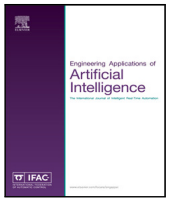
Engineering Applications of Artificial Intelligence 127 (2024) 107246

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Machine learning and data augmentation methods for multispectral capacitance images of nanoparticles with nanoelectrodes array biosensors

Federico Lombardo [a,b], Federico Pittino [a], Daniele Goldoni [b], Luca Selmi [b,*]

[a] Silicon Austria Labs GmbH, Europastraße 12, Villach 9524, Austria
[b] Department of Engineering "Enzo Ferrari", Universitá degli Studi di Modena e Reggio Emilia, Via Vivarelli 10, 41125, Modena, Italy

ARTICLE INFO

Keywords:
Data augmentation
Efficient nano-particles metrology
Impedance spectroscopy numerical simulations
Machine learning
Nano-scale size estimation
Nano-capacitors biosensor

ABSTRACT

A large number of technology applications still remain where Artificial Intelligence techniques, carefully tailored to the specific application needs, could provide performance benefits to hardware technologies. One such area is biosensing with innovative complementary-metal–oxide–semiconductor nanocapacitor arrays. These sensors operate as powerful imaging platforms but, despite the advancements in the field, the knowledge necessary for precise and robust interpretation of their response to analytes is still largely lacking.

In this work, we leverage the ability of Machine Learning methods for computer vision to construct precise and robust models in different operation scenarios. By recognizing the similarity between multifrequency capacitance maps and multispectral images, we identified optimal Machine Learning algorithms to accurately estimate the size of analytes measured by the nanoelectrode array biosensor.

As a relevant case study, we focus on measurements of the radius of dielectric spherical nano-particles dispersed in deionized water and phosphate buffer saline. The performance of large, established image-processing neural networks is compared to that of less complex, purposely developed ones. Sizable training data sets are generated by accurate finite element simulations of the sensor response combined with measured data. An excellent accuracy, comparable to traditional sizing technology, is achieved for the task of providing a quantitative measure of the nano-particle radius when the latter is comparable to the pitch of the pixels in the array. We report a size median error below 15% in all scenarios when a few percent of measured data samples is added to the simulation-based training data set.

1. Introduction

Machine learning (ML) and Deep learning (DL) techniques have enabled tremendous progress in several domains, such as computer vision, speech recognition, natural language processing, and generative models (Dong et al., 2021; Otter et al., 2020; Croitoru et al., 2023). In some other fields, however, their adoption is still in its infancy, while showing tremendous potential. One such field is that of biosensors capable of accurate and reliable measurements on biological objects dispersed in liquid solutions, ranging from small biomolecules to cells. So far, only limited applications of ML to biosensing have been demonstrated (see for instance Cui et al., 2020; Meiler et al., 2023; Oh et al., 2021; Massarelli et al., 2021; Lorenzo-Navarro et al., 2020), and in most cases, it is difficult to obtain large and reliable training data sets covering with fine granularity the entire operational space of the sensors, as required to achieve high accuracy and robustness.

Empowerment of conventional biosensors with ML-based data analysis appears as a promising strategy to advance measurement accuracy and reliability of data interpretation (Ansah et al., 2023; Gomes et al., 2021; Pennacchio et al., 2022; Hamedi et al., 2023). Measurements, however, are affected by variability, uncertainty, drifts, and noise, which hamper accuracy, and can be difficult and/or expensive. Accurate physics-based simulations, which could enable the derivation of larger and cleaner data sets, are not yet established, nor of widespread usage. Moreover, most biosensing platforms have small parallelism, which prevents to obtain large data sets for several different analytes.

In recent years, several complementary-metal–oxide–semiconductor (CMOS) micro/nano-electrode array systems have been proposed that overcome one or more of the limitations above (Abbott et al., 2022; Hu et al., 2021; Jung et al., 2021; Lai et al., 2023b; Senevirathna et al., 2019; Lai et al., 2023a). Among these, high-frequency impedance spectroscopy nano-electrode (NE) array biosensors as the one in (Widdershoven et al., 2010; Laborde et al., 2015) offer numerous advantages, such as miniaturization, massive parallelization, mitigation of

* Corresponding author.
E-mail address: luca.selmi@unimore.it (L. Selmi).

https://doi.org/10.1016/j.engappai.2023.107246
Received 18 July 2023; Received in revised form 26 September 2023; Accepted 1 October 2023
Available online 11 October 2023

the spatial sensitivity limits imposed by Debye screening, and label-free detection of nano-scale objects (hereafter defined as having dimensions $\leq$ 1 μm), and it is therefore used in this work. The potentiality of this platform has been proven with measurements on micro-particles (Stadlbauer et al., 2019; Widdershoven et al., 2018), nano-particles (Goldoni et al., 2023), oil droplets (Renault et al., 2021), cells (Laborde et al., 2015), and recently for continuous water monitoring of nano-pollutants (Goldoni et al., 2023).

Besides these examples, the detection and quantification of nanoscale objects are also of great interest to environmental and health studies. In fact, nanopollutants and contaminants in soil and water pose serious and not yet fully understood health hazards on animals and plants (Jan et al., 2022; Sangkham et al., 2022; Lee and Chae, 2021). In this regard, state-of-the-art technologies for monitoring and sizing nano-particles are mainly based on optical principles and suffer several limitations in terms of costs, complexity, and detection of nanoscale particles (Lee and Chae, 2021; Enfrin et al., 2020). For these reasons, several techniques have been recently proposed for micro and nano-particles detection and sizing, also integrating machine learning methods. As an example, electrochemical impedance spectroscopy enhanced by machine learning was used to detect and size large micro-plastics (2–4 mm) (Meiler et al., 2023); microwave block lop-gap resonator and support vector machine (SVM) were proposed for detection (not sizing) of metallic pollutants in water (Oh et al., 2021). Additionally, computer vision techniques (RGB images) have been reported for plastic debris counting and classification (Massarelli et al., 2021; Lorenzo-Navarro et al., 2020). However, sizing and morphology measurements were performed after the sampling procedure for water analysis, which is complex and time-consuming (Lee and Chae, 2021) and the plastic debris analyzed had a large diameter (1–5 mm) (Massarelli et al., 2021; Lorenzo-Navarro et al., 2020).

Thanks to the huge throughput, massively parallel CMOS nanobiosensors are perfect companions of ML algorithms. In addition, these biosensors share an architectural similarity with conventional camera sensors, such that their output resembles a multispectral image. This, in turn, implies that state-of-the-art computer vision DL methods, modified for multispectral images (as, for instance, in Jameel et al., 2020; Khan et al., 2022), are promising candidates for the processing and interpretation of the sensor's output. For these reasons, building upon the largely unexplored potentialities of the NE array platforms (Widdershoven et al., 2018), and on accurate simulations of the NE array response to analytes (Pittino and Selmi, 2014; Cossettini et al., 2021), we investigate if this combination enables the derivation of ML/DL models trained on simulations and/or measurements, for accurate estimation of key features of nano-particles (NPs) dispersed in aqueous solution. As a challenging demonstrator, we target the empowerment of nanoelectrode array biosensors with particles' radii estimation algorithms; however, the approach may be extended to estimate additional features, e.g. shape, dielectric constant or position w.r.t. the electrodes. We also evaluate the Neural Network (NN) complexity requirements in terms of the number of parameters, in order to address low-memory and low-power requirements for edge devices sensing applications.

To the best of our knowledge, this work reports for the first time the application of multiple ML algorithms (originally derived for computer vision of RGB images), to multifrequency capacitance images of nanoscale objects in water solution and for metrology purposes.

The manuscript is organized as follows. Section 2 describes the settings of both measurements (Section 2.1) and simulations (Section 2.2) of the dielectric beads on the nanoelectrode array. Moreover, Section 2.3 discusses the algorithms developed to estimate the particle radius. Subsequently, Section 3 shows and discusses the results, both in deionized water (MilliQ) and phosphate buffer saline (PBS) solutions. Finally, Section 4 draws the conclusions.

