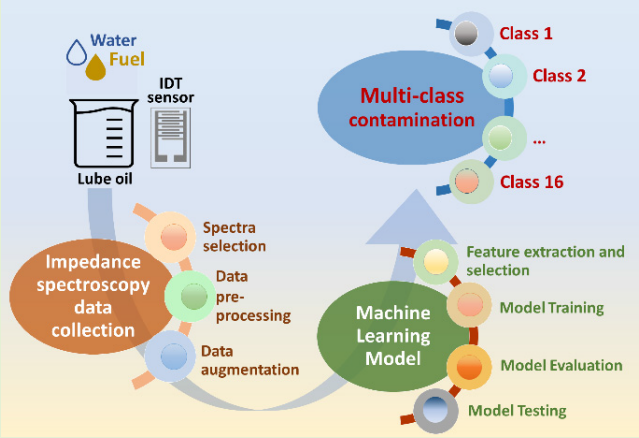
A Rapid Classification of Cross-Contaminations in Aviation Oil Using Impedance-Driven Supervised Machine Learning

Chiara De Pascali, Daniele Bellisario, Maria Assunta Signore, Elisa Sciurti, Antonio Vincenzo Radogna, *Member, IEEE*, and Luca Nunzio Francioso

***Abstract*—Real-time monitoring of lube oil plays a crucial role in ensuring optimal machinery performance, preventing failures, and facilitating timely maintenance strategies. The approach proposed in this work, based on impedance spectroscopy and supervised machine learning (ML), addresses this need by a novel solution to a multiclass classification problem of cross-contaminations in aviation lubricant. Impedance measurements were performed at room temperature by immersing a microfabricated sensor in 16 aged oil samples containing increasing concentrations of water and aviation fuel. Two datasets were constructed: the first based on impedance components spectra and the second based on dissipation factor spectra. A data pre-processing and augmentation method was proposed for generating synthetic examples from the measured data. Both datasets were independently used to train three supervised classifiers, whose performance was evaluated based on three different approaches of dataset split ratio and *k*-fold cross-validation. The 1-nearest neighbors (NN) classifier proved to be the most effective in reducing false positives (FPs), false negatives (FNs), and computational running time. The best results were obtained by employing a split ratio of 60:40 and threefold cross-validation scheme on the impedance components-based dataset, yielding an accuracy of 99.8%.**



***Index Terms*— Contamination, data augmentation, impedance spectroscopy, lubricant, machine learning, monitoring.**

I. Introduction

LUBRICANTS in motor vehicles and industrial machinery are mainly used to reduce friction and wear due to heating and mechanical contact between interfaces of moving components. In addition to lubrication, oil also performs other functions, such as cooling and cleaning. A lubricant is a balanced mixture of base oils and additives, whose combination impacts the overall performance and lifetime of the equipment. Base oil, generally mineral or synthetic, constitutes the highest percent of the lubricant composition, while the rest comprises chemical additives (detergents, dispersants, anti-wear agents, oxidation inhibitors, viscosity index improvers, etc.), with different formulations designed for specific applications [1].

During its operation, lubricant tends to degrade its chemo-physical properties and its lubricity, due to several problems, including thermal stress, external contamination, oxidation, and depletion of additives. Contamination is a primary cause of oil deterioration, which can cause corrosion and excessive wear of moving parts, culminating in failure [2]. Fuel and water are among the most dangerous contaminants; the former can derive by blow-by processes or incomplete burning of exhaust gases [3], [4], the latter from condensation in combustion processes, from seal losses of defective head gaskets, or cracks in the block [5], [6].

Received 24 July 2024; revised 5 September 2024; accepted 16 September 2024. Date of publication 30 September 2024; date of current version 14 November 2024. This work was supported in part by the Italian Ministry for Universities and Research (MUR) by the project SMEA: Metodologie diagnostiche e prognostiche e sviluppo di sensori per il monitoraggio di integrità funzionale applicato al settore aeronautico e dei trasporti under Grant PON03PE_00067_5. The associate editor coordinating the review of this article and approving it for publication was Dr. Xiaojin Zhao. *(Corresponding author: Daniele Bellisario.)*

Chiara De Pascali, Daniele Bellisario, Maria Assunta Signore, Elisa Sciurti, and Luca Nunzio Francioso are with the Institute for Microelectronics and Microsystems (IMM), National Research Council (CNR), 00185 Rome, Italy (e-mail: chiara.depascali@cnr.it; daniele.bellisario@imm.cnr.it; mariaassunta.signore@cnr.it; elisa.sciurti@imm.cnr.it; lucanunzio.francioso@cnr.it).

Antonio Vincenzo Radogna is with the Department of Experimental Medicine, University of Salento, 73100 Lecce, Italy (e-mail: antonio.radogna@unisalento.it).

Digital Object Identifier 10.1109/JSEN.2024.3463878

Continuous and reliable monitoring of oil condition represents a desirable and effective control strategy to provide an early warning of impending problems. The most common methods used to diagnose oil condition are offline and require time-consuming and often expensive procedures, both for collecting and handling samples for accredited laboratory analysis, and for adopting safety precautions for use of solvents and chemicals (crackle-test, calcium hydrate test, distillation, Karl Fischer titration, flash point, gas chromatography, and Fourier transform infrared spectroscopy) [7], [8], [9], [10], [11].

Great interest of industry and scientific community is currently directed toward the development of online and inline approaches to monitor in a continuous manner the lubricant quality and machine health condition, while overcoming the drawbacks of offline methods. Various commercial solutions for direct and indirect measurements of oil condition are available, including Hydac AS2000 by Aqua Sensor, WSPS 05 by EATON, WS10 Series Water Sensor by Pall Corporation, WIO200 sensor by PAJ Group, OQSx-G2 by Tan Delta, and Trident DM4500-DM4600 by Poseidon Systems. However, these systems typically lack the capability to identify the type and concentration of contaminating fluid, providing only a general quality score.

The dielectric permittivity is an important quality index of oil that can be easily measured online. Several products present in the market allow to measure the dielectric constant and the dielectric dissipation factor over a very narrow frequency range, in compliance with operative procedures indicated by international standards, including IEC and ASTM [12], [13]. However, several works published in the literature show that broadening the spectrum of frequencies under investigation can be more effective for analysis purposes [14], [15], [16], [17].

Impedance spectroscopy (IS) is non-destructive characterization technique able to evaluate the electrical and dielectric properties of solid materials and liquids at various frequencies. It is effective in the in-line monitoring of lubricants because it can measure several properties related to the oil condition [18]. In recent years, even machine learning (ML) has had a significant impact on condition monitoring and predictive maintenance tasks, for which proactive and data-driven strategies can be effective to foresee potential failures. An interesting survey on condition monitoring using ML methods reviewed recent developments in this field [19].

The success of a predictive maintenance strategy using an IS data-driven ML model depends not only on the performance of the sensing technology employed to monitor physical and environmental conditions, but also on the data processing and ML methods employed. The performance of an ML algorithm is frequently tied to the quality and quantity of data used to train the model. Even though impedance measurements offer access to a large amount of information and a variety of frequency-dependent data patterns [20], employing highly correlated variables to train an ML algorithm may lead to problems of redundancy and multicollinearity that can complicate the classification task [21]. On the other hand, collecting a great amount of data for the model training can be challenging in certain real-world scenarios, such as the oil condition monitoring in (near) real-time. In this context, it is reasonable to expect that an optimized timing of the sensor's readings can be crucial to promptly executing monitoring actions devoted to detect potentially dangerous changes in the oil condition, such as those deriving from contamination.

Achieving excellent performance in training models using a limited number of labeled samples is the aim of few-shot learning. This ML technique attempts to replicate the cognitive process of humans, who can efficiently learn a new object by exploiting the knowledge accumulated through a few examples. Nowadays, few-shot learning is used to solve different tasks in various fields, including fault diagnosis and structural health monitoring [22], [23], medical image recognition [24], defective products detection [25], and so on.

Inspired by these examples, this article presents a novel and effective approach for classifying impedance measurements related to cross-contamination of water and fuel in aged oil. The combined use of ML and IS has been already proposed in the literature by other authors for regression purposes, specifically to identify the equivalent circuit and the circuit parameters that fit the impedance data. In such an approach, the traditional human-assisted data analysis is replaced with artificial intelligence-based methods [26], [27], [28], [29]. In this work, IS and ML were used for classification purpose. Few steady-state impedance measurements were gathered by a proprietary capacitive sensor immersed in the oil sample under test. These measurements were then pre-processed and augmented to create two datasets for training three ML models independently. The first dataset employed the spectral distribution of the real and imaginary components of impedance, and the second one utilized the spectral variation of their ratio, known as dissipation factor or $\tan\delta$. The most relevant features extracted from the two datasets were selected to train the models. This strategy allowed to reduce redundant information, simplifying the model and reducing computational cost and training time. In this work, a classification-based approach was adopted for the following reasons: 1) regression models are more complex and require larger amounts of labeled data; 2) regression models tend to be more sensitive to noise, making data augmentation more challenging to apply; and 3) a classification-based approach allowed for a more immediate and robust assessment of the effectiveness of the proposed data augmentation method.

The comparison of the results obtained using the two datasets (impedance components- and $\tan\delta$-based) aimed to determine which properties of the oil (electrical, dielectric, or both) are most significantly affected by the contamination, and which contributions (resistive, reactive, or both, in terms of dissipation factor) best emphasize such an effect.

