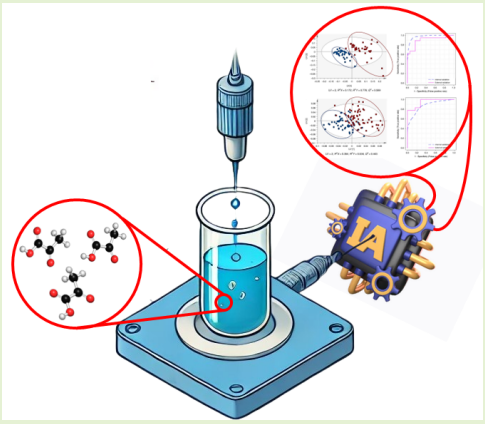
Boosting the Performance of Planar CSRR Sensors in Biomedical Applications Through Machine Learning

Javier Alonso-Valdesueiro, Luis Fernández, Agustín Gutiérrez-Gálvez, and Santiago Marco-Colás

***Abstract*— Complementary Split Ring Resonators (CSRRs) have been widely explored as planar sensors over the last decade. Despite their proven performance in laboratory settings — where high-end Vector Network Analyzers (VNAs) are employed, environmental conditions are strictly controlled, and the Sample Under Test (SUT) is meticulously prepared — their applicability outside controlled environments remains limited. Most studies have evaluated CSRR-based sensors in terms of linearity, sensitivity, and accuracy, often overlooking their practical feasibility in real-world applications.**



In this contribution, a novel approach is proposed to assess the practical applicability of CSRR sensors in biomedical research. Unlike conventional studies, the S-parameters are measured using an inexpensive, portable VNA, the SUT is placed inside standard test vials commonly used in biomedical laboratories, and no environmental control is applied during sampling. Ethanol solutions diluted in deionized water, ranging from 10% to 96% in concentration, were prepared and poured into commercial vials, with 10 replicates per concentration. Measurements were randomized, yielding 50 repetitions per concentration and approximately 500 measurements in total — a sample size significantly larger than typically reported in the literature, aiming to enhance the robustness and reliability of the Machine Learning models.

Principal Component Analysis (PCA) was applied to the acquired dataset for visualization and to explore the underlying structure of the data. The large number of measurements enabled a more comprehensive representation of the experimental variability. Subsequently, a Partial Least Squares Regressor (PLSR) was trained and validated using a Leave-One-Group-Out Cross-Validation scheme. The model achieved a Root Mean Square Error in Prediction (RMSEP) of approximately 3.7% and a Limit of Agreement of approximately 15% within a 95% Confidence Interval — without exhibiting signs of underfitting or overfitting. These results highlight the potential of CSRR-based sensors for the quantification of bi-component liquid mixtures under realistic, uncontrolled conditions for biomedical applications.

***Index Terms*— CSRR Sensors, Machine Learning, RF Biomedical, Resonators.**

I. Introduction

COMPLEMENTARY Split Ring Resonators (CSRR) were introduced as metasurfaces and metamaterials [1], [2]. They were designed to act as microwave stopband and notch filters, transmission lines, and antennas [3]–[8]. However, due to their sharp behavior as filters and the influence of surface surroundings on their Scattering Parameters (S-Parameters) [4], [9], [10], they started to be used as sensors a few years after their first appearance [11]. Since then, several applications have focused on measuring the complex permittivity of different materials [12]–[17].

In the last seven years, CSRR sensors have been increasingly used in biomedical applications to measure the dielectric properties of different solutions and concentrations of solutes in solvents [18]–[20]. Some applications have attracted particular attention during this period (such as glucose concentration quantification in water [19], [21]), but most studies have focused on developing CSRR sensors for microfluidic devices applied to various fields [22]–[25].

The use of statistical learning tools (or Machine Learning algorithms) to enhance the performance of CSRR-based devices has also been explored over the past five years [26]–[29]. These algorithms have been employed as fitting tools to generalize the dependence of certain features of the CSRR S-

J. Alonso-Valdesueiro is with University of Barcelona, Carrer Martí i Franquès,1. 08028, Barcelona, Spain (e-mail: javier.alonsov@ub.edu).

L. Fernández, is with University of Barcelona, Carrer Martí i Franquès,1. 08028, Barcelona, Spain (e-mail: lfernandez@ub.edu) and the Institute for Biomedical Engineering of Catalonia (IBEC), Carrer Baldiri i Reixac, 4, Torre R. 08028, Barcelona, Spain (email: lfernandez@ibecbarcelona.eu).

A. Gutiérrez-Gálvez, is with University of Barcelona, Carrer Martí i Franquès,1. 08028, Barcelona, Spain (e-mail: agutierrez@ub.edu).

S. Marco-Colás, is with University of Barcelona, Carrer Martí i Franquès,1. 08028, Barcelona, Spain (e-mail: santiago.marco@ub.edu) and the Institute for Biomedical Engineering of Catalonia (IBEC), Carrer Baldiri i Reixac, 4, Torre R. 08028, Barcelona, Spain (email: smarco@ibecbarcelona.eu).

parameters (such as frequency resonances, magnitude values, and phase distortion) on the dielectric properties of the samples [21]. Moreover, in recent years, their use has expanded to include Multivariate Analysis (MVA) for correlating solute concentrations in water [30].

However, the full potential of ML algorithms applied to data from CSRR sensors has not yet been fully explored. These algorithms can generate quantification models inmune (up to certain point) to the variability (or noise) present in sensor readings [31]–[34]. ML models are capable of identifying common features in the data that allow correlation with specific characteristics of the sample under test (SUT) [35].

In this contribution, the proposed benchtop device is designed to be portable and low-cost, which introduces uncertainties due to the CSRR sensor, the Vector Network Analyzer (VNA) used to measure the S-parameters, and the sampling methodology. Enhancing the performance of this device is achieved by the appliying ML algorithms, specifically Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR). Therefore, a massive effort on building a comprehensive database was carried out. That allowed to use training schemes such as Leave-One-Group-Out Cross-Validation (LOGO-CV) to train the PLSR model which resulted in a quatification model that neither underfits nor overfits the training data.

