, different kind of electrodes have been used in literature. When electrodes are directly in contact with activated carbons the overall electrical resistance is very low (tens of ohms) because of the high conductivity of carbons. In this case DC measurements are usually used for resistance evaluation [31] [32], while AC measurements carried out through a sinusoidal stimulus are adopted when electrodes are insulated (through polyurethane or nylon) as in [32] to enhance the capacity variation with overall resistance values higher than tenth of kilo-ohms [32]. A procedure based on AC measurements with sinusoidal stimulus frequency in the range 10 mHz to 300 kHz has been adopted by Muriel et al. in [34]. Insulated electrodes configuration and EIS measurement technique allows to analyze the frequency response of the material providing the possibility to model an equivalent circuit. In such a way, the contribution of different phenomenon could be analyzed through the analysis of the behavior of specific electrical parameters. The work carried out for this Ph.D. thesis has been based on the usage of insulated electrodes and AC sinusoidal measurements as deeply detailed in following paragraphs.

As regards biofilters, multiple scientific works highlight the relationship between the presence of biological microorganisms in a medium and its electrical impedance behavior [35][36]. The frequency range from 100 Hz up to 1 MHz has been widely exploited to detect biological organisms in a material as reported in [35][37]. Furthermore, electrical circuit

modeling is adopted to analyze the specific parameters related to bacteria or, more in general, biological compounds state of health. Multiple aspects of biofilters relationship with different pollutants substances have been studied and analyzed in the scientific literature. In [38] and [39] the authors studied biofilter impedance dependence on acetone injection and its capability to remove it from air. More in particular, the biofilter electrical impedance has been analyzed with respect to the variation of both temperature and acetone concentration of the biological film [38].

2.3 Developed prototypes

Two hardware prototypes of the ACF state of health monitoring device have been designed and developed in this research activity. The design and development processes have been carried out considering the points listed in the following as the main application requirements:

- Electrical Impedance Spectroscopy measuring capability;
- Low power consumption and small sizes;
- Wireless communication capabilities;

2.3.1 Prototype based on BLE board

The BLE board, whose details are provided in section 1.4.1, has been the first prototype developed for this application. A preliminary experimental phase, exploiting the BLE board, has been conducted in order to prove the feasibility of the research activity. To this aim, the testing and demonstration setup shown in Figure 2.1 has been prepared. The BLE board has been directly integrated inside the filter and the SP chip connected to the electrodes. The PC, running an ad hoc developed software, has been used to control and communicate with the BLE board. Furthermore, a very first prototype of the RLI algorithm has been developed and tested in this activity. The 7-segment led digit exposed through a drill in the ACF container has been used, as shown in Figure 2.2, to provide the filter state during gas filtering. The algorithm, described in section 2.5.4, is mainly based on an initial calibration phase, indicated with the letter C on the 7-segment led, and ten different residual life levels before reaching the saturation, indicated with letter F.

2.3.2 Prototype based on the USB and Wi-Fi board

Stricter application requirements, mainly in terms of miniaturization and power consumption, have brought to design and develop the Wi-Fi/USB board, whose details are provided in section

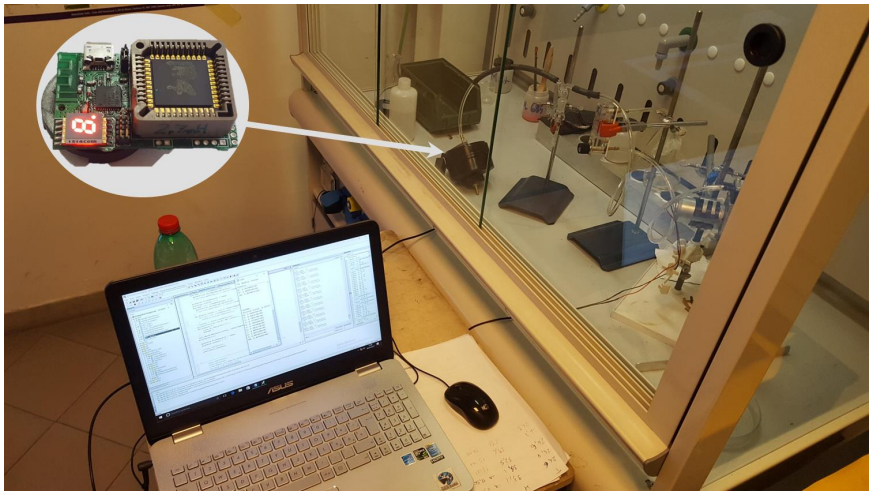


Fig. 2.1 Measurement setup adopted for the feasibility demonstration

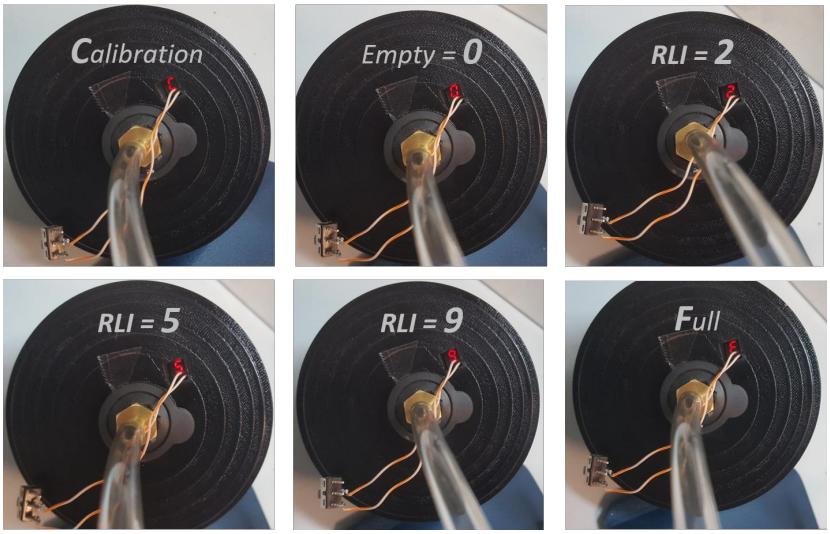


Fig. 2.2 Some pictures of a sensorized (BLE board) ACF running a first (very simple) version of the RLI estimation algorithm.

1.4.2. The new release of SP (RUN5), characterized by improved measurement capability, has been integrated in this board. Furthermore, the modularity characterizing this device allows to singularly use the STM32L0 shield in scenarios where ultra-low dimensions and power consumption are needed and wireless communication capability is not required. Most of the research activity carried out for this application has been conducted with this secondly developed board. In order to meet the application requirement regarding the long lifetime of such device in a battery-operated mode, a smart usage of its modules has been defined. Exploiting the sleep modalities of the three main elements (i.e. SP, STM32L0 and ESP8266) of

the device, an operating cycle based on different phases has been designed. Each phase and relative power consumption are shown in Figure 2.3 and described in the following:

- Initialization: The unique operation performed here is to set the ESP8266 MCU in deep sleep mode. Since all blocks are in their normal mode this phase is characterized by the maximum power consumption of about 600 mW. The duration is very short and negligible with respect to the other phases. This first phase is performed only at the device start-up, the real operating cycle starts with the second phase.

- Measurements: the STM32L0 MCU sets the SP chip to perform an electrical impedance measurement, read the acquisition and calculate the impedance module and phase values (v[i]). This operation is repeated a parametric number of times (N), which can be set considering the trade-off between measurement precision and time, enqueuing calculated values in a RAM stored buffer. The overall duration of this phase is equal to N x $T_{measure}$ (100 ms) and the relative power consumption is about 6 mW.

- Elaboration: mean and standard deviation of impedance module and phase values are computed by the STM32L0 according to 2.1 and 2.2. This operation is performed in a negligible time with respect to the other phases and 6 mW is its power consumption.

$$\mu_b = \frac{1}{N}\sum_{i=1}^{N} v[i] \tag{2.1}$$

$$\sigma_b = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(v[i]-m)^2} \tag{2.2}$$

- Data transmission: the ESP8266 MCU is woken up to transmit the mean and standard deviation values for both impedance module and phase. The Wi-Fi communication technology and the TCP-IP protocol are adopted for this phase. Once data have been sent, the ESP8266 and the SP chip are set in their low power mode. The duration of this phase is the sum of T_{wakeUp} (about 1 s) and T_{send} (negligible). The maximum power consumption of about 600 mW characterizes this phase since all modules are in their normal mode.

- Pause: this last cycle phase starts with the STM32L0 entering its low power mode (i.e. Stop Mode) and is held for T_{wait} seconds. This last is a parametric value which can be chosen according to the specific application needs. This phase is characterized by the minimum power consumption of about 0.33 mW. Finally, the STM32L0 wakes up after the T_{wait} time interval and turns the SP chip in its normal mode for another cycle.

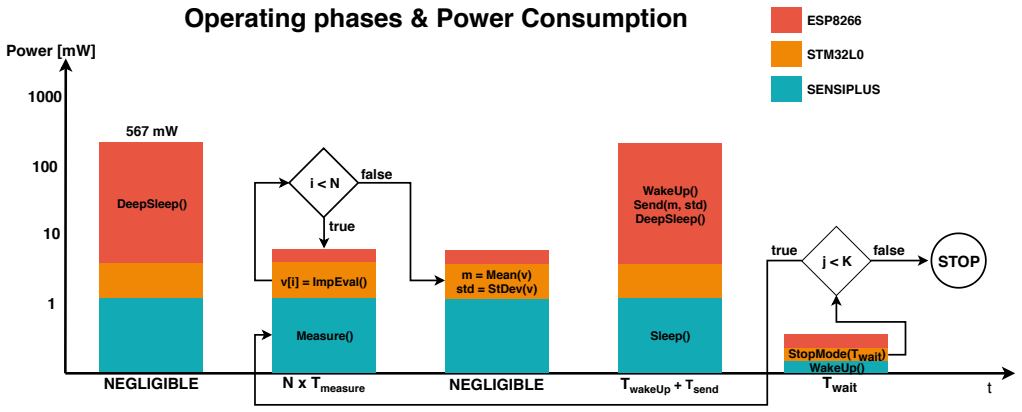


Fig. 2.3 Diagram of the operating phases designed for the ACFs monitoring.

As shown in Figure 2.3, the operating cycle, composed of the phases 2 to 5, can be repeated K times. In such a way a continuous monitoring can be implemented and the update rate can be controlled through the T_{wait} parameter. The different parameters introduced in the designed operating cycle allows to properly customize the monitoring device in terms of:

- Measurement precision and update rate;
- Lifetime;
- Physical dimensions;

More in detail, these requirements can be controlled by N and T_{wait} parameters, and the battery sizes. Considering as application scenarios the respirator mask cartridges (used by military and civil operators) and industrial air filtering systems, two specific combinations of such parameters are provided in the following. A high update rate is needed for respirator masks, since rapid flow of high toxic gases concentration can rapidly saturate the ACF. Furthermore, tiny batteries need to be selected for proper integration inside the cartridge and 24 hours is the minimum operating time [24].

Table 2.1 Selected parameters and estimated lifetime for two specific applications

	Respirator Mask	**Industrial Air Systems**
N	100	100
Twait	10s	60s
Overall current adsorption	Coin cell, 3 V, 500 mAh	Saft-Litio, 3.6V, 17000 mAh
Estimated lifetime	20 days	more than 5 years

Weaker requirements characterize the industrial air purification filters in terms of battery dimension and update rate, but much longer lifetime (e.g. years) is needed. Specific details are provided for both the applications in Table 2.1.

2.4 Metrological Characterization

The SP platform has been subjected to a metrological characterization to test and verify its measuring capability. To this aim, the experimental setup shown in Figure 2.4 has been prepared.

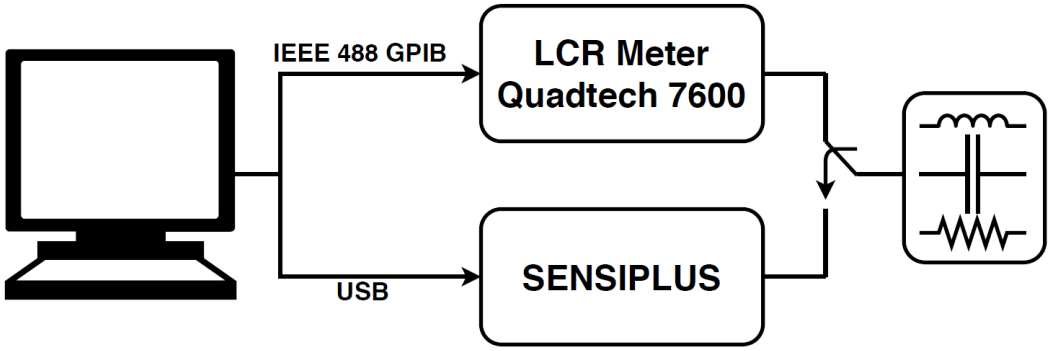


Fig. 2.4 Experimental laboratory setup for metrological characterization

It is mainly composed of:

- An electrical circuit, emulating the ACF electrical load, composed of discrete components (resistors, capacitors);
- An LCR Meter Quadtech 7600 adopted as reference measuring instrument. Its highly accurate measuring capability allows to perform a reliable comparison with the SP platform;
- The SP platform, which is hereafter also referred as Device Under Test (DUT);
- A PC running ad hoc developed software for measurement acquisition and storage. It is connected to the reference instrument and DUT through different communication technologies, IEEE 488 GPIB for the Quadtech 7600 and USB for the SP. More in detail, the PC software deals with the measuring instrument configuration, by setting needed parameters, and gathering data once they have been measured by the instruments for storage and elaboration.

Table 2.2 Discrete components values

Component	Range	Step
Resistor [kΩ]	[350 - 440]	10
Capacitor [pF]	[70 - 79]	1

Measurements acquisition has been conducted alternating the two devices (reference instrument and DUT) and keeping constant the laboratory environmental conditions. Moreover, the same wires have been used to perform measurements with both instruments to neutralize the parasite effects.

The electrical discrete components composing the impedance load for the emulation of the ACF have been chosen according to a preliminary characterization of few real filters' samples. Their electrical impedance variability range has been analyzed and resistance and capacitance values have been calculated modeling a real filter according to an RC parallel equivalent circuit. To this aim, variables decades resistors and capacitors, whose values range and steps are reported in Table 2.2, have been used.

Ten different loads have been generated combining the adopted decade resistor and capacitor. Starting from the minimum values (350 kΩ for resistance and 70 pF for capacitance), further 9 loads have been created increasing the resistance and capacitance values of the relative step. DUT comparison with respect to the reference instrument has involved 3 different sinusoidal frequency (13, 20 and 78 kHz). They have been selected according to the most suitable frequency highlighted in literature, as reported in section 2.2, and the SP chip signal generation capability. Multiple repetitions of the same measurement have been conducted for each frequency and load to compensate effects due random phenomena. More in detail the operating cycle, described in section 2.3.2, has been configured by setting $N = 16$, $K = 20$ and $T_{wait} = 15$ s. At each iteration of the cycle a burst of N measurements is performed and mean and standard deviation values (μ_b, σ_b) are calculated. For each load and frequency, the same operation is repeated K times. The metrological characterization has been addressed to the analysis of following main three factors:

- Repeatability comparison between the reference instrument and the DUT;
- Metrological linearity of the DUT;
- Measuring error;

In order to analyze the above reported metrological aspects of the DUT, different figures of merit have been adopted. The Mean and standard deviation values have been calculated over the K burst measurements repetitions according to equations 2.3 and 2.4, as parameters

Table 2.3 Worst case Standard deviation values (worst case) evaluated for reference instrument (REF) and DUT

Frequency [kHz]	$\sigma_{\|Z\|,\mathbf{ref}}[\Omega]$	$\sigma_{\theta,\mathbf{ref}}[rad]$	$\sigma_{\|Z\|,\mathbf{DUT}}[\Omega]$	$\sigma_{\theta,\mathbf{DUT}}[rad]$
13	5.5	2.9×10^{-5}	120	4.4×10^{-4}
20	3.0	2.6×10^{-5}	110	6.6×10^{-4}
78	1.9	6.1×10^{-5}	39	1.1×10^{-4}

for the repeatability comparison assessment. The DUT linearity has been analyzed mainly through the building of Calibration Curve (CC) and the evaluation of the Non-Linearity Error (NLE), calculated according to equation 2.5. CC is an analytical relationship between the reference quantity (in our case it corresponds to the measurement performed with the reference instrument) and the measurement carried out with the DUT. The NLE is calculated as the maximum distance between the raw measured value ($y_R(x)$) and the relative point on the linear least-square approximation ($y_A(x)$), normalized with respect the approximating value in the maximum location (x_{max})

$$\mu = \frac{1}{K}\sum_{j=1}^{K}\mu_{b,j} \tag{2.3}$$

$$\sigma = \sqrt{\frac{1}{K^2}\sum_{j=1}^{K}\sigma_{b,j}^2} \tag{2.4}$$

$$\mathrm{NLE}_{\%} = \frac{max_x(|y_R(x) - y_A(x)|)}{|y_A(x_{max})|} * 100\% \tag{2.5}$$

2.4.1 Repeatability comparison between reference instrument and DUT

As first step of the characterization, the compliance of the Quadtech 7600 to be adopted as reference instrument has been proved. Since the DUT metrological performance in electrical measurements are not characterized, a first comparison between the Quadtech 7600 and the DUT has been addressed to determine if the adopted LCR meter can be used as a reference instrument. To this aim, the standard deviation of the obtained measurements for the 3 frequencies have been analyzed. As reported in Table 2.3, the DUT standard deviation values are always (for all 3 frequencies and for both impedance module and phase) much greater than the Quadtech 7600 ones. For this reason, the LCR meter Quadtech 7600 can be used in this activity as reference instrument for the DUT.

2.4.2 DUT linearity analysis

As aforementioned, the DUT linearity analysis has been carried out through the observation of the CCs and NLEs. More in detail, the CCs highlight the relationship between the reference instrument (x-axis) and DUT (y-axis) measurements. A linear approximation fitting curve has been used to evaluate the analytical relationship. The validity of usage of a linear fitting has been analyzed through the NLE figure of merit. Obtained NLEs results for impedance module and phase at the 3 adopted frequency, shown in Table 2.4, are lower than 2%. The most linear trend has been achieved with the 78 kHz measurements, with 0.6 % and 0.2 % as NLEs for impedance module and phase, respectively.

Table 2.4 NLE evaluated with respect the reference instrument

Frequency [kHz]	**NLE$_{|Z|}$ [%]**	**NLE$_{\theta}$ [%]**
13	1.2	0.3
20	1.9	0.5
78	0.6	0.2

The low values obtained for NLEs allows to reliably use the linear approximation function. As regards CCs, the ideal case is represented by the bisector line (y=x), which is obtained if the DUT measurements are the same as for the reference instrument. CCs obtained for impedance module and phase at each frequency are reported in Figures 2.5-2.7. The slope of the best fitting linear approximation obtained for impedance module becomes lower with the frequency increase. As for the impedance phase, CCs results to be closer to the ideal case, especially for the 20 kHz and 78 kHz frequencies, in terms of both slope and offset. The impedance phase at the 3 analyzed frequencies highlight a different behaviour in terms of points distribution. More in detail, 13 kHz and 20 kHz measurements show equally distributed points on the linear approximation, while the 78 kHz ones are thickened around the 1.4 rad point. This phenomenon is probably due to the different impedance phase behavior at various frequencies. Finally, the CCs slope and offsets have been used to adjust the DUT measurements approaching the ideal behavior.

2.4.3 Measuring error evaluation

The CCs slopes and offsets compensations has brought to correct the DUT systematic errors. A further measurements campaign has been conducted and the developed correction function has been applied to DUT measurements. Corrected data have been used to evaluate the residual random measuring errors, according to equations 2.6 and 2.7, where $y_{|Z|,c}(x)$ and $y_{\theta,c}(x)$ are

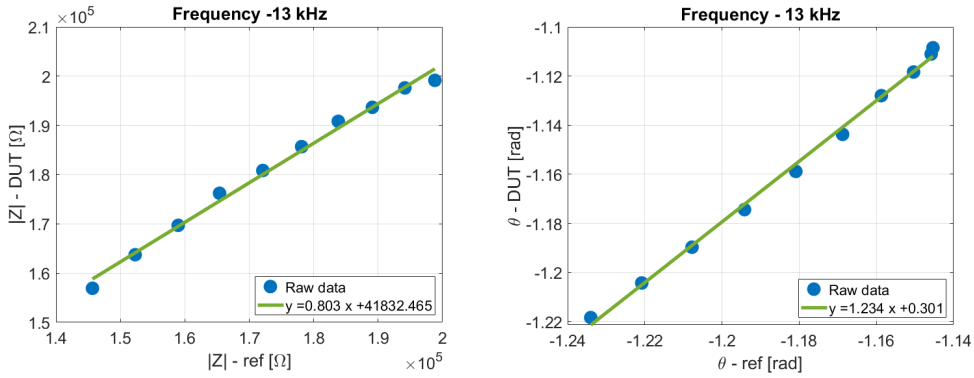


Fig. 2.5 Calibration curves (for impedance module and phase) obtained for 13 kHz frequency.