The data augmentation method proposed in this work prevented overfitting and improved model accuracy, addressing the challenge of having few but significant measured data. This advantage can be particularly beneficial in applications where gathering large datasets to predict oil condition is time-consuming and costly, posing a significant limitation, especially for onsite monitoring applications.

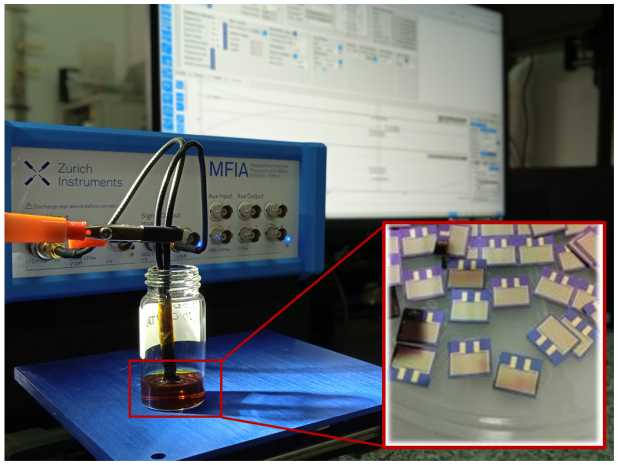


Fig. 1. Photograph of the microfabricated sensors, and the experimental setup used for the IS measurements.

Three schemes of dataset split ratio and K-fold cross-validation were evaluated for each of three investigated models. The comparative analysis of performance between models and datasets demonstrated that the three evaluated algorithms can effectively discriminate different oil conditions even with a small dataset and a limited number of features. The comparison between the two investigated datasets suggests that the spectral content of real and imaginary impedance components can be more informative and useful than $\tan\delta$ for monitoring oil condition. This result stimulates greater interest in the development of impedance-based monitoring systems for industrial applications. The successful identification of 16 aged oil samples cross-contaminated with water and fuel was demonstrated.

II. Sensor Design and Fabrication

Fig. 1 shows a photograph of the microfabricated sensors, and the experimental setup used for the impedance spectroscopy measurements, with the sensor immersed in the aged oil sample under test.

The sensor was designed by using finite element analysis to maximize the sensitivity to changes in the dielectric constant of oil due to oxidation and contamination. The simulation results were discussed in a previous work, showing how the sensitivity varies as a function of the finger width and gap size in the range from 20 to 80 μm, for increasing values of dielectric constant [30]. The sensor has an active area of 24.3 mm^2 and comprises a pair of interdigitated electrodes of 46 fingers per electrode, with width and gap size of 20 μm and length of 6.56 mm. It was realized on 500-μm-thick (100)-silicon substrate, with 0.5 μm of thermal oxide, by using 365-nm UV optical lithography, RF sputtering deposition, and lift-off processes. The electrodes comprise 20-nm-thick titanium adhesion layer and 200-nm-thick gold layer.

III. Impedance Spectroscopy: Basic Concepts

Interdigitated electrodes are attractive transduction platforms for capacitive sensor devices, whose impedance can provide an insight into the dielectric properties of the medium in which the sensor operates. Lubricating oil is a non-ideal dielectric that exhibits resistive and reactive losses when exposed to an alternating electrical field. When immersed in a lossy dielectric medium and excited by an alternating voltage at frequency ω, an interdigitated electrodes sensor has an impedance that can be expressed as the sum of resistive (Z_R) and reactive (Z_I) contributions

$$Z(\omega) = Z_R + jZ_I \tag{1}$$

where Z_R and Z_I are the real and imaginary components of the impedance. The measured impedance is the result of bulk and interfacial processes related to the interaction of the lubricant with the electric field generated between the electrodes. The main mechanisms, such as bulk relaxations, adsorption on the electrode interface, mass transport from the bulk solution, and charge transfer at the electrode, are typically represented by equivalent circuits constituted by simple parallel/series combinations of resistances, capacitors, and constant phase elements [31]. The major effect of lubricant contamination with fuel is on the reactive impedance at low frequencies associated with the electrode–solution interface [15]. The contamination with water affects both the resistive and reactive components of impedance, with a marked increase of electrical conductivity at low frequencies and permittivity at medium and high frequencies [32].

Generally, the resistive impedance is associated with the lubricant capacity to dissipate electrical energy as heat; the reactive impedance is associated with the lubricant ability to store electrical energy. The contamination modifies the capacity of energy dissipation and storage of the lubricant, and such an effect is overall evaluated by the dissipation factor. Tanδ represents the ratio between the energy dissipated and stored in the lubricant. It is a frequency-dependent parameter that can be calculated by the complex impedance, as the ratio of the resistive (real) and reactive (imaginary) impedance components

$$\tan\delta = Z_R/Z_I. \tag{2}$$

IV. Materials and Methods

The first part of this section presents the main properties of the aviation lubricant and contaminants used and describes the experimental procedures for thermal aging and contamination of the lube oil. In the second part, the proposed supervised ML-based classification model is explained in detail, from the impedance data acquisition to the model implementation, including data pre-processing, augmentation, features extraction, and selection.

A. Main Properties of Lubricant and Contaminants

The aviation lubricant used in this work is AeroShell Turbine Oil 500, produced by Shell for aviation; the aviation fuel is AeroShell Calibrating Fluid 2. Their main properties are reported in Table I [33], [34]. It is expected that the dielectric constant and the dissipation factor of the oil change in the presence of contaminants. These are two important parameters that several international standards, including IEC and ASTM, recommend to consider as indicators of oil quality [12], [13].

TABLE I
MAIN PROPERTIES OF OIL AND FUEL USED IN THIS WORK

Property	AeroShell Turbine Oil 500	AeroShell Calibrating fluid 2
Density (kg/m^3)	1005 @15°C	770 @15 °C
Viscosity, kinematic	25,26 mm^2/s @40°C	0,95 mm^2/s @40°C
	5,17 mm^2/s @100°C	1,15 mm^2/s @25°C
		1.46 mm^2/s @10°C
Flash point	256°C	43°C
Auto-ignition temp.	> 320°C	> 200°C

The greater the difference in dielectric constant between oil and the contaminant, the easier it is to detect contamination by observing changes in the dielectric properties of the oil. However, a multidimensional data analysis, possibly based on ML techniques, may be necessary to distinguish between different types and amounts of contaminants. Water dilution in oil causes an increase in the dielectric constant of the mixture, since water has a much higher dielectric constant than oil (78.8 at 25 °C). Most lubricants containing non-polar hydrocarbons have dielectric constant in the range between 2 and 4 [35], [36], [37], while for most fuels it is in the range 1.7–2.8 [38], [39], [40]. The dielectric similarity of oil and fuel makes difficult to detect small amounts of fuel in oil using capacitive sensors. Such difficulty may be more prominent in aged oil, since some of the effects of the fuel dilution in oil can be blended with others deriving from aging processes of the oil (e.g., additive depletion, oxidation, changes in viscosity). These considerations motivated the interest to face the detection and classification of fuel and water contamination in aged oil by using capacitive sensors.

B. Thermal Aging and Contamination of Oil

The cooling temperature is 20 °C (twenty Celsius degree) of AeroShell Turbine Oil 500 was done in laboratory, by subjecting the fresh oil sample to repeated cycles of heating (up to a maximum temperature of 140 °C for 14 h) and cooling in air (up to 20 °C for 10 h), for a duration of 200 h at high temperature.

The levels of contamination investigated in this work fall within a range of values considered alarming or dangerous. The impact of lubricant dilution with water and fuel varies significantly depending on the application. To the best of our knowledge, there are no specific thresholds related to the maximum admissible amount of fuel in lubricating turbine oils. A fuel content exceeding 3%–5% is considered excessive in various scenarios [3], [41]. As regards the content of water in lubricating oils, ASTM D4378 sets 1000 ppm or 0.1% as a warning level, while some gas and steam turbines original equipment manufacturer (OEMs), such as solar[1] turbines, consider 2000 ppm of water in oil as the maximum limit before requiring an oil change or reconditioning. Nevertheless, a water content of 3% is still tolerable in wind turbines applications [42].

Based on the above, in this work, the aged oil was contaminated with distilled water and fuel, with weight concentrations comprised between 0.2% and 2% for water, and 1% and 5% for fuel, with the intent to detect the most critical contamination thresholds (low, moderate, and high) within the range of interest for each contaminant. The mass of oil was fixed equal to 5 g for all samples. An analytical balance (ABT 120-5DNM, manufactured by KERN and SOHN GmbH) was used to weight the aliquots of each substance, with an uncertainty of ±0.0001 g. Fuel was gently mixed with oil by a magnetic stirrer for 2 min at low speed. Water was emulsified with oil by direct probe ultrasound sonication (SONICATOR 700 W by Fisher Scientific) at 70 W power and 12 consecutive ON/OFF cycles with a ratio 5:30 for a duration of 60 s, for minimizing the excessive heating of the emulsions. Table II lists the 16 classes of cross-contamination prepared and investigated in this work.

[1]Registered trademark.