2. Materials and methods

2.1. Measurements: experimental setup, procedure and data preparation

In this work we use the High Frequency Impedance Spectroscopy (HFIS) nano-electrode (NE) array chip described in Widdershoven et al. (2018), which was designed and provided by NXP Semiconductors. For the sake of a self-contained manuscript, we report in the following its key features. The CMOS chip embeds an array of 256 × 256 individually addressable NEs (pixels) with 90 nm radius, 600 nm × 720 nm pitch, fabricated with 90 nm CMOS technology. It employs Charged-Based Capacitance Measurement (CBCM) and *ad hoc* calibration procedures to deliver spectrally resolved measurement of each electrode's capacitance in the 1–70 MHz frequency range (Widdershoven et al., 2010), with potential extension up to 500 MHz (Cossettini et al., 2020). A capacitance image is constructed sequentially activating one row of electrodes (working electrodes, WEs) while all other rows are grounded to create a large counter electrode (CE). The capacitance change caused by analyte(s) is measured as the difference between the mean capacitance (100 values) after ($C_{mean,w/analyte}$) and before ($C_{mean,\ w/o\ analyte}$) the analyte arrival event:

$$\Delta C = C_{mean,\ w/\ analyte} - C_{mean,\ w/o\ analyte} \tag{1}$$

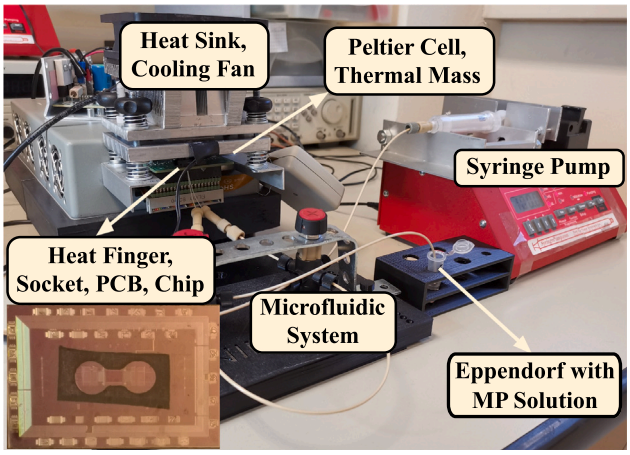
ΔC is due to the perturbation of the electric field caused by the analyte and related to its conductivity, dielectric permittivity, dimensions, and position with respect to the electrodes (Cossettini et al., 2019; Pittino et al., 2015; Widdershoven et al., 2018). The HFIS NE array chip is integrated in an experimental setup (Fig. 1(a)) that implements the chip initialization and communication, the temperature control, and the microfluidic system (Widdershoven et al., 2018). An example of analyte arrival event (nano-particle with radius 500 nm in MilliQ) is shown in Fig. 1(b).

Polystyrene (PS) nano-particles (NPs) with nominal radius (mean ± std) of (275 ± 8) nm and (500 ± 5) nm suspended in 2.74%, and 2.66% (w/v) 15 mL of H_2O, respectively, were used in this work. The NPs were analyzed in MilliQ and PBS. Each solvent was filtered with 0.2 μm pores syringe filter to remove possible impurities; then each NP solution was diluted (≈ 1:40), filtered across 3 μm pores membrane, and sonicated.

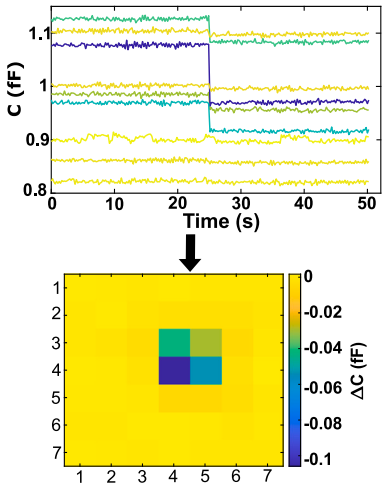
Before proceeding with the measurement with the NE array chip, the NPs solution is measured with Dynamic Light Scattering (DLS), which is a common sizing technology for nanoparticles (Jia et al., 2023). The DLS system used in our work employs a 633 nm HeNe Laser coupled to an optical fiber as a light source, and a single-photon counting module by PerkinElmer as a light detector. The setup was mounted on an optical table and equipped with a PC with Bi-9000AT Digital Autocorrelator by Brookhaven Instruments for data processing. The results with DLS are (mean value ± standard deviation): 278.3 ± 65 nm and 500.4 ± 68 nm for 275 and 500 nm nominal values respectively.

Once the NP dispersion was prepared, the sensor platform and the temperature controller were turned on, the measurement conditions initialized, and the system was allowed to reach thermal equilibrium (temperature set to 22 °C) before starting the measurements. The system was then flushed with ≈ 3 mL isopropyl alcohol (IPA, filtered with 0.2 μm pores syringe filter), at a flow rate of 100 μL/min. Afterward, the NPs were introduced into the microfluidics. At the end of the measurements, the NE array and the microfluidics system were flushed with ≈ 3 mL of IPA at a flow rate of 200 μL/min for cleaning.

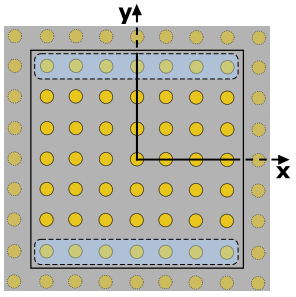
Concerning data preparation, the raw-data were elaborated according to (Widdershoven et al., 2018) and to Eq. (1). Additionally, a pre-processing of the measurements was necessary to exclude outliers and to match the data to the simulations. The first step has been the identification of 7 × 7 electrodes ΔC maps, hereafter denoted (data) samples, in which only a single bead (no clusters) is present around



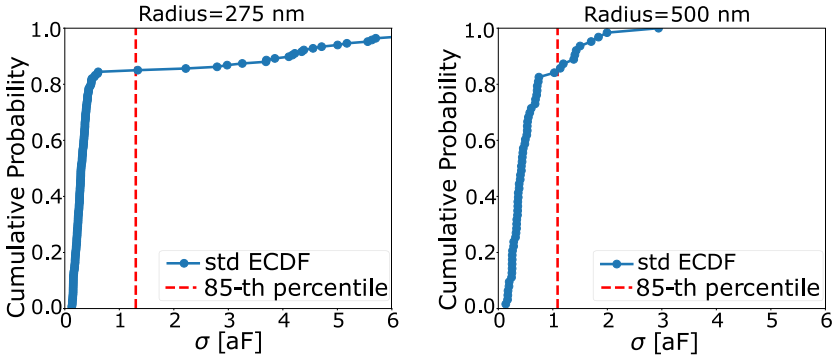
(a) The experimental setup.



(b) Example of analyte arrival event and 7×7 ΔC image.



(c) 7×7 electrode matrix highlighting (in light-blue) electrodes used to estimate noise.



(d) ECDF of the standard deviation values of all measured samples.

Fig. 1. (a) picture of the experimental setup and magnification of the chip with the PDMS seal ring around the NE array on top. 1(b) example of capacitance traces upon arrival on electrodes of a 500 nm radius NP in MilliQ, and corresponding 7 × 7 ΔC map, as per Eq. (1). 1(c)) 7 × 7 electrode matrix highlighting electrodes used to estimate noise. 1(d)) Empirical Cumulative Distribution Functions (ECDF) of the ΔC standard deviation of all the samples in the measurement dataset. Radius=275 nm (left) and 500 nm (right) in MilliQ environment. All samples above the 85th percentile vertical lines are removed from the dataset.

the central electrode. These maps are complete of all 49 capacitance values and exclude outliers.