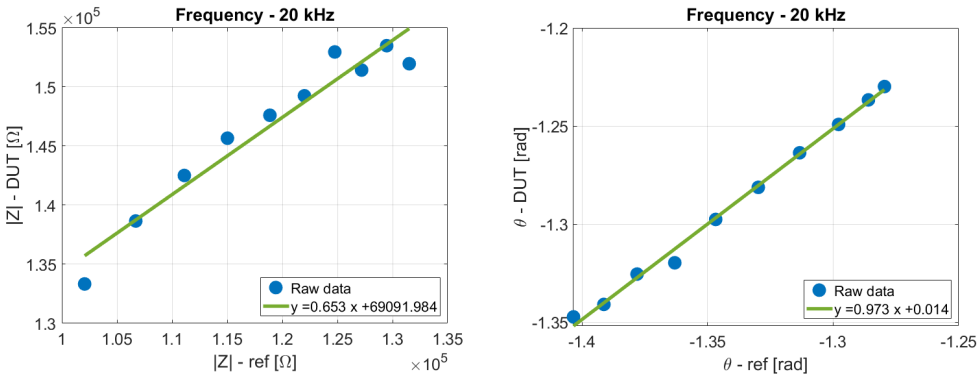


Fig. 2.6 Calibration curves (for impedance module and phase) obtained for 20 kHz frequency.

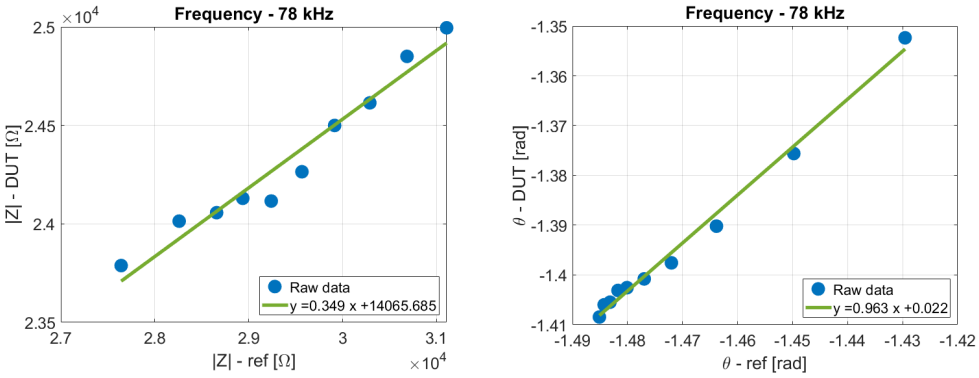


Fig. 2.7 Calibration curves (for impedance module and phase) obtained for 78 kHz frequency.

the DUT measurements for module and x the actual (measured by the reference instrument) load value.

$$\Delta_{|Z|}\% = \frac{|y_{|Z|,c}(x) - x_{|Z|}|}{x_{|Z|}} * 100\% \tag{2.6}$$

Table 2.5 Mean errors after correction

Frequency [kHz]	$\Delta_{\mathbf{|Z|,mean}}[\%]$	$\Delta_{\boldsymbol{\theta}\mathbf{,mean}}[\%]$
13	0.692	0.125
20	1.616	0.145
78	0.635	0.097

$$\Delta_\theta\% = \frac{|y_{\theta,c}(x) - x_\theta|}{x_\theta} * 100\% \tag{2.7}$$

Obtained error percentages are shown in Figure 2.8 and reported, in a more synthetic view, in Table 2.5. Impedance module and phase are for all 3 frequencies lower than 4 % and 0.6 %, respectively. More in detail, worst peaks are obtained for 20 kHz measurements while best performance have been achieved with 78 kHz ones.

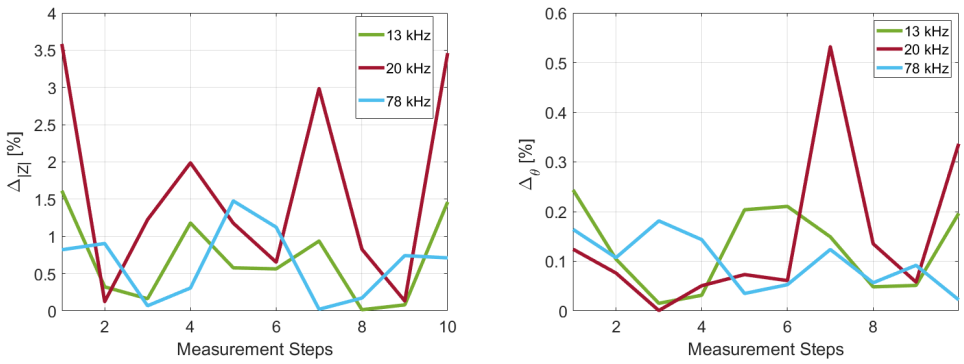


Fig. 2.8 Impedance module and phase errors percentages after correction

The metrological characterization results have allowed to state the capability to use the SP chip for the reference application. Although no comparisons with other on-chip LCR meters have been conducted, the SP platform can be considered a promising technology for a such application given the metrological results the availability of further on-chip sensors. The SP chip is endowed with on-chip temperature, humidity and gas sensors which can be simultaneously measured with the analog port connected to the ACF electrodes. The sensors' array provides a more comprehensive feature vector which can be exploited by an RLI algorithm. For instance, a correlation between temperature, humidity and the filter impedance values can be analyzed and the presence of gases flowing inside the filter can be immediately detected.

2.5 Experimental campaign on real scenario for ACF

An experimental campaign on real scenario has been conducted after the validation of the measuring capability of the SP platform, whose description is provided in section 2.4. More in detail, prototypes of ACF have been assembled and testes during the injection of different gases. This part of the activity aims to analyze the sensitivity of the SP platform in real ACF and with the presence of gas. The electrical impedance behavior of the ACF in the following 3 states has been analyzed with different prototypes and gases:

- Clean region: filter is clean and still uncontaminated;
- Working region: filter is working during polluting injection;
- Exhaust region: filter has reached the saturation level.

Before starting the analysis of the obtained impedance measurements, it is relevant to underline and consider that the ACF have been prepared with a hand-made process, involving both the electrodes' insulation through nylon and the cartridge filling with activated carbons. This has caused different baseline values at the beginning of the impedance measurements. For this reason, a relative representation, with respect to initial baseline, of obtained impedance module and phase measurements has been considered.

2.5.1 Measurement setup and procedure

The experimental measurement setup shown in Figure 2.9 has been prepared to test the ACF impedance behavior during its usage to filter contaminants. To this aim, the following main elements (reported in the left picture of Figure 2.9) have been involved:

- A gas source which pumps the chemical inside the filter under test;
- The assembled ACF under test;
- The SP chip connected to the ACF for impedance measurements;
- A gas detector to reveal the filter breakthrough moment.

As regards the contaminants, the experiments have been conducted using the following gases: ethanol, isooctane, methyl ketone (MEK) and chlorine (diluted at 1%). All of them can be found in both industrial and military scenarios. Considering the metrological characterization results, the 78 kHz frequency has been selected for impedance measurements. As for the measurement procedure, 3 phases for each experiment have been pursued:

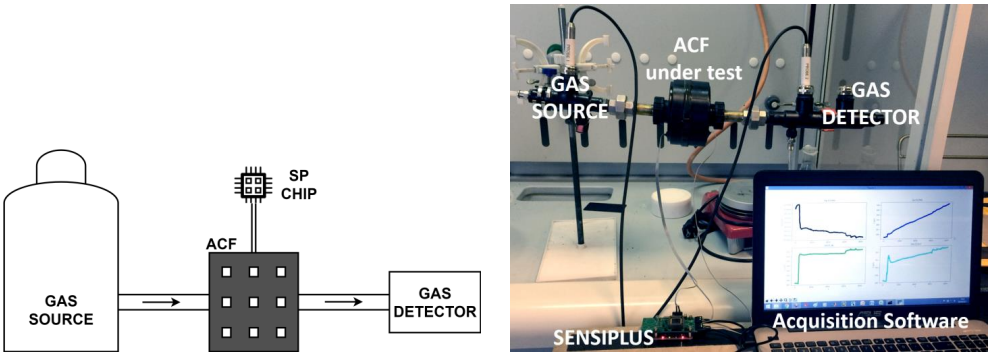


Fig. 2.9 Experimental measurement setup adopted for real scenario measurements. A logical representation with main involved elements is reported on the left while a picture of the real laboratory setup is provided on the right

- Phase #1: clean air is pumped inside the filter. During this phase the ACF remains unaffected and, consequently, it is expected that the impedance module and phase values do not change.
- Phase #2: the chemical substance is injected inside the ACF. The gas pumping continues until the gas detector reveals traces of chemicals. The impedance is expected to change during this phase since activated carbons are adsorbing the substance molecules modifying their electrical properties.
- Phase #3: the impedance module and phase continue to be monitored in order to verify the achievement of a steady state.

It is worth noting that the developed adjustment procedure has been applied to these measurements in order to correct them from systematic errors.

2.5.2 Obtained results

A first test has been carried out injecting only clean air inside the filter in order to analyze the activated carbon impedance behavior. In this experiment, whose results are reported in Figure 2.10, a mean value of 48.3 kΩ and -1.59 rad and a standard deviation of 109.8 Ω and 0.0002 rad have been obtained for impedance module and phase, respectively. Finally, a relative variation lower than 1% for impedance module and lower than 0.1% for impedance phase characterize the obtained trend in more than 2 hours of measurements. For all the other experiments, for which the gases reported in previous section have been used, the 3 phases are highlighted on each plot on Figures 2.11-2.14. The green dotted lines mark the moment in which the gas injection is started, while the red dashed-dotted lines correspond to the breakthrough moment, revealed by

the gas detector. Figures 2.11 report the impedance measurements acquired during the injection of isooctane. During the phase #1, standard deviation values of 0.23% and 0.008% have been obtained for modulus and phase. The variations observed during the phase #2, equal to 16% for modulus and 0.6% for phase, are much greater than the relative repeatability experimented for clean air. Finally, during phase #3 the steady state is reached and standard deviation values of 0.3% and 0.01% are observed. In the second test, whose results are depicted in Figures 2.12, ethanol has been used as contaminant. Impedance module and phase measurements highlight a very similar behavior with respect to isooctane experiment for all 3 phases. Chlorine (diluted at 1%) has been adopted as contaminant for the third experiment. As depicted in Figures 2.13, also for this experiment the modulus and phase values variability of the second phase is much greater than the one obtained during the first and third ones. Furthermore, relative repeatability of phase #1 and phase #3 are very similar to the one observed for the previous experiments. Also for MEK, which has been the last tested substance, an appreciable variation characterizes the middle phase with respect to the starting and ending zones (Figures 2.14). Summarizing consideration are listed in the following:

- The relative impedance repeatability during clean air injection is very similar among the different experiments (carried out with different hand-crafted ACF);
- A steady state of the impedance measurements is reached when the filter is exhaust and no more capable to trap chemicals' molecules, although the gas continues to be injected;
- The phase #2 variability is always greater than the ones obtained for the initial and final phases, highlighting the dependencies of the impedance measurements on the adsorption phenomenon of the activated carbons.

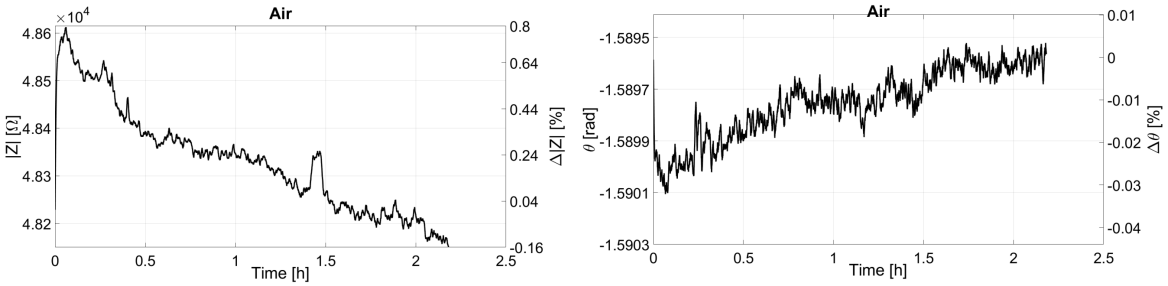


Fig. 2.10 Impedance module and phase during clean air injection

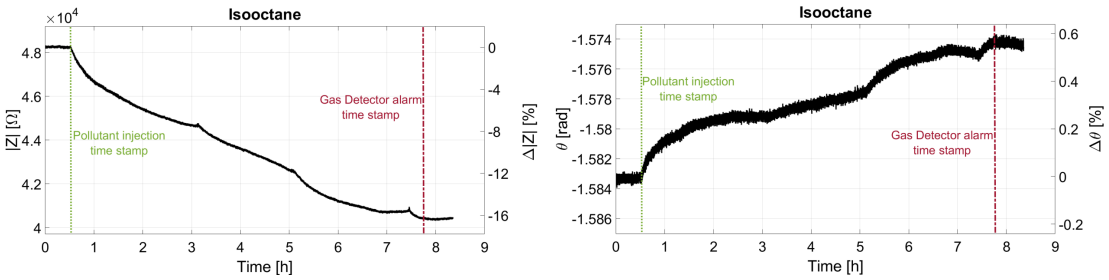


Fig. 2.11 Impedance module and phase during isooctane injection

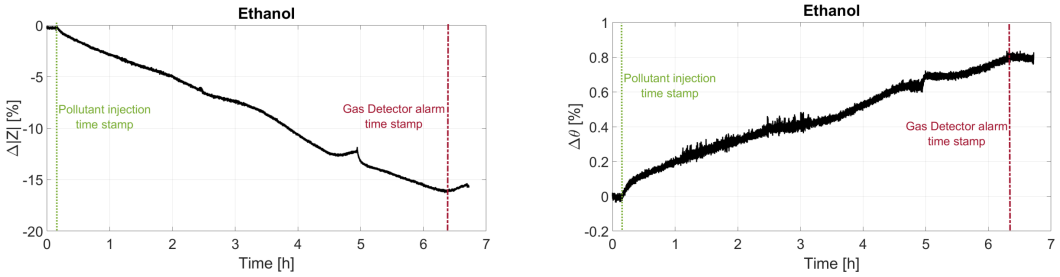


Fig. 2.12 Impedance module and phase during ethanol injection

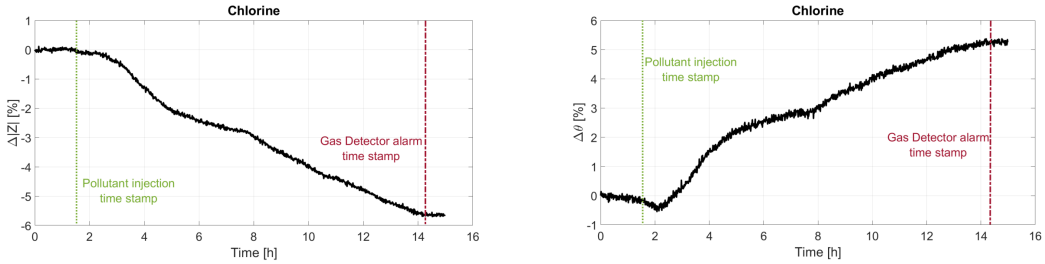


Fig. 2.13 Impedance module and phase during chlorine injection

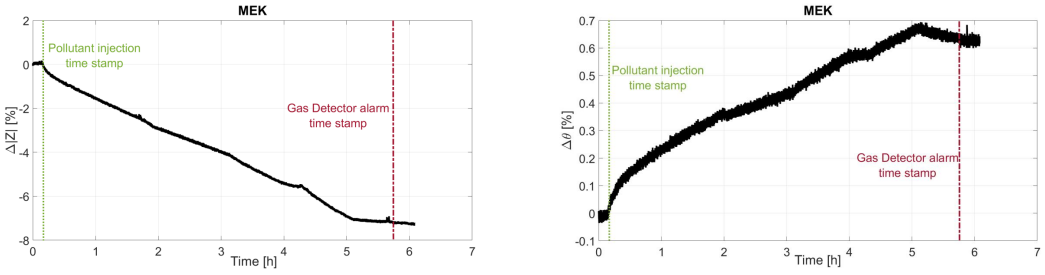


Fig. 2.14 Impedance module and phase during mek injection

2.5.3 Equivalent circuit modeling

Considering the ACF hardware structure, which, as expressed in previous sections, is composed of activated carbons surrounded by insulated electrodes, a ohmic-capacitive impedance load

behavior is expected. For this reason, the ACF has been modeled according to a parallel resistance–capacitance (R–C) equivalent circuit, as shown in Figure 2.15. Resistance (Rp) and capacitance (Cp) values have been calculated according to equations 2.8 and 2.9. Finally, the acquired impedance module and phase values relative to ethanol and isooctane experiments have been transformed in resistance and capacitance, as shown in Figures 2.16-2.17. As regard the relative variation, a greater percentage is obtained for resistance values for both experiments. More in detail, resistance percentage variation are 30% and 24% for ethanol and isooctane, respectively, while it is lower than 10% for capacitance. As expected considering the impedance module and phase behavior, an opposite trend characterized the resistance and capacitance values. Both the conductivity and the dielectric constant of the carbons increase with the adsorption of injected chemical molecules.

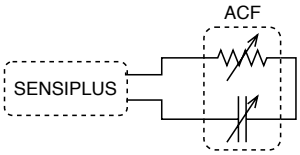
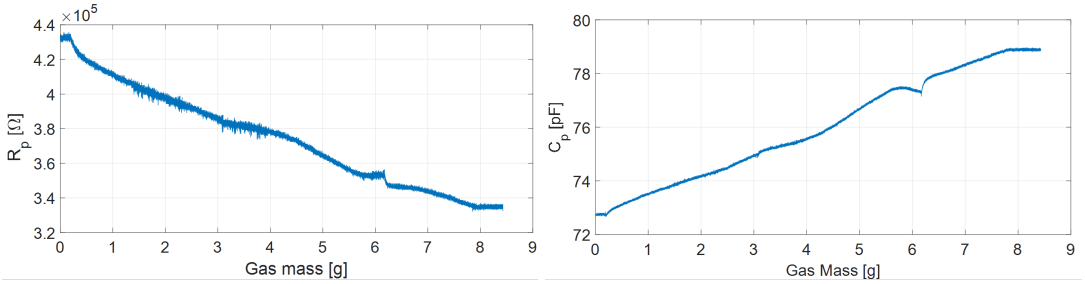


Fig. 2.15 Activated Carbon Filter equivalent circuit

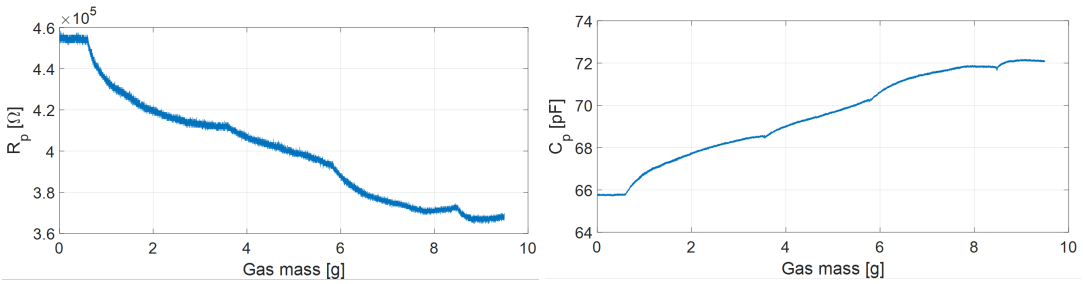
$$R = \frac{|Z|}{cos(\phi_Z)} \tag{2.8}$$

$$C = -\frac{sin(\phi_Z)}{|Z|} \tag{2.9}$$



(a) Resistance (b) Capacitance

Fig. 2.16 Resistance and Capacitance values evaluated according to the parallel equivalent circuit during ethanol injection.