TABLE II
SAMPLES OF OIL CONTAMINATED WITH WATER AND FUEL

Class of contamination	Water-in-oil (wt.%)	Fluid-in-oil (wt.%)	Dispersion method
1	0	0	None
2	0	5	Stirring
3	0	3	Stirring
4	0	1	Stirring
5	2	5	Sonication + Stirring
6	2	3	Sonication + Stirring
7	2	1	Sonication + Stirring
8	2	0	Sonication
9	0.5	5	Sonication + Stirring
10	0.5	3	Sonication + Stirring
11	0.5	1	Sonication + Stirring
12	0.5	0	Sonication
13	0.2	5	Sonication + Stirring
14	0.2	3	Sonication + Stirring
15	0.2	1	Sonication + Stirring
16	0.2	0	Sonication

C. Supervised ML-Based Classification Model

Supervised classification is a learning technique that uses statistical methods to categorize a set of labeled data in distinct classes, by predictions based on learned patterns and features. Typically, a supervised classification approach involves the following steps: 1) data acquisition and pre-processing, followed by features extraction and human-assisted data labeling, for assigning the correct target to each feature; 2) dataset splitting into training set, validation set, and test set, based on the training strategy; 3) model training on the training set, by which the model understands the patterns and relationships within data and learns to make predictions on new inputs; 4) model validation, in order to identify and prevent important issues, such as overfitting or underfitting; 5) evaluation of the model's performance by several metrics, often averaged over multiple cross-validation rounds; and 6) model testing on the test set (unseen data) to evaluate the predictive performance of prediction and generalization to new data. Fig. 2 depicts a schematic representation of the impedance-driven ML model developed in MATLAB and proposed in this work. Each step of the algorithm is discussed in detail as follows.

1) Impedance Data-Acquisition: For identifying contamination rapidly, impedance measurements were recorded in a

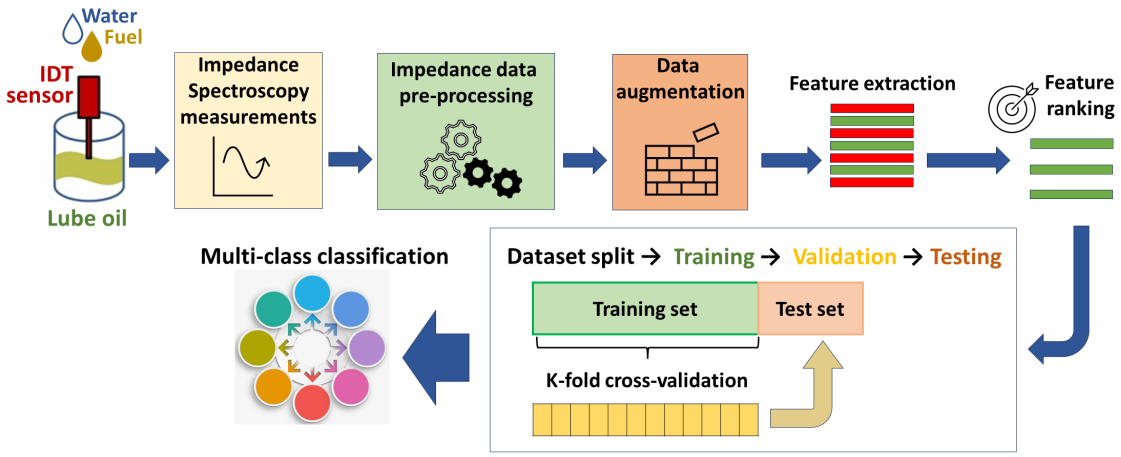


Fig. 2. Schematic representation of the impedance-driven ML algorithm developed to classify fuel and water cross-contaminations in aged oil (color in print).

short time after the contamination of oil. Impedance spectra were measured at room temperature using MFIA impedance analyzer by Zurich Instruments. The sensor was immersed in a beaker containing the oil sample under test and excited by an AC voltage of amplitude 2 V and frequency in the range between 5 Hz and 1 MHz (see Fig. 1). Impedance spectra were acquired after immersing the sensor in the oil sample under test, at intervals $t = 1$, 2, 5, 10, and 15 min to assess the measurement repeatability and obtain steady-state data readings for classification purposes. For aged oil (class 1) and fuel-in-oil samples (classes 2-3-4), four steady-state spectra were recorded at $t = 2$, 5, 10, and 15 min. In water-in-oil samples, water affected the time needed for the sensor to achieve steady-state conditions, probably due to the heterogeneous structure of the emulsion. Therefore, three steady-state spectra were recorded at $t = 5$, 10, and 15 min for the remaining classes. Only steady-state measurements were utilized to populate the training dataset (four spectra for each of the classes 1-2-3-4, and three spectra for each of the remaining classes). The mean coefficient of variation (percentage ratio of standard deviation to mean) of steady-state impedance spectra in the investigated frequency range was lower than 4%.

2) Impedance Data Pre-Processing and Augmentation: When the amount of labeled data is limited (e.g., due to expensive or time-limited experimental conditions), data augmentation becomes an efficient method for increasing the dataset size. A possible approach consists in generating synthetic data by utilizing the statistical characteristics of the real data [43], [44]. This allows to enhance the existing dataset, improves the model generalization to unseen data, and reduces the risk of overfitting. In this work, a statistical-based method was proposed for the generation of synthetic spectra to be added to the measured ones.

a) Data pre-processing: The steady-state impedance data were interpolated on evenly spaced frequency points and normalized on a common range [0, 1] to make them scale-invariant, while preserving the relationships between data points. Then, two distinct training datasets were constructed from the impedance data: the first dataset included, for each class, a set of measured and synthetic spectra of the real and imaginary parts of the impedance (Z_R, Z_I) defined as in (1). The second dataset included, for each class, a set of measured and synthetic spectra of $\tan\delta$ defined as in (2).

All the normalized data were truncated to the more spectrally significant frequency range, between 5 and 10^3 Hz. Fig. 3 depicts the normalized mean impedance components and $\tan\delta$ for some classes of contamination. The graphs illustrate the effect of increasing fuel content in oil samples containing 2% water on the spectral variation of the impedance components [Fig. 3(a)] and dissipation factor [Fig. 3(b)].

b) Data augmentation: The proposed method of data augmentation not only prevented overfitting and improved model accuracy, but also addressed the challenge of having few but significant measured data needed for model training. This advantage can be particularly beneficial in applications where gathering large datasets to predict oil condition is time-consuming and costly, posing a significant limitation, especially for onsite monitoring applications where timely detection of hazardous conditions is crucial. For the impedance components-based dataset, the following procedure was adopted for each class of contamination. The mean values of the two components, $\widehat{Z_R}(\omega)$ and $\widehat{Z_I}(\omega)$, and their standard deviations, $\sigma_R(\omega)$ and $\sigma_I(\omega)$, calculated in each frequency point, were used to generate n synthetic examples per class, as follows:

$$\left[Z_R^{S_j}(\omega_i), Z_I^{S_j}(\omega_i)\right] = \left[c_i^j \cdot \sigma_R(\omega_i) \cdot \widehat{Z_R}(\omega_i), c_i^j \cdot \sigma_I(\omega_i) \cdot \widehat{Z_I}(\omega_i)\right]. \quad (3)$$

The term in the left side of the previous equation represents the couple of impedance components at the i-frequency point and for the j-synthetic example S_j. The term c_i^j is a random value comprised in the range $[-1, 1]$ that was used for perturbing the i-couple of components of the mean impedance $[\widehat{Z_R}(\omega_i), \widehat{Z_I}(\omega_i)]$, within the range defined by the respective standard deviation.

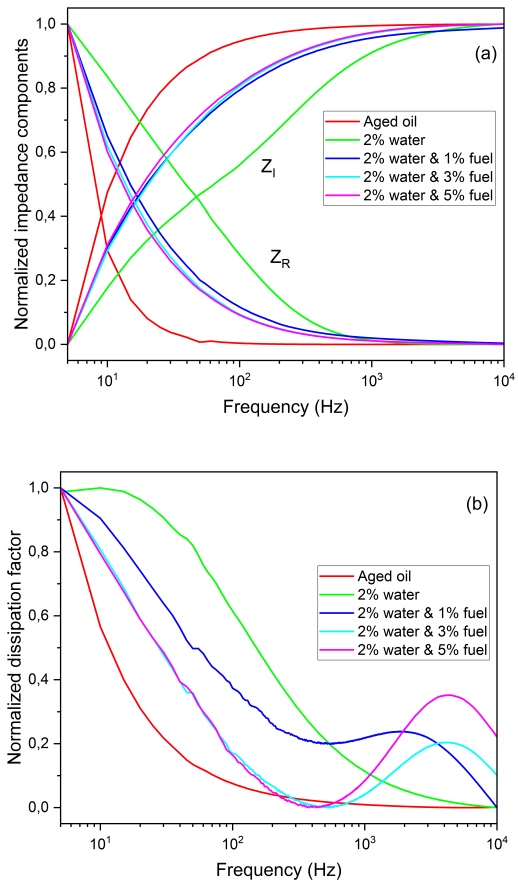


Fig. 3. Normalized spectral variation of the mean impedance components (a) and tanδ (b) of oil samples contaminated with 2% water and increasing fuel content (color in print).

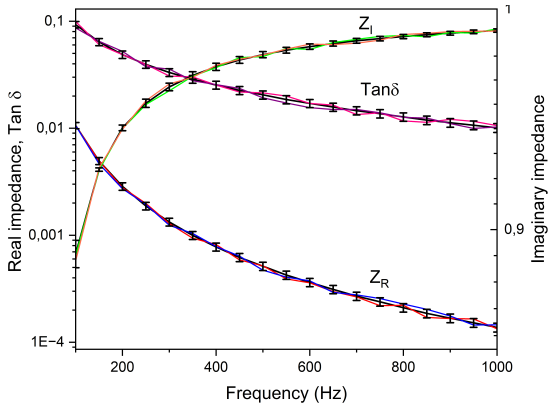


Fig. 4. Mean real data (black lines) and synthetic examples (colored lines) generated using (3) and (4) of aged oil contaminated with 0.2% water and 1% fuel (color in print).