The text is organized as follows: In Section II, the measurement system is described, including the main characteristics of the CSRR sensor and the low-cost VNA used to measure the on of the S-Parameters (S_{21}), the sampling methodology and the ML workflow to train the quantification PLSR model which. Section III presents the visualization of the extensive database generated for model training and the main results obtained with traditional modeling of CSRR sensor performance. This section also presents the data exploration through PCA and the the intermediate results of the PLSR training process. At the end the performance on the trained PLSR is reported. In Section IV, the performance of the CSRR system in quantifying the concentration of components in a binary solution is analyzed and discussed in comparisson with the traditional curve fitting models. Finally, Section V summarizes the capabilities of CSRR-based systems as benchtop devices working in biomedical applications.

II. CSRR BENCHTOP SYSTEM

A. System Block Diagram

Figure 1 (a) shows the block diagram of the CSRR based system. The importance of describing the measurement system comes from the fact that, as it has been discussed in section I, in the literature, measurement systems are optimized for extracting the best performance of the sensors.

In this contribution, the CSRR system is thought to be used as a bench-top tool in a biomedical laboratory. Therefore, the Sample Under Test (SUT) is presented to the sensor in a commercial vial (Chromacol 20-HSV). Its position on top of the sensor is ensured by a PLA holder 3D-printed. This holder keeps the CSRR still during measurements and ensures that the commercial vial is placed always centred with respect to the resonant structure of the CSRR sensor.

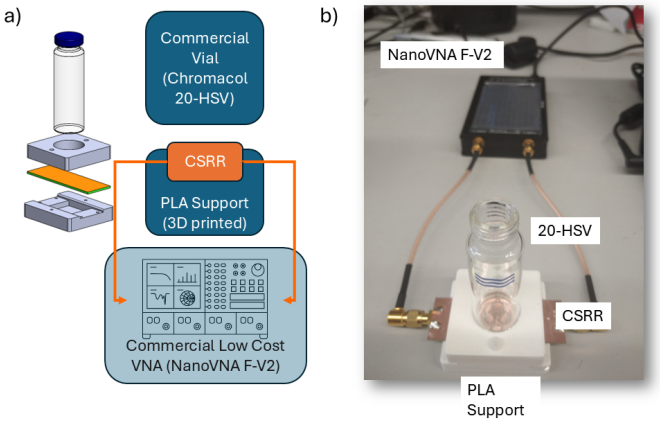


Fig. 1. Proposed bench-top CSRR system. a) The commercial vial (Chromacol 20-HSV) containing the SUT is placed on the CSRR using a PLA 3D printed support. The CSRR is connected to the low cost VNA (NanoVNA F-V2) by two sma cables. The NanoVNA can be operated via graphical interface running in the device or serial commands coming from an application running in a common laptop. b) Real setup with a vial on top of the CSRR.

The measurement device in a bench-top equipment, must be small, portable, and, if possible, affordable for being placed at any laboratory. In this case, the equipment must measure the S-parameters, in particular, the S_{21}. In some contributions, specific electronics were developed for the purpose of measuring this parameter [19], [36]. In many other contributions, commercial, high quality and expensive devices are used [22]–[24]. However, in this study, a commercial low-cost VNA was chosen as measurement platform (NanoVNA F-V2). This choice is based on the device affordability, performance in the 1 to 3 GHz band, and operability with an external laptop via serial commands.

Finally, the VNA is connected to a controlling GUI developed in python and running in a commercial lap-top. The GUI allows to acquire data, record it in a comprehensive structure for generating a useful database and run in a continuous measurement mode. Also, it includes a quantification module to accommodate the resulting ML models after training and make predictions of concentration from measurements de S-parameter.

B. CSRR Sensor

Figure 2 (a) shows the structure of the CSRR used in this contribution and the PCB manufactured by Eurocircuits NV. Its structure is based on the honeycomb CSRR presented in [36]. However, in order to present some extra features in the S_{21} parameter, some of its dimensions were modified. In particular, the symmetry of the vertical and horizontal combination of honeycombs was modified by 100 μm. This modification provided to CSRR with an extra feature ($\sim$ 3.5 GHz) close to the main resonance (3.2 GHz). This extra resonance comes from the fact that, the resonator presented here shows asymmetries in the horizontal and vertical plane. Therefore, it is expected to observe some directionality when measuring the S_{21} with a SUT placed on top of the resonator.

The coupling transmission line and the dimensions of each honeycomb are as defined in [36], so the CSRR does not

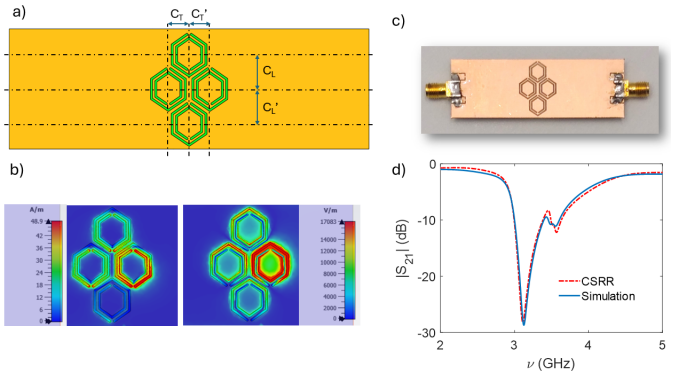


Fig. 2. Modified CSRR. a) Structure of the measured CSRR with certain modifications. The honeycombs are not placed equidistantly to the centrer of the PCB ($\mathbf{22} \times \mathbf{66}$ mm). The longitudinal honeycombs are placed at $C_L = \mathbf{6.3}$ and $C_L' = \mathbf{6.4}$ mm away from the centrer of the PCB and the transverse honeycombs are placed $C_T = \mathbf{3.7}$ and $C_T' = \mathbf{3.8}$ mm away from the centrer of the PCB. b) Simulated electric ($\vec{E}$ V/m) and magnetic field ($\vec{H}$ A/m) in the surface of the multiple honeycomb structure. c) CSRR mounted with sma connectors. d) S_{21} comparison between simulations and measurements when the CSRR is unloaded (Keisight E5071C-240).

introduce further uncertainty in the measurement.

C. Sampling Methodology

Sampling for training ML algorithms is a well established field in Statistical Learning. The main idea behind any sampling technique for ML model training is to randomize repetitions of each sample. This randomization mitigates undesired effects on the obtained dataset such as badge effects, repetition correlation and instrumental error propagation [37].

In this study, samples were organized by ethanol concentration in solution. The concentrations ranged from clean water (Ethanol at 0%) to Ethanol (96% purity) considering the steps depicted in figure 5 (a). Badge solutions were prepared with the mentioned concentrations and poured in 10 commercial vials, chosen randomly from a pool of 100 vials.