(a) Resistance (b) Capacitance

Fig. 2.17 Resistance and Capacitance values evaluated according to the parallel equivalent circuit during isooctane injection.

2.5.4 Preliminary prototype for RLI method

A preliminary design of the RLI method has been developed. It is based on the experimental activity documented in previous sections, where the relation between ACF electrical impedance behavior with contaminants filtering has been analyzed. According to the specific chemical contaminant injected, different absolute saturation levels and percentage variation have been observed on both impedance modulus and phase. A further variability, characterizing the different absolute values measurement in the experimental activity, is introduced by environmental condition (e.g. temperature and humidity) and hand-crafted filters preparation. While these last aspect can be compensated with temperature and humidity sensors (available on SENSIPLUS chip) and automatizing the filters preparation process, the impedance values dependence on the specific chemical is still an open issue. The designed method, explained in the following, is based on the minimum percentage variation observed in the experimental phase (i.e. worst case). The method is organized according to following points:

- Baseline evaluation: a buffer of impedance measurements is acquired when the filter seal is opened and the mean value and standard deviation is calculated;
- Saturation level evaluation: a maximum threshold is calculated considering a percentage of the baseline;
- Residual life levels evaluation: considering the baseline as level 0 and the saturation level, N intermediate equally separated steps are calculated;
- Operative phase: Impedance measurements are acquired and elaborated directly on the MCU connected to the SP chip. A notification is provided when a new level is reached, informing the operator for the military mask and the maintenance in an industrial air purification system.

2.6 Experimental campaign on real scenario for Biofilters

The preliminary experimental activity conducted for the development of a state of health monitoring for biofilters is organized according to two main phases:

- Test with reference instrument: impedance measurements of filtering material at different concentrations have been performed through a laboratory LCR meter. More in detail, two soils at different bacteria concentrations have been exploited for this phase and a frequency sweep (from 1 kHz up to 100 kHz) has been performed for measurements acquisition.
- Test with SP system: for this phase three different SP devices have been adopted and a sensitivity analysis of the impedance dependence on soil relative humidity and bacteria concentration has been conducted.

As introduced, both phases have involved the soil used as filtering material provided by CDB. More in detail, soils with different concentration of bacteria have been adopted for the experiments reported in the following.

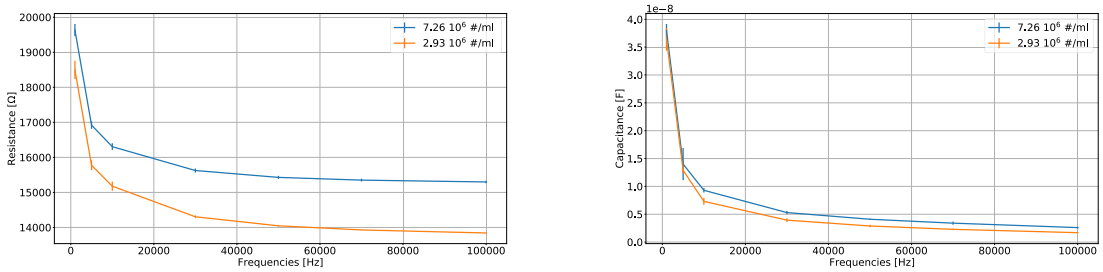
2.6.1 Test with reference instrument

Two soils at different bacteria concentrations (7.26×10^6 #/ml and 2.93×10^6 #/ml) and at 15% of relative humidity have been used for this phase and the LCR meter Quadtech 7600 has been adopted as reference instrument. Resistance and capacitance values, calculated according to a parallel equivalent circuit, are reported in Figure 2.18. Both for resistance and capacitance a decreasing trend with respect to frequency has been obtained. Furthermore, the higher the bacteria concentration the greater the resistance and capacitance values. According to this result, the presence of bacteria decrease the soil conductivity allowing, in such a way, to monitor the filtering capability. An analysis of the dispersion index (DI), calculated according to equation 2.10, has been conducted and obtained results are reported in Figure 2.19. As shown, DI values lower than 1.5% are obtained for resistance and lower than 10% for capacitance except for the first frequency (1 kHz).

$$DI[\%] = \frac{\sigma}{\mu} \times 100 \tag{2.10}$$

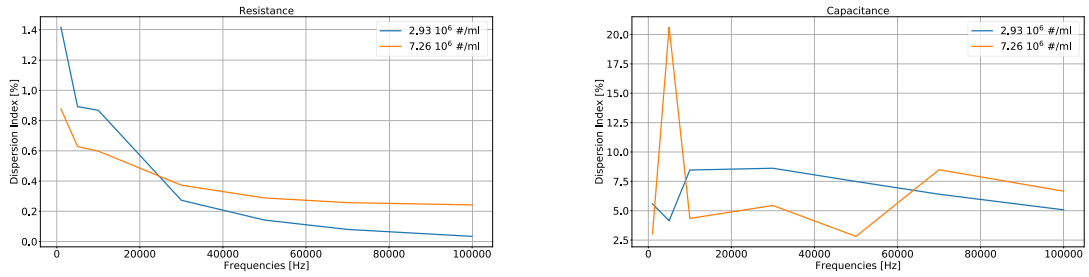
2.6.2 Test with SP system

In order to analyze the behavior of the impedance measurements performed with SENSIPLUS chip, three phases have been considered:



(a) (b)

Fig. 2.18 Measurements performed with reference LCR meter with 2 different bacteria concentrations. a) Resistance values; b) Capacitance values



(a) (b)

Fig. 2.19 Dispersion index values: a) Resistance; b) Capacitance

- Impedance measurements with zero bacteria concentration soil;
- Sensitivity analysis of electrical impedance to relative humidity;
- Sensitivity analysis of electrical impedance to different concentration of bacteria.

As regards the adopted sensors to be immersed in soil, electrodes endowed with two copper plates, as the one shown in Figure 2.20, have been used. More in detail, three electrodes of the same typology have been connected to three distinct SP chips.

Tests with soil containing no bacteria

This test has been executed immersing the three sensors in the same soil, as shown in Figure 2.21, in order to analyze the impedance behavior over the whole measurement time interval (8 hours). No soil variations in terms of temperature, humidity and composition have been applied during this test and an acquisition rate of one measure every 30 seconds has been adopted.

According to expectations, a constant trend has been obtained for both resistance and capacitance values as shown in Figure 2.22. More in particular, although a constant value

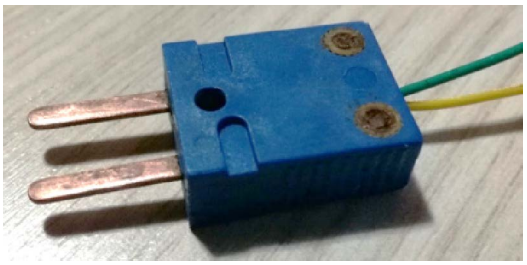


Fig. 2.20 Electrode used as sensors for impedance measurements in soil mixture

Fig. 2.21 Setup for impedance measurements with electrodes immersed in biofilter soil without bacteria

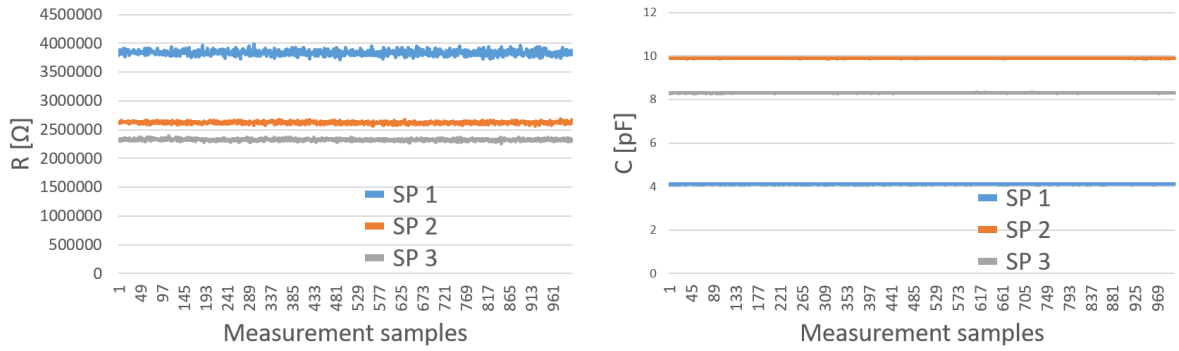


Fig. 2.22 Resistance and capacitance values obtained with 3 different SP chips and electrodes on the same soil.

is obtained for both quantities, resistance measurements are characterized by a greater noise. Different absolute values have been obtained since no–calibration and adjustment of SP chips was adopted.

Tests with soil at different relative humidity

A sensitivity analysis of the relationship between soil relative humidity and electrical impedance

has been carried out. Since humidity is a fundamental parameter for the microorganism life in the biofilm and for this reason it can be used to handle their existence, it is worth to study the dependence of the proposed measuring technique (electrical impedance) on this parameter. To this aim, three different SP chip connected to the electrodes, as the one shown in Figure 2.20, have been used for three different soils. Distinct containers with the same soil conditions, in terms of bacteria concentration (0 %) and relative humidity, have been adopted. Soil humidity has been changed during measurements acquisition in two of the three containers, introducing water to have a final relative humidity of 40% and 20% (measured with a hygrometer). Obtained measurements, reported in Figure 2.23, highlight an higher variation for the 40% humidity change (gray line) than the on observed with 20% variation (orange line). A constant trend has been obtained for the unchanged soil (blue line). Furthermore, it is important to underline the lack of a linear relationship between the humidity change and the electrical impedance variation.

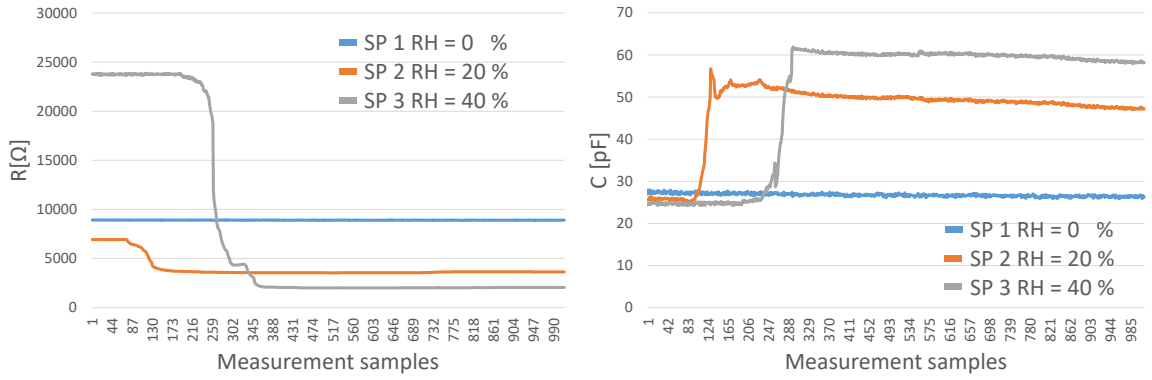


Fig. 2.23 Resistance and capacitance values obtained for 3 different soil and relative humidity.

Tests with soil at different bacteria concentrations

This last test has been conducted exploiting 3 containers filled soil at different bacteria concentrations ([0, 11675, 80350] #/ml) and a relative humidity equal to 40%. The same 3 SP chips and electrodes have been sequentially used to perform measurements on the 3 containers, as the one shown in Figure 2.24.

Obtained impedance values are reported in Figure 2.25 in terms of resistance and capacitance values. As highlighted by the plots, different impedance values have been measured in the 3 soil containers and, as obtained with the LCR meter, higher resistance and capacitance values correspond to greater bacteria concentrations. Furthermore, very close values have been obtained for each container by the different SP chips, confirming the metrological repeatability of this devices and its validity to be used in a network of monitoring sensors in the same biofilm material.

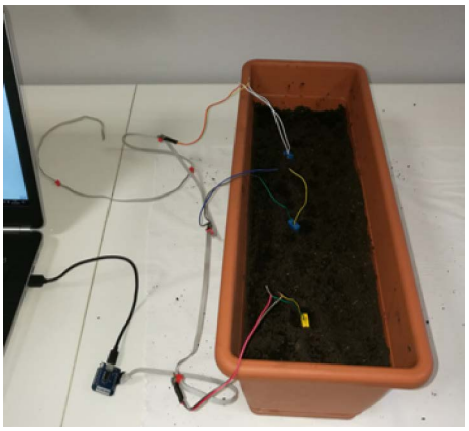


Fig. 2.24 Measurement setup adopted for data acquisition with 3 SP chips on the same soil for different concentrations.

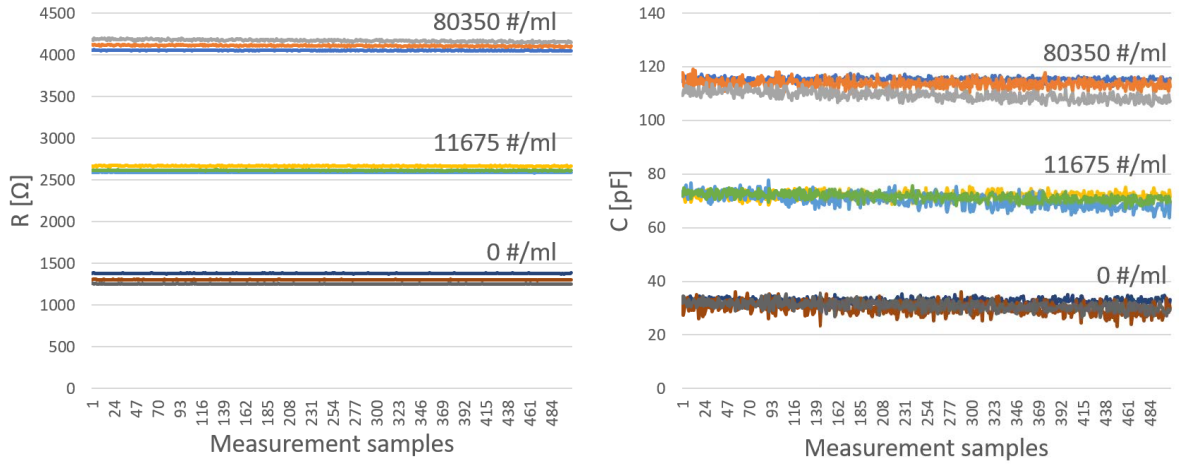


Fig. 2.25 Resistance and capacitance values acquired for the 3 different bacteria concentrations.

2.7 Conclusion

The research activity reported in this chapter has been conducted as part of two research projects involving Sensichips and Arescosmo for activated carbon filters and Centro Diagnostico Baronia for biofilters. The Embedded SENSIPLUS system has been customized for this application and a metrological characterization has been performed to analyze and validate its measuring capability. A calibration and adjustment procedure has been performed and real scenario (activated carbon filters and biofilters) tests have been conducted. Tests with the injection of contaminants typically used in industrial processes and military field have been conducted for activated carbon filters underlining the dependence of electrical impedance on the filter saturation. As for biofilters, the SP system has been analyzed to validate its sensitivity to

different concentrations of bacteria present in the biofilm. Obtained results for both filtering typologies constitute the basis for the future development of a residual life evaluation algorithm capable to evaluate the activated carbon and biological filters state of health, allowing a smart and safer usage.

Chapter 3

Detection and recognition of indoor air contaminants

3.1 Introduction

Indoor air quality monitoring represents a diffused concern since different dangerous pollutants can be found in daily attended places as household environments, offices and schools. According to US Environmental Protection Agency, the concentration of some pollutants in indoor environments are often 2 to 5 time higher than typical outdoor concentrations [40] [41], with negative consequences on human health. More in detail, eye and skin irritation, allergic pathologies, respiratory diseases and cancer are some of the indoor pollution effects on health. Combustion generated through fireplaces, coal heating and cooking appliances [42], and Volatile Organic Compounds (VOCs) [43] contained in cleaning substance or building material are among the most diffused sources of indoor air contaminants. For these reasons, there is a great need for pervasive and continuous monitoring of indoor air. As highlighted in the introduction of this Ph.D. thesis, many research activities are addressed to the development of such technologies [4][44], both from the sensing and data processing points of view. As regards the sensors, common main goal is the development of compact and low-cost devices capable to sensitively and reliably detect pollutants. In addition, a multiple sensors based approach is usually followed to increase the sensing complementarity of the whole system. The large amount of data gathered by sensors array and their limited metrological performance, with respect to laboratory instruments, furthers the development of complex data analysis methods capable to deal with raw sensors data. To this aim, statistical methods and AI techniques are commonly used. As part of the research activity of the Ph.D. thesis work, a first prototype of an integrated system for indoor air monitoring based on the SP chip has been

carried out. The integrated system, whose generic architecture description has been provided in chapter 1, has been specifically customized for this application selecting proper sensors and data manipulation and analysis techniques. More in detail, a special focus has been addressed to chemicals substances detection and recognition through the adoption of multiple sensors and AI techniques, as ML and DL architectures. Both the sensing and processing capabilities of the developed system have been addressed in this work. Good sensitivity, high recognition capability and lightweight processing solutions for an embedded implementation have been the main considered requirements over low power and miniaturization, already reported in section 1. A very first experimental activity for air contaminants detection and classification has been reported in [45], where the SP chip and a DL technique have been involved. Starting from this experience, an important advance has been computed through the employment of a sensors array, according to the concept that multiple and complementary sensors increase the capability of providing discriminating patterns. Aiming to minimize the computational burden, a lightweight Multi Layer Perceptron (MLP) solution has been selected and compared with more complex DL architectures as Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM). Furthermore, the contaminants recognition problem has been faced according to a multi–class classification approach, considering the undesired substances as mutually exclusive. Indeed, no experimental activities of measurements acquisition have been conducted submitting multiple substances simultaneously.

3.2 State of the Art

Many research activities aiming to the development of miniaturized, low cost and smart solutions for air monitoring, underline the sensing technological limitations in terms of sensitivity, selectivity and dependence on the environmental conditions [46][44]. In [5], low cost commercial platforms performance are compared with CEN (European Committee for Standardization) reference instruments, highlighting that although the lower accuracy, stability and selectivity such technologies allow to perform large data aggregation through the deployment of large sensors networks. High resolution pollution indexes mapping and contaminants sources identification algorithms can be developed exploiting this sensors' typologies. Methods and techniques for chemicals sensing devices performance improvement and their exploitation in classification and regression problems is reported in [47]. Furthermore, a widespread adoption of AI, ML and DL techniques to deal with sensor data characterizes the scientific literature scenario. ML models are adopted for both classification (gases recognition) and regression (estimation of gases concentration) problems. Both classification and quantification are addressed in [49], where an embedded implementation of an Artificial Neural Network (ANN) for pattern recognition

model is used to recognize the specific gas (between hydrogen, methane and carbon monoxide) and estimate the relative concentration. De Vito et al. develop a tapped delay architecture to estimate gas concentration in [50] while a carbon monoxide and methane quantification system based on a sensor array and an ANN is proposed in [51]. Furthermore, the sensors drift problem is analyzed in [48] and a solution based on Dynamic Neural Network for on field calibration through a low cost technology is proposed. The born of DL has been a technological breakthrough in the ML field. It has been adopted in an increasing number of applications, outperforming other classical ML techniques. Considering the key capabilities to automatically extract and learn useful features and to profit from huge dataset, it has been adopted also in the field of sensors based applications. In [52] and [53], the authors highlight a meaningful improvement of performance with respect to other classical ML techniques in applications as human activity recognition and car tracking. A CNN based approach is proposed in [54] for gas classification and a performance comparison with MLP and Support Vector Machine (SVM) architecture is performed obtaining higher accuracy with the DP technique. As underlined in [55] by LeCun et al., CNN are a well suited model for different fields as image or time series analysis and have represented an important step forward in the ML field. LSTM neural network is a further DL architecture specifically appropriate for time series analysis [56]. As for Recurrent Neural Network (RNN) architecture, the LSTM is based on loops networks. This kind of architecture allows to conserve useful information inside the network, as a memory, providing the capability to extract data time dependencies. More in detail, the complex hidden structure of LSTM nodes counteracts the long short term dependency of more classical RNN [57]. RNN and LSTM models are adopted in [58] and [59] for different applications characterized by important data time dependencies.