TABLE III
STATISTICAL SIMILARITY OF REAL AND SYNTHETIC DATA, EVALUATED BY THE FRAC SCORE (%)

Impedance components-based dataset				
	No water	**0.2% water**	**0.5% water**	**2% water**
No fuel	99.99	99.37	94.12	94.40
1% fuel	99.98	99.53	94.42	95.50
3% fuel	99.98	98.16	99..07	98.18
5% fuel	99.99	99.09	99.74	98.36
Tanδ-based dataset				
	No water	**0.2% water**	**0.5% water**	**2% water**
No fuel	99.74	99.98	99.97	93.43
1% fuel	99.68	99.95	99.23	99.99
3% fuel	99.84	98.80	99.82	99.98
5% fuel	99.85	99.78	99.33	99.99

For the tan δ-based dataset, the synthetic examples were generated using the same approach, as follows:

$$\left[\tan\delta^{S_j}(\omega_i)\right] = \left[c_i^j \cdot \sigma_{\tan\delta}(\omega_i) \cdot \widehat{\tan\delta}(\omega_i)\right]. \tag{4}$$

The term at the left side in (4) represents the value of tan δ at the i-frequency point and for the j-synthetic example S_j. Again, for each j-synthetic example, in each i-frequency point, the random value c_i^j is in the range $[-1, 1]$ and modulates the mean dissipation factor $\widehat{\tan\delta}$, within the range defined by the respective standard deviation $\sigma_{\tan\delta}$.

The augmented dataset included 160 labeled instances, with 10 instances per class, 52 steady-state real measurements, and 112 synthetic examples. Each instance is represented by a matrix of 2000 rows, with each row containing the i-frequency point and the respective real and imaginary impedance components (for the first dataset), or the i-frequency point and the respective tan δ value (for the second dataset). Fig. 4 propones a comparison between normalized mean real data (black lines) and synthetic examples (colored lines) generated using (3) and (4), for aged oil contaminated with minimum cross-contamination (class 15). For better visualization, the frequency range was limited between 100 Hz and 1 kHz.

The statistical similarity of the real and synthetic data was evaluated using the Frequency Response Assurance Criterion (FRAC), that is, a representative metric in the frequency domain defined as [45]

$$\text{FRAC} = \frac{\left|\sum_{i=1}^{N} \boldsymbol{\chi}_{\text{augm}}(\omega_i)^H \boldsymbol{\chi}_{\text{meas}}(\omega_i)\right|^2}{\left[\sum_{i=1}^{N} \boldsymbol{\chi}_{\text{augm}}(\omega_i)^H \boldsymbol{\chi}_{\text{augm}}(\omega_i)\right]\left[\sum_{i=1}^{N} \boldsymbol{\chi}_{\text{meas}}(\omega_i)^H \boldsymbol{\chi}_{\text{meas}}(\omega_i)\right]} \tag{5}$$

where $\boldsymbol{\chi}_{\text{augn}}$ and $\boldsymbol{\chi}_{\text{meas}}$ are the synthetic and measured frequency functions, respectively. The superscript H refers to the Hermitian, that is, the transpose of complex conjugate. Table III reports the statistical similarity score calculated by (5) for each class of contamination in the two datasets. The high score indicates that the synthetic data preserve the essential patterns and relationships of the real ones.

3) Feature Extraction and Selection: The Python library tsfresh [46] was used to identify and extract relevant features from the dataset. Tsfresh is typically used for features extraction from evenly spaced time series data; however, under some conditions, it can be also used with frequency series data [29]. In particular, the latter has to be stationary, and have an adequate sampling rate, with evenly spaced data points. In this work, both these conditions were satisfied, as documented in the two previous subparagraphs dedicated to the acquisition and pre-processing of impedance data.

TABLE IV
BRIEF DESCRIPTION OF THE TOP-*N* FEATURES RANKED BY CHI-SQUARE SCORE FOR THE TWO INVESTIGATED DATASETS [50]

Feature name and description	Chi-square score
Impedance components-based dataset	
• Z_I _agg_linear_trend_attr_"slope"_chunk_len_10_f_agg_"var" It quantifies the linear trend of the imaginary part of impedance, as variance of the slopes of the linear least-squares regression function calculated over chunks of 10 frequency points.	409
• Z_I _agg_linear_trend_attr_"slope"_chunk_len_50_f_agg_"max" It quantifies the linear trend of the imaginary part of impedance as maximum slope of the linear least-squares regression function between those calculated over chunks of 50 frequency points.	392
• Z_I _ratio_beyond_r_sigma_r_3 It quantifies the proportion of data points in the spectrum of the imaginary part of impedance that fall beyond three standard deviations away from the mean.	381
Tanδ-based dataset	
• Tanδ_change_quantiles_f_agg_"mean"_isabs_True_qh_0.2_ql_0.0 It evaluates the spectral distribution of tanδ, by calculating the mean of the absolute changes of tanδ failing beyond the 20th percentile.	266
• Tanδ_energy_ratio_by_chunks_num_segments_10_segment_focus_8 It splits the spectrum of tanδ into 10 segments, then it computes the ratio of the sum of the squared values of tanδ belonging in the 8th segment to the overall sum calculated across all the segments.	265
• Tanδ_fft_coefficient_attr_"real"_coeff_10 It extracts the amplitude of the 11th frequency component of the Fourier Transform of the tanδ spectrum, as if it was a time series.	263
• Tanδ_cwt_coefficients_coeff_4_w_2_widths_(2, 5, 10, 20) It extracts the 5th coefficient of the continuous wavelet transform of tanδ calculated with a sampling period equal to 2 and a shift on four different scales.	261
• Tanδ_energy_ratio_by_chunks_num_segments_10_segment_focus_6 It splits the spectrum of tanδ into 10 segments, then it computes the ratio of the sum of the squared values of tanδ belonging in the 6th segment to the overall sum calculated across all the segments.	259

Feature selection was performed based on importance to reduce the number of extracted features used in training, avoiding overfitting, and enhancing model accuracy and generalizability. To this aim, the extracted features were normalized by Z-score to achieve a normal distribution (a mean of 0 and a standard deviation of 1) [47]. Then, feature ranking was performed by assigning a weight to each feature based on its inherent predictive capability, aside from the chosen classification model. The MATLAB function "*fscchi2*" was used to rank features by increasing importance using Chi-square test. This method neglects the interactions between features and ranks them according to their correlation strength based on the following statistical value [48], [49]:

$$\chi^2 = \sum_{i=1}^{n} (E_i - O_i)^2 / E_i \tag{6}$$

where E_i represents the expectation value of the i-feature appearance in a certain class, and O_i is the actual appearance value of the i-feature in a certain class. Continuous predictors were binned into 10 categories, and the Chi-square statistic was computed for each feature using (6). The resulting score helps to evaluate the statistical significance of these associations: a higher Chi-square value indicates a strong correlation with the target variable and a potentially great predictive capability.

Table IV reports a brief description of the top-N features selected by decreasing Chi-square score, for the two investigated datasets. Regarding the impedance components-based dataset, it is worth noting that the three most relevant features refer to the imaginary part of impedance. The latter represents a capacitive reactance, as described in (1), since the sensor forms a capacitor with the oil acting as dielectric. This suggests that the contamination affects mainly the dielectric properties of the lubricant, rather than the electrical ones. Among the top-5 features selected from the second dataset, the most important one evaluates the mean variation of $\tan\delta$ beyond the 20th percentile, while the remaining features calculate the energy ratio of the segmented series or extract frequency components from data transformations. By comparing the results in Table IV, it can be noticed that the highest Chi-square scores were obtained for the impedance components-based dataset.

4) Model Implementation: The performance of three supervised classifiers was evaluated as a function of the training dataset, using different dataset split ratios and K-fold cross-validation approaches: k-nearest neighbors (k-NN), support vector machine (SVM), and ensemble bagged tree (EBT). The choice was made by considering previous results obtained using these models in combination with impedance spectroscopy measurements for classification purposes [51], [52], [53], [54], [55].

TABLE V
MAIN HYPERPARAMETERS OF THE ML MODELS USED IN THIS WORK

Classifier	Hyperparameters	Value
SVM	Kernel function	Gaussian
	Kernel scale	2.2
	Box constraint level	1
	Multiclass method	One Vs One
	Standardize data	True
KNN	Number of neighbors	1
	Distance metric	Euclidean
	Distance weight	Equal
	Standardize data	True
EBT	Ensemble method	Bag
	Learner type	Decision tree
	Maximum number of splits	127
	Number of learners	30

The k-NN algorithm can be used for both classification and regression of small or moderately sized dataset. The predictions are based on the distance that each data point has from the k-nearest data points in the set. In this work, the number of k-NN was set to 1. The SVM algorithm can be used for both classification and regression of not linearly separable data in a multidimensional space. It works for finding the best hyperplanes that separate data points into different classes while maximizing the distance between them. The EBT reaches a final decision by combining the predictions obtained by multiple decision trees trained on randomly selected subsets of the original dataset (bootstrap samples). This algorithm reduces the overfitting problem and reaches better performance than individual decision trees, exploiting the diversity existing among the trees in the ensemble. Table V summarizes the main hyperparameters of the ML models used in this work.