The 90 vials were labelled for randomization and identification, and placed together for a day in a fridge at 5 °C. The day after, the vials were leaved out the fridge for an hour at 23 °C (temperature at the laboratory) and 5 rounds of measurements were performed randomizing the order in which the vials were taken. Each round, consisted in ten repetitions of each concentration, where each repetition is a different vial. This provided a total of 100 measurements. Before each round, the response of the CSRR sensor unloaded (S_{21}) was measured and used for baseline correction of the data. In total, 455 measurements were taken for training and test of the ML model. Each measurement consisted in 201 points of recording the $20\log\left(|\dot{S}_{21}|\right)$ from 1.6 GHz to 3 GHz using the NanoVNA F-V2.

D. Machine Learning Workflow

Pincipal Component Analysis (PCA) and Partial Least Squares Regressor (PLSR) are the two algorithms used for data analysis and model training in this contribution. The workflow is depicted in figure 3 and described in the following sections.

1) Principal Component Analysis: Principal Component Analysis (PCA) was performed on the database generated as explained in section II-C. In this analysis, the dataset was prepared and pre-processed as depicted in figure 3 and the PCA was performed using the Statistics and Machine Learning Toolbox from MATLAB.

As it is shown in figure 3, the dataset is cleaned by extracting the metadata (labels, date of acquisition, etc.) of the database and adjusting the number of features of each repetition if necessary. Then, the data matrix (450×201) is pre-processed in two steps. First, each round of measurements is corrected in its baseline by using the $20\log\left(|\dot{S}_{21}|\right)$ of the unloaded CSRR. Second, each $20\log\left(|\dot{S}_{21}|\right)$ is auto-scaled by calculating the average, μ_C, and de standard deviation, σ_C, for each concentration at each round.

Up to 20 PCs were considered for the analysis. The EV of the dataset was calculated for each PC and the cumulative EV was plotted. The distribution of the repetitions in the reduced vectorial space was also plotted for the first three PCs.

2) Partial Least Squares Regressor: The results from PCA analysis suggested that using every feature in the $20\log\left(|\dot{S}_{21}|\right)$ spectrum could lead to good results in quantification. The most extended predictor in this cases is the Partial Least Squares Regressor (PLSR), especially suitable for datasets with short number of repetitions per sample [33].

Before starting the PLSR training, the feature matrix was pre-processed in the same way it was done for pca analysis (see figure 3). The pre-processed data is then split in two different subsets. The data belonging to 6 concentrations is used for training (training dataset) and the rest is left aside for final testing (test dataset). Therefore, the training dataset consisted in a 300×201 feature matrix and the test dataset in a 150×201.

A leave-one-block-out cross-validation (LOGO-CV) scheme is used for model training [38]. This approach uses data from 4 rounds for model training and reserves one round for validation, resulting in 5 evaluations of different PLSR models. The model with best performance, in terms of Root Mean Square Error, is then selected. The LOGO-CV scheme is run for different number of Latent Variables up to 31.

Once the PLSR model is trained, its evaluation is performed by introducing the test dataset and making predictions of the concentrations. In this case, predictions for 40%, 60% and 80% concentrations were calculated with the resulting model. These concentrations were chosen intentionally for compensating the presence of outliers (in 40% for example) without having into account the both ends of the considered range.

III. CSRR System Performance

In this section, the results from measurements, traditional characterization and ML analysis are presented. The performance of the CSRR system is evaluated in terms of the quantification of the concentration of ethanol in water. The results are discussed in terms of the traditional characterization of the system, the PCA analysis of the dataset and the performance of the PLSR model.

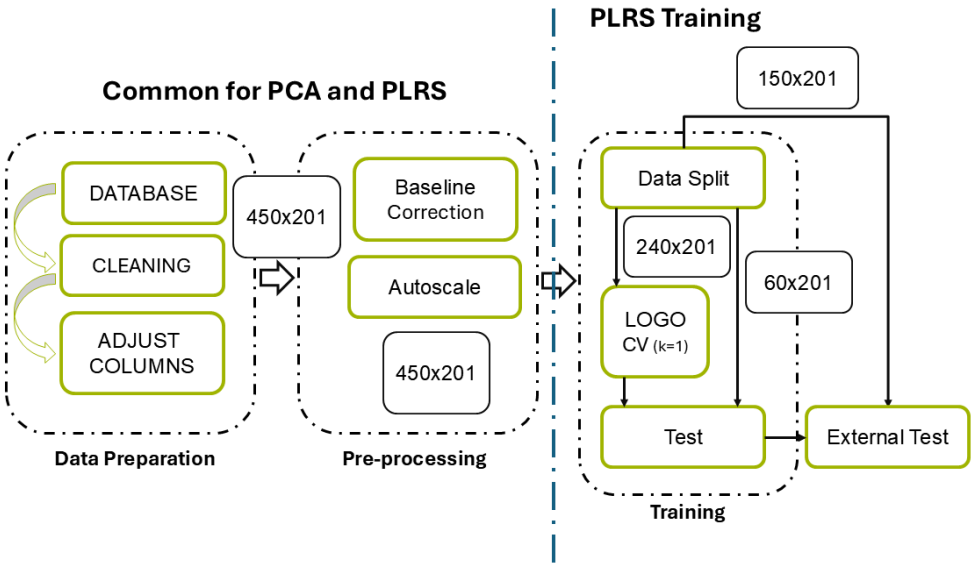


Fig. 3. Block diagram of the developed workflow for training the bench-top CSRR system. In Data Preparation step, the database is cleaned by extracting the actual data (features) from metadata, providing a $\mathbf{450 \times 201}$ feature matrix plus a $\mathbf{5 \times 201}$ baseline matrix. In the Pre-Processing step the feature matrix baseline is correcte using the baseline matrix applied to each of the rounds separately. In the Training step, the feature matrix is split into the train dataset ($\mathbf{300 \times 201}$ matrix) and the test dataset ($\mathbf{150 \times 201}$ matrix). These two step are common for PCA and PLSR training algorithms. Then the train dataset is introduced in a leave-one-block-out cross-validation (LOGO-CV) training scheme which produces and optimized PLSR model. In the External Test step, the performance of the optimized model is evaluated by obtaining predictions from the test dataset.