Moreover, since measurements are intrinsically noisy (unlike simulations), and outliers may fool the training algorithm reducing the accuracy, training the prediction model on ideal simulations only is not recommended for robustness against real-world fluctuations.

To mitigate this issue:

1. for each measurement data sample i, we extracted the mean (μ_i) and the standard deviation (σ_i) of the ΔC of the outermost electrodes (see light blue sets in Fig. 1(c)), where the impact of the NP is negligible;
2. we removed from the datasets outlier samples having a standard deviation higher than the 85th percentile of the σ_i Empirical Cumulative Distribution Function (ECDF) of the measurement samples, see Fig. 1(d);
3. we subtracted μ_i to all the elements of the ΔC matrix of sample i;
4. we computed the median $\hat{\sigma}$ of the remaining σ_i, and used it to generate a zero-mean Gaussian noise to perturb the simulated ΔC_{eff} at each electrode of the simulated samples.

This procedure removes from measurements possible residual offsets deriving from drifts and supports the generation of realistic (hence, noisy) simulated samples, thus enabling direct comparison of measurements and simulations.

This procedure yielded 194 data samples for the MilliQ environment (141 and 53 for the 275 and 500 nm radii, respectively) and 169 data samples for the PBS environment (116 and 53 for the 275 and 500 nm radii, respectively), excluding the outliers as described below.

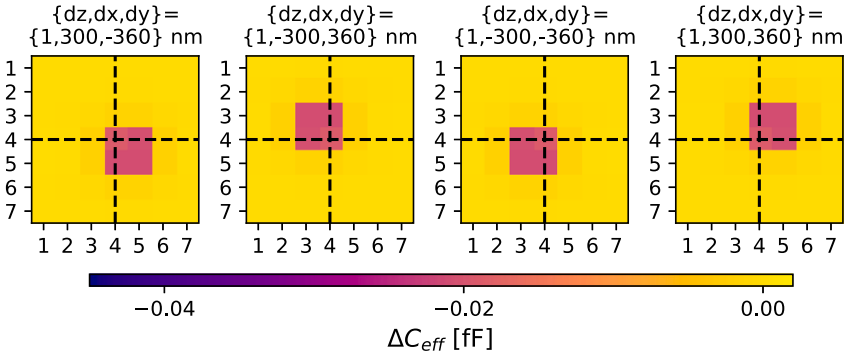
2.2. Numerical simulations: calculations and data preparation

The response of the nanoelectrode array described in Section 2.1 has been simulated with the ENBIOS software (Pittino and Selmi, 2014), which is capable to estimate the ΔC with good accuracy, as previously demonstrated in (Laborde et al., 2015; Widdershoven et al., 2018; Cossettini et al., 2021). The simulation model solves the so called Poisson–Boltzmann and Poisson–Drift–Diffusion set of equations (in DC and AC, respectively) to calculate the change in electrode capacitance due to the nanoparticle, i.e., the same quantity that is also measured by the sensor. Similar results could be achieved with general-purpose multiscale-multiphysics simulation platforms, e.g. COMSOL (COMSOL Inc., 2022). The simulation domain comprises a matrix of 13 × 17 electrodes around the particle to accurately represent the electric field distribution at the innermost electrodes. Simulations are carried out at the same frequencies as in the experiments, i.e.:

- a unique $f = 50$ MHz for the MilliQ case. Since it is essentially deionized, similarly to air and IPA, the NE response is independent of frequency (Widdershoven et al., 2018; Cossettini et al., 2020).

Key	Description	Value
dx	In plane displacement on x of the particle with respect to the central electrode [nm]	{0, 75, 100, 150, 200, 250, 300}
dy	In plane displacement on y of the particle with respect to the central electrode [nm]	{0, 75, 90, 180, 240, 300, 360}
dz	Vertical distance between the plane of the sensor and the bottom of the particle [nm]	{1, 5, 10, 20, 50, 100}
f	Frequencies at which simulation is performed. [Hz]	50 MHz (MilliQ), 2-70 MHz log-spaced (PBS)
R_i	NP radius [nm]	{200, 275, 400, 500, 750, 1000, 1500}
Solution	Simulation Environment	MilliQ or PBS
ΔC_{eff}	$7 \times 7 \times 1$ or $7 \times 7 \times 10$ matrices for MilliQ and PBS respectively of ΔC_{eff} [fF]	

(a) Summary of the simulations parameters



(b) Sample augmentation strategy by flipping one simulationalong x and y

Fig. 2. Parameters used for the simulations 2(a) and schematic representation of how the dataset has been augmented by flipping each simulated image along the x and y axes (example with 500 nm radius nano-particle, d_z=1 nm, in MilliQ) 2(b).

- ten equally log-spaced frequencies in the range 2–70 MHz for the PBS case.

Two cases are simulated at each frequency: (i) the reference condition without NP, and (ii) the detection condition, where the particle is placed at more than 100 locations above and around the central electrode. In the construction of the simulated images, only data from the innermost 7×7 electrodes out of the 13×17 ones in the domain is considered for improved stability and accuracy. The simulated response is computed according to (Laborde et al., 2015) as $\Delta C_{eff} = |\Delta Y_A|/(2\pi f)$, where ΔY_A is the AC admittance change upon particle arrival; this expression is a good approximation of the variation of the *measured* capacitance. Throughout this work, we will refer to the ΔC_{eff} matrix as a *sample*. Taking advantage of the array symmetry, only particle displacements in the positive x and y direction are considered. Additional capacitance images are easily constructed with a simple data augmentation technique; namely: by flipping the image along the x and y axis (see Fig. 2(b)). These data augmentation transformations are necessary to render the model insensitive to the exact position of the particle in the whole horizontal plane around the electrode. The radius of the particle is also varied, in the range [200, 1500] nm. A summary of the simulations parameter is shown in Fig. 2(a). The number of simulated samples in the training data set are 1065 and 770; the total including test and validation sets is 1578 and 1133, for MilliQ and PBS, respectively.

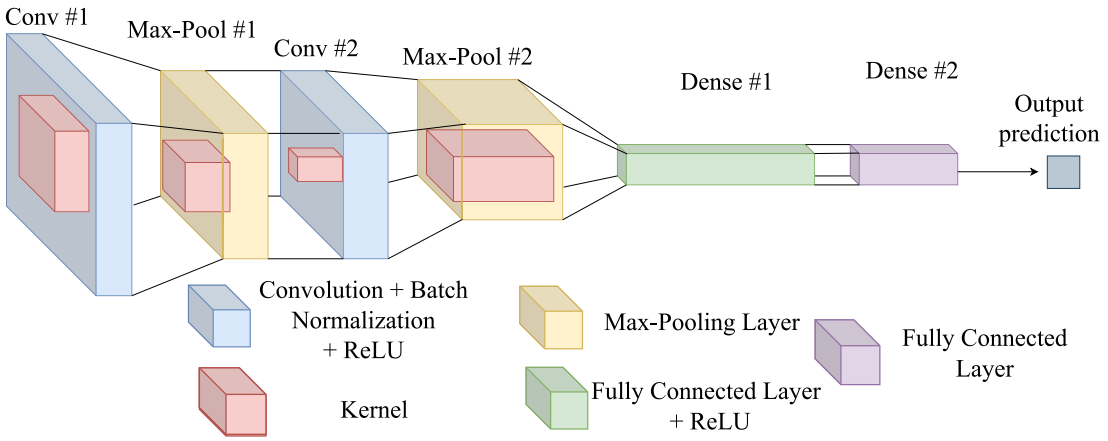
All the data sets (measured and simulated data samples) that are used in this work are available online (Lombardo et al., 2023).