3.3 Data acquisition

3.3.1 Experimental measurement setup

An experimental measurement campaign has been carried out to acquire a set of raw measurements which have been used for the training of ML and DL models. To this aim, the setup shown in Figure 3.1 has been prepared. Focusing this research activity phase for the assessment and testing of the integrated system sensitivity and classification capability, a very compact measurement setup has been adopted. A specific implementation of the SP integrated system (presented in chapter 1) has been developed, involving:

- a PC collecting data acquired through the ESP32 MCU running the SP API;
- a SP chip positioned on the internal surface of a transparent glass box;

- three gas sensors;
- a little bowl containing the chemical substance in liquid state, exploiting the evaporation process for the submission of gases to sensors.

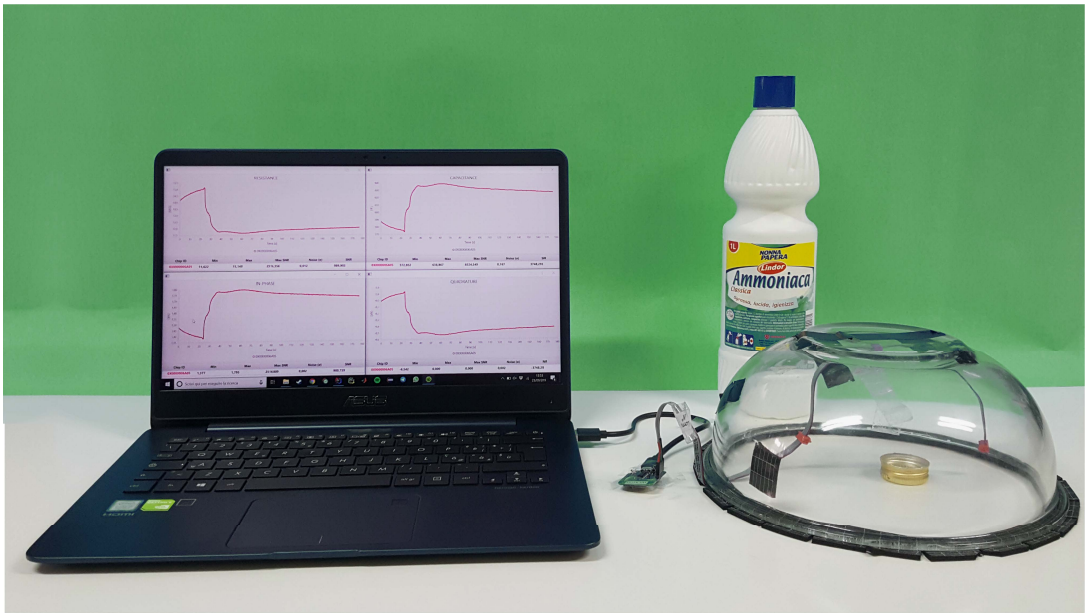


Fig. 3.1 Experimental measurement setup for data acquisition

As for sensing, the three sensors listed in the following have been used:

- The SP built-in gas sensor, namely ONCHIP_ALUMINUM_OXIDE, based on interdigitated electrodes covered by aluminum oxide. It is a generic gas sensor, operating at room temperature, which quickly react to different substances' typologies. The adsorption phenomenon of the contaminants molecules between the very close electrodes tracks (less than 2 μm) causes a variation of the sensor electrical properties.
- The external commercial humidity sensor, namely OFFCHIP_HUMIDITY, manufactured by IST ([60]). It is based on a capacitive operating principle involving a polymeric material where polluting molecules are adsorbed. The dielectric constant of the polymer changes causing a variation of the measured capacitance. According to manufacturer datasheet, typical values' range is 140 ± 40 pF measured at 1 kHz - 100 kHz as frequency range and with 23 °C and 30% RH as environmental conditions.
- The external sensor based on gold interdigitated electrodes covered by graphene, namely OFFCHIP_GRAPHENE. The research activity carried out by the ENEA Portici research center on graphene as gas sensing material [61] and the collaboration with some

researchers of the same center have allowed the adoption of such sensor. The functionalization of the gold electrodes with the deposition of a graphene film enables a selectively sensitivity to oxidizing and reducing gases (e.g. nitrogen dioxide).

Electrical impedance measurements have been performed through the SP chip on the three sensors. A sensitivity optimization of the adopted sensors has been carried out with a preliminary campaign where SP measuring settings have been properly tuned and the electrical quantities providing the best trade off between measurement noise and response variation have been selected. This testing campaign has allowed to experience a meaningful dependency of the sensors on the environmental condition variation during their usage. As further detailed in section 3.4 a reference value for each sensor has been calculated through the Exponential Moving Average (EMA) to compensate the sensors drift due to environmental condition variation.

3.3.2 Measurement Procedure

A systematic measurement strategy has been defined for the raw dataset collection. For each substance the same procedure has been repeated 10 times performing a measurement sweep from all three sensors at each acquisition timestep. The phases listed in the following have been pursued:

- Exposure to clean air for 120 seconds;
- Injection of chemical contaminant inside the box for 600 seconds;
- Exposure to clean air for further 120 seconds.

An acquisition rate providing a sweep measurement from the 3 sensor in 2 seconds has been empirically chosen. In such a way 420 samples are acquired for each sensor during a single experiment. The clean air phases designed as starting and ending parts of the measurement strategy aim to let the sensors reach a stable steady state before submitting the substance and to analyze the recovery capability. Ten different repetition have been computed for each substance in distinct days and times in order to enhance the dataset variability, in terms of environmental condition dependency, and further the generalization capability of the ML models' training. Acetone, alcohol, ammonia and bleach have been the adopted substances. They are commonly used and present in indoor environments and represents meaningful hazards for people health conditions. Further 10 experiments have been carried out also with both clean air and water vapour, representing the background of the integrated system working scenario. In this way, the classification technique training is allowed to learn to distinguish chemicals from environmental background variation, minimizing false positives.

3.4 Data pre–processing

The developed software for the pre-processing phase of the SDM(1.3) is mainly characterized by the following operations: evaluation of the environmental reference, normalization, preparation of data sample for successive classification techniques and labeling. The environmental reference is computed filtering each sensor response through EMA. This operation is carried out to normalize the sensors responses baseline variation due to environment conditions changing, which are expected to be very slow. The environmental reference has been calculated according to equation 3.1, where s_t is a 3–dimensional vector containing measurement acquired from the 3 sensors at time instant t, $\alpha = 10^{-4}$ the degree of weighting decrease and r_t the environmental reference vector computed at the same time instant.

$$r_t = \alpha s_t + (1-\alpha) r_{t-1} \tag{3.1}$$

The α coefficient is chosen to have a high relevance to an initial calibration phase where clean air is submitted to the integrated system. It is worth noting that a long term dependency on the response to contaminants characterizes this approach since the environmental reference is affected by the response to chemicals. No counter measurements have been adopted yet for this problem since the primary aim of the research activity is to provide an early detection and recognition signal which can be used by the system itself to reset the internal variable status before continuing the monitoring operations. As for normalization, the ratio between each sensor response and its r_t parameter is computed, as in equation (3.2), generating the feature vector $f.v.(t) = [f.v.(1,t), f.v.(2,t), f.v.(3,t)]$ for each timestep (t). Finally, a time window with parametric size (n) of feature vectors is composed ($F(t_n) = [f.v.(t_1), .., f.v.(t_n)]$) and labeled.

$$f.v.(i,t) = s(i,t)/r(i,t) \tag{3.2}$$

3.5 Data analysis

A lightweight solution for target gases classification has been adopted for an embedded implementation according to the edge computing paradigm. To this aim, a single hidden layer MLP architecture has been chosen and training and testing experiments have been conducted on the collected dataset. Furthermore, in order to compare the MLP performance with respect to more complex and powerful techniques, two DL architectures, CNN and LSTM, have been implemented and tested on the same dataset. A more detailed presentation of each architecture is provided in following sections. For each adopted technique, a preliminary tuning phase

of the networks' hyper-parameters has been performed, defining the networks' structure (e.g. number of layers and internal neurons) and training parameters (e.g. learning rate, drop–out and batch normalization).

3.5.1 Multi Layer Peceptron

As introduced, a single hidden layer MLP whose schematic representation is provided in Figure 3.2 has been designed and implemented. It is composed of an input layer with 3 neurons, 64 in the hidden layer and 6 in the output. The Rectified Linear Unit (ReLU) has been chosen as activation function and applied on the hidden layer output, while the softmax has been adopted for the output layer. In order to minimize the resources usage of this classification architecture, a single timestep window (n=1) has been adopted as network input sample (F). In such a way, a 3 x 1 feature vector (f.v.), computed according to equation 3.2, is submitted to the network at each acquisition timestep. Since 3x40 sample structure (F) has been chosen for the DL architectures to exploit and test their capability to find hidden time dependencies, a further implementation of the MLP has been made for a fairer comparison. In such a way, a 3 x 40 input buffer has been flattened in a 120 x 1 input sample. For this reason the MLP input layer has been modified to have 120 input neurons, while all the other network parameters have been preserved. Classification results of the two implemented MLP, hereafter referred as MLP_3 and MLP_{120}, are provided in section 3.6.2.

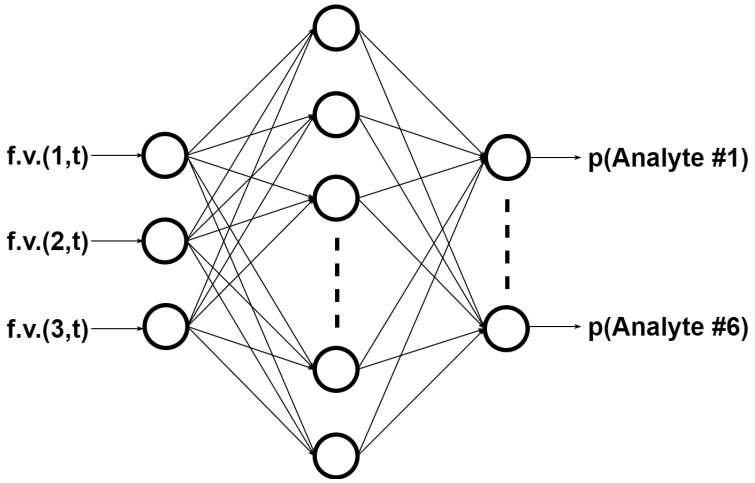


Fig. 3.2 MLP_3 architecture

3.5.2 Convolutional Neural Network

The designed and implemented CNN architecture, whose representation is provided in Figure 3.3, is mainly based on three stages: input, features' extraction and classification. As regards the input, the bi-dimensional input (F) has been chosen in order to exploit and test the network capability to find correlations on both dimension (inter sensors relationships considering the rows and time dependencies for the columns). To this aim, 2 consecutive Convolutional Layers (CLs) have been used for the automatic features' extraction stage. Both the CLs are characterized by the following operations: convolution, batch normalization and ReLU activation function. The adopted convolutional kernels (CL_1 and CL_2) sizes are 16x3x3 and 14x16x3x3. The 3-dimensional output matrix (14x3x40) of the CLs is flattened in a mono-dimensional array and then passed through the final fully connected layer. It represents the hidden layer of the classification stage and is composed of 128 neurons which are fully connected to the 6 output neurons.

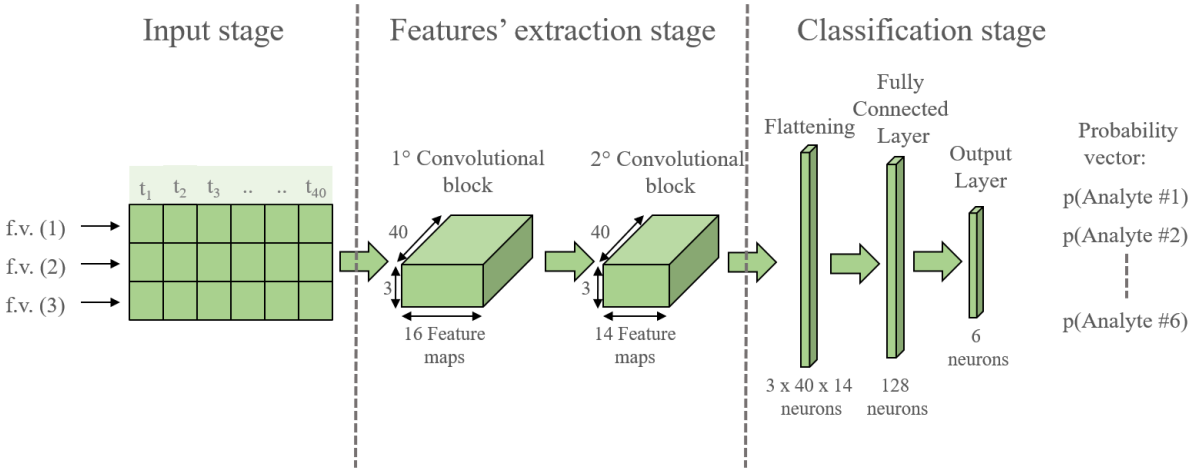


Fig. 3.3 Convolutional Neural Network architecture

3.5.3 Long Short Term Memory neural network

A LSTM neural network based on 8 layers has been implemented and the same input structure, as for the CNN, has been used. It is a particular type of LSTM capable to deal with multivariate dataset (3-dimensional in this specific case). A schematic representation of the adopted LSTM architecture is provided in Figure 3.4. Here, un unrolled view of the architecture is shown and both the multi layers structure and the number of hidden units (corresponding to the trainable parameters) for cell are highlighted.

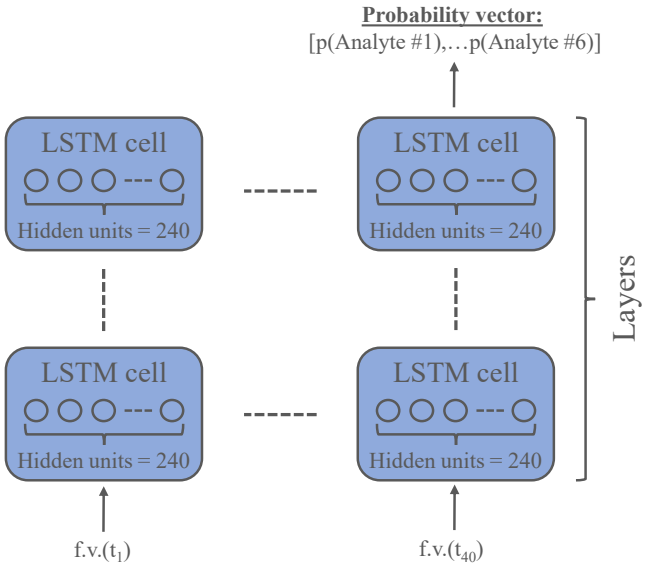


Fig. 3.4 Long Short Term Memory architecture

3.5.4 Network training strategy

A 10–fold cross–validation approach has been performed in order to have a statistical analysis of the results over the whole dataset. It is based on the idea of rotating the data samples used for training, validation and test sets in order to assess the generalization capability of the model trained on different sets of samples. Furthermore, it is a useful measure which allows to easily highlight any data dependent overfitting or biases. In our case, as reported in section 3.3.2, the whole dataset is composed of 10 different experiments for each class. For each iteration of the adopted 10–fold cross–validation procedure data samples of a distinct single experiment are withdrawn and used as test set. Two further experiments are randomly selected among remaining and all the other seven experiments used for the training set. As regards the model learning phase, cross entropy and adam optimizer have been adopted as loss function and optimizing algorithm, respectively. Finally, an early stopping strategy based on a maximum number of epochs has been used to avoid the model overfitting All the training experiments have been computed exploiting the GPU acceleration through a NVIDIA®TITAN X Pascal. It was mounted on a Linux based desktop computer endowed with an Intel(R) Xeon(R) E5-2609 v4 (8 cores) CPU and 256 GB RAM.

3.6 Results

The data pre–processing and analysis SDM stages results are provided in this section. More in detail, the visualization of the sensors responses for each substance and for all 10 experiments are reported in section 3.6.1 while the classification results obtained with each involved architecture follows in section 3.6.2.

3.6.1 Collected dataset

A summary of the acquired measurements normalized as explained in section 3.4 for each sensor is reported in Figures 3.5-3.7. As first graph of each figure, the measurements relative to clean air exposure are reported, where the mean value, corresponding to the unitary value, is set as baseline and repeatability is analyzed in terms of standard deviation. The sensor intrinsic dispersion range is expressed considering two times the standard deviation and is depicted through the horizontal lines. In such a way, the sensitivity of the sensors can be analyzed proving that the variation is greater that the reported variability range. It can be highlighted that ammonia and bleach substances cause meaningful variations (higher than the defined intrinsic variability range) for all three sensors. As for alcohol and acetone, only the generic humidity sensor shows a decreasing trend, while very low responses are obtained for the other sensors. Finally, water vapour is detected by both aluminum oxide and humidity.