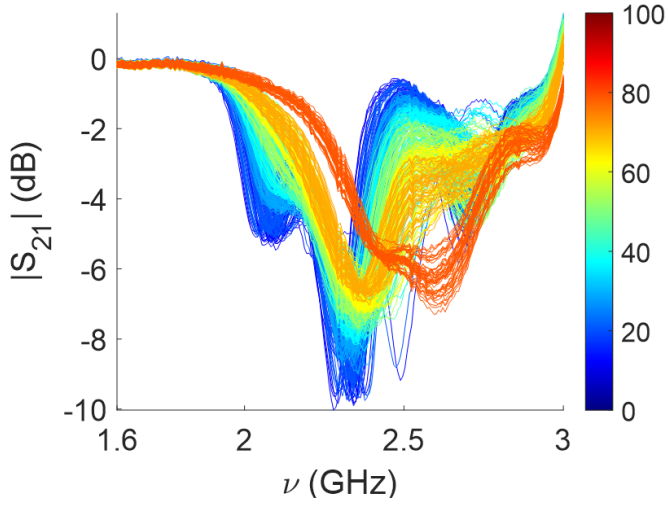


Fig. 4. $\mathbf{20log\,(|\dot{S}_{21}|)}$ of each measurement contained in the generated dataset. Sampled concentrations of Ethanol in clean water ranges from $\mathbf{0\%}$ and $\mathbf{96\%}$. $\mathbf{10}$ random commercial vials for each concentration were taken from a pool of $\mathbf{100}$. $\mathbf{5}$ rounds of measurements were performed. Each round consisted on measuring the $\mathbf{20log\,(|\dot{S}_{21}|)}$ of the CSRR when a random vial is picked and placed on top of the sensor, continuing and pass through the $\mathbf{90}$ vials. Before each round, the $\mathbf{20log\,(|\dot{S}_{21}|)}$ of the CSRR unloaded was recorded for baseline correction of the data.

A. Acquired Database

Figure 4 shows the $20\log(|\dot{S}_{21}|)$ of each measurement contained in the dataset used for ML model training and test. Some features can be spotted by simple inspection. First, a wide dispersion of the curves is observed for each concentration. In particular, dispersion of deepest resonance seems to decrease with the concentration of ethanol.

However, features around 2.1 GHz and 2.5 GHz show better discrimination between samples. Also, for 20% of ethanol, two repetitions show an outlier behaviour with a $20\log(|\dot{S}_{21}|)$ out of the trend shown by rest of the dataset.

Second, some outliers for certain concentrations (0%, 20% and 40%) were spotted in rounds 1, 2, 3 and 5. The repetitions corresponding to 40% showed one outlier at each round, spotting a very different characteristics of the container vial with respect of the rest of the pool. However, the outliers from 0% and 20% where spotted at different rounds which points to different manipulation of the vials during measurement.

B. Traditional System Characterization

Following the studies presented in the literature, figure 5 (a) shows the averaged $20\log(|\dot{S}_{21}|)$ for each concentration. It shows how the evolution of the shape of the S_{21} parameter with the concentration implies a non-linear response of the system with ethanol. Evaluating the frequency displacement of the most prominent resonance in Figure 5 (a) [29], a rational fitting of the deepest resonance results in the following expression:

$$C_{\%}(\nu) = \frac{p1 \cdot \nu + p2}{\nu + q} \tag{1}$$

when the frequency,ν is in GHz and p1$=110.7274\ \%/GHz^2$, p2$=-257.0246\ \%/GHz$ and q$=-2.2893\ GHz$. The R^2 of the fit is 0.9422 and the RMSE$=7.9208\ \%$. The fitting was performed using the Curve Fitting Toolbox from MATLAB. This model is presented in Figure 5 (b).

C. Principal Component Analysis

Figures 6 (a), (b) and (d), show the distribution of the repetitions in the reduced vectorial space of the extracted PCs. In both figures, the first PC shows its dominance which can be related with the ethanol concentration in solution. The spread in this component shows that, despite the careful preparation of

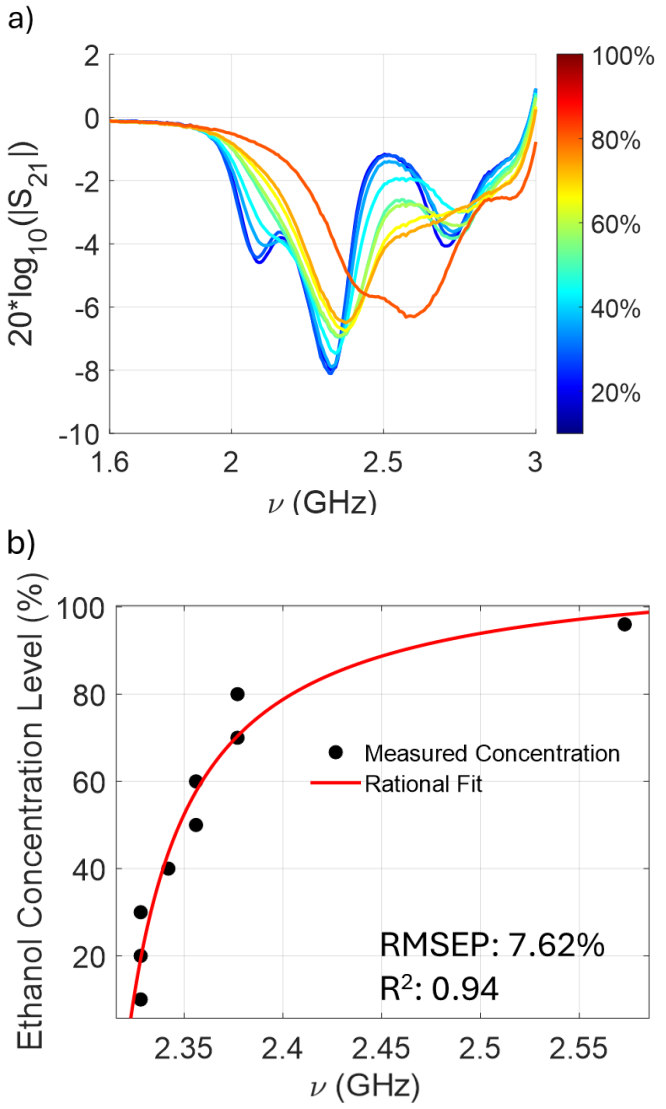


Fig. 5. Traditional characterization of the CSRR system when measuring concentrations of ethanol diluted in clean water and poured in commercial vials. a) Averaged $\mathbf{20log}\,(|\dot{S}_{21}|)$ response of the generated database. b) Rational model for the frequency of the minimum observed in the averaged $\mathbf{20log}\,(|\dot{S}_{21}|)$ when every concentration of ethanol in clean water is observed ($\mathbf{10}$, $\mathbf{20}$, $\mathbf{30}$, $\mathbf{40}$, $\mathbf{50}$, $\mathbf{60}$, $\mathbf{70}$, $\mathbf{80}$, and $\mathbf{96}$ %)

each stock solution, the system is sensitive enough to observe relatively small variations in concentration.