Two different ratios were considered to split the dataset in training, validation, and test sets. K-fold cross-validation was used to assess the generalization performance of the model with respect to unseen data. The training set was divided into K-folds, $K-1$ folds were used for training, while the remaining fold was used for validation. This process was repeated K times, until every fold was used for the validation once. The performance of each model was calculated as the mean of K iterations. Table VI summarizes the different cases of dataset splitting and K-fold cross-validation investigated in this work.

TABLE VI
SCHEMES USED IN THIS WORK FOR DATASET SPLIT RATIO AND K-CROSS-VALIDATION

Split ratio	Cross validation (K)	Dataset partition	No. of spectra per set
60:40	3	Training set	96
		Training subset	64
		Cross-validation subset	32
		Test set	64
80:20	4	Training set	128
		Training subset	96
		Cross-validation subset	32
		Test set	32
	8	Training set	128
		Training subset	112
		Cross-validation subset	16
		Test set	32

V. RESULTS AND DISCUSSION

A. Contamination Sensing Performance

The sensing performance of the proposed capacitive sensor was evaluated by analyzing the steady-state spectra of the dissipation factor and calculating the mean value of $\tan\delta$ at 60 Hz, as suggested by the standard test method for dissipation factor and relative permittivity of insulating liquids [13]. The limit of detection (LOD) of fuel and water in aged oil was derived by using the equation in [56]

$$\mathrm{LOD} = \left| \frac{3\sigma_{\mathrm{oil}} c_{\mathrm{min}}}{\widehat{\tan\delta}_{\mathrm{min}}} \right|. \tag{7}$$

σ_{oil} is the standard deviation of the mean $\tan\delta$ measured for the blank solution, intended as the aged oil (class 1); c_{min} is the smallest measured analyte concentration; and $\widehat{\tan\delta}_{\mathrm{min}}$ is the mean $\tan\delta$ measured at c_{min}. A LOD of 0.16% was obtained for fuel with $\sigma_{\mathrm{oil}} = 0.00136$, $c_{\mathrm{min}} = 0.01$, and $\widehat{\tan\delta}_{\mathrm{min}} = 0.0242$. A LOD of 0.003% was obtained for water, with $c_{\mathrm{min}} = 0.002$ and $\widehat{\tan\delta}_{\mathrm{min}} = 0.238$.

B. Performance of the Supervised Models

The performance of the supervised algorithms was compared using four metrics derived from the confusion matrices for cross-validation and test results: accuracy (percentage of the total and correctly classified instances), sensitivity or recall (percentage of the correctly classified positive instances), precision (proportion of the correctly classified positive instances to the total number of positive predicted observations), and $F1$-score. The latter is the harmonic mean between precision and recall, and takes into account both the false positive (FP) and false negative (FN) observations. These metrics were calculated as a function of the true positive (TP), true negative (TN), FP, and FN observations, as follows:

$$\mathrm{Accuracy} = (\mathrm{TP}+\mathrm{TN})/(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}) \tag{8}$$

$$\mathrm{Sensitivity} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \tag{9}$$

$$\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FP}) \tag{10}$$

$$F1 = 2 \times (\mathrm{Precision} \times \mathrm{Sensitivity}) / (\mathrm{Precision} + \mathrm{Sensitivity}). \tag{11}$$

Table VII reports a comparison between the cross-validation and test performance achieved by the three classifiers for each case listed in Table VI, using the two datasets with the respective top-N features ranked in Table III. All classifiers achieved high performance, demonstrating their ability to effectively discriminate 16 different oil conditions even with a small dataset and a limited number of features. In all the cases, misclassifications have never involved the class 1 (aged oil), ensuring that contamination conditions were always detectable.

TABLE VII
MACRO-AVERAGED PERFORMANCE METRICS (ACCURACY, SENSITIVITY, PRECISION, $F1$-SCORE, AND RUNNING TIME) CALCULATED ON THE VALIDATION AND TESTING SET, FOR THE TWO DATASETS

Training dataset	Model	k	Cross-Validation results (%)				Testing results (%)				Time (s)
			Accuracy	Recall	Precision	F1	Accuracy	Recall	Precision	F1	
Impedance components-based	1-NN	3	99.2	93.8	95.1	93.5	**99.8**	**98.4**	**98.8**	**98.4**	**2.14**
		4	99.6	96.9	97.4	96.8	99.6	96.9	97.9	96.7	1.61
		8	99.6	96.9	97.4	96.8	99.6	96.9	97.9	96.7	2.06
	SVM	3	99.2	93.8	95.5	93.5	99.6	96.9	97.9	96.7	8.67
		4	99.4	95.3	95.6	95.3	99.6	96.9	97.9	96.7	11.93
		8	99.7	97.7	98.1	97.6	99.6	96.9	97.9	96.7	20.46
	EBT	3	99.2	93.8	95.5	93.4	99.4	95.3	96.7	95.1	6.08
		4	99.8	98.4	98.6	98.4	99.6	96.9	97.9	96.7	3.23
		8	100	100	100	100	99.6	96.9	97.9	96.7	5.69
Tand-based	1-NN	3	98.7	89.6	89.7	89.4	97.5	79.7	77.8	77.8	2.60
		4	97.7	81.3	80.9	80.6	**99.6**	**96.9**	**97.9**	**96.7**	**1.36**
		8	97.9	82.8	85.1	82.1	99.6	96.9	97.9	96.7	2.42
	SVM	3	99	91.7	92.5	91.8	99.2	93.8	95.4	93.8	8.77
		4	98.9	91.4	91.8	91.5	99.6	96.9	97.9	96.7	11.79
		8	98.9	91.4	91.8	91.5	99.6	96.9	97.9	96.7	19.89
	EBT	3	99.6	96.9	96.8	96.8	99.2	93.8	94.4	93.4	6.50
		4	99.2	93.8	94	93.5	99.6	96.9	97.9	96.7	3.39
		8	99.4	95.3	95.8	95.3	99.6	96.9	97.9	96.7	6.62

The best classification using the impedance components-based dataset was obtained by the 1-NN classifier with a three-cross-validation (accuracy of 99.8%, recall of 98.4%, precision of 98.8%, $F1$ of 98.4%, running time of 2.14 s). The algorithm classified 15 contaminations correctly, except for the class 10 (containing 0.5% water mixed to 3% fuel) that was confused with the class 13 (containing 0.2% water mixed to 5% fuel) with a chance of 25%, as shown by the confusion matrix in Fig. 5(a).

The best classification using the $\tan\delta$-based dataset was performed by the 1-NN classifier using a four-cross-validation (accuracy of 99.6%, recall of 96.9%, precision of 97.9%, $F1$ of 96.8%, and running time of 1.36 s). In this case, the algorithm classified 15 contaminations correctly, except for the class 8 (containing 2% water) that was confused with the class 12 (containing 0.5% water) with a chance of 50%, as shown by the confusion matrix in Fig. 5(b). By comparing the performance of the best cases for both datasets, it is worth noting that, even though the two algorithms provided very similar accuracy values, the first one (1-NN with threefolds) achieved higher values of recall and precision, demonstrating a superior ability to reduce FP and FN rates.

Overall, the results in Table VII suggest that the spectral content of the real and imaginary impedance components may be more informative and useful than $\tan\delta$ for monitoring oil condition, thereby stimulating greater interest in the development of impedance-based monitoring systems for industrial applications. Nevertheless, it is important to remind that the classification results presented in the article are referred to specific brands of lubricating oil and fuel, namely AeroShell Turbine Oil 500 and AeroShell Calibrating Fluid 2. However, the approach proposed in this study remains valid and can be adapted to different types of oils and fuels by appropriately retraining the ML models to account for the unique physical and chemical characteristics of the involved fluids.

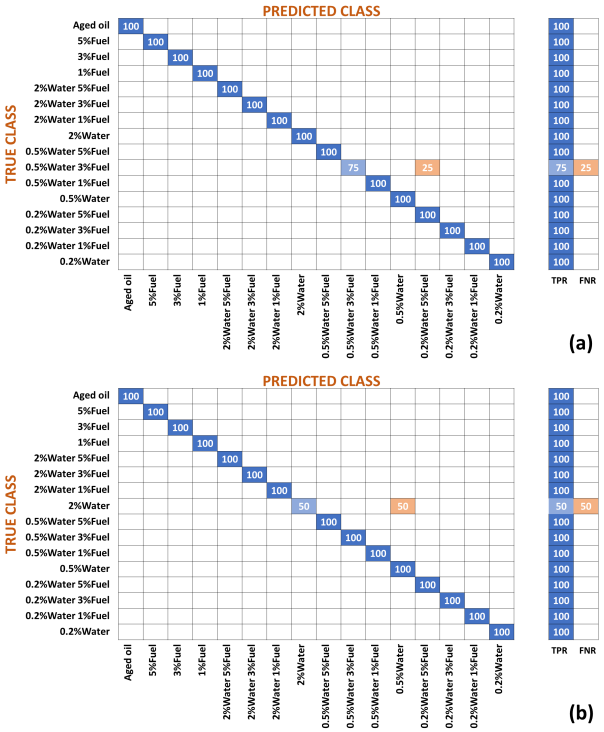


Fig. 5. Confusion matrix on the test results obtained by the 1-NN classifier trained by (a) three-cross-validation using the impedance components-based dataset and (b) four-cross-validation using the tanδ-based dataset (color in print).