2.3. Radius estimation model development

The estimation of the particles radii takes advantage of the structural similarity of the multifrequency ΔC maps to common (RGB) images, thereby employing successful machine learning (ML) algorithms developed for Computer Vision applications. In particular, the following ML model architectures have been investigated:

- popular Deep Learning models for image recognition available in the PyTorch library (e.g., MobileNet-V3Small, ResNet18) (Paszke et al., 2019), and their adaptations to the problem at hand.
- Two simple Convolutional Neural Networks (CNN) specifically designed for this work (see below), where the layers are the ones originally proposed in (LeCun et al., 1995) as the foundations of all modern CNNs.
- The popular distributed gradient-boosted decision tree XGBoost (Chen and Guestrin, 2016).

Since the number of training samples is limited in our application, and so are the memory and computational resources within ultra-low power edge devices, we focused our attention on small size models to avoid overfitting the training data. MobileNet-V3 Small (Howard et al., 2019) has one of the smallest number of parameters ($\approx 1 \times 10^6$) among the Neural Networks (NNs) in the PyTorch Library, since it targets embedded applications. Nevertheless, it was still necessary to further reduce the model's dimension, by creating two variants: (I) the first one denoted Tiny Mobile Net (also shortly referred as Tiny MoNet in the following) ($\approx 1.5 \times 10^4$ parameters), where eight of the last



Solution	Net	Conv #1		Max-pool #1		Conv #2		Max-pool #2		Dense #1	Dense #2
		input size	kernel size	input size	kernel size	input size	kernel size	input size	kernel size	input size	input size
MilliQ	Simplest	1×7×7	3×3	12×5×5	2×2	-	-	-	-	108×1	32×1
	Two Blocks	1×7×7	3×3	16×5×5	2×2	16×5×5	1×1	32×4×4	2×2	128×1	32×1
PBS	Simplest	10×7×7	3×3	32×5×5	2×2	-	-	-	-	216×1	32×1
	Two Blocks	10×7×7	3×3	40×5×5	2×2	40×5×5	1×1	80×4×4	2×2	320×1	32×1

Fig. 3. Schematics and list of parameters for the neural network (NN) models designed in this work. The input is the sequence of 7 × 7 simulated (by ENBIOS) or measured (by the biosensor chip) capacitance maps at one (milliQ) or multiple (PBS) frequencies. The output is the predicted radius of the nanoparticle.

inverted residual blocks are replaced by Identity layers (Learning Rate (LR) 4×10^{-3}, epochs 500); (II) the second one denoted SuperLite MoNet ($\approx 4\times10^4$ and $\approx 5.5\times10^4$ parameters for MilliQ (LR 5×10^{-4}, epochs 1000) and PBS case (LR 5×10^{-4}, epochs 1000), respectively), removing the last seven inverted residual blocks. The number of parameters and complexity of both were dramatically reduced. Other models from the PyTorch library, such as Resnet18 (He et al., 2016) have been investigated in a preliminary phase, but soon discarded due to unsatisfactory performance in our application.

In order to further optimize the radius estimation models for low-memory, low-power application, in terms of number of model parameters, two simple *ad-hoc* CNNs have been designed as sketched in Fig. 3 and they are:

- The Two Blocks Network ($\approx 5\times10^3$ and $\approx 17\times10^3$ parameters for MilliQ (LR 5×10^{-4}, epochs 500) and PBS case (LR 5×10^{-4}, epochs 1000), respectively) (Fig. 3) consists of two convolutional blocks and two fully-connected layers separated by a Rectified Linear Unit (ReLU) activation function. Each convolutional block consists of:
  - A convolutional layer with: n input and k output channels, kernel size of 3 or 1, and stride of 1.
  - A batch normalization layer, to improve overall stability and speed-up the architecture (Ioffe and Szegedy, 2015).
  - A ReLU used as a non-linearity.
  - A pooling layer, to reduce the dimensions of the feature maps, and in turn select the most significant features of the image.
- The Simplest Network ($\approx 1.5\times10^3$ parameters for MilliQ with LR 1×10^{-3}, epochs 500 and $\approx 9\times10^3$ parameters for the PBS case with LR 1×10^{-3}, epochs 1000 respectively), a simplification of the Two Blocks Network, featuring only one convolutional block.

The input is the sequence of 7 × 7 simulated (by ENBIOS) or measured (by the biosensor chip) capacitance maps at one (milliQ) or multiple (PBS) frequencies. The output is the predicted radius of the nanoparticle. The number of channels and kernel sizes depend on the architecture and on the case it was designed for (MilliQ or PBS), see Fig. 3.

All models have been trained to predict the radius of the particle in units of micrometers, in order to manage data values close to unity which improves the stability of the training.

The chosen loss function is the Gaussian Negative Log-Likelihood loss (GNLL) (Nix and Weigend, 1994), which assumes that the quantity to estimate is a random variable with Gaussian distribution, thereby allowing the model to estimate both the mean value and its variance (a key feature for uncertainty quantification). The training procedure is carried out using the Adamax (Kingma and Ba, 2017) algorithm implemented in PyTorch. Note that the whole procedure has to be unambiguously considered as supervised learning, since the simulated and experimental samples are labeled with the known (in simulation) and nominal (in experiments) particle radius, respectively.

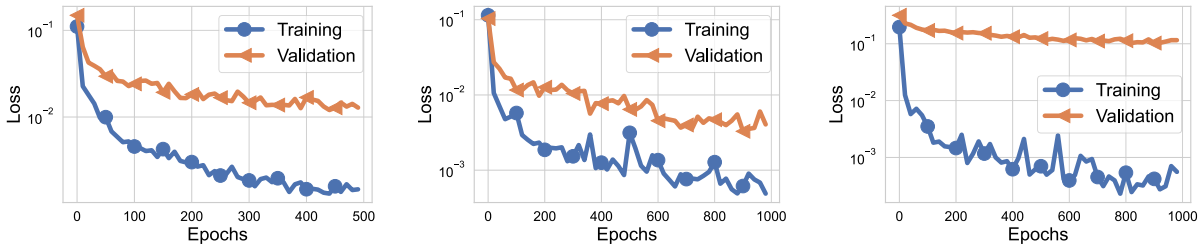
2.4. Model evaluation criteria

In this section, we detail the optimization of the NN algorithms with two main objectives: (i) to obtain an average estimated radius as close as possible to the nominal (in the case of measurements) and set (in the case of simulations) radius, (ii) to obtain as low as possible spread of the estimated radii. For these reasons, the results are mostly presented in boxplots, which report both the average estimated radius and the spread of the results. We aim to achieve these objectives with *simple* NNs, having a small number of parameters, to meet the low-memory and low-power specifications of sensing applications with edge devices.

The results are compared to a common and consolidated nanoparticle sizing method; namely: dynamic light scattering (see Section 2.1). The following strategies have been adopted and investigated.

The models have been trained and tested in both MilliQ and PBS environments. The traits common to both conditions are:

- as a first attempt, the models have been trained and validated (hyperparameters tuning) on data sets of simulations only. The



(a) Two Blocks network for MilliQ (b) Two Blocks network for PBS (c) Superlite Mobile Network for PBS

Fig. 4. Plots of the loss functions associated with two of the NNs developed. (a) Two Blocks Network tuned for the MilliQ environment with $LR = 5 \times 10^{-4}, epochs = 500$. (b) Two Block Network tuned for the PBS environment with $LR = 5 \times 10^{-4}, epochs = 1000$. (c) SuperLite tuned for the PBS environment with $LR = 5 \times 10^{-4}, epochs = 1000$.

Table 1
Summary of the number of samples (NS) for each data set. The definition of a sample is given in Section 2.