As shown by figure 6 (c), the 95% of the Explained Variance (EV) is contained in this first three Principal Components (PCs). Most of this variance is accumulated in the first PC (73%), and the rest show an EV lower than 15%.

Also, these figures show a spread in the second and third components which can be related to the variability introduced by the commercial vial and the actual measurement action (see figure 6 (d)). Is in this variability where the outliers are observed. In particular, plot in figure 6 (b) leads identify the outliers in concentrations of 0%, 20% and 40% shown in figure 4.

D. Partial Least Squares Regressor Training

Once the workflow presented in figure 3 is performed, there are certain intermediate results that provide ideas about the

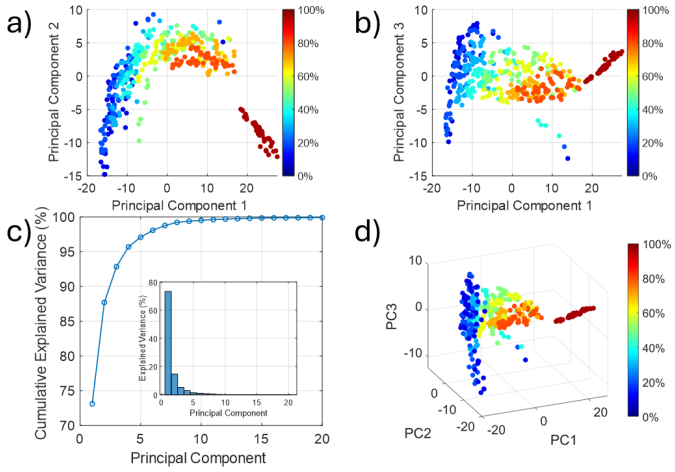


Fig. 6. Principal Component Analysis (PCA) of the dataset generated as described in section II-C. a) Distribution of each repetition in the reduced vectorial space when representing the second Principal Component (PC) with respect to the first PC. b) Distribution of each repetition in the reduced vectorial space when representing the third Principal Component (PC) with respect to the first PC. c) Cumulative Explained Variance of the dataset (in %) with respect to the considered PCs and Explained Variance distribution over the set of PCs considered. d) 3D distribution of each repetition in the reduced vectorial space.

performance of the obtained model. Figure 7 (a) shows the evolution of the Root Mean Square Error (RMSE) with the number of components of the PLSR considered during the LOGO-CV scheme iteration.

It is observed that with 6 Latent Variables (LVs), the MSE of the best of the 5 models coming out from the LOGO-CV scheme is minimized up to ~ 6.8 which leads to a Root Mean Square Error in Cross Validation (RMSECV) of $\sim 3.62\%$.

Also, the most important features in the training dataset were observed by using the Variable Importance in Projection (VIP) scores per each of the features [39]. The calculated scores provide information of the averaged importance of each feature in the projection of the repetition in the restricted vectorial space that handles the PLSR. This information can be used to trim the feature matrix columns to the most relevant part of the $20\log\,(|\dot{S}_{21}|)$. There are many ways of defining the threshold that delimits which feature is more important than the rest [15]. However, a good approach is to consider that every mean VIP score higher than 1 belongs to a relevant feature.

Figure 7 (b) shows the VIP scores for the training dataset once the PLSR is optimized to 6 LVs. In this figure, pink areas (▭) show scores > 1. It is observed that most important parts of the $20\log\,(|\dot{S}_{21}|)$ correspond to data close to 2.1 GHz, 2.45 GHz and 2.5 GHz. These frequency spans correspond to the most relevant differences between concentrations observed in figure 4.

1) Performance in Cross Validation: As mentioned in section III-D, the LOGO-CV training scheme produces 5 PLSR models per LVs considered during the training. From these 5 models, the one with lower Root Mean Square Error (RMSE) is selected as representative of the LV set considered.

According with figure 7 (a), the PLSR model with 6 LVs provides better performance during the LOGO-CV scheme. Therefore, this model was used to produce predictions on