Monitoring the oil condition using complex impedance signals from interdigitated capacitive sensors is a promising approach currently absent on the market. However, integrating

TABLE VIII
COMPARISON OF THIS WORK WITH RECENT LITERATURE WHERE IMPEDANCE SPECTROSCOPY-DRIVEN ML MODELS WERE USED IN OIL MONITORING APPLICATIONS

Application	ML Method	Best performance	Ref
Prediction of contamination by water and aging degree of lubricant oils from Attenuated Total Reflectance-Fourier-Transform Infrared Spectroscopy (ATR-FTIR) spectra	Three-layer-feed-forward based-Artificial Neural Network, Linear Discriminant Analysis	Accuracy: 97.3% (aging week) 97.7% (for viscosity and water presence). Recall, Precision, F1: N.A.	[57]
Assessing the condition of lubricating oil in automotive diesel engines	Logistic Regression, Classification by Decision Tree, Random Forest and Gradient Boosting	Accuracy: N.A. Recall 93% Precision 96% F1-score 94%.	[58]
Condition monitoring of oil regime in spur gear based on acoustic emission data	Feed-Forward Back-Propagation Neural Networks, Elman network	Accuracy: **99.7 – 100%** as a function of different load and speed conditions	[59]
Predictions of lubrication regimes of hydrodynamic journal bearings using high-speed torque measurements	Deep neural networks, Logistic regression	Accuracy: 99.25% Recall, Precision, F1: N.A.	[60]
Monitoring of multiple properties of the diesel powered machines' lubricants	Artificial Neural Network with a stochastic global optimization method	Prediction of water, diesel, base, and soot concentrations with an average error of 12.1%, 7.8%, 5.7%, and 3.9% respectively, and maximum errors of 15.2%, 12.3%, 6.8%, and 6.0% respectively.	[61]
Classification and prediction of oil brands and engine types using elemental composition	Discriminant analysis	Accuracy: 92.1% for oil brand prediction 82.9% for engine type prediction 92.1% for vehicle prediction Recall, Precision, F1: N.A.	[62]
Engine condition predictions in presence of metal and wear contamination	Support vector machine (SVM), Radial basis function (RBF)	Accuracy: 97% (SVM) 99% (RBF) Recall, Precision, F1: N.A.	[63]
Prediction of solid particle contaminants in the lubricant of a Spur Gearbox using Vibration, Acoustic Emission, and Sound Signature features	Classification and Regression Tree and Support Vector Machine	SVM Accuracy: 99.7% F1-score: 99.6% Recall, Precision: N.A	[64]
This work	1-NN SVM EBT	**Accuracy: 99.8%** **Recall: 98.4%** **Precision: 98.8%** **F1-score: 98.4%**	

the sensor into existing industrial components requires to address some technical and operational challenges to maximize performance and preserve sensor functionality in hazardous environments. Proper implementation, supported by ML algorithms for data analysis, can promote enabling technology for enhancing equipment maintenance and lifespan. Table VIII reports a list of recent works published in the literature, in which impedance spectroscopy was used with ML for oil monitoring and classification purposes. To the best of our knowledge, these represent state-of-the-art results on the prediction and classification with high performance of oil condition across multiple lubricant monitoring tasks. All the prediction models developed in such works were trained by using large dataset. The main novelties of our work are as follows: 1) use of few but significant impedance measurements measured in a short time from the onset of the contamination condition to classify cross-contamination of water and fuel in aged oil by a proprietary capacitive sensor; 2) use of an effective method of data augmentation to address the challenge of having few measured data for model training; and 3) rapid and effective identification of various cross-contamination conditions of lube oil. The promising results obtained in this work suggest that the proposed approach can be extended to other contexts where rapid recognition of anomalies may enhance the effectiveness of timely predictive and preventive maintenance strategies.

Future work aims to refine our approach to include non-steady-state spectra measured shortly after the onset of anomalies, for extracting valuable features from the transient regime that could be related to the nature of the anomalies.

VI. CONCLUSION

This work presented a solution to a multiclass classification problem of cross-contaminations in aviation lubricant by an

effective approach of impedance spectroscopy and data-driven supervised ML. The successful classification of 16 aged oil samples cross-contaminated with water and fuel was demonstrated. This opportunity is in line with the industrial need to perform an effective (near) real-time monitoring of oil condition that allows to foresee potential failures and improves the effectiveness of timely maintenance plans.

Impedance measurements were performed at room temperature by immersing a microfabricated sensor in 16 samples of aged aviation oil containing increasing concentrations of water and aviation fuel. Few steady-state spectra were used for creating two datasets for training separately three supervised ML models. The first dataset included real and imaginary impedance components spectra; the second dataset included tan δ spectra. A data pre-processing and augmentation method was proposed for generating synthetic examples from the measured data. Three schemes of dataset split ratio and K-fold cross-validation were evaluated for each of three investigated models. The comparative performance analysis between models and datasets allowed to evaluate which properties of the oil (electrical, dielectric, or both) were most significantly affected by the contamination, and which contributions (resistive, reactive, or both, in terms of dissipation factor) best emphasized such an effect. The successful identification of 16 aged oil samples cross-contaminated with water and fuel was demonstrated. In all the cases, misclassifications have never involved the class related to uncontaminated oil, allowing always to detect contamination conditions. The 1-NN classifier demonstrated to be the most effective to reduce FPs, FNs, and computational running time. The best results were obtained by employing a dataset split ratio of 60:40 and threefold cross-validation scheme on the impedance components-based dataset, yielding an accuracy of 99.8%.

Based on the obtained results, the following conclusions can be drawn for the proposed approach: 1) it allowed for an effective model learning using few training data, with a good generalization capability to new data; 2) its effectiveness and robustness were confirmed on different configurations and subsets of data; and 3) it allowed to develop a balanced classification model that could be successfully applied to the contamination monitoring in lubrication systems in real-world scenarios, and expanded to other contexts where timely predictive and preventive maintenance strategies are crucial.

REFERENCES

[1] S. Lee, D. C. P. Lozano, H. E. Jones, K. Shin, and M. P. Barrow, "Characterization of mineral and synthetic base oils by gas chromatography–mass spectrometry and Fourier transform ion cyclotron resonance mass spectrometry," *Energy Fuels*, vol. 36, no. 22, pp. 13518–13525, Nov. 2022, doi: 10.1021/acs.energyfuels.2c02437.

[2] S. Rațiu, A. Josan, V. Alexa, V. G. Cioată, and I. Kiss, "Impact of contaminants on engine oil: A review," *J. Phys., Conf. Ser.*, vol. 1781, no. 1, Feb. 2021, Art. no. 012051, doi: 10.1088/1742-6596/1781/1/012051.

[3] R. I. Taylor, "Fuel-lubricant interactions: Critical review of recent work," *Lubricants*, vol. 9, no. 9, p. 92, Sep. 2021, doi: 10.3390/lubricants9090092.

[4] P. Kaminski, "Experimental investigation into the effects of fuel dilution on the change in chemical properties of lubricating oil used in fuel injection pump of pielstick PA4V185 marine diesel engine," *Lubricants*, vol. 10, no. 7, p. 162, Jul. 2022, doi: 10.3390/lubricants10070162.

[5] X. Liu et al., "Direct observation of the impact of water droplets on oil replenishment in EHD lubricated contacts," *Friction*, vol. 10, no. 3, pp. 388–397, Mar. 2022, doi: 10.1007/s40544-020-0466-0.

[6] J. J. Benner, F. Sadeghi, M. R. Hoeprich, and M. C. Frank, "Lubricating properties of water in oil emulsions," *J. Tribol.*, vol. 128, no. 2, pp. 296–311, Apr. 2006, doi: 10.1115/1.2164464.

[7] R. Niculescu, V. Iorga-Simăn, A. Trică, and A. Clenci, "Study on the engine oil's wear based on the flash point," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 147, Aug. 2016, Art. no. 012124, doi: 10.1088/1757-899x/147/1/012124.

[8] R. Tang et al., "Detailed speciation of semi-volatile and intermediate-volatility organic compounds (S/IVOCs) in marine fuel oils using GC × GC-MS," *Int. J. Environ. Res. Public Health*, vol. 20, no. 3, p. 2508, Jan. 2023, doi: 10.3390/ijerph20032508.

[9] C. Winterfield and F. R. van de Voort, "Automated acid and base number determination of mineral-based lubricants by Fourier transform infrared spectroscopy: Commercial laboratory evaluation," *SLAS Technol.*, vol. 19, no. 6, pp. 577–586, Dec. 2014, doi: 10.1177/2211068214551825.

[10] T. Holland, A. Abdul-Munaim, D. Watson, and P. Sivakumar, "Importance of emulsification in calibrating infrared spectroscopes for analyzing water contamination in used or in-service engine oil," *Lubricants*, vol. 6, no. 2, p. 35, Apr. 2018, doi: 10.3390/lubricants6020035.

[11] A. Wolak, W. Krasodomski, and G. Zając, "FTIR analysis and monitoring of used synthetic oils operated under similar driving conditions," *Friction*, vol. 8, no. 5, pp. 995–1006, Oct. 2020, doi: 10.1007/s40544-019-0344-9.

[12] *Insulating Liquids-Measurement of Relative Permittivity, Dielectric Dissipation Factor (Tan D) and DC Resistivity*, Standard IEC 60247. Accessed: Jan. 2024. [Online]. Available: https://webstore.iec.ch/publication/1150

[13] *Standard Test Method for Dissipation Factor (or Power Factor) and Relative Permittivity (Dielectric Constant) of Electrical Insulating Liquids*, Standard ASTM D924-15. Accessed: Jan. 2024. [Online]. Available: https://www.astm.org/d0924-15.html

[14] S. R. Chowdhury, R. Kumar, R. Kaur, A. Sharma, and A. P. Bhondekar, "Quality assessment of engine oil: An impedance spectroscopy based approach," in *Proc. IEEE Uttar Pradesh Sect. Int. Conf. Electr., Comput. Electron. Eng. (UPCON)*, Varanasi, India, Dec. 2016, pp. 608–612, doi: 10.1109/UPCON.2016.7894724.