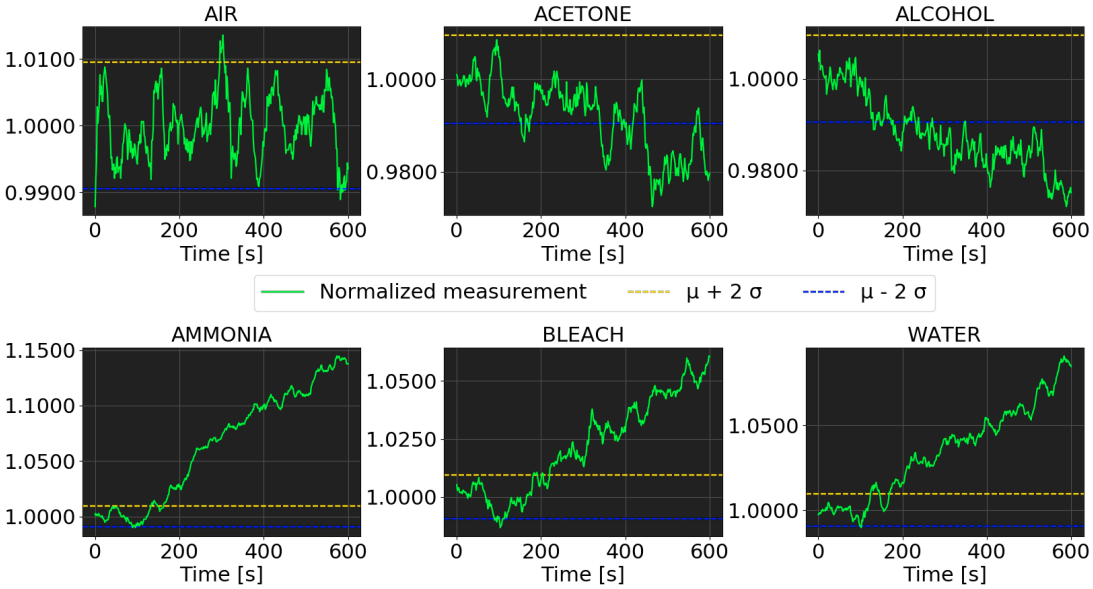


Fig. 3.5 Normalized ONCHIP_ALUMINUM_OXIDE sensor responses to contaminants

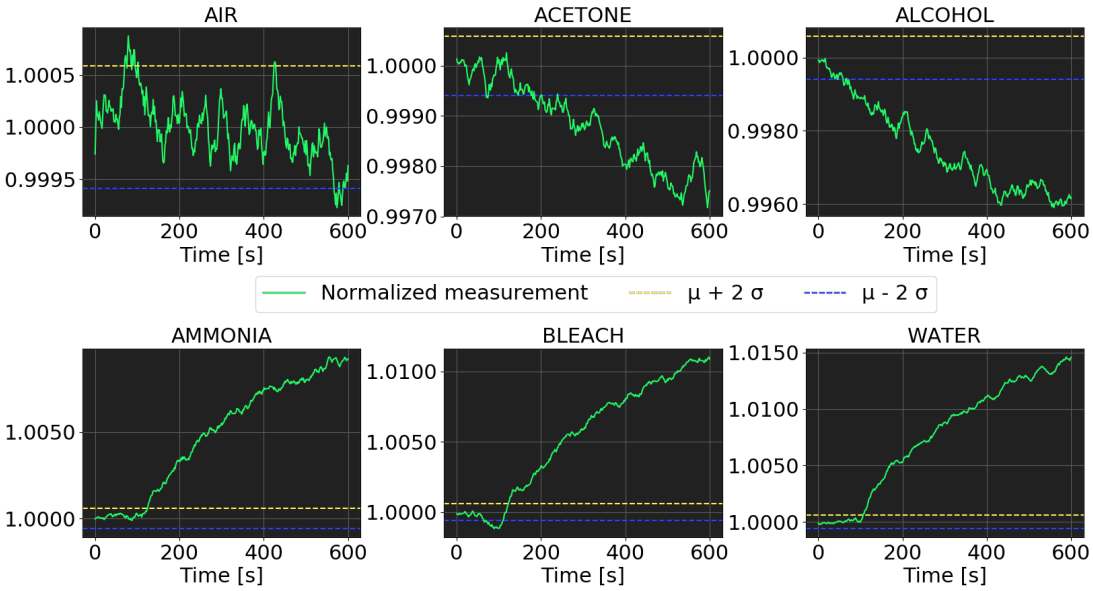


Fig. 3.6 Normalized OFFCHIP_HUMIDITY sensor responses to contaminants

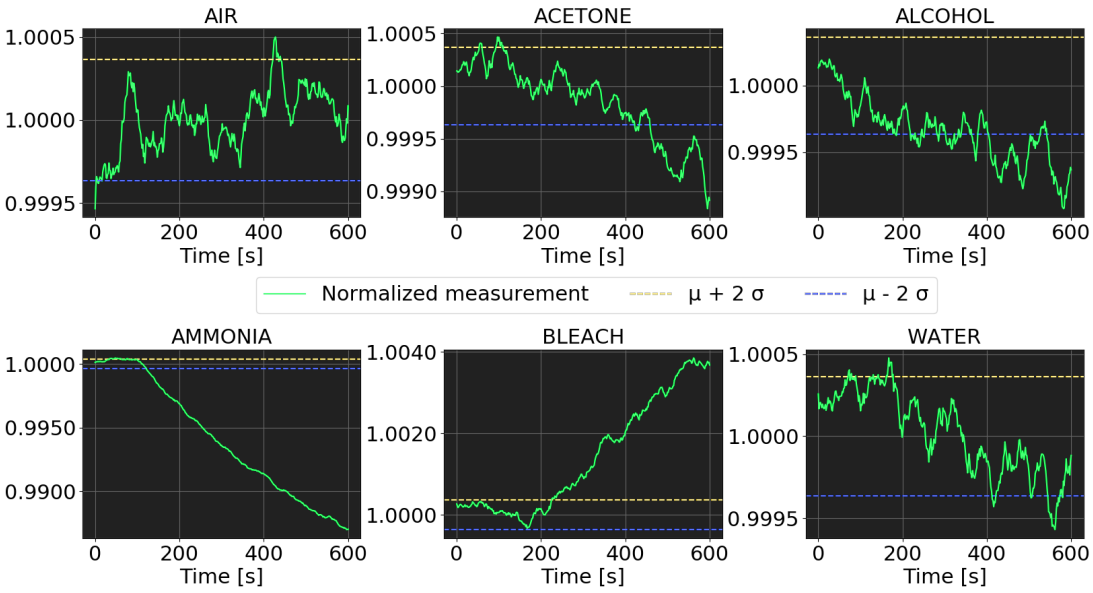
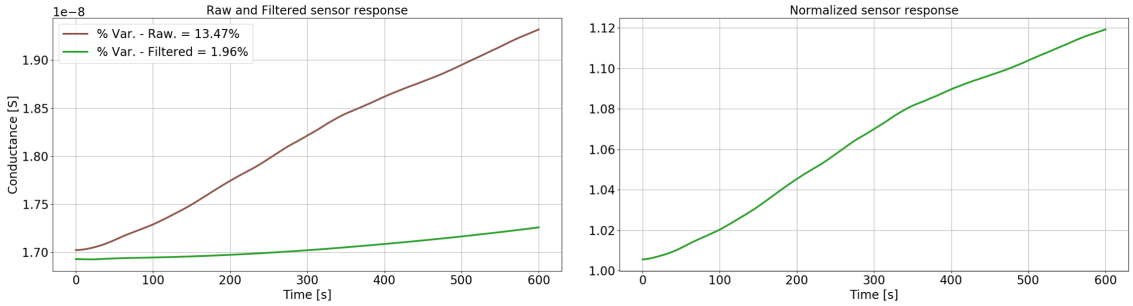


Fig. 3.7 Normalized OFFCHIP_GRAPHENE sensor responses to contaminants

As for instance, the output process of the EMA filtering and normalization is shown in Figures 3.8(a) 3.8(b) for one sensor response. As reported in Figure 3.8(a) a 13.47% and 1.96% have been obtained as percentage variation for raw and filtered responses, respectively. The



(a) (b)

Fig. 3.8 EMA filtering and normalization examples. a) Raw sensor response and its EMA filtered version. b) Normalization computer on measurements shown in a.

reported percentages' variations prove the negligible effect of the EMA long term dependency problem on a short observation time interval.

Finally, Figures 3.9—3.11 report a synthetic overview of the normalized sensors responses for each experiment. The vertical black dotted lines separate the sensors responses to the different contaminants alphabetically ordered (acetone, air, alcohol, ammonia, bleach and water). All the clean air phases (initial and final) of each experiment are removed in this graphs to highlight and focus only the response to the contaminant. It is worth noting that, as expected according to the sensors nature and operating principle, the ONCHIP_ALUMINUM_OXIDE and OFFCHIP_HUMIDITY sensors show a very similar behavior. Contrary, a meaningful selectivity and complementarity is observable for the OFFCHIP_GRAPHENE sensor responses which, for this reason, is expected to help the recognition capability of the classification stage.

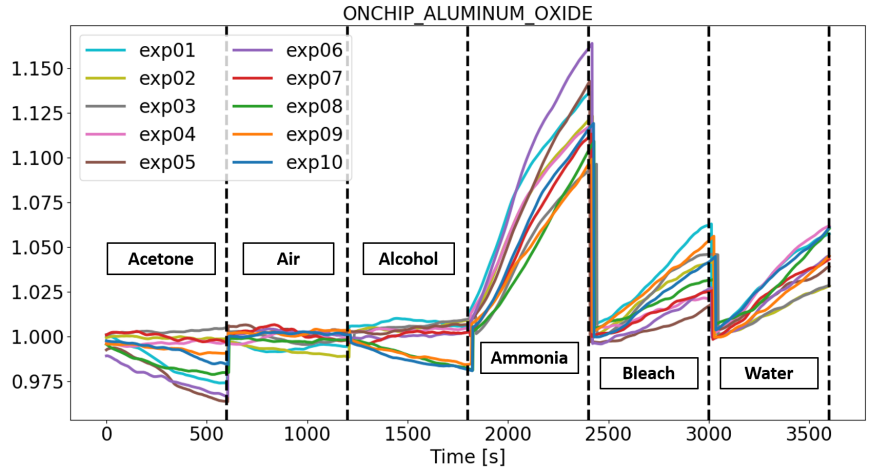


Fig. 3.9 Responses to contaminants acquired with the ONCHIP_ALUMINUM_OXIDE sensor. Substances order: acetone, air, alcohol, ammonia, bleach and water.

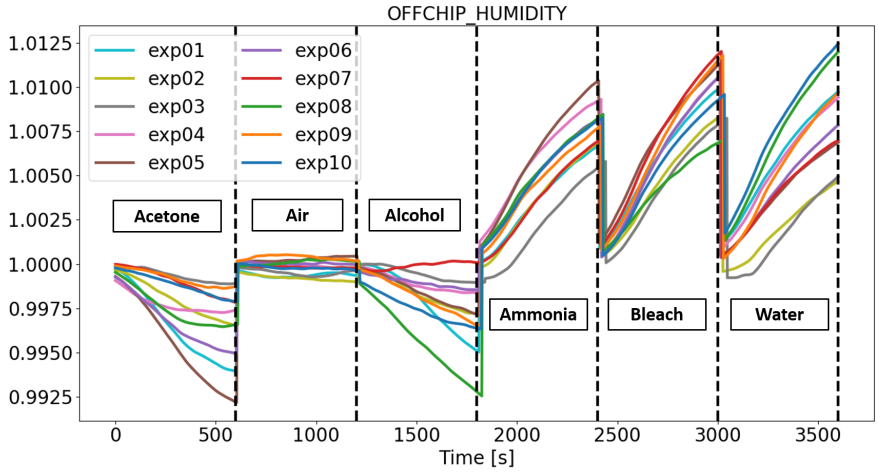


Fig. 3.10 Responses to contaminants acquired with the OFFCHIP_HUMIDITY sensor. Substances order: acetone, air, alcohol, ammonia, bleach and water.

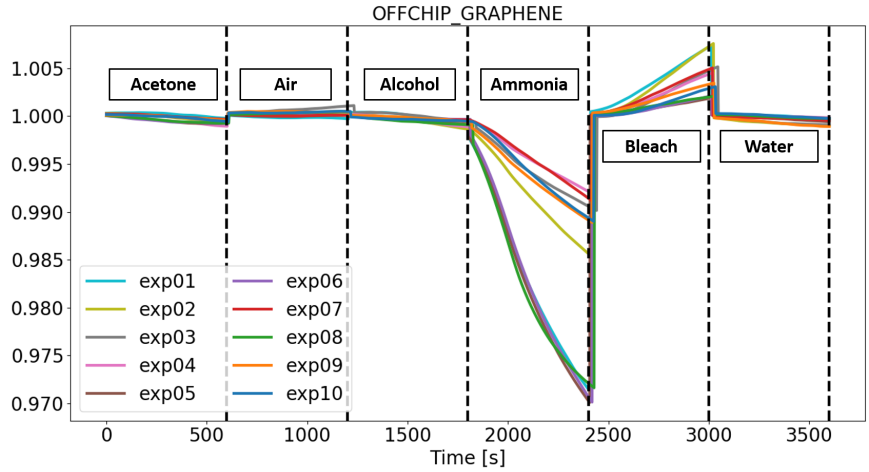


Fig. 3.11 Responses to contaminants acquired with the OFFCHIP_GRAPHENE sensor. Substances order: acetone, air, alcohol, ammonia, bleach and water.

3.6.2 Classification results

The different architectures model training and testing performance have been evaluated according to the recognition accuracy figure of merit, calculated as reported in equation 3.3. A very synthetic overview of the obtained results is provided in table 3.1 in terms of mean (μ) and standard deviation (σ) computed on the recognition accuracy obtained at each iteration of the cross–validation procedure.

Table 3.1 Mean and accuracy values evaluated for each ML architecture

Architecture	**Mean (μ)**	**Standard Deviation (σ)**
MLP_3	71.1 %	7.9 %
MLP_{120}	62.4 %	10.8 %
CNN	**75.1** %	**5.6** %
LSTM	69.9 %	12.1 %

$$Accuracy\ [\%] = \frac{Correctly\ Classified\ Samples}{Total\ Samples} \times 100 \tag{3.3}$$

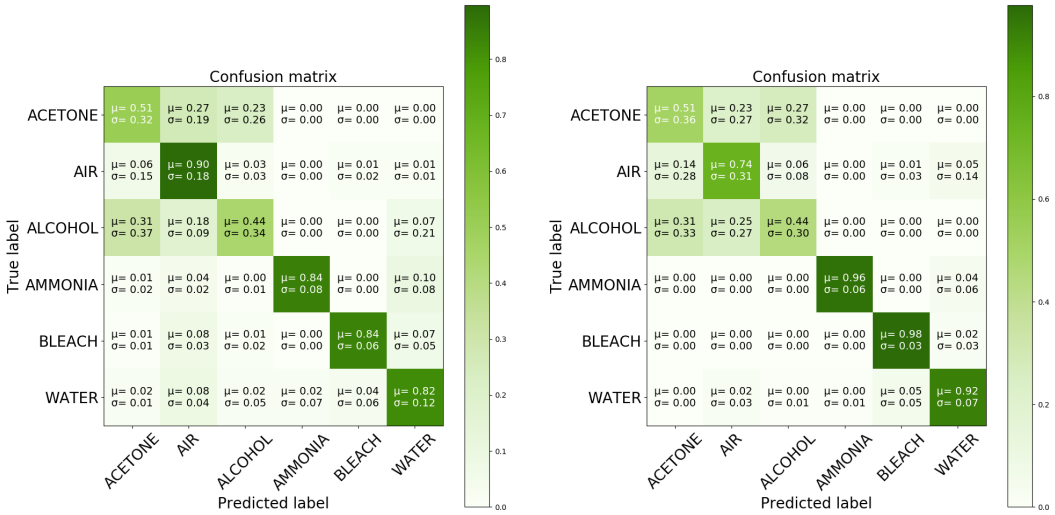
Overall results show that the CNN has brought to best performance both in terms of mean and standard deviations. By a general point of view, considering the concept of measurement compatibility, no considerable improvements have been reached with the two DL techniques. Furthermore, worse results have been obtained with the MLP_{120} architecture. In this case, the accuracy mean is about 9% lower than the one had with the MLP_3, probably due too low network complexity with respect to the input layer size. More detailed results are provided for each class always in term of accuracy mean and standard deviation in the form of confusion matrices in Figures 3.12(a)-3.12(c).

The confusion matrix is a figure of merit showing the ground truth (*true label*) on the rows and the predictions (*Predicted label*) on the columns. The correct recognition accuracy values are placed on the main diagonal and the best ideal case is represented by the identity matrix. The μ and σ values shown in a *(i,j)* matrix cell of Figure 3.12, represent the mean and the standard deviation (computed on the 10-fold cross validation testing results) of the samples that belong to class *i* and have been classified as *j*.

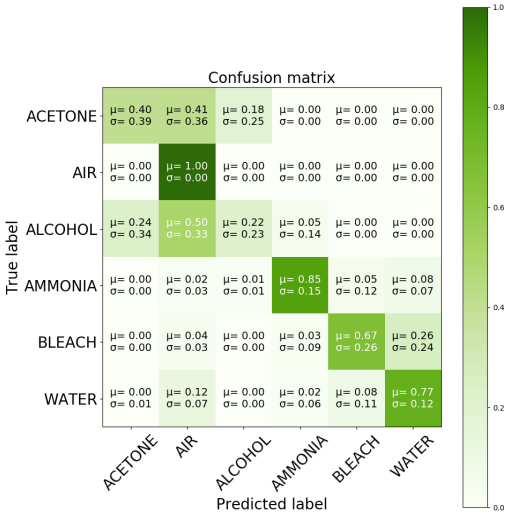
The analysis of confusion matrices results highlights three main observation:

- The integrated system has shown good sensitivity and recognition capability for ammonia bleach and water vapor. Indeed, high accuracy values with relative small standard deviations have been obtained for these classes with all three architectures.
- Although CNN architecture has provided best overall results, a lower performance characterizes the recognition of air samples with respect to MLP and LSTM.
- All three architectures confuse acetone and alcohol samples. These substances chemical nature, as the low boiling points, makes it hard to be detected by the adopted sensors array.

For this reason, an analysis of the confusion matrices unifying the acetone and alcohol in a unique class, namely acet–alc, has been carried out. Obtained results are shown in Figures



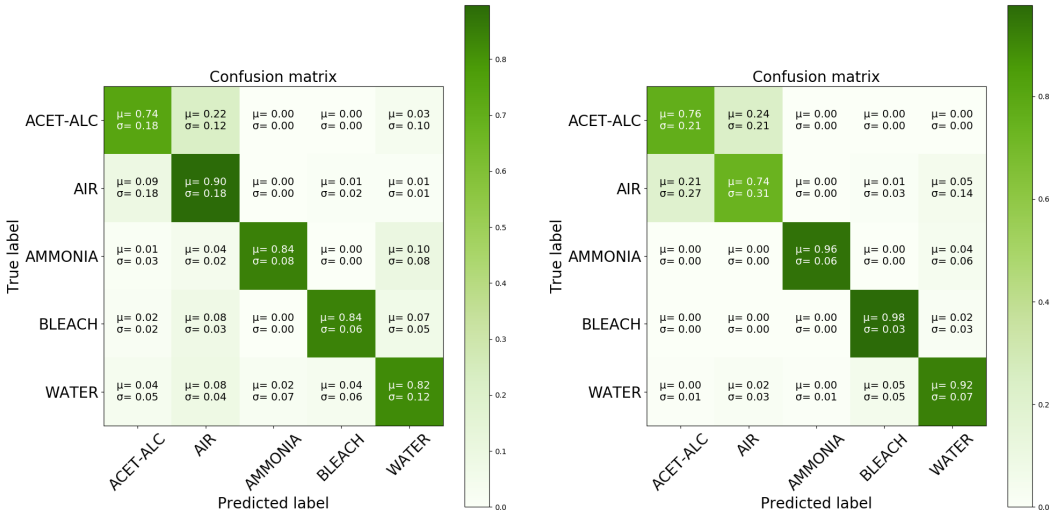
(a) MLP (b) CNN



(c) LSTM

Fig. 3.12 Confusion matrices obtained with MLP (a), CNN (b) and LSTM (c).

3.13(a)–3.13(c). In such a way, the element (1,1) corresponding to the acet-alc samples recognized as acet–alc is characterized by a high recognition rate, comparable with those belonging to the other classes.



(a) MLP (b) CNN

(c) LSTM

Fig. 3.13 Confusion matrices obtained for MLP (a), CNN (b) and LSTM (c) unifying acetone and alcohol as acet–alc class.

3.7 Computational analysis of MLP and CNN

An evaluation of the computational complexity, in terms of FLoating point OPerationS (FLOPS), both for the best performance and the lightest architectures has been conducted in this research activity. Needed FLOPS for both MLP_3 and CNN have been evaluated according to equation 3.4, where *CL1-CL2* and *FC* represent convolutional layer and fully connected flops, respectively. Reported equations adopt the notation symbols detailed in Table 3.2.

Table 3.2 Notation symbols used for FLOPS evaluation

Symbol	**Description**	**Value (MLP)**	**Value (CNN)**
w	time window	1	40
s	sensors	3	3
k_1	1^{st} layer kernels	N/A	16
k_2	2^{nd} layer kernels	N/A	14
k	1^{st}-2^{st} L kernel size	N/A	3
n_0	input neurons	3	$w\cdot s\cdot k_2$
n_1	hidden neurons	64	128
n_2	output neurons	6	6

$$FLOPS = \begin{cases} wsk_1(2k^2+1) & \text{CL1} \\ wsk_2(2k^2k_1+1) & \text{CL2} \\ 2n_1(n_0+n_2) & \text{FC} \end{cases} \tag{3.4}$$

Obtained FLOPS for each layer of the two architectures are reported in Table 3.3. As highlighted by these results, the MLP is clearly much lighter than CNN because of its lower complexity. Indeed, the lack of features' extraction layers and the reduced number of neurons used as FC input (s vs wsk_2) and hidden layer sizes (64 vs 128) required a much lower number of operations for the MLP_3 than CNN.

Table 3.3 Evaluated number of floating point operations for CNN and MLP.

Technique	**CL1**	**CL2**	**FC**	**Tot.**
CNN	36480	485520	462848	984848
MLP	N/A	N/A	1152	1152

3.8 Embedded implementation of MLP

An embedded implementation of the whole SDM operations chain has been carried out. The MLP architecture has been adopted and implemented as the data analysis block. To this aim, the ESP32 MCU has been used. Experimental testing campaigns have been carried out with the developed SENSIPLUS Embedded System. Execution times and power consumption have been assessed and reported in 3.4. Here times and power consumption relative to each phase are reported. As regards the acquisition, 372 ms are required to perform a single timestep measurement from the 3 sensors. Elapsed time for data analysis, based on MLP, has been equal

Table 3.4 Execution time and power consumption

	Acquisition	**MLP**	**Sleep**
T [ms]	372	1	1627
P [mW]	201	218	2.6

to 1 ms, which can be considered negligible with respect to the acquisition phase. Considering an overall rate of 1 classification every two seconds, as the one used during dataset acquisition, the ESP32 is set in its light–sleep mode to minimize the power usage during the inactive phase (1626 ms). A power consumption of 39.6 mW is obtained as weighted mean. Aiming to propose a complete prototype of this system the two battery models, reported in the following, have been considered and their times of life have been estimated:

- Coin cell battery 3 V, 500 mAh - Expected time of life: 41.5 h.
- Saft Litio battery 3.6 V, 17 000 mAh - Expected time of life: 59 days.