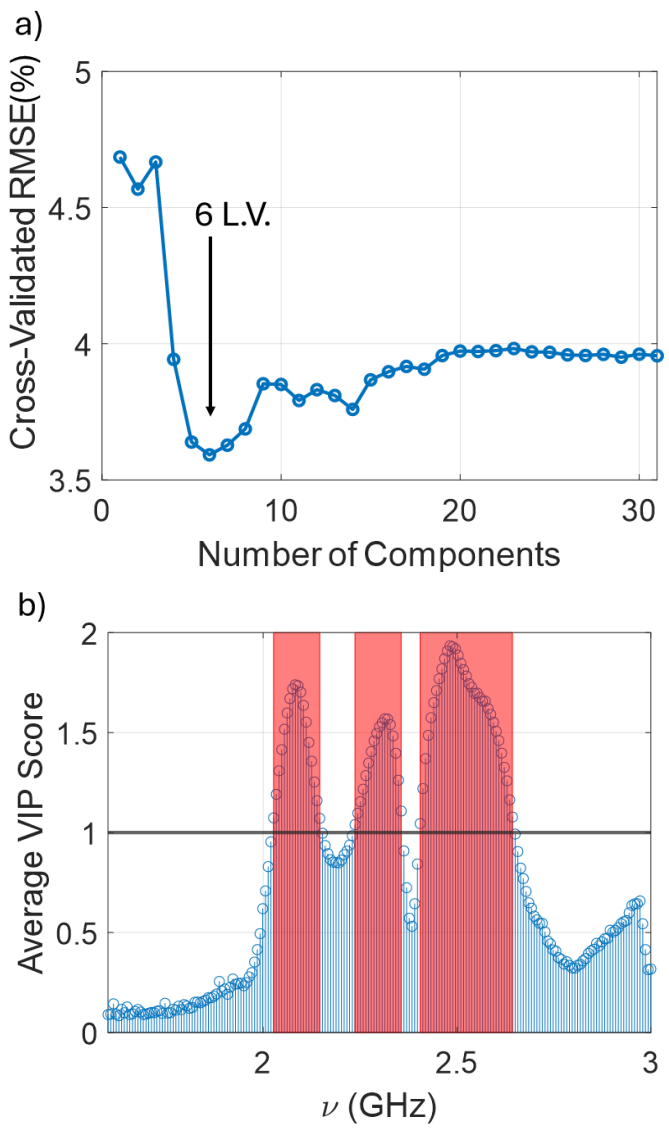


Fig. 7. Statistics of the Partial Least Squares Regressor (PLSR) trained with the generated training dataset. a) Evolution of the Mean Square Error (MSE) with the number of Latent Variables (LV) during the Leave-One-Block-Out Cross Validation (LOGO-CV) training scheme. b) Mean Variable Importance in Projection (VIP) scores for each feature from **1.6** GHz to **3** GHz. The scores higher than **1** have been highlighted with ■.

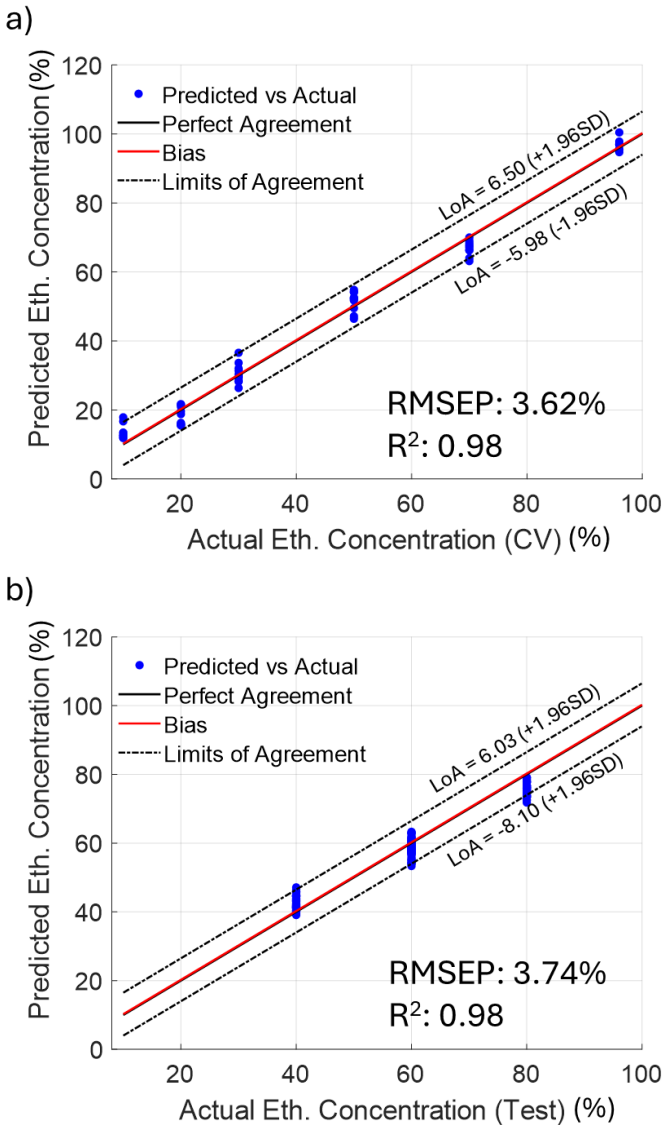


Fig. 8. Statistics of the Partial Least Squares Regressor (PLSR) trained with the generated training dataset. a) Evolution of the Mean Square Error (MSE) with the number of Latent Variables (LV) during the Leave-One-Block-Out Cross Validation (LOGO-CV) training scheme. b) Mean Variable Importance in Projection (VIP) scores for each feature from **1.6** GHz to **3** GHz. The scores higher than 1 have been highlighted with ■.

the training dataset. In this process, repetitions for 20%, 30%, 50%, 70% and 96% samples were part of the training dataset and the 20% of these repetitions were kept aside for performance evaluation.

Figure 8 (a) shows the predictions of the PLSR model for the repetitions of the training dataset kept aside. The RMSE in CV (RMSECV) got close to 3.62 % with R^2 coefficient of 0.98. The Limits of Agreement (LoA) were calculated using Blant-Altmand method for the 95% of Confident Interval resulting in 13.5% deviation from linear prediction.

2) Performance in Test: Despite the test during training is performed with repetitions that the LOGO-CV scheme had no access to, it is interesting to test if the PLSR model is able to predict concentrations that the training process have not seen before.

In section II-D, the database was divided into training dataset and test dataset. The test dataset contained the repetitions of 40%, 60% and 80% samples. This dataset was used to obtain predictions with the PLSR model and evaluate its performance against "blind" samples.

Figure 8 (b) shows the predictions obtained for the test dataset. These predictions produced a RMSE in Prediction (RMSEP) about 3.7 %, very close to the RMSECV, with a R^2 coefficient of 0.98. The LoAs were calculated in the same way presented in previous section, resulting in a LoA $\sim$ 14.2% with a 95% of Confident Interval.

The PLSR models showed a 1.5% bias when predicting the concentrations of the test dataset and the LoAs increased also $\sim$ 1% with respect to the LoAs obtained in training.

These performance was a bit worse than the performance

shown during training, but expected. The model was never put in contact with the data in the test dataset, therefore the variability of these samples (with their repetitions) might be different than the represented by the training dataset.

IV. CSRR System Evaluation

In this contribution, the impact of using Machine Learning Algorithms when measuring liquids with a CSRR based bench-top system is demonstrated. If properly applied, ML algorithms not only show the properties of the generated databases, but enhance the performance of low-cost bench top systems predicting concentrations of SUTs in samples.

For a bench-top system like the presented in section II the measurement procedure and the commercial vials used regularly in laboratories, introduce a considerable variability across repetitions. In the literature, this variability is managed, either enhancing the hardware of the system, which might be impossible in many of the cases due to the increase in hardware cost, or designing ad-hoc solutions specifically designed for reducing variability, as it happens with the vials used in the experiments [19], [36].