[15] C. Ulrich, H. Petersson, H. Sundgren, F. Björefors, and C. Krantz-Rülcker, "Simultaneous estimation of soot and diesel contamination in engine oil using electrochemical impedance spectroscopy," *Sens. Actuators B, Chem.*, vol. 127, no. 2, pp. 613–618, Nov. 2007, doi: 10.1016/j.snb.2007.05.014.

[16] H. Liu, X. Tang, H. Lu, W. Xie, Y. Hu, and Q. Xue, "An interdigitated impedance microsensor for detection of moisture content in engine oil," *Nanotechnol. Precis. Eng.*, vol. 3, no. 2, pp. 75–80, Jun. 2020, doi: 10.1016/j.npe.2020.04.001.

[17] I. Halalay and E. Schneider, "In-situ monitoring of engine oils through electrical AC impedance measurements," SAE Tech. Rep. 2007-01-4092, 2007, doi: 10.4271/2007-01-4092.

[18] Poseidon Systems. *Trident QM3100, Oil Quality Monitor*. Accessed: Jan. 2024. [Online]. Available: https://www.poseidonsys.com/products-and-services/products/oil-quality-sensors/trident-qm3100/

[19] O. Surucu, S. A. Gadsden, and J. Yawney, "Condition monitoring using machine learning: A review of theory, applications, and recent advances," *Expert Syst. Appl.*, vol. 221, Jul. 2023, Art. no. 119738, doi: 10.1016/j.eswa.2023.119738.

[20] W. H. Hunter Woodward, "Broadband dielectric spectroscopy—A practical guide," in *Broadband Dielectric Spectroscopy: A Modern Analytical Technique*. Washington, DC, USA: American Chemical Society (ACS), Jan. 2021, ch. 1, pp. 3–59, doi: 10.1021/bk-2021-1375.ch001.

[21] J. Y.-L. Chan et al., "Mitigating the multicollinearity problem and its machine learning approach: A review," *Mathematics*, vol. 10, no. 8, p. 1283, Apr. 2022, doi: 10.3390/math10081283.

[22] H. Tao, L. Cheng, J. Qiu, and V. Stojanovic, "Few shot cross equipment fault diagnosis method based on parameter optimization and feature mertic," *Meas. Sci. Technol.*, vol. 33, no. 11, Nov. 2022, Art. no. 115005, doi: 10.1088/1361-6501/ac8368.

[23] H. V. Luong, N. Deligiannis, R. Wilhelm, and B. Drapp, "Few-shot classification with meta-learning for urban infrastructure monitoring using distributed acoustic sensing," *Sensors*, vol. 24, no. 1, p. 49, Dec. 2023, doi: 10.3390/s24010049.

[24] W. Wang et al., "Medical tumor image classification based on few-shot learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 21, no. 4, pp. 715–724, Jul. 2023, doi: 10.1109/TCBB.2023.3282226.

[25] X. Shi, S. Zhang, M. Cheng, L. He, X. Tang, and Z. Cui, "Few-shot semantic segmentation for industrial defect recognition," *Comput. Ind.*, vol. 148, Jun. 2023, Art. no. 103901, doi: 10.1016/j.compind.2023.103901.

[26] V. Bongiorno, S. Gibbon, E. Michailidou, and M. Curioni, "Exploring the use of machine learning for interpreting electrochemical impedance spectroscopy data: Evaluation of the training dataset size," *Corrosion Sci.*, vol. 198, Apr. 2022, Art. no. 110119, doi: 10.1016/j.corsci.2022.110119.

[27] S. Zhu, X. Sun, X. Gao, J. Wang, N. Zhao, and J. Sha, "Equivalent circuit model recognition of electrochemical impedance spectroscopy via machine learning," *J. Electroanal. Chem.*, vol. 855, Dec. 2019, Art. no. 113627, doi: 10.1016/j.jelechem.2019.113627.

[28] B.-Y. Chang, "A novel analysis method for electrochemical impedance spectra using deep learning," *Electrochim. Acta*, vol. 462, Sep. 2023, Art. no. 142741, doi: 10.1016/j.electacta.2023.142741.

[29] J. Schaeffer et al., "Machine learning benchmarks for the classification of equivalent circuit models from electrochemical impedance spectra," *J. Electrochem. Soc.*, vol. 170, no. 6, Jun. 2023, Art. no. 060512, doi: 10.1149/1945-7111/acd8fb.

[30] A. V. Radogna, C. De Pascali, E. Sciurti, M. A. Signore, S. D'Amico, and L. Francioso, "A 18-ms measurement-time MLS-based system for moisture assessment in lubricant oil," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023, doi: 10.1109/TIM.2023.3291775.

[31] V. F. Lvovich and M. F. Smiechowski, "Impedance characterization of industrial lubricants," *Electrochim. Acta*, vol. 51, nos. 8–9, pp. 1487–1496, 2006, doi: 10.1016/j.electacta.2005.02.135.

[32] N. Perini, A. R. Prado, C. M. S. Sad, E. V. R. Castro, and M. B. J. G. Freitas, "Electrochemical impedance spectroscopy for in situ petroleum analysis and water-in-oil emulsion characterization," *Fuel*, vol. 91, no. 1, pp. 224–228, 2012, doi: 10.1016/j.fuel.2011.06.057.

[33] *Safey Data Sheet of AeroShell Turbine Oil 500*. Accessed: Jan. 2024. [Online]. Available: https://webstore.iec.ch/publication/1150.https://epcprodstorage.blob.core.windows.net/epcblobstorage/GSAP_msds_02625099.PDF

[34] *Safey Data Sheet of AeroShell Calibrating Fluid 2*. Accessed: Jan. 2024. [Online]. Available: https://www.alphapioneer.com/doc/Fluids/Calibrating%20Fluid%202/MSD_Aeroshell_Calibrating_2_wef_2020.PDF

[35] H. Sun, Y. Liu, and J. Tan, "Research on testing method of oil characteristic based on quartz tuning fork sensor," *Appl. Sci.*, vol. 11, no. 12, p. 5642, Jun. 2021, doi: 10.3390/app11125642.

[36] D. Coronado and J. Wenske, "Monitoring the oil of wind-turbine gearboxes: Main degradation indicators and detection methods," *Machines*, vol. 6, no. 2, p. 25, Jun. 2018, doi: 10.3390/machines6020025.

[37] D. Lovrec and V. Tič, "The importance of the electrical properties of hydraulic fluids," *Ind. Lubrication Tribology*, vol. 74, no. 3, pp. 302–308, Mar. 2022, doi: 10.1108/ilt-06-2021-0218.

[38] J. E. De Souza, M. D. Scherer, J. A. S. Cáceres, A. R. L. Caires, and J.-C. M'Peko, "A close dielectric spectroscopic analysis of diesel/biodiesel blends and potential dielectric approaches for biodiesel content assessment," *Fuel*, vol. 105, pp. 705–710, Mar. 2013, doi: 10.1016/j.fuel.2012.09.032.

[39] D. Miljković, "Misfueling detection with two offseted capacitive fuel senders," *HDKBR INFO Magazin*, vol. 3, no. 2, pp. 32–38, 2013. [Online]. Available: https://hrcak.srce.hr/148798

[40] S. Bucko, P. Havran, J. Leško, and D. Megyesi, "Dielectric frequency spectroscopy as potential tool for the designation water content in aviation fuel," in *Proc. Mod. Saf. Technol. Transp. (MOSATT)*, Košice, Slovakia, 2019, pp. 25–28, doi: 10.1109/MOSATT48908.2019.8944095.

[41] D. Ljubas, H. Krpan, and I. Matanović, "Influence of engine oils dilution by fuels on their viscosity, flash point and fire point," *NAFTA*, vol. 61, no. 2, pp. 73–79, 2010.

[42] M. Hossain, A. Abu-Siada, and S. Muyeen, "Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review," *Energies*, vol. 11, no. 5, p. 1309, May 2018, doi: 10.3390/en11051309.

[43] Z. Wang, B. Draghi, Y. Rotalinti, D. Lunn, and P. Myles, "High-fidelity synthetic data applications for data augmentation," in *Artificial Intelligence*. London, U.K.: IntechOpen, Jan. 2024, doi: 10.5772/intechopen.113884.

[44] A. Figueira and B. Vaz, "Survey on synthetic data generation, evaluation methods and GANs," *Mathematics*, vol. 10, no. 15, p. 2733, Aug. 2022, doi: 10.3390/math10152733.

[45] W. Heylen and S. Lammens, "FRAC: A consistent way of comparing frequency response functions," in *Proc. Conf. Identificat. Eng. Syst.*, 1996, pp. 48–57.

[46] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series FeatuRe extraction on basis of scalable hypothesis tests (tsfresh—A Python package)," *Neurocomputing*, vol. 307, pp. 72–77, Sep. 2018, doi: 10.1016/j.neucom.2018.03.067.

[47] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the influence of normalization/transformation process on the accuracy of supervised classification," in *Proc. 3rd Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Tirunelveli, India, Aug. 2020, pp. 729–735, doi: 10.1109/ICSSIT48917.2020.9214160.