	MilliQ		PBS	
Initial simulation data set	1578		1133	
Initial measurement data set	194		169	
Initial total data set	1772		1302	
	Test on Simulations	**Test on Measurements**	**Test on Simulations**	**Test on Measurements**
Training set	1065	1065 (sim.)/ 1085 (sim.+meas)	770	654 (sim.)/ 674(sim.+meas)
Validation set	267	267	193	164
Test set	246	174	170	149
Particles excluded from training and validation sets	750 nm	750 nm	750 nm	1000, 1500 nm

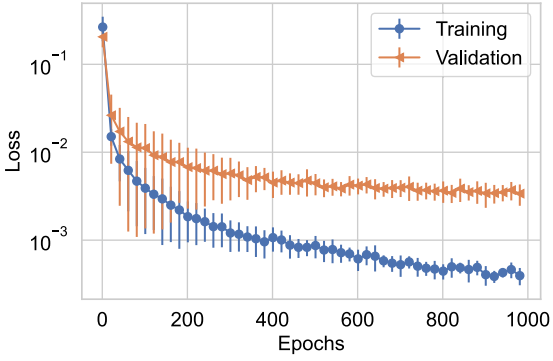


Fig. 5. Training and validation loss functions for the 5-fold cross validation of Simplest Net for the MilliQ environment. The solid lines are the average; the error bars represent the spread between the five different curves.

samples (excluding the 750 nm particles used for testing) were randomly split between a training and a validation set with a 80–20 ratio, respectively. Fig. 4 shows three representative loss functions optimization examples on training and validation sets. In addition, a 5-fold cross-validation has been performed, in order to confirm the validity of the hyper-parameters tuning. Fig. 5 shows the training and validation losses at each epoch for the Simplest Net trained on MilliQ simulation samples, the solid lines are the average and the error bars are the spread between the different 5 curves. We see that, as the number of epochs increases, the trained models are very consistent with each other, also meaning that the hyperparameters are appropriate to reduce overfitting.

Then, the models trained on simulations are tested:

- on simulated data samples only (samples of NPs with 750 nm radius, excluded from previous training and validation). Figs. 7(a) and 8(a) show the results of this first step for MilliQ and PBS case, respectively;
- on measured data samples. The results are reported in Figs. 7(b) and 8(b) for MilliQ and PBS case, respectively.

- Additionally, the models have been trained and validated on data sets combining simulations with a small portion of the measurements samples not previously used (10 samples for each radius). The performances are then tested on the remaining measurement samples. To this end, we consider the smallest portion of added measurement that led to an improvement in the performance (10 samples for each radius, i.e. ≈ 2% and 3% in MilliQ and PBS environments, respectively). These results are reported as well in Figs. 7(b) and 8(b) for MilliQ and PBS case, respectively.

For the NN testing on measured samples in PBS, the training data set excluded 1000 nm and 1500 nm radii, since we found that simulations of these large particles, otherwise not present in the experiments produced larger errors.

This procedure has a twofold objective. On one hand, to show that it is possible to achieve good model performance in this challenging application of nanoscale metrology by training the networks on simulations data set, as opposed to relying solely on measurements, which are more difficult to control and to reliably gather in large quantities. On the other hand, to understand whether the addition of a small number of measurements to the training data sets is or not beneficial for improving the performance beyond what simulations data sets can achieve.

Table 1 summarizes the number of samples (NS) in the data sets. Notice that the NS of the Training set and Validation set always respect a ≃ 80:20 ratio, as stated before. Additionally (see row "Training set"), the "Sim. + Meas." data set always has 20 samples more than the "Sim. Only" data set, which is the number of measured data samples (10 for each radius) included. Finally, notice that the "Test set" used for the test on simulations is removed from the total initial data set in both MilliQ and PBS cases. Indeed, by summing the number of samples (NS) of "Test set", "Validation set", and Training set" for the "Test on Simulations" case ($NS_{training/validation/test,sim-test}$) in both MilliQ and PBS ($NS_{training,sim-test} + NS_{validation,sim-test} + NS_{test,sim-test}$):

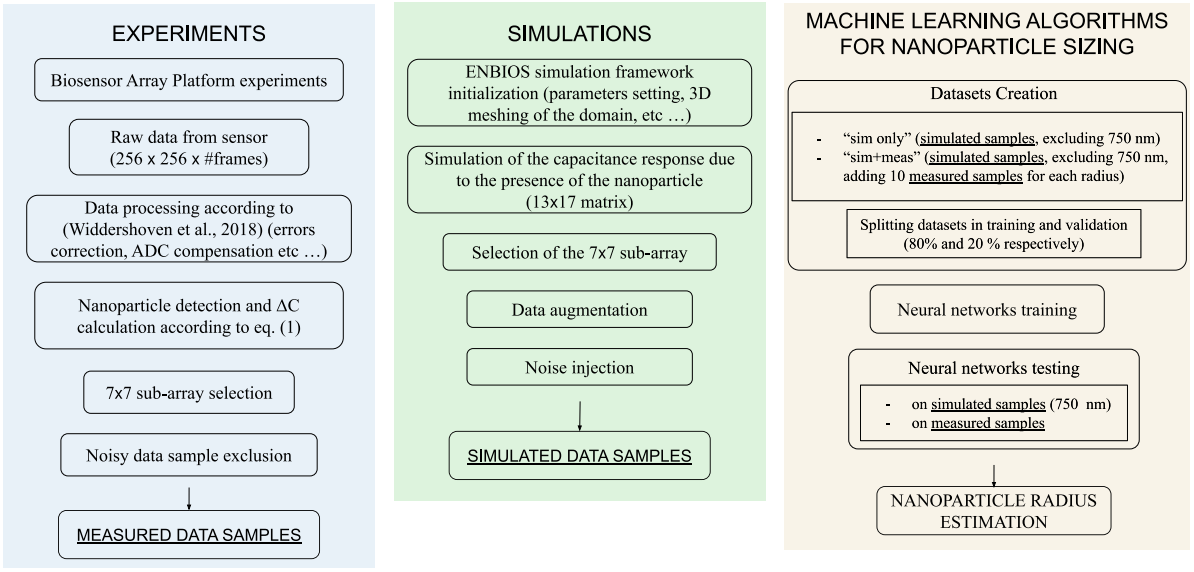


Fig. 6. Schematic flowchart of the steps made for dataset generation from simulations (left column), for measurements (middle column), and for the NN model development and test (right column). Please consider the specificity of the NN testing on measured samples in the case of PBS reported in Section 2.4. Table 1 provides the sizes of all data sets.

- in the MilliQ case: 1065 + 267 + 246 = 1578.
- in the PBS case: 770 + 193 + 170 = 1133.

which are exactly the initial simulation data set in MilliQ and PBS, respectively.

For the "Test on Measurements" case the sum is the following: $NS_{test,sim-test} + NS_{test,meas-test} + NS_{validation,meas-test} + NS_{training,meas-test}$, which for MilliQ is simply: 246 + 174 + 267 + 1085 = 1772, which is exactly the initial total dataset. The PBS case is less straightforward because, as stated before, the 750 nm samples (NS = 170) remained in the training/validation set (i.e. NS = 170 reported in the third column is actually included in the training/validation numerosity), whilst the 1000 nm and 1500 nm (NS = 315) were removed. This is also the reason why the training and validation set for "Test on Simulations" and "Test on Measurements" in the case of PBS are different, while in the case of MilliQ are equal. In fact: 149 + 164 + 674 + 315 = 1302, which is exactly the initial total dataset for PBS environment.

3. Results and discussion

Fig. 6 provides a schematic summary of the steps made to generate the simulated and measured datasets used to train and test the NNs either individually or in a mixed fashion. For a more detailed description of the data sets employed in this work see Table 1.

In the following, we will discuss in detail the performance achieved in the task of measuring the radius of nanoparticles in different representative scenarios.

3.1. Results in MilliQ environment

3.1.1. Training and test on simulated samples only

The models were firstly trained on the set containing only simulation samples of all NP radii except the 750 nm one. The added white Gaussian noise estimated from measurements was $\hat{\sigma}_{MQ} = 0.3$ aF. Its addition to the simulation datasets was important to achieve more stable results. The models' hyperparameters were then tuned on the validation set, which comes from the same distribution as the training set. Finally, the model was applied to the test data set (simulations of 750 nm radius particles only) to check its robustness and accuracy.