However, it has been demonstrated in section III-B that repeating measurements of samples, even considering different vials, results in datasets with large standard deviations and it is expected that a simple curve fitting will fail in predicting the concentration of the SUT. In particular, the presented system, measurement procedure and commercial vials resulted in a curve-fitting based model with an RMSE of $7.62\ \%$. These result shows that an univariant analysis of the S_{21} parameter of the CSRR sensor is not enough for predicting the concentration of the SUT in these type of systems. If the system is used for biomedical applications in laboratories working as benchtop device a multivariate analysis is needed.

As demonstrated in section III-D, the application of ML algorithms might mitigate the observed lack of accuracy when using univariate models. Furthermore, an oriented and extensive sampling (455 repetitions), that provides useful datasets (see section III-A), and a previous exploration of the obtained datasets, as explained in section III-C, lead to a better understanding of the behaviour of CSRR bench-top systems. It also orientate to the proper prediction model useful for the application and the final use.

In particular, for the benchmark problem of ethanol concentration diluted in clean water, the PCA shows the variability of the dataset concentrated around 4 Principal Components (PCs), as it is shown in figure 6. The first PC is clearly related to the ethanol concentration in the solution, while the rest of the PCs might be related to the variability introduced by the commercial vials and the measurement procedure. Furthermore, the training evolution of the PLSR model showed that 6 LVs minimized the RMSECV and the VIP scores analysis confirmed the multivariate nature of the problem.

In fact, the trained PLSR model trained with the workflow presented in section II-D, shows a better performance than the univariate model. In particular, a reduction by a factor of ~ 2 in the RMSE at test (RMSEP $= 3.74\ \%$). Also, the LoAs are stable over the range of concentrations considered in the test dataset and very similar at training and test. This behaviour in training and test shows the robustness of the PLSR model compared with the quantification models presented in the literature.

V. Conclusions

AS a result of the presented study, it has been demonstrated the enhancement of low-cost, benchtop, ready ot use in laboratories device performance by using Machine Learning algorithms, techniques and workflows. This demonstration was performed over a representative example using the benchmark problem of quantifying ethanol concentration diluted in clean water. The results show that and extensive sampling campaign and the use of ML algorithms, in particular PCA and PLSR, can improve the performance of traditional curve fitting by a factor of 2. These result overcome the limitations of the univariate analysis of the S_{21} parameter of the CSRR-based system when working in real scenarios as biomedical laboratories. This opens the opportunity to use CSRR-based systems in biomedical applications as benchtop devices and move forward their development as usefull tools in screening, diagnosis or monitoring scenarios.

Acknowledgment

Author want to thank Professor Antonio Pardo and Gema Guedes, Ph.D. for their collaboration during experiments in the laboratory and their understanding on the increasing entropy inside it.

References

[1] F. Falcone, T. Lopetegi, J. Baena, R. Marques, F. Martin, and M. Sorolla, "Effective negative-/spl epsiv/ stopband microstrip lines based on complementary split ring resonators," *IEEE Microwave and Wireless Components Letters*, vol. 14, no. 6, pp. 280–282, 2004.

[2] J. Baena, J. Bonache, F. Martin, R. Sillero, F. Falcone, T. Lopetegi, M. Laso, J. Garcia-Garcia, I. Gil, M. Portillo, and M. Sorolla, "Equivalent-circuit models for split-ring resonators and complementary split-ring resonators coupled to planar transmission lines," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 4, pp. 1451–1461, 2005.

[3] J. Garcia-Garcia, F. Martin, F. Falcone, J. Bonache, J. Baena, I. Gil, E. Amat, T. Lopetegi, M. Laso, J. Iturmendi, M. Sorolla, and R. Marques, "Microwave filters with improved stopband based on sub-wavelength resonators," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 6, pp. 1997–2006, 2005.

[4] J. Bonache, I. Gil, J. Garcia-Garcia, and F. Martin, "Novel microstrip bandpass filters based on complementary split-ring resonators," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 1, pp. 265–271, 2006.

[5] M. Mandal, P. Mondal, S. Sanyal, and A. Chakrabarty, "Low insertion-loss, sharp-rejection and compact microstrip low-pass filters," *IEEE Microwave and Wireless Components Letters*, vol. 16, no. 11, pp. 600–602, 2006.

[6] M. Gil, J. Bonache, J. Selga, J. Garcia-Garcia, and F. Martin, "Broadband resonant-type metamaterial transmission lines," *IEEE Microwave and Wireless Components Letters*, vol. 17, no. 2, pp. 97–99, 2007.

[7] A. Velez, J. Bonache, and F. Martin, "Varactor-loaded complementary split ring resonators (vlcsrr) and their application to tunable metamaterial transmission lines," *IEEE Microwave and Wireless Components Letters*, vol. 18, no. 1, pp. 28–30, 2008.

[8] H. Zhang, Y.-Q. Li, X. Chen, Y.-Q. Fu, and N.-C. Yuan, "Design of circular/dual-frequency linear polarization antennas based on the anisotropic complementary split ring resonator," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 10, pp. 3352–3355, 2009.

[9] T. Grzegorczyk, C. Moss, J. Lu, X. Chen, J. Pacheco, and J. A. Kong, "Properties of left-handed metamaterials: transmission, backward phase, negative refraction, and focusing," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 9, pp. 2956–2967, 2005.

[10] I. Stevanovic, P. Crespo-Valero, K. Blagovic, F. Bongard, and J. Mosig, "Integral-equation analysis of 3-d metallic objects arranged in 2-d lattices using the ewald transformation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 10, pp. 3688–3697, 2006.

[11] M. S. Boybay and O. M. Ramahi, "Material characterization using complementary split-ring resonators," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 11, pp. 3039–3046, 2012.

[12] K. Song and P. Mazumder, "Design of highly selective metamaterials for sensing platforms," *IEEE Sensors Journal*, vol. 13, no. 9, pp. 3377–3385, 2013.

[13] C. S. Lee and C. L. Yang, "Thickness and permittivity measurement in multi-layered dielectric structures using complementary split-ring resonators," *IEEE Sensors Journal*, vol. 14, no. 3, pp. 695–700, 2014.

[14] ——, "Complementary split-ring resonators for measuring dielectric constants and loss tangents," *IEEE Microwave and Wireless Components Letters*, vol. 24, no. 8, pp. 563–565, 2014.

[15] M. A. H. Ansari, A. K. Jha, and M. J. Akhtar, "Design and application of the csrr-based planar sensor for noninvasive measurement of complex permittivity," *IEEE Sensors Journal*, vol. 15, no. 12, pp. 7181–7189, 2015.