These two battery models can be considered to address two distinct scenarios: the smaller coin cell battery can be used as a backup where the device is reachable by electric network, while the Saft Litio can be used in battery operated applications given its longer lifetime.

3.9 Conclusion

The development of an integrated system endowed with sensing and computing capabilities for indoor air monitoring and detection of dangerous substances has been developed. Experimental activities for both sensing and elaboration have been presented and multiple data analysis techniques have been compared for the classification of chemicals in air. Finally, a fully embedded implementation of the complete system, involving a MLP as classification technique, has been performed adopting the ESP32 as MCU running all the required software. Good sensitivity and classification capabilities have been proven and reported in the results section for most of substances, while confusion has regarded acetone and alcohol, which have been finally unified as a unique class given their chemical similarity. As regards research activity follow–up, further low–cost and low–power sensors with high sensitivity to acetone and alcohol are going to be adopted. Furthermore, an improvement in terms of experimental setup placing sensors in larger glass containers or, possibly, entire testing rooms where environmental conditions (i.e. temperature and humidity) can be properly changed is very important to validate and advance the developed technology.

Chapter 4

Detection and Recognition of contaminants in Water

4.1 Introduction

Water quality is a worldwide concerning topic since it is increasingly threatened by the intensive industrial and agricultural activities. Accidental or premeditated contamination can rapidly cause environmental disasters with severe consequences for human health. For this reason, affordable and reliable water monitoring solutions represent the key to enable more pervasive control to reduce the risk of diffusion of chemical, biological and radiological contaminants in drinking water, rivers, lakes and sea. Current controls based on laboratory instruments and expert professionals can not provide continuous examinations and rapid detection of contaminants emissions. Furthermore, their high costs are the main cause for very low spatiotemporal samplings. Although this is a global scale phenomenon, worse conditions characterizes poorer and developing countries. Water pollution levels represent huge hazards for people life quality and planet safeguard [62–64]. This is a very sensitive topic for WHO, which defines the guidelines to accurately assess and control water and for its safe management [65]. Low-cost sensors technologies capable to provide continuous monitoring are the basis to guarantee water safety. Many efforts, both in the industrial and academic research field, are addressed to the development of on–line monitoring water sensor technologies [66][67] and to perform on field early analysis and detection [68, 69]. Considering the very complex environment and the many discriminating water parameters as pH, conductivity and turbidity (as documented in [70] by US Environmental Protection Agency) to be analyzed, statistical and AI techniques are often adopted [71] for the analysis of the huge amount of acquired data. Since low-cost sensing technologies heavily suffer the poor selectivity problem since no specific sensors can be found

for each contaminant of interest, more general parameters are analyzed and sensors fusion algorithms are commonly developed.

This research activity, carried out as part of this Ph.D. thesis work, is based on two specific sub-projects requirements, characterized by different objectives in terms of substances to be detected and recognized. As further described in this chapter, it has been necessary to involve two different measuring and data analysis techniques for the two substances subsets. As regard data acquisition, Electrical Impedance Spectroscopy (EIS) and the Cyclic Voltammetry have been the two involved techniques, while different ML techniques have been adopted for data analysis. More in detail, a MLP based approach has been used to deal with the EIS measurements and Principal Component Analysis (PCA) has been applied to reduce the feature space dimensionality, lightening the MLP complexity. As for voltammetry data, Random Forest has been adopted as classifier and the statistical method Linear Discriminant Analysis (LDA) has been exploited to reduce the dataset dimensionality. As for the air monitoring application, a multi–class classification solution has been adopted here for both EIS and voltammetry dataset. Experimental activities have been conducted submitting a single substance for every experiment, providing mutually exclusive classes in the obtained dataset. This research activity has been first proposed in [72], where a preliminary prototype of the SENSIPLUS Embedded System (section 1) for on–line water monitoring and for the detection and recognition of chemical contaminants is proposed.

4.2 State of the Art

Many research activities are conducted for the study and development of novel low-cost sensor technologies and integrated system for on–line and near real time water analysis. Different approaches are commonly pursed both for sensing and data analysis. A review of the emerging technologies for water quality parameters monitoring and their relevant role in nowadays needed large scale deployable sensor networks is provided in [73]. Electrodes made of different materials [74], optical sensors [6] and electrodes covered by sensing films [75] are the most adopted solutions in the scientific scenario. More in detail, a low cost monitoring system integrating both optical and electrochemical sensors is proposed in [6]. The authors develop a solution for pervasive in–pipe drinking water monitoring exploiting commercially available low-cost technologies. Solutions based on optical approaches are proposed also in [76, 77], where a colorimeter and fluorescence spectroscopy are exploited, respectively. Electrochemical sensing devices are usually adopted with impedance spectroscopy and cyclic voltammetry measuring techniques. While the former is based on analyzing the sensor response in the frequency domain, the latter performs current measurements through a voltage sweep. Impedance spectroscopy is

the measurement method adopted in [78] for the detection of Escherichia coli in river water samples. As for cyclic voltammetry, a solution for the detection of harmful substances in water based on this technique is proposed in [74]. Further voltammetry methods as the Anodic Stripping Voltammetry (ASV) and the Differential Pulse Voltammetry (DPV) have been used in [79] to detect metal ions in water samples. DPV, which is mainly based on performing differential current measurements between two consecutive voltage potentials, is studied in [80] for the determination of nitrite in river water samples. Here, the authors prove a sensitivity improvement using DPV with respect to linear sweep voltammetry. Functionalization of electrodes through the deposition of specific sensing films is often adopted to analyze the capability to improve the sensitivity to specific contaminants. Such solutions are mainly affected by the problem of fast deterioration as highlighted in [75, 81]. As regards the analysis of the large amount of acquired sensors data, novel approaches and methods based on AI techniques are widely adopted in literature. Artificial Neural Network and Principal Component Regression techniques are used in [82] for nitrate concentration estimation in ground water. In [83] and [74] the authors propose ML based solution to detect and classify contaminants in water. More in detail, in [74] the authors propose a pattern recognition approach, based on partial least-square discriminant analysis (PLS-DA), for the contaminant prediction. LSTM and Fully Convolutional Networks (FCNs) have been adopted in [84] as ML technique to detect and identify chemicals in sea water. Some further works can be found extending the search to contaminant detection in other liquids, not specifically water. That is the case of [85], where a sensor array, composed of a pH measuring device and an electrical impedance sensor, and a K-nearest neighbors (KNN) based algorithm have been proposed for the detection of Bovine Milk Adulteration. Electrical impedance spectroscopy is adopted as measuring technique in [86] in order to evaluate and analyze the presence of chemical additives in juices. Here the authors propose an approach based on equivalent electrical circuits, analyzing the relationship between circuit parameters and the chemical additive presence.

4.3 Data acquisition for water monitoring

4.3.1 Experimental measurement setup

The experimental measurement setup shown in Figure 4.1 for dataset acquisition has been designed and developed. It is composed of the items listed in the following:

- two interdigitated gold and copper electrodes (IDEs) used as sensors for EIS measurements;
- an IDE covered with platinum for cyclic voltammetry measurements;

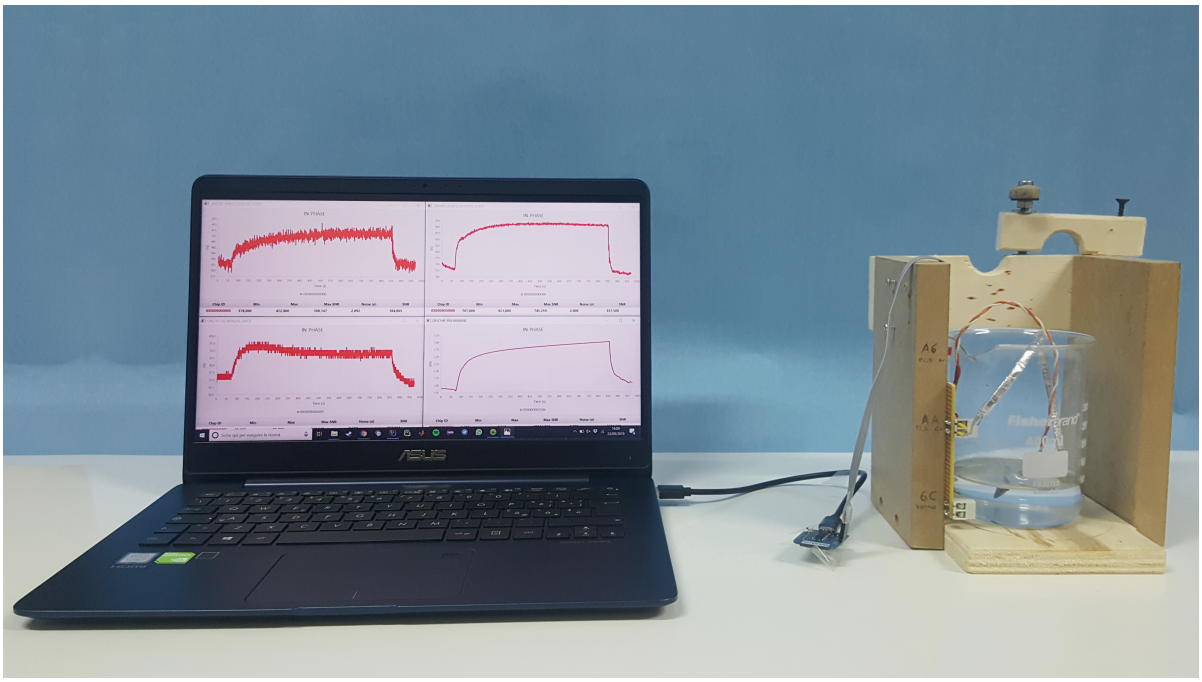


Fig. 4.1 Experimental measurement setup

- a beaker containing the water and contaminants where the IDEs are immersed;
- the SP chip glued on the beaker external surface and connected to the IDEs;
- the ESP32 MCU running the SP API and a PC for data storage.

4.3.2 Electrochemical Impedance Spectroscopy (EIS) experiments

As introduced, the EIS and DPV measuring techniques have been exploited for the measurement acquisition phase. As regards EIS experiments, the measuring parameters have been optimized in a preliminary phase where proper sinusoidal stimulus frequencies and other measuring parameters have been selected. More in detail, electrical impedance measurements have been performed at two distinct stimulus frequencies (2 kHz, 78 kHz) and resistance–capacitance values have been calculated considering a R–C parallel equivalent circuit, represented in Figure 4.2, according to equations 2.8 and 2.9 (the measurement device provide impedance values in terms of modulus and phase). The two adopted frequencies have been chosen to acquire and explore different work-points of the sensors' frequency responses. In such a way, an improvement of the discriminating potentialities of the system is furthered since distinct electrochemical behaviors are expected in different frequency points of the sensors response. Five contaminants have been adopted for this experimental measuring campaign: alcohol,

bleach, sulfuric acid, detergent, NaCl. As regards detergent and NaCl, they are cleaning

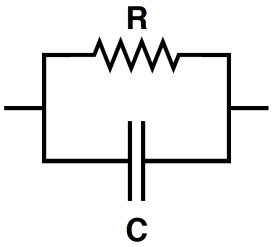


Fig. 4.2 Equivalent electrical circuit adopted for the gold and copper IDEs

substances and salt which can be considered as contaminants emitted in water. More in detail, the adopted detergent is composed of 20% of Sodium Carbonate, 20% Oxygen–Based Bleaching and less than 5% of non–ionic tensioactive substances.

A systematic measurement procedure has been designed and performed for each experiment in order to obtain comparable raw measurements. It is composed of the following points:

- impedance baseline assessment for 350 seconds with the sensors immersed in tap water;
- chemical contaminant injection;
- impedance measurements for further 1300 seconds, acquiring both transient and regime periods of the sensor response.

For each substance 9 different experiments have been carried out. Further 9 acquisitions have been performed with only tap water inside the beaker in order to capture samples of the working background. Finally, computed resistance and capacitance values for both frequencies and for the 2 IDEs have been collected in a 1400 x 8 dataset for each experiment, where 8 is the number of variables and 1400 the acquisition timesteps.

4.3.3 Differential Pulse Voltammetry (DPV) experiments

As regards the cyclic voltammetry experiments, the DPV method has been involved for the second set of substances. From a general point of view, a voltage stimulus, as depicted in Figure 4.3, is applied to the electrodes immersed in solution and at each step two current measurements are performed. A first measure is acquired after the voltage application while the second is performed after the stimulus is updated. Finally, a unique current values is reported as the difference between the measured quantities. In such a way, it is expected to obtain current peaks in correspondence of the chemicals oxidation/reduction potentials and low currents for the remaining voltage range. Preliminary experiments have been conducted to tune the waveform

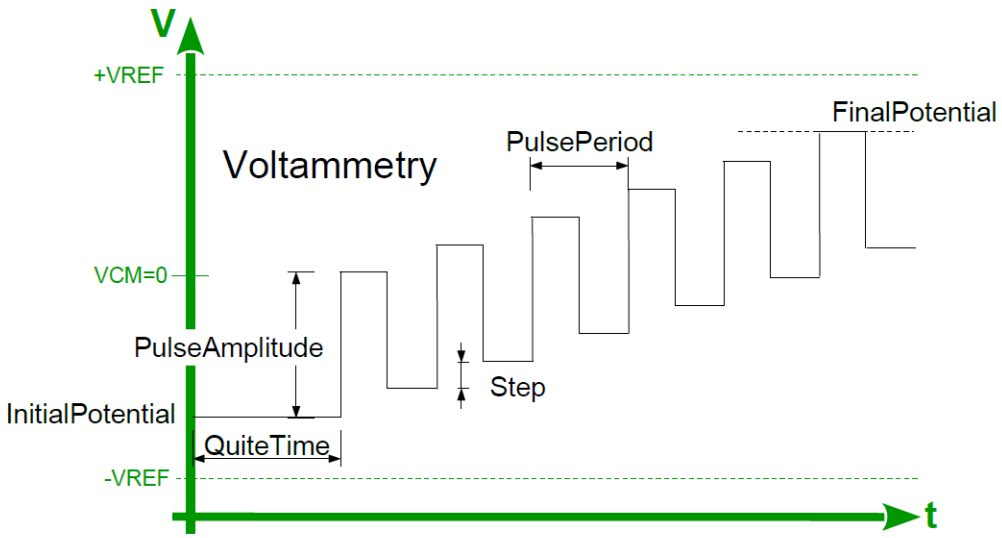


Fig. 4.3 General representation of Differential Pulse Voltammetry waveform

settings, reported in Figure 4.3, in order to optimize the acquired current measurements cycle, which is defined as the sweep of voltage both for the increasing (initial to final potentials) and decreasing (final to initial potentials) phases. The parameter listed in the following have been finally selected:

- Voltage range: [-1250, 1250] mV
- Pulse Amplitude: 100 mV
- Pulse Period: 50 ms
- Step: 40 mV

The analysis of best parameters has been carried out to maximize the current peak amplitude while minimizing its width. In such a way, a more defined and specific pattern characterizes the measured currents with respect to the contaminants. A trade–off has been met since the reported optimization inversely depend on the voltage sweep speed.

Formic acid, phosphoric acid and sodium hydroxide have been the chemicals tested with DPV. The reason for this particular substances set is related to the specific scenario, which is the industrial waste water monitoring. Indeed, since these substances are often used in industrial processes, their detection is the proof that filtering systems are not properly working and early warning notification can prevent environmental disasters.

The measurement procedure adopted for DPV experiments has been carried out performing measurements once the chemical was dissolved in the water. Multiple measurement cycles have

been performed with a specific solution in the same experiment in order to observe the system repeatability. Furthermore, for the same chemical different experiments have been conducted cleaning the beaker and changing the solution to test the system capability to generalize with respect to the specific solution. Finally, different platinum electrodes have been used for data acquisition. An overall dataset of 120 cycles for each chemical contaminant has been produced, where each of them is composed of 164 current measurements.

4.4 Data Pre-processing

As regards the pre-processing block, the same approach as for the indoor air monitoring application (chapter 3) has been applied to raw measurements acquired through EIS experiments. The environmental reference r_t evaluated for each measured quantity through EMA filtering with formula 3.1 has been used to perform the sensor response normalization (formula 3.2). As for air, the adopted normalization solution has resulted to properly work in this specific scenario, where early detection and recognition capabilities are required. As output of this SDM block the features vector ($f.v.(t) = [f.v.(1,t), f.v.(2,t), .., f.v.(8,t)]$) is computed and provided for each acquisition timestep. It is composed of the normalized version of the 8 measured quantities. Finally, the labeling process is performed and a ML suitable dataset is provided for the data analysis phase. It has been splitted into training, validation and test sets, obtaining the number of samples reported in Table 4.1.

As for the DPV technique, measurement cycles have been labeled according to the substance used in the specific experiments and properly collected for the analysis and classification phases. The whole dataset has been divided into three subset for classification purposes. Considering a whole DPV cycle as a sample of the final dataset, a total amount of 480 samples has been produced (120 for class). Finally, it has been divided in training (80 samples), validation (20 samples) and test (20 samples) sets

Table 4.1 EIS dataset: number of samples for training, validation and test sets

Classes	Training	Validation	Test	Total
Alcohol (1)	5183	3174	1014	9371
Bleach (2)	5059	3073	990	9122
H2SO4 (3)	5079	3071	1012	9162
Detergent (4)	5078	3081	1013	9172
NaCl (5)	5078	3071	1014	9162
Water (6)	5079	3072	1014	9165
Total	30556	18542	6056	55154

4.5 Data analysis for EIS dataset

In order to minimize the computational burden of the SDM data analysis block, the statistical method PCA [87], has been adopted. It has been used to reduce the features space dimensionality simplifying and lightening the MLP architecture with a decreased input layer size. An acceptable trade–off between dimensionality reduction and information loss has been reached analyzing the Principal Components (PC) scores. The projection of the original dataset into a 3 dimensional data–space is computed applying the transformation of the PCA model which has been fitted on the training set and used for test sets. The transformed dataset for one of the 9 experiments carried out for each substance is reported in Figure 4.4. Since the projection of the whole sensors responses (baseline, transient and steady state according to section 4.3) is depicted, all contaminants curves start from the same region which corresponds to clean tap water. Different regions are reached by the regime zones (depicted with darker colors tones) which is a proof of discriminability. Loading coefficients of the fitted PCA model are depicted in Figure 4.5.

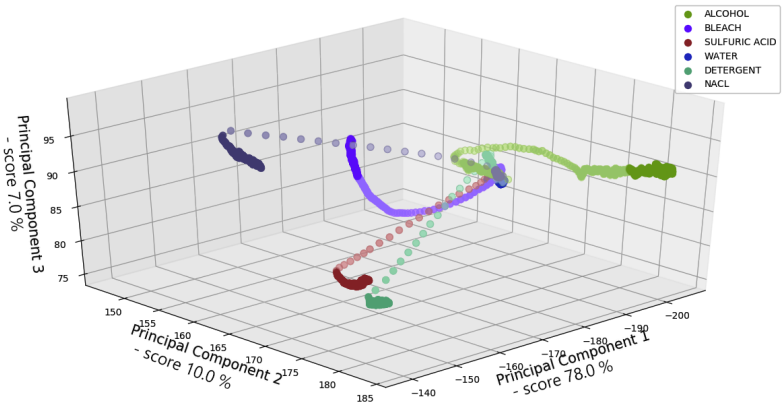


Fig. 4.4 Raw dataset projected on the first 3 PCs.