Fig. 7(a) (Table and Boxplot) shows the test results for the MilliQ environment and for all the NNs taken into consideration. We observe that the models with the less parameters achieve the best performances, since they are more robust to overfitting. The best models, Simplest Net and Two Blocks Net, exhibit a remarkably low (below 10%, see Fig. 7(a)) relative median error. On the other hand, the heaviest model of all, MobileNet, dramatically suffers from overfitting, and does not reach a satisfactory performance. Therefore it will be excluded from the following analysis.

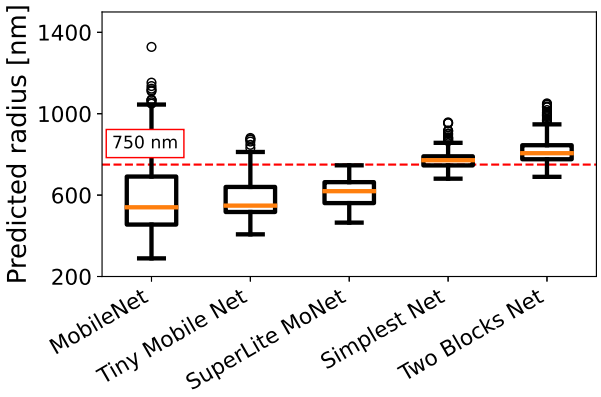
3.1.2. Test on measurement samples (different trainings)

The models described above have also been tested on the measurements data set. As discussed in Section 2.1, these measurements refer to 275 nm and 500 nm radius NPs. Since the two measurements were done separately, and in order to better identify the models strengths and weaknesses, their performance is evaluated separately for each radius. Due to its poor performance, MobileNet has been replaced with the popular algorithm XGBoost in this investigation.

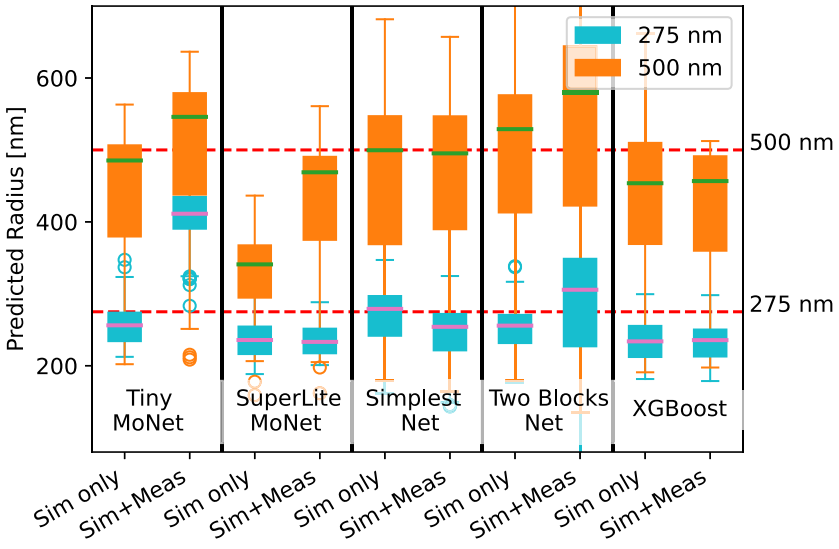
Fig. 7(b) shows the results produced by each model on the measurement test set including or not in the training set the small amount of measurements previously excluded from the test data set (ten samples per radius). We observe that all models except Tiny MoNet are able to generate two separated clusters for the two radii. Simplest Net exhibits very good performance with the lowest median error (highest accuracy), and a spread of predictions only slightly larger than that of XGBoost. Except for the very simple SuperLite MoNet model, the inclusion of a few measured data samples during training, rather than improving the model predictions, either degrades or leaves unchanged the accuracy.

The most likely explanation for the degradation is that, since Tiny MoNet and Two Blocks Net (the ones showing the largest degradation) are the ones with the most parameters, their training remains unstable due to overfitting. On the other hand, the more accurate Simplest Net and XGBoost have probably reached their maximum performance given their capacity, and therefore a small amount of additional data does not significantly improve their accuracy. In fact, the measurements are quite sensitive to uncontrolled variables (particle displacement and height w.r.t. the electrode), and therefore the radius estimation is particularly challenging, especially for the NPs with dimensions smaller than the electrode pitch. The frequency-independent spectrum of NPs in MilliQ deprives the NNs of a source of information that can, instead, facilitate a more precise estimation, as we will see in the next sections.

Model name	Error @95 Percent. [%]	Median Error [%]	Prediction Mean [nm]
MobileNet	53%	31%	597.4
Tiny Mobile Net	36%	27%	588
SuperLite MoNet	32%	17%	610.2
Simplest Net	17%	3%	777.3
Two Blocks Net	32%	7%	822.1



(a) Table and boxplot results for training and test on simulations (only 750 nm particles)



(b) Boxplot results for test on measurements (training on simulations or simulations and few measurements)

Fig. 7. Comparison among all models for the simulations test set (table and 7(a)) and the measurements test set 7(b) in MilliQ solution. The targets are highlighted with a dashed horizontal line in the boxplot.

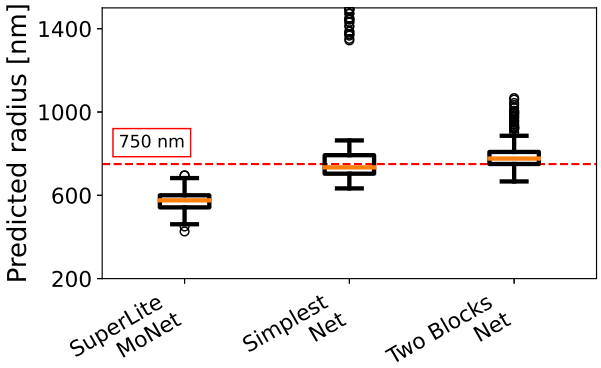
As a final note, Fig. 7(b) (and as well Fig. 8(b) for the PBS case) provides in itself also the results of an ablation study. In fact, the two simplest model (Simplest Net and Two Blocks Net) have been derived through a bottom-up approach, by starting from the simplest conceivable model and adding complexity. On the other hand, Tiny MoNet and SuperLite Monet come from repeated simplification of a larger model (MobileNet). This ablation study points in the direction of using the simplest available models.

3.2. Results in PBS environment

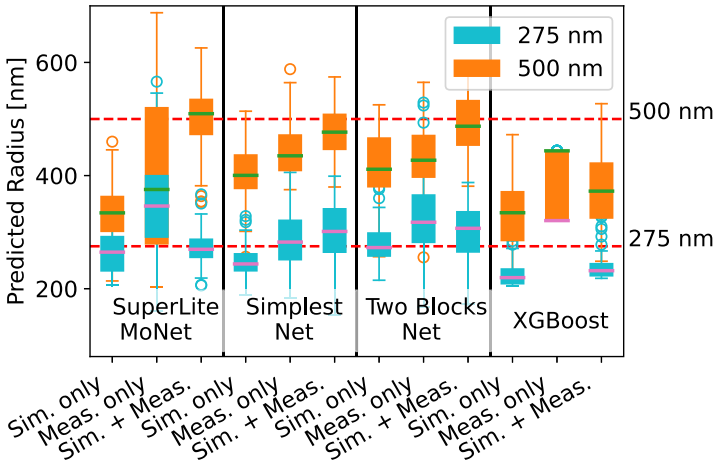
3.2.1. Training and test on simulated samples only

Similarly to Section 3.1.1, also in this PBS case the models have been firstly trained and tested on simulations. Tiny Mobile Net has been excluded from the comparison due to its very poor performance in the MilliQ environment.