[48] Y. Zhai, W. Song, X. Liu, L. Liu, and X. Zhao, "A chi-square statistics based feature selection method in text classification," in *Proc. IEEE 9th Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Beijing, China, Nov. 2018, pp. 160–163, doi: 10.1109/ICSESS.2018.8663882.

[49] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," in *Proc. 7th IEEE Int. Conf. Tools Artif. Intell.*, Herndon, VA, USA, Nov. 1995, pp. 388–391, doi: 10.1109/TAI.1995.479783.

[50] *Overview on Extracted Features*. Accessed: Jan. 2024. [Online]. Available: https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html

[51] B. McDermott, M. O'Halloran, E. Porter, and A. Santorelli, "Brain haemorrhage detection using a SVM classifier with electrical impedance tomography measurement frames," *PLoS ONE*, vol. 13, no. 7, Jul. 2018, Art. no. e0200469, doi: 10.1371/journal.pone.0200469.

[52] M. A. Shini, S. Laufer, and B. Rubinsky, "SVM for prostate cancer using electrical impedance measurements," *Physiol. Meas.*, vol. 32, no. 9, pp. 1373–1387, Sep. 2011, doi: 10.1088/0967-3334/32/9/002.

[53] M. Tiitta, V. Tiitta, J. Heikkinen, R. Lappalainen, and L. Tomppo, "Classification of wood chips using electrical impedance spectroscopy and machine learning," *Sensors*, vol. 20, no. 4, p. 1076, Feb. 2020, doi: 10.3390/s20041076.

[54] R. Onanena, L. Oukhellou, E. Come, D. Candusso, D. Hissel, and P. Aknin, "Fault-diagnosis of PEM fuel cells using electrochemical spectroscopy impedance," *IFAC Proc. Volumes*, vol. 45, no. 21, pp. 651–656, 2012, doi: 10.3182/20120902-4-fr-2032.00114.

[55] M. Lee, D. Han, K. Yoo, and J. Kim, "Impedance technique combined with supervised algorithms-based internal degradation state classification and its economic analysis for safety in retired battery pack," *J. Energy Storage*, vol. 73, Dec. 2023, Art. no. 109037, doi: 10.1016/j.est.2023.109037.

[56] R. A. Potyrailo et al., "Multivariable electrical resonant sensors for independent quantitation of aging and external contaminants in lubricating oils," *IEEE Sensors J.*, vol. 19, no. 4, pp. 1542–1553, Feb. 2019.

[57] C. Chimeno-Trinchet, C. Murru, M. E. Díaz-García, A. Fernández-González, and R. Badía-Laíño, "Artificial intelligence and Fourier-transform infrared spectroscopy for evaluating water-mediated degradation of lubricant oils," *Talanta*, vol. 219, Nov. 2020, Art. no. 121312, doi: 10.1016/j.talanta.2020.121312.

[58] R. Malaguti, N. Lourenço, and C. Silva, "A supervised machine learning model for determining lubricant oil operating conditions," *Expert Syst.*, vol. 40, no. 5, Jun. 2023, Art. no. e13116, doi: 10.1111/exsy.13116.

[59] Y. H. Ali, R. A. Rahman, and R. I. R. Hamzah, "Artificial neural network model for monitoring oil film regime in spur gear based on acoustic emission data," *Shock Vib.*, vol. 2015, no. 1, pp. 1–12, 2015, doi: 10.1155/2015/106945.

[60] J. Moder, P. Bergmann, and F. Grün, "Lubrication regime classification of hydrodynamic journal bearings by machine learning using torque data," *Lubricants*, vol. 6, no. 4, p. 108, Dec. 2018, doi: 10.3390/lubricants6040108.

[61] A. Urban and J. Zhe, "A microsensor array for diesel engine lubricant monitoring using deep learning with stochastic global optimization," *Sens. Actuators A, Phys.*, vol. 343, Aug. 2022, Art. no. 113671, doi: 10.1016/j.sna.2022.113671.

[62] Y. Kim, N. Y. Kim, S. Y. Park, D.-K. Lee, and J. H. Lee, "Classification and individualization of used engine oils using elemental composition and discriminant analysis," *Forensic Sci. Int.*, vol. 230, nos. 1–3, pp. 58–67, Jul. 2013, doi: 10.1016/j.forsciint.2013.01.013.

[63] M. Rahimi, M.-R. Pourramezan, and A. Rohani, "Modeling and classifying the in-operando effects of wear and metal contaminations of lubricating oil on diesel engine: A machine learning approach," *Expert Syst. Appl.*, vol. 203, Oct. 2022, Art. no. 117494, doi: 10.1016/j.eswa.2022.117494.

[64] K. Rameshkumar, R. Sriram, M. Saimurugan, and P. Krishnakumar, "Establishing statistical correlation between sensor signature features and lubricant solid particle contamination in a spur gearbox," *IEEE Access*, vol. 10, pp. 106230–106247, 2022, doi: 10.1109/ACCESS.2022.3210983.

Chiara De Pascali received the bachelor's degree in information engineering, the master's (Hons.) degree in telecommunication engineering, and the Ph.D. degree in energetic systems and environment from the University of Salento, Lecce, Italy, in 2009 and 2013, respectively.

Since 2013, she has been working with the Institute for Microelectronics and Microsystems (IMM), Italian National Research Council (CNR), Lecce, where her research activity is mainly focused on finite element method (FEM) design, fabrication, and characterization of microscale sensors and micro electro mechanical systems (MEMS) devices for wearable electronics, environmental monitoring, and energy harvesting applications. She has been involved as component in many national projects. She has authored more than 50 peer-reviewed papers on national and international journals, 1 book chapter, and several communications to national and international conferences.

Daniele Bellisario received the bachelor's degree in experimental and applied biology from the University of Salento, Lecce, Italy, in 2022.

Since 2022, he has been a Research Fellow with the Institute for Microelectronics and Microsystems (IMM), Italian National Research Council (CNR), Lecce. His research interests focus around signal processing and development of machine learning models designed for spectrum classification across various domains, including industrial sectors, biomedical applications, and organ-on-chip devices.

Maria Assunta Signore received the master's (Laurea) degree in physics from the University of Lecce, Lecce, Italy, in 2002.

In 2004, she joined the National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Brindisi, Italy, working in the field of thin films deposition by physical vapor deposition (PVD) processes. Since 2010, she has been with the Sensors and Microsystems Group, Institute for Microelectronic and Microsystems (IMM), Italian National Research Council (CNR), Lecce. Since 2018, she has been a permanent Researcher with IMM, CNR. The research line is mainly dedicated in the field of growth and functional characterization of environmentally sound lead-free piezoelectric materials to integrate into micro electro mechanical systems (MEMS) devices as active layer for a wide range of energy harvesting and sensing applications, including biosensors and implantable sensors. She has presented the results of her research, as contributed talk, in more than 20 domestic and international congresses. She has been involved in many national projects, as component and as a scientific coordinator. She has authored more than 50 peer-reviewed papers on national and international journals.

Ms. Signore is a member of the examining board for research grant and regularly acts as a reviewer for numerous peer-reviewed journals.

Elisa Sciurti was born in Zurich, Switzerland, in 1991. She received the M.Sc. degree in biology from the University of Salento, Lecce, Italy, in 2016, and the Ph.D. degree in materials and structural engineering and nanotechnology from the University of Salento and Italian Institute of Technology (IITCBN), Genova, Italy, in 2020.

Since 2021, she has been a Postdoctoral Fellow with the Institute for Microelectronics and Microsystems (IMM), National Research Council (CNR), Lecce. Her research interests include the development of electrochemical sensors for biomedical applications, laboratory-on-a-chip devices, and energy storage systems.

Antonio Vincenzo Radogna (Member, IEEE) received the M.Eng. degree in electronic engineering from the University of Parma, Parma, Italy, in 2012, and the Ph.D. degree in complex systems engineering from the University of Salento, Lecce, Italy, in 2021.

From 2012 to 2017, he was an Electronic Engineer mainly with the Ambient Assisted Living (AAL) and renewable energies industries. From 2017 to 2023, he was a Research Fellow with the Institute for Microelectronics and Microsystems (IMM), National Research Council (CNR), Lecce. Since 2023, he has been an Assistant Professor with the Department of Experimental Medicine, University of Salento. His current research interests include the design of CMOS integrated readout interfaces for sensors, biomedical electronic circuits for healthcare, e-noses development, and industrial sensing technologies.

Dr. Radogna is a Publication Chair of the committee of the 2023 International Conference on IC Design and Technology (ICICDT).

Luca Nunzio Francioso received the M.Sc. degree in physics from the University of Lecce, Lecce, Italy, in 2001, and the Ph.D. degree from the University of Lecce, in 2006, with a thesis on application of miniaturized gas sensors for combustion and cabin air quality assessment in the automotive field.

Since 2001, he has been with the Institute for Microelectronic and Microsystems (IMM), Italian National Research Council (CNR), Lecce, working in the field of silicon micromachined systems and solid state thin film chemical sensors, where he has been a permanent Staff Researcher since 2003, devoted to silicon technology development and micro electro mechanical systems (MEMS) devices fabrication. He is currently the Head of the Multifunctional Devices Design and Characterization Laboratory (M2DCL), CNR-IMM. His current research interests include flexible thermoelectric and piezoelectric generators, design and manufacturing of semiconductor-based chemical sensors, and organ-on-chip platforms. He has co-authored many peer-reviewed papers on national and international journals, two book chapters, and several communications to national and international conferences as regular and invited talks.

Open Access funding provided by 'Consiglio Nazionale delle Ricerche-CARI-CARE-ITALY' within the CRUI CARE Agreement