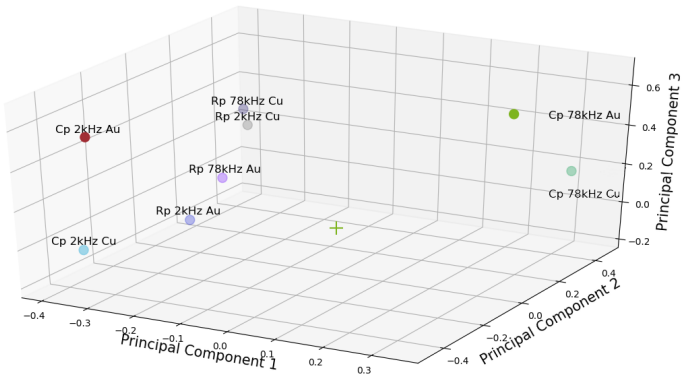
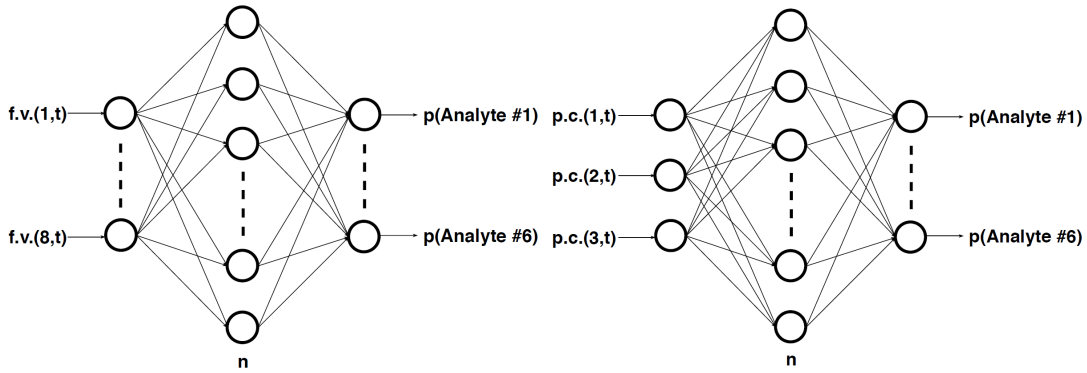


Fig. 4.5 Loadings coefficients used for the PCA transformation

This plot allows to state that all original features have a meaningful weight in the transformation. Finally, an overall information score equal to 95% is obtained as the sum of the PC scores reported in Figure 4.4 for each axis.

4.5.1 MLP architectures

As regards the classification phase, two MLP with a different input layer size have been implemented to deal with the feature vector ($f.v.(t)$) and the PCA transformation output. A schematic representation of both architectures is pictured in Figures 4.6(a) and 4.6(b), where a different input layer is reported. Furthermore, aiming to minimize the network complexity different hidden layer size have been tested. Starting from 64 neurons in the hidden layer, further experiments have been conducted halving the hidden layer size. A meaningful worsening of performance has been obtained with hidden layer size lesser than 16 neurons and for this reason not reported in section 4.7. Finally, the *ReLU* activation function has been chosen for all the tested architectures in order to minimize the vanishing gradient problem and to provide an easy embedded programming.



(a) MLP 8 input (b) MLP with 3 input

Fig. 4.6 MLP architecture with 8 (feature vector) and 3 (principal components) input and 6 output neurons

4.5.2 Computational Analysis

Since the final research activity aim is to provide an embedded implementation of the whole system, an analysis of the computational costs of the proposed solutions for data analysis has been carried out. To this aim, a study of the needed mathematical operations for the involved block has been performed in terms of FLoating point OPerationS (FLOPS). Obtained results are shown in Table 4.2, where n_0, n_1 and n_2 correspond to the number of neurons

Table 4.2 FLoating Point Operations for tested configurations with and without PCA pre–processing

n_0	n_1	n_2	FLOPS MLP	FLOPS PCA	Total FLOPS
3	16	6	288	45	333
3	32	6	576	45	621
3	64	6	1152	45	1197
8	16	6	448	0	448
8	32	6	896	0	896
8	64	6	1792	0	1792

in the input, hidden and output MLP layers, respectively. FLOPS for PCA transformation and MLP classification are separately reported. As regards the MLP, an increasing FLOPS value is related to more complex network architectures (i.e. greater number of neurons both in the input and hidden layer), calculated according to equation 4.1. Needed FLOPS for PCA transformation are always the same (equal to 45) since the projection of a vector from 8 to 3 dimensional data space is performed in all cases. Furthermore, as expressed in previous sections the PCA transformation is not applied when the MLP input layer is endowed with 8 neurons, corresponding to the original feature vector.

$$\text{FLOPS MLP} = 2n_1(n_0 + n_2); \tag{4.1}$$

Lighter solutions are obtained with the PCA transformation and using the MLP with 3 input only. Despite of the additional operations required for the data–space dimensionality reduction through the PCA based solutions, an overall reduced total FLOPS values is obtained. For this reason, a lighter implementation of the data analysis block can be reached involving the PCA based solution and the 32 hidden neurons MLP since it can be considered the best trade off between performance and complexity.

4.5.3 Network training

Implemented networks' hyper–parameters have been initially tuned to avoid overfitting and to let training converge in a relatively short time. A batch size equal to 32, a learning rate equal to 10^{-5} and a momentum equal to 0.9 have been chosen. Furthermore, the cross entropy has been used as loss function, Adam Optimizer as optimizing algorithm and dropout has been set to 1%. The early stopping strategy, based on a maximum number of epochs without improvements on the validation set accuracy, has been used to stop the model training avoiding overfitting. Finally, in order to statistically validate the model on the whole dataset a 9-fold cross validation strategy has been adopted. The test set has been selected as an entire experiment among the 9

total ones and the 8 remaining have been randomly splitted in training and validation according to 80% and 20%, respectively. At each of the 9 fold of cross validation the same approach has been performed rotating the experiment used as test. The same hardware devices, as for indoor air monitoring application, have been involved for the training and testing phases: a Linux computer endowed with an Intel® Xeon® E5-2609 v4 (8 cores) CPU, 256 GB RAM and a NVIDIA® TITAN X Pascal (exploited for training acceleration).

4.6 Data Analysis for DPV dataset

As for the EIS dataset, a features' space dimensionality reduction has been applied to the DPV dataset. To this aim, the Linear Discriminant Analysis (LDA) [88] has been adopted and used to project the original dataset into a 3 dimensional space. The result of such transformation is depicted in Figure 4.7, where, as shown, points belonging to different classes form separate clusters. As regards the substances recognition task, the Random Forest [89] classifier composed

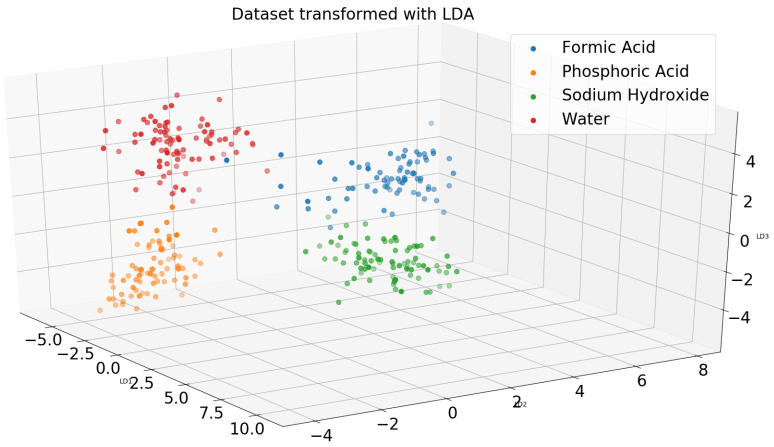


Fig. 4.7 DPV Dataset projected through LDA

of 10 trees with a maximum depth equal to 2 has been trained and tested. Also in this case a cross validation strategy based on 3-fold has been exploited. Results regarding the classification performance are reported in section 4.8.

4.7 Results for EIS dataset

For each of the 6 implemented architectures, accuracy (evaluated according to 3.3) mean and standard deviation values have been calculated over the 9-fold testing experiments. Global results are provided in Table 4.3, where the maximum mean value equal to 87.3% has been

Table 4.3 Accuracy mean (μ) and standard deviation (σ) values for each MLP architecture.

Hidden Layer neurons	**PCA**	**Mean** (μ)	**Standard Deviation** (σ)
64	no	0.8226	0.11651
64	yes	0.8009	0.12908
32	no	0.7795	0.10832
32	yes	**0.8732**	0.13132
16	no	0.8264	**0.10541**
16	yes	0.8366	0.12749

obtained with the 32 hidden neurons MLP based on PCA and the minimum standard deviation of 10.5% has resulted for the 16 hidden neurons MLP with the 8 original feature vector as input. For a quantitative comparison of the networks performance, a unique figure of merit, namely dispersion index (DI), has been calculated as in formula 4.2. Obtained results, reported in Figure 4.8, show a higher variability for the classification accuracy results obtained with MLP based on PCA transformation.

$$DI[\%] = \frac{\sigma}{\mu} \times 100 \tag{4.2}$$

A further and more detailed representation of the obtained results is provided in the 6 confusion matrices depicted in Figure 4.9. For each of them, the generic $\hat{CM}_{i,j}$ cell corresponds to the percentage of samples belonging to class i and predicted as class j. As shown, the alcohol, bleach and water classes are the ones providing best results while lower accuracy are obtained for the others in most of cases. A common lack of repeatability of classification results can be highlighted observing the standard deviation values, underlining training result dependence on the specific set of data used in each fold.

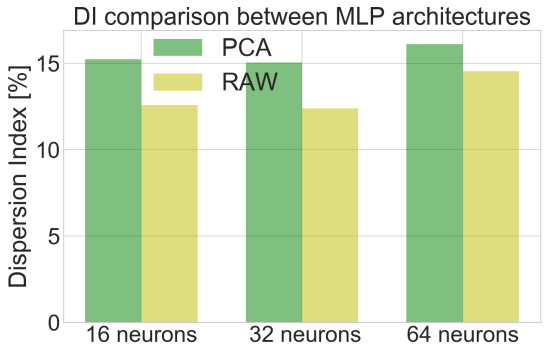
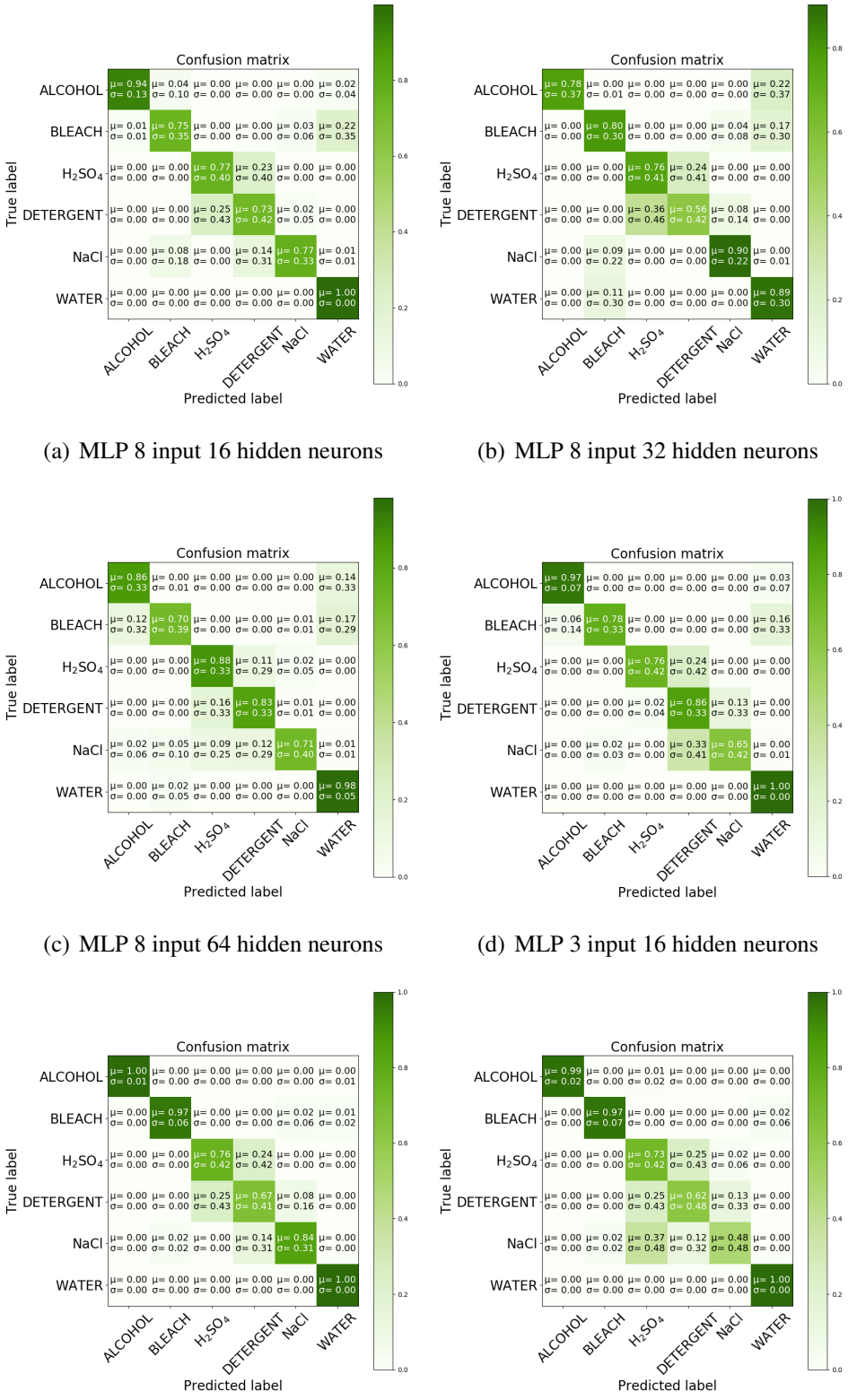


Fig. 4.8 Dispersion Index evaluated for the 6 ANN architectures



(a) MLP 8 input 16 hidden neurons

(b) MLP 8 input 32 hidden neurons

(c) MLP 8 input 64 hidden neurons

(d) MLP 3 input 16 hidden neurons

(e) MLP 3 input 32 hidden neurons

(f) MLP 3 input 64 hidden neurons

Fig. 4.9 Confusion matrices obtained with MLP architectures

4.8 Results for DPV dataset

The acquired data with the DPV method are reported in Figure 4.10. A DPV cycle sample for each substance is reported in this graph, where the x–axis corresponds to the applied voltage and the y–axis to measured currents.

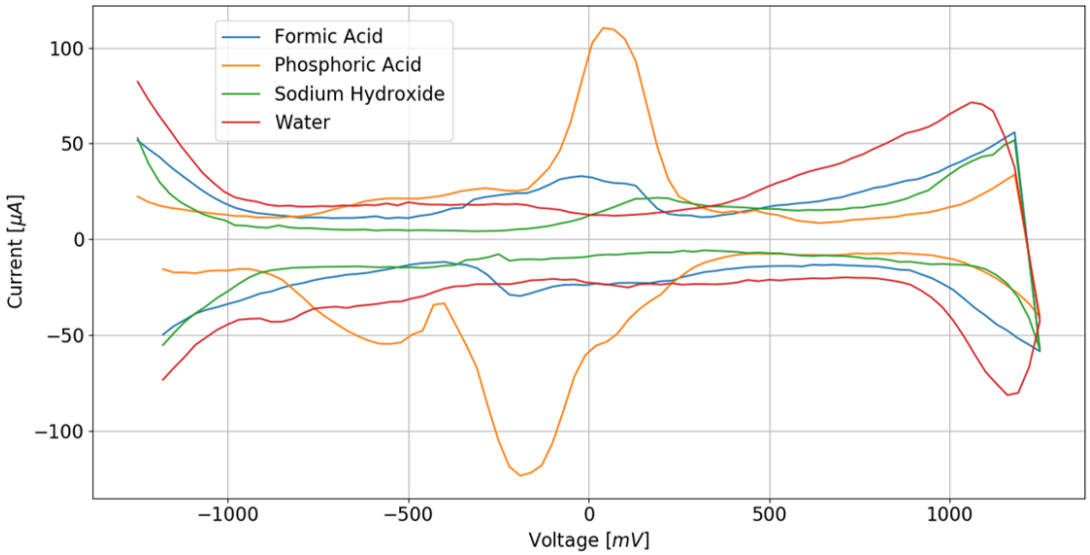


Fig. 4.10 DPV cycles for the adopted substances

The closed form of each line is due to the fact that the voltage sweep starts from -1250 mV and once reached the 1250 mV value comes back to the starting point performing the same steps. Furthermore, the reason why each cycle is characterized by positive and negative current measurements is due to the inverse trend of voltage variation during the two semi–cycles. While during the increasing phase the current measured at higher voltage is the first term of the difference, it becomes the second term during the descending part. Patterns not clearly evident to human eye are highlighted in this plots, except for Phosphoric Acid, where clear current peaks are obtained. Although raw data visualization does not provide clear discriminating potentialities among classes, the application of the cascade LDA and Random Forest has brought to high recognition capabilities as reported in the following. Before analyzing classification results, a further observation relative raw measurements is represented by current peaks obtained at the voltage range limits independently of the chemical substance. According to electrochemical theory, water electrolysis occurs about at the adopted voltage range limits. As result of the this reaction, more electrons are produced in the solution and an increase of measured current is obtained.

As regards classification tests, results obtained with the Random Forest classifier and through the cross validation strategy are summarized in the form of confusion matrix which is

shown in Figure 4.11. True and predicted labels are reported in the form of indexes according to the mapping: (0, Formic Acid) - (1, Phosphoric Acid) - (2, Sodium Hydroxide) - (3, Water).

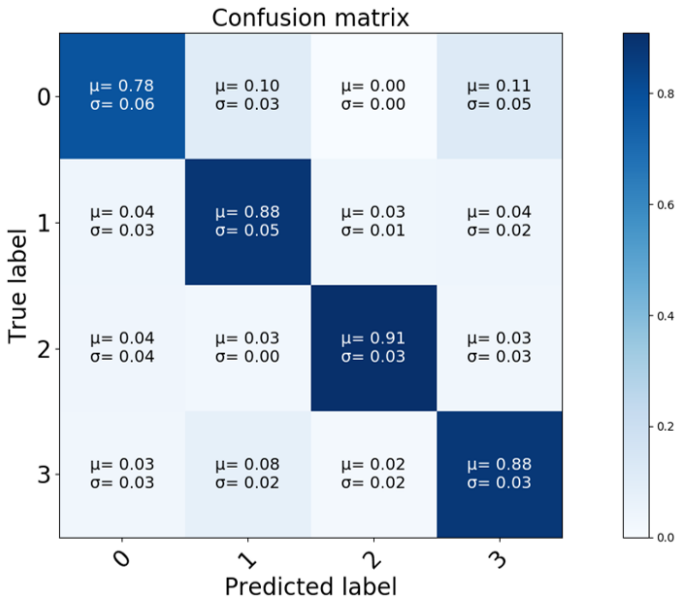


Fig. 4.11 Confusion Matrix obtained with the Random Forest classifier

For each cell the accuracy mean and standard deviation, evaluated on the 3-fold results, is reported. The best performance is obtained for the Sodium Hydroxide class (2) with 91% and 3% as accuracy mean and standard deviation, respectively, while the worst result is related to Formic Acid class (0). From a global point of view, considering all classes, a mean accuracy equal to 85% and standard deviation equal to 2.5% have been obtained by the classification tests.

4.9 Conclusion

The research activity conducted for detection and recognition of contaminants in water has brought to the development of a customized version of the SENSIPLUS embedded system for this application. The proposed solutions are based on the adoption of electrochemical sensors and lightweight data analysis techniques suitable for embedded low-power and low-cost devices. The entire chain of the integrated system, composed of both hardware and software components, has been designed and developed. Finally, testing experimental campaigns have been conducted to validate the system detecting and recognizing capabilities. Positive results in terms of accuracy have been obtained, encouraging the research activity follow–up. A wider set of contaminants is the first target for future works involving also further data analysis

techniques. For instance, according to the study of scientific literature, an approach to be tested regards the transformation of measurements buffer (e.g. Differential Pulse Voltammetry cycles or time interval of Electrochemical Impedance Spectroscopy (EIS) measurements) to images through processing techniques as Gramian Angular Field and the adoption of Convolutional Neural Network for the classification. Furthermore, a wider and more dense stimulus frequency range can be involved for EIS experiments, to deeply analyze the sensor response through Nyquist and Bode diagrams.