Model name	Error @95 Percent. [%]	Median Error [%]	Prediction Mean [nm]
SuperLite MoNet	36%	23%	571.3
Simplest Net	99%	6%	833.6
Two Blocks Net	31%	4%	794.4



(a) Table and boxplot results for Test on simulations (only 750 nm particles)



(b) Boxplot results for test on measurements (training on simulations, measurements, or simulations and few measurements).

Fig. 8. Comparison on the simulations test set 8(a) and the measurements test set 8(b) between all models for the case with a PBS solution. The target radius for each case is highlighted with a dashed horizontal line in the boxplot.

Fig. 8(a) (Table and Boxplot) shows results in PBS on the simulations test set. Consistently with previous observations in Section 3.1.1, the most accurate models are the simplest ones, i.e., Simplest Net and Two Blocks Net. They achieve low median errors of a few percent, consistent with those in MilliQ. SuperLite MoNet exhibits much worse figures. Comparing these results to Fig. 7(a) puts in evidence a larger number and broader distribution of the outliers, especially for Simplest Net. On the other hand, SuperLite MoNet has the lowest accuracy, but it does not suffer from as many outliers.

A closer inspection of the outliers for Simplest Net and Two Blocks Net revealed that they correspond to particles at the largest distance from the electrodes ($dz = 100$ nm). Indeed, this is the most challenging detection condition in PBS because, due to strong Debye screening (especially at low frequency), the electrode response decays rapidly from its maximum value when dz increases and/or the particle centroid is not vertically aligned to the center of the electrode. This limitation could probably be alleviated by extending the measurement/simulation range to even higher frequencies.

3.2.2. Test on measurements samples (different trainings)

The same procedure of Section 3.1.2 was applied also in this case. However, due to the salinity of the solution, and the use of multi-frequency data, a few differences emerged which led to a slightly different procedure:

- Due to the weakened Debye screening, especially at high frequency, the NPs affect the capacitance of much more distant electrodes than in the MilliQ case. Therefore, to increase the accuracy of the noise estimation, only measurements of the 275 nm radius particles have been considered when determining the noise from the response of the electrodes on the edge.

Table 2
Summary of the best results achieved by the Neural Network models on the measurement test set. The models with best average performance in MilliQ and PBS, respectively, are highlighted in bold.

Solution	Model Name	Training Dataset	Radius [nm]	Predicted mean [nm]	Median rel. error [%]	95th percent error [%]
MilliQ	SuperLite MoNet	sim.	275	237	14	27
			500	326	31	58
		sim. + meas.	275	235	15	24
			500	419	7	58
	Simplest Net	**sim.**	**275**	**264**	**8**	**33**
			500	**454**	**14**	**60**
		sim. + meas.	275	241	8	40
			500	449	11	61
	Two Blocks Net	sim.	275	246	7	34
			500	473	15	62
		sim. + meas.	275	287	22	53
			500	525	28	63
	XGBoost	sim.	275	234	14	31
			500	421	12	60
		sim- + meas.	275	232	14	30
			500	413	8	59
PBS	**SuperLite MoNet**	sim.	275	263	10	24
			500	333	33	49
		sim. + meas.	**275**	**274**	**6**	**21**
			500	**505**	**7**	**29**
	SimplestNet	sim.	275	248	12	26
			500	403	20	37
		sim. + meas.	275	298	15	38
			500	477	6	18
	Two Blocks Net	sim.	275	278	7	23
			500	413	18	40
		sim. + meas.	275	300	16	37
			500	487	7	19
	XGBoost	sim.	275	227	20	25
			500	334	33	52
		sim. + meas.	275	238	16	21
			500	370	25	48

- Different percentiles of the noise standard deviation $\hat{\sigma}$ have been considered. After the model hyperparameters tuning on the validation set, we decided to use the median of the σ distribution at each frequency as in Section 3.1.2, hereafter denoted as $\hat{\sigma}_{PBS}$. The noise variance is calculated separately at each frequency (and values between $\sigma_{PBS,min} = 2$ aF and $\sigma_{PBS,max} = 6$ aF) have been found and used in the generation of noisy simulated images.
- one additional processing step was introduced in the training set to increase the sample diversity, hence the accuracy; namely: the addition of a random Gaussian noise to the ground-truth value of the radius. In this way, for each sample and at each training epoch, the value of the radius R_i has been perturbed as:

$$R_i^* = R_i + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma_{R_i}) \tag{2}$$

where $\mathcal{N}(0, \sigma_{R_i})$ is a Gaussian distribution with zero mean and standard deviation $\sigma_{R_i} = 0.05 \times R_i$. This value is marginally larger than the ideal nominal dimensional spread of the NP radius, see Section 2.1.

Fig. 8(b) and Table 2 show the performance of the models on the measurements test set. If only simulations are used for training (left box-plot for each model), we see that SuperLite MoNet and XGBoost are unable to create two separate radius clusters, while Simplest Net and Two Blocks Net provide quantitatively unsatisfactory results for the largest particle. The inclusion of a small number of measurements (10 data samples for each radius) during training dramatically increases the models performance, especially for SuperLite MoNet which becomes the most accurate (right box-plot for each model). Training the models only on measurements, instead, yields largely unsatisfactory performance (middle box-plot for each model), most probably because the data is too scarce, thereby proving the importance of expanding the training data set using simulations.

Table 2 summarizes the most relevant accuracy metrics for the best NNs explored in this work. We notice that after the improvements described above, the performance in PBS became actually better than in the MilliQ environment. This is probably due to the wealth of additional information embodied in the frequency sweep, that helps to overcome the detrimental impact of the uncontrolled position of the particles. This highlights once again the advantage of the high frequency impedance spectroscopy sensing paradigm implemented by the platform, which was specifically designed to operate in physiological solutions (Widdershoven et al., 2018).

4. Conclusions

This work investigated strategies to empower innovative nanoelectrode array biosensors with optimized Machine Learning algorithms, to quantitatively and robustly estimate physical parameters of nanoparticles in liquid solutions from multi-frequency capacitance measurements. The approach leverages the structural similarity of the measurements to regular multi-spectral images and employs *ad-hoc* and remarkably simplified versions of state-of-the-art computer vision algorithms. The radius estimation case study is particularly challenging because the NPs have sub-micron size and a much smaller radius than the electrode pitch, i.e. the image pixel size. We highlight the following as unique contributions of this study (to the best of our knowledge):

- the use of simulated data samples as a promising data-augmentation technique for ML-based advanced nanoparticle metrology with High Frequency Impedance Spectroscopy biosensor arrays.
- optimization strategies and investigation of traditional computer vision algorithms, usually applied on RGB images, to capacitance images of NPs taken with an advanced biosensor array.
- the application of ad-hoc developed machine-learning methods for sizing nanoparticles in water solution at the resolution limit of the imaging biosensor array, which gives a concrete and unique contribution in the rising field of nano-plastics monitoring in water, overcoming limitations of current monitoring technologies.

For the optimized training protocol (based on less than 1100 and 700, mostly simulated samples for the MilliQ and PBS case, respectively), the lean Simplest and SuperLite nets exhibit an excellent accuracy, with a median error below 15% in all scenarios (MilliQ and PBS, training on simulations or simulations and measurements, see Table 2). Furthermore, the *ad-hoc* developed Simplest Net shows $\leq 20\%$ median error in all scenarios with both pure simulations and mixed training sets. The results reported in this work with Simplest Net and SuperLite MoNet are in agreement and with an even lower spread with respect to the DLS sizing benchmark.

While investigating strategies for ML-enhanced biosensor array applications, the study has highlighted a few important points to consider:

- it is advantageous to derive the training data from a large number of accurate simulations, covering the operational space of the biosensor and providing accurate labels to the samples;
- for robust algorithms, the simulations must be diversified with a realistic random noise, which can be estimated from a much smaller set of measurements;
- the size of the ML models has to be carefully tailored to the amount of available data, the size of the capacitance map and the complexity of the task, in order to avoid overfitting. In particular, two-layers network have a size appropriate to the estimation of spherical NP radius comparable to the array pixel pitch.