[16] A. Standaert, M. Rousstia, S. Sinaga, and P. Reynaert, "Permittivity measurements in millimeter range of ptfe foams," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 8, pp. 766–768, 2017.

[17] L. Su, J. Mata-Contreras, P. Vélez, and F. Martín, "Splitter/combiner microstrip sections loaded with pairs of complementary split ring resonators (csrrs): Modeling and optimization for differential sensing applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 12, pp. 4362–4370, 2016.

[18] P. Vélez, K. Grenier, J. Mata-Contreras, D. Dubuc, and F. Martín, "Highly-sensitive microwave sensors based on open complementary split ring resonators (ocsrrs) for dielectric characterization and solute concentration measurement in liquids," *IEEE Access*, vol. 6, pp. 48 324–48 338, 2018.

[19] A. E. Omer, G. Shaker, S. Safavi-Naeini, K. Ngo, R. M. Shubair, G. Alquié, F. Deshours, and H. Kokabi, "Multiple-cell microfluidic dielectric resonator for liquid sensing applications," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6094–6104, 2021.

[20] K. Zhang, R. K. Amineh, Z. Dong, and D. Nadler, "Microwave sensing of water quality," *IEEE Access*, vol. 7, pp. 69 481–69 493, 2019.

[21] M. Martinic, M. Mertens, D. Schreurs, B. Nauwelaers, G. A. E. Vandenbosch, and T. Markovic, "Highly sensitive impedance-matched microwave dielectric sensor for glucose concentration measurements," *IEEE Sensors Journal*, pp. 1–1, 2025.

[22] S. Patel, S. Das, D. Mitra, S. Sarkar, and C. Koley, "Biological liquid monitoring using microwave resonator," in *2022 URSI Regional Conference on Radio Science (USRI-RCRS)*, 2022, pp. 1–7.

[23] S. Jiang, G. Liu, M. Wang, Y. Wu, and J. Zhou, "Design of high-sensitivity microfluidic sensor based on csrr with interdigital structure," *IEEE Sensors Journal*, vol. 23, no. 16, pp. 17 901–17 909, 2023.

[24] S. Liu, P. Cheong, C. Yang, Y. Ye, and W.-W. Choi, "Microfluidic sensor for simultaneous liquid classification and concentration detection," *IEEE Microwave and Wireless Technology Letters*, vol. 34, no. 3, pp. 358–361, 2024.

[25] Q. Zhang, Z. Kou, J. Yang, Y. Li, Z. Yin, and G. Deng, "Highly sensitive and robust metasurface-inspired microfluidic sensor for oil detection," *IEEE Sensors Journal*, vol. 24, no. 24, pp. 40 847–40 854, 2024.

[26] D. Prakash and N. Gupta, "High-sensitivity grooved csrr-based sensor for liquid chemical characterization," *IEEE Sensors Journal*, vol. 22, no. 19, pp. 18 463–18 470, 2022.

[27] L. Harrsion, M. Ravan, D. Tandel, K. Zhang, T. Patel, and R. K. Amineh, "Material identification using a microwave sensor array and machine learning," *Electronics*, vol. 9, no. 2, 2020.

[28] N. Kazemi and P. Musilek, "Enhancing microwave sensor performance with ultrahigh q features using cyclegan," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 12, pp. 5369–5382, 2022.

[29] M. Abdolrazzaghi, N. Kazemi, V. Nayyeri, and F. Martin, "Ai-assisted ultra-high-sensitivity/resolution active-coupled csrr-based sensor with embedded selectivity," *Sensors*, vol. 23, no. 13, 2023.

[30] S. Trovarello, O. Afif, A. Di Florio Di Renzo, D. Masotti, M. Tartagni, and A. Costanzo, "A non-invasive, machine learning assisted skin-hydration microwave sensor," in *2024 54th European Microwave Conference (EuMC)*, 2024, pp. 932–935.

[31] T. M. Mitchell, *Machine learning*. McGraw-hill New York, 1997, vol. 1, no. 9.

[32] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987, proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

[33] S. Wold, M. Sjöström, and L. Eriksson, "Pls-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001, pLS Methods.

[34] I. Nirmal, A. Khamis, M. Hassan, W. Hu, and X. Zhu, "Deep learning for radio-based human sensing: Recent advances and future directions," *IEEE Communications Surveys Tutorials*, vol. 23, no. 2, pp. 995–1019, 2021.

[35] P. Loutchanwoot and S. Harnsoongnoen, "Microwave microfluidic sensor for detection of high equol concentrations in aqueous solution," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 2, pp. 244–251, 2022.

[36] A. E. Omer, G. Shaker, and S. Safavi-Naeini, "Portable radar-driven microwave sensor for intermittent glucose levels monitoring," *IEEE Sensors Letters*, vol. 4, no. 5, pp. 1–4, 2020.

[37] C. Wu and M. E. Thompson, *Sampling theory and practice*, 2020th ed., ser. ICSA Book Series in Statistics. Cham, Switzerland: Springer Nature, may 2020.

[38] P. Filzmoser, B. Liebmann, and K. Varmuza, "Repeated double cross validation," *Journal of Chemometrics*, vol. 23, no. 4, pp. 160–171, 2009.

[39] I. Chong and J. C.H., "Performance of some variable selection methods when multicollinearity is present," *Chemometrics and Intelligent Laboratory Systems*, vol. 78, no. 1, pp. 103–112, 2005.

Javier Alonso-Valdesueiro was born in Madrid, Spain, in 1980. He earned a Telecommunications Engineering degree from the University of Alcalá, Madrid, in 2006, and a Ph.D. from the Polytechnic University of Catalonia, Barcelona, in 2011. From 2012 to 2020, he worked on developing electronic instrumentation for NMR experiments, initially at the C.E.A Center in France, followed by the University of Southampton in the UK, and later as a Marie Skłodowska-Curie Fellow at the University of the Basque Country (UPV/EHU). Over the past four years, he has advanced his career as an RF and instrumentation engineer in both private companies and public research centers, including TECNALIA Innovation Foundation and the Institute of Biomedical Engineering of Catalonia. Recently, he was appointed Assistant Professor in the Department of Electronic and Biomedical Engineering at the University of Barcelona. His current research focuses on radio-frequency (RF) devices for MRI, instrumentation electronics for gas sensing applications, and advancements in RF sensor technologies.