Conclusions

In this Ph.D. thesis the SENSIPLUS Embedded System has been developed and tested for health and safety applications. It is an integrated solution, designed and developed according to Internet of Things (IoT) and Edge Computing paradigms, endowed with sensing, elaboration and communication capabilities. The sensing part is based on the SENSIPLUS (SP) chip, which is a micro analytical tool capable to integrate heterogeneous sensors typologies, developed by Sensichips s.r.l. and the Department of Information Engineering of the University of Pisa. As for the elaboration, proper data analysis methods, based on statistical and Machine Learning techniques, have been involved and implemented for embedded low resources devices as Micro Controller Unit (MCU). As regards the transmission of the on–field elaborated results, the developed prototypes have been endowed with different wireless and wired communication technologies as Bluetooth Low Energy, Wi–Fi and USB. Laboratory and real scenarios experimental activities of the SENSIPLUS system have been performed for the three addressed applications obtaining positive and encouraging results which prove the system capability and validity for such tasks. As for the state of health monitoring of Activated Carbon Filters (ACF) and Biofilters, the SP chip metrological characterization has been performed and experimental activities on real filters samples have been carried out proving the system capability to monitor their state of health. As regards the two remaining applications, contaminants detection in air and in industrial waste water, different sensors technologies, measuring techniques and data analysis methods have been involved obtaining high sensitivity and recognition rates.

The advantages of the developed SENSIPLUS Embedded System are mainly based on the development and integration of different technologies (from sensing to elaboration) in miniaturized, low–power and low–cost devices. In such a way, continuous and on–line environmental monitoring is enabled, allowing to implement real–time routines which provide early warning proper notification. Although this kind of technology is still far from the reliability of laboratory instruments, it represents a huge step forward to guarantee people health and to prevent environmental disasters. As regards the main limitations, the sensing technologies are characterized by low–selectivity and fast deterioration. The former is due to the absence of highly specific sensors and requires the adoption of multiple sensing typologies and data analysis

techniques capable to find out correlation among acquired data. The latter is mainly due to the continuous interaction with environment and with high concentrations of contaminants. Sensors electrodes deterioration and acquired signals drift have been observed during experimental activities. The most worrying result of the drift phenomenon is represented by meaningful deviations of the sensors signals responses which highly confuse the classification phase. For this reasons, improvements in terms of the sensing technology and their stability over time represent an important part for future developments of such technology. Moreover,advances in terms of methods capable to on–line recognize and compensate the drift phenomenon have to be accomplished since it is a natural process and so hard to be completely eliminated.

An important aspect characterizing the addressed application, complicating the role of data analysis techniques, is represented by the environment itself. It is a very difficult background for classification algorithms because of its natural and continuously variable behavior. A novel concept, based on machine learning, which aims to face and tackle problems regarding applications where the background keeps moving over time is represented by the so called continual learning. According to this concept, a continuous upgrading of the classification model has to be accomplished without forgetting past information. In such a way, the algorithm is capable to understand the system deviations by learning and compensating them.

References

[1] Dean T. Jamison, Joel G. Breman, Anthony R. Measham, George Alleyne, Mariam Claeson, David B. Evans, Prabhat Jha, Anne Mills, and Philip Musgrove. *Disease Control Priorities in Developing Countries*. The International Bank for Reconstruction and Development / The World Bank, 2006.

[2] Andrew J. Whelton, LaKia McMillan, Matt Connell, Keven M. Kelley, Jeff P. Gill, Kevin D. White, Rahul Gupta, Rajarshi Dey, and Caroline Novy. Residential tap water contamination following the freedom industries chemical spill: Perceptions, water quality, and health impacts. *Environmental Science & Technology*, 49(2):813–823, jan 2015.

[3] Malek Bentayeb, Verene Wagner, Morgane Stempfelet, Marie Zins, Marcel Goldberg, Mathilde Pascal, Sophie Larrieu, Pascal Beaudeau, Sylvie Cassadou, Daniel Eilstein, Laurent Filleul, Alain Le Tertre, Sylvia Medina, Laurence Pascal, Helene Prouvost, Philippe Quénel, Abdelkrim Zeghnoun, and Agnes Lefranc. Association between long-term exposure to air pollution and mortality in france: A 25-year follow-up study. *Environment International*, 85:5 – 14, 2015.

[4] Lidia Morawska, Phong K. Thai, Xiaoting Liu, Akwasi Asumadu-Sakyi, Godwin Ayoko, Alena Bartonova, Andrea Bedini, Fahe Chai, Bryce Christensen, Matthew Dunbabin, Jian Gao, Gayle S.W. Hagler, Rohan Jayaratne, Prashant Kumar, Alexis K.H. Lau, Peter K.K. Louie, Mandana Mazaheri, Zhi Ning, Nunzio Motta, Ben Mullins, Md Mahmudur Rahman, Zoran Ristovski, Mahnaz Shafiei, Dian Tjondronegoro, Dane Westerdahl, and Ron Williams. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment International*, 116:286 – 299, 2018.

[5] Nuria Castell, Franck R. Dauge, Philipp Schneider, Matthias Vogt, Uri Lerner, Barak Fishbain, David Broday, and Alena Bartonova. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International*, 99:293 – 302, 2017.

[6] 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*, 14(8):2765–2772, Aug 2014.

[7] K. Gopavanitha and S. Nagaraju. A low cost system for real time water quality monitoring and controlling using iot. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pages 3227–3229, Aug 2017.

[8] Michael V. Storey, Bram van der Gaag, and Brendan P. Burns. Advances in on-line drinking water quality monitoring and early warning systems. *Water Research*, 45(2):741 – 747, 2011.

[9] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer Networks*, 54(15):2787 – 2805, 2010.

[10] A. Martín-Garín, J.A. Millán-García, A. Baïri, J. Millán-Medel, and J.M. Sala-Lizarraga. Environmental monitoring system based on an open source platform and the internet of things for a building energy retrofit. *Automation in Construction*, 87:201 – 214, 2018.

[11] W. Shi and S. Dustdar. The promise of edge computing. *Computer*, 49(5):78–81, May 2016.

[12] M. Satyanarayanan. The emergence of edge computing. *Computer*, 50(1):30–39, Jan 2017.

[13] Albert Greenberg, James Hamilton, David A. Maltz, and Parveen Patel. The cost of a cloud: Research problems in data center networks. *SIGCOMM Comput. Commun. Rev.*, 39(1):68–73, December 2008.

[14] Manuel Aleixandre and Michel Gerboles. Review of small commercial sensors for indicative monitoring of ambient gas. 2012.

[15] Emily G. Snyder, Timothy H. Watkins, Paul A. Solomon, Eben D. Thoma, Ronald W. Williams, Gayle S. W. Hagler, David Shelow, David A. Hindin, Vasu J. Kilaru, and Peter W. Preuss. The changing paradigm of air pollution monitoring. *Environmental Science & Technology*, 47(20):11369–11377, 2013. PMID: 23980922.

[16] Sensichips. learning microsensors. https://www.sensichips.com/, 2019.

[17] Zeba Idrees, Zhuo Zou, and Lirong Zheng. Edge computing based iot architecture for low cost air pollution monitoring systems: A comprehensive system analysis, design considerations and development. *Sensors*, 18(9), 2018.

[18] Y. Chen. Challenges and opportunities of internet of things. In *17th Asia and South Pacific Design Automation Conference*, pages 383–388, Jan 2012.

[19] L. D. Xu, W. He, and S. Li. Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, 10(4):2233–2243, Nov 2014.

[20] Radislav A. Potyrailo, Nandini Nagraj, Cheryl Surman, Hacene Boudries, Hanh Lai, Joseph M. Slocik, Nancy Kelley-Loughnane, and Rajesh R. Naik. Wireless sensors and sensor networks for homeland security applications. *TrAC Trends in Analytical Chemistry*, 40:133 – 145, 2012.

[21] H. M. Hashemian and W. C. Bean. State-of-the-art predictive maintenance techniques*. *IEEE Transactions on Instrumentation and Measurement*, 60(10):3480–3492, Oct 2011.

[22] J. Yan, Y. Meng, L. Lu, and L. Li. Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance. *IEEE Access*, 5:23484–23491, 2017.

[23] H. Kim, B. Han, C. G. Woo, Y. Kim, G. Lim, and W. G. Shin. Air cleaning performance of a novel electrostatic air purifier using an activated carbon fiber filter for passenger cars. *IEEE Transactions on Industry Applications*, 53(6):5867–5874, Nov 2017.

[24] North Atlantic Treaty Organization. CBRN DEFENCE ON OPERATIONS - VOLUME I. I(atp-3.8.1):1–464, Jan 2010.

[25] Marc A Deshusses. Biological waste air treatment in biofilters. *Current Opinion in Biotechnology*, 8(3):335 – 339, 1997.

[26] Joseph S. Devinny, Marc A. Deshusses, and Todd S. Webster. *Biofiltration for air pollution control*. Lewis Publishers, Boca Raton, Fla, 1999.

[27] 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)*, pages 1–5, Sep. 2017.

[28] 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*, 67(10):2504–2515, Oct 2018.

[29] 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 (Accepted for publication)*.

[30] A Mason, S Wylie, A Shaw, A I Al-Shamma'a, A Thomas, and H Keele. Determination of activated carbon residual life using a microwave cavity resonator. *Journal of Physics: Conference Series*, 307:012041, aug 2011.

[31] Measurement of the impedance change of impregnated activated carbon during exposure to so2 vapors at ambient temperatures. *Carbon*, 47(15):3566 – 3573, 2009.

[32] Glenn O. Rubel Gregory W. Peterson Thomas M. Ball Jingjing Bao, Victor Giurgiutiu. Active carbon filter health condition detection with piezoelectric wafer active sensors, 2011.

[33] G W Peterson, D Friday, and M Shrewsbury. *Residual Life Indicator for Physical Adsorption Capacity of NBC Filters. Part 1. Acetone Vapor Pulses and the Effect of Moisture Content on Retention Characteristics*. Defense Technical Information Center, 2008.

[34] Marcelis L. Muriel, Rajaram Narayanan, and Prabhakar R. Bandaru. Increasing energy storage in activated carbon based electrical double layer capacitors through plasma processing. *MRS Proceedings*, 1773:15–20, 2015.

[35] M. S. Webster, I. V. Timoshkin, S. J. Macgregor, and M. Mattey. Computer aided modelling of an interdigitated microelectrode array impedance biosensor for the detection of bacteria. *IEEE Transactions on Dielectrics and Electrical Insulation*, 16(5):1356–1363, October 2009.

[36] S. M. Radke and E. C. Alocilja. Design and fabrication of a microimpedance biosensor for bacterial detection. *IEEE Sensors Journal*, 4(4):434–440, Aug 2004.

[37] R. Dev Das, C. RoyChaudhuri, and S. Das. Macroporous silicon as an interdigitated electrode-less platform for impedance based bacteria detection. In *2010 International Conference on Systems in Medicine and Biology*, pages 121–125, Dec 2010.

[38] Wu-Chung Chan and Liang-Yuan Chang. Effects of temperature and inlet concentration on acetone biofiltration in a composite bead biofilter. *Journal of Polymers and the Environment*, Dec 2006.

[39] Pranas Baltrėnas, Alvydas Zagorskis, and Antonas Misevičius. Research into acetone removal from air by biofiltration using a biofilter with straight structure plates. *Biotechnology & Biotechnological Equipment*, 29(2):404–413, 2015. PMID: 26019659.

[40] ORD US EPA. Indoor Air Quality, November 2017.

[41] Nellie Brown. Indoor Air Quality. *Manuals and User Guides*, January 2019.

[42] G. Buonanno, L. Morawska, and L. Stabile. Particle emission factors during cooking activities. *Atmospheric Environment*, 43(20):3235 – 3242, 2009.

[43] Godwin A. Ayoko and Hao Wang. *Volatile Organic Compounds in Indoor Environments*, pages 69–107. Springer Berlin Heidelberg, Berlin, Heidelberg, 2018.

[44] Alexandre Caron, Nathalie Redon, Patrice Coddeville, and Benjamin Hanoune. Identification of indoor air quality events using a k-means clustering analysis of gas sensors data. *Sensors and Actuators B: Chemical*, 297:126709, 2019.

[45] 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)*, pages 470–475, June 2018.

[46] Frank Rock, Nicolae Barsan, and Udo Weimar. Electronic nose: Current status and future trends. *Chemical Reviews*, 108(2):705–725, 2008. PMID: 18205411.

[47] S. Marco and A. Gutierrez-Galvez. Signal and data processing for machine olfaction and chemical sensing: A review. *IEEE Sensors Journal*, 12(11):3189–3214, Nov 2012.

[48] E. Esposito, S. De Vito, M. Salvato, V. Bright, R.L. Jones, and O. Popoola. Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. *Sensors and Actuators B: Chemical*, 231:701 – 713, 2016.

[49] B. Mondal, M.S. Meetei, J. Das, C. Roy Chaudhuri, and H. Saha. Quantitative recognition of flammable and toxic gases with artificial neural network using metal oxide gas sensors in embedded platform. *Engineering Science and Technology, an International Journal*, 18(2):229 – 234, 2015.

[50] Saverio De Vito, Anna Castaldo, Fausta Loffredo, Ettore Massera, Tiziana Polichetti, Ivana Nasti, Paolo Vacca, Luigi Quercia, and Girolamo Di Francia. Gas concentration estimation in ternary mixtures with room temperature operating sensor array using tapped delay architectures. *Sensors and Actuators B: Chemical*, 124(2):309 – 316, 2007.

[51] G. Huyberechts, P. Szecówka, J. Roggen, and B.W. Licznerski. Simultaneous quantification of carbon monoxide and methane in humid air using a sensor array and an artificial neural network. *Sensors and Actuators B: Chemical*, 45(2):123 – 130, 1997.

[52] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 351–360, Republic and Canton of Geneva, Switzerland, 2017. International World Wide Web Conferences Steering Committee.

[53] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 2018.

[54] Pai Peng, Xiaojin Zhao, Xiaofang Pan, and Wenbin Ye. Gas classification using deep convolutional neural networks. *Sensors*, 18(1), 2018.

[55] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 255–258. MIT Press, Cambridge, MA, USA, 1998.

[56] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[57] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, March 1994.

[58] A. Graves, A. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649, May 2013.

[59] Xiaolei Ma, Zhimin Tao, Yinhai Wang, Haiyang Yu, and Yunpeng Wang. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54:187 – 197, 2015.

[60] IST. Humidity modules and sensors. https://www.ist-ag.com/en.

[61] Filiberto Ricciardella, Sten Vollebregt, Tiziana Polichetti, Mario Miscuglio, Brigida Alfano, Maria L. Miglietta, Ettore Massera, Girolamo Di Francia, and Pasqualina M. Sarro. Effects of graphene defects on gas sensing properties towards no2 detection. *Nanoscale*, 9:6085–6093, 2017.

[62] P.K. Goel. *Water Pollution: Causes, Effects and Control*. New Age International, 2006.

[63] Werner Brack et al. Towards the review of the european union water framework directive: Recommendations for more efficient assessment and management of chemical contamination in european surface water resources. *Science of The Total Environment*, 576:720 – 737, 2017.

[64] Kai Zhang, Huahong Shi, Jinping Peng, Yinghui Wang, Xiong Xiong, Chenxi Wu, and Paul K.S. Lam. Microplastic pollution in china's inland water systems: A review of findings, methods, characteristics, effects, and management. *Science of The Total Environment*, 630:1641 – 1653, 2018.

[65] R. Helmer, I. Hespanhol, WHO et al. *Water pollution control: a guide to the use of water quality management principles*. London: E & FN Spon, 1997.

[66] Shuming Liu, Han Che, Kate Smith, and Tian Chang. A real time method of contaminant classification using conventional water quality sensors. *Journal of Environmental Management*, 154:13 – 21, 2015.

[67] John Hall, Alan D. Zaffiro, Randall B. Marx, Paul C. Kefauver, E. Radha Krishnan, Roy C. Haught, and Jonathan G. Herrmann. On–line water quality parameters as indicators of distribution system contamination. *Journal - AWWA*, 99(1):66–77, 2007.

[68] Jinfeng Li and Shun Cao. A low-cost wireless water quality auto-monitoring system. *International Journal of Online Engineering (iJOE)*, 11(3):37–41, 2015.

[69] Wiebke Schmidt, David Raymond, David Parish, Ian GC Ashton, Peter I Miller, Carlos JA Campos, and Jamie D Shutler. Design and operation of a low-cost and compact autonomous buoy system for use in coastal aquaculture and water quality monitoring. *Aquacultural engineering*, 80:28–36, 2018.

[70] Environmental Protection Agency (EPA). Parameters of Water Quality - Interpretation and Standards.

[71] Naga Siva Gunda, Siddharth Gautam, and Sushanta Mitra. Artificial intelligence for water quality monitoring. *Meeting Abstracts*, MA2018-02(56):1997, 2018.

[72] M. Ferdinandi, M. Molinara, G. Cerro, L. Ferrigno, C. Marroco, 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)*, pages 186–191, June 2019.

[73] Serge Zhuiykov. Solid-state sensors monitoring parameters of water quality for the next generation of wireless sensor networks. *Sensors and Actuators B: Chemical*, 161(1):1 – 20, 2012.

[74] Cloé Desmet, Agnes Degiuli, Carlotta Ferrari, Francesco Saverio Romolo, Loïc Blum, and Christophe Marquette. Electrochemical sensor for explosives precursors' detection in water. *Challenges*, 8(1), 2017.

[75] 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*, April 2013.

[76] Jun-Chao Yan, Jun Ren, Lin-Lin Ren, Yi Yang, Si-Fan Yang, and Tian-Ling Ren. A novel structure design and fabrication method for low liquid consumption and high precision device of colorimeter in water quality detection. *Sensors and Actuators A: Physical*, 289:1 – 10, 2019.

[77] Joseph Wasswa, Natalie Mladenov, and William Pearce. Assessing the potential of fluorescence spectroscopy to monitor contaminants in source waters and water reuse systems. *Environ. Sci.: Water Res. Technol.*, 5:370–382, 2019.

[78] Ping Geng, Xinai Zhang, Weiwei Meng, Qingjiang Wang, Wen Zhang, Litong Jin, Zhen Feng, and Zirong Wu. Self-assembled monolayers-based immunosensor for detection of escherichia coli using electrochemical impedance spectroscopy. *Electrochimica Acta*, 53(14):4663 – 4668, 2008.

[79] Prerna Sonthalia, Elizabeth McGaw, Yoshiyuki Show, and Greg M. Swain. Metal ion analysis in contaminated water samples using anodic stripping voltammetry and a nanocrystalline diamond thin-film electrode. *Analytica Chimica Acta*, 522(1):35 – 44, 2004.

[80] Sonia Maria da Silva and Luiz Henrique Mazo. Differential pulse voltammetric determination of nitrite with gold ultramicroelectrode. *Electroanalysis*, 10(17):1200–1203, 1998.

[81] Ramón Martínez-Máñez, J. Camino, Luis Gil-Sánchez, Eduardo García-Breijo, J. Civera, and E. Morant. System for determining water quality with thick-film multisensor. volume 2005, pages 607 – 610, 03 2005.

[82] 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*, 10(6):128, Mar 2017.

[83] Tehila Asheri Arnon, Shai Ezra, and Barak Fishbain. Water characterization and early contamination detection in highly varying stochastic background water, based on machine learning methodology for processing real-time uv-spectrophotometry. *Water Research*, 155:333 – 342, 2019.

[84] Scott N. Dean, Lisa C. Shriver-Lake, David A. Stenger, Jeffrey S. Erickson, Joel P. Golden, and Scott A. Trammell. Machine learning techniques for chemical identification using cyclic square wave voltammetry. *Sensors*, 19(10), 2019.

[85] G. Durante, W. Becari, F. A. S. Lima, and H. E. M. Peres. Electrical impedance sensor for real-time detection of bovine milk adulteration. *IEEE Sensors Journal*, 16(4):861–865, Feb. 2016.

[86] Anna Nakonieczna, Bartosz Paszkowski, Andrzej Wilczek, Agnieszka Szypłowska, and Wojciech Skierucha. Electrical impedance measurements for detecting artificial chemical additives in liquid food products. *Food Control*, 66:116 – 129, 2016.

[87] Ian Jolliffe. *Principal Component Analysis*, pages 1094–1096. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[88] Petros Xanthopoulos, Panos M. Pardalos, and Theodore B. Trafalis. *Linear Discriminant Analysis*, pages 27–33. Springer New York, New York, NY, 2013.

[89] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.