The proposed methodology supports accurate and label-free metrology of nanoparticles featuring size comparable to the spatial resolution of the array by purely electrical means, and with accuracy in line with the more complex DLS method. Furthermore, it is amenable to extensions serving additional tasks (such as estimation of the particle dielectric constant, shape, and orientation), possibly by processing the 256×256 array with techniques developed for object detection in computer vision, with the goal of automatically identifying the locations of all particles, and then quantitatively characterize them.

We finally observe that, while numerous radii have been analyzed by simulations, only two are experimentally measured. Furthermore, We use spherical nanoparticles, but real-world nanoplastic monitoring technologies might deal also with other shapes. Notwithstanding the above observations, our promising new results reinforce the interest toward nanoelectrode array biosensor technology. Analysis of functionalized particles with different shapes and materials represents an interesting future development but falls outside the scope of this pioneering work. Future work is expected to push these highly innovative HFIS nanobiosensors to evolve from demonstrators in prototypical lab conditions to powerful and reliable tools for a wide variety of industrial and commercial applications.

CRediT authorship contribution statement

Federico Lombardo: Investigation, Software, Data curation, Writing – review & editing. **Federico Pittino:** Conceptualization, Methodology, Supervision, Software, Validation, Funding acquisition, Writing – review & editing. **Daniele Goldoni:** Investigation, Software, Data curation, Validation, Writing – review & editing. **Luca Selmi:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All the data sets (measured and simulated data samples) are available online Lombardo et al. (2023).

Acknowledgments

The authors acknowledge partial support of this work by Silicon Austria Labs (SAL), owned by the Republic of Austria, the Styrian Business Promotion Agency (SFG), Austria, the federal state of Carinthia, the Upper Austrian Research (UAR), and the Austrian Association for the Electric and Electronics Industry (FEEI), by ESF REACT-EU National Operational Programme (PON) Research and Innovation 2014–2020, and by the European Union, NextGenerationEU initiative via Linea di Investimento 1.3 – Partenariati Estesi Area 6 the PNRR HEAL ITALIA Project PE 0000019. The authors would also like to thank F. Widdershoven (NXP Semiconductor) for providing the chips and for fruitful discussion, Dr. Cristina Murari (UniMORE) for help with the setup of the HPC simulation framework, Dr. C. Ongaro (UniMORE) for the design and fabrication of the PDMS microchamber sealing, Prof. Luigi Rovati (UniMORE) for providing the nano-particles, Prof. Claudio Fontanesi (UniMORE) for the sonication equipment.

References

Abbott, J., Mukherjee, A., Wu, W., Ye, T., Jung, H.S., Cheung, K.M., Gertner, R.S., Basan, M., Ham, D., Park, H., 2022. Multi-parametric functional imaging of cell cultures and tissues with a CMOS microelectrode array. Lab Chip 22, http://dx.doi.org/10.1039/d1lc00878a.

Ansah, I.B., Leming, M., Lee, S.H., Yang, J.-Y., Mun, C., Noh, K., An, T., Lee, S., Kim, D.H., Kim, M., Im, H., Park, S.G., 2023. Label-free detection and discrimination of respiratory pathogens based on electrochemical synthesis of biomaterials-mediated plasmonic composites and machine learning analysis. Biosens. Bioelectron. 227, 115178. http://dx.doi.org/10.1016/j.bios.2023.115178, URL: https://www.sciencedirect.com/science/article/pii/S0956566323001203.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD, ACM, pp. 785–794.

COMSOL Inc., 2022. Comsol multiphysics v. 6.0.

Cossettini, A., Brandalise, D., Palestri, P., Bertacchini, A., Ramponi, M., Widdershoven, F., Benini, L., Selmi, L., 2020. Ultra-high frequency (500 MHz) capacitance spectroscopy for nanobiosensing. In: Proceedings of IEEE Sensors, Vol. 2020-October. http://dx.doi.org/10.1109/SENSORS47125.2020.9278583.

Cossettini, A., Laborde, C., Brandalise, D., Widdershoven, F., Lemay, S.G., Selmi, L., 2021. Space and frequency dependence of nanocapacitor array sensors response to microparticles in electrolyte. IEEE Sens. J. 21 (4), http://dx.doi.org/10.1109/JSEN.2020.3032712.

Cossettini, A., Stadlbauer, B., Morales, J.A., Taghizadeh, L., Selmi, L., Heitzinger, C., 2019. Determination of micro- and nano-particle properties by multi-frequency Bayesian methods and applications to nanoelectrode array sensors. In: Proceedings of IEEE Sensors, Vol. 2019-October. http://dx.doi.org/10.1109/SENSORS43011.2019.8956529.

Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M., 2023. Diffusion models in vision: A survey. IEEE Trans. Pattern Anal. Mach. Intell..

Cui, F., Yue, Y., Zhang, Y., Zhang, Z., Zhou, H.S., 2020. Advancing biosensors with machine learning. ACS Sensors 5 (11), 3346–3364. http://dx.doi.org/10.1021/acssensors.0c01424.

Dong, S., Wang, P., Abbas, K., 2021. A survey on deep learning and its applications. Comp. Sci. Rev. 40, 100379.

Enfrin, M., Lee, J., Gibert, Y., Basheer, F., Kong, L., Dumée, L.F., 2020. Release of hazardous nanoplastic contaminants due to microplastics fragmentation under shear stress forces. J. Hard Mater. 384, 121393. http://dx.doi.org/10.1016/j.jhazmat.2019.121393, URL: https://www.sciencedirect.com/science/article/pii/S0304389419313470.

Goldoni, D., Rovati, L., Selmi, L., 2023. Toward continuous nano-plastic monitoring in water by high frequency impedance measurement with nano-electrode arrays. IEEE Sens. J. 23 (17), 20180–20188. http://dx.doi.org/10.1109/JSEN.2023.3296158.

Gomes, J.C.M., Souza, L.C., Oliveira, L.C., 2021. SmartSPR sensor: Machine learning approaches to create intelligent surface plasmon based sensors. Biosens. Bioelectron. 172, 112760. http://dx.doi.org/10.1016/j.bios.2020.112760, URL: https://www.sciencedirect.com/science/article/pii/S0956566320307478.

Hamedi, S., Jahromi, H.D., Lotfiani, A., 2023. Artificial intelligence-aided nanoplasmonic biosensor modeling. Eng. Appl. Artif. Intell. 118, 105646. http://dx.doi.org/10.1016/j.engappai.2022.105646, URL: https://www.sciencedirect.com/science/article/pii/S0952197622006364.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., Adam, H., 2019. Searching for MobileNetV3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV.

Hu, K., Arcadia, C.E., Rosenstein, J.K., 2021. A large-scale multimodal CMOS biosensor array with 131,072 pixels and code-division multiplexed readout. IEEE Solid-State Circuits Lett. 4, http://dx.doi.org/10.1109/LSSC.2021.3056515.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.

Jameel, S.M., Rehman Gilal, A., Hussain Rizvi, S.S., Rehman, M., Hashmani, M.A., 2020. Practical implications and challenges of multispectral image analysis. In: 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies. ICoMET, pp. 1–5. http://dx.doi.org/10.1109/iCoMET48670.2020.9073821.

Jan, N., Majeed, N., Ahmad, M., Ahmad Lone, W., John, R., 2022. Nano-pollution: Why it should worry us. Chemosphere 302, http://dx.doi.org/10.1016/j.chemosphere.2022.134746.

Jia, Z., Li, J., Gao, L., Yang, D., Kanaev, A., 2023. Dynamic light scattering: A powerful tool for in situ nanoparticle sizing. Colloids Interfaces 7 (1), http://dx.doi.org/10.3390/colloids7010015, URL: https://www.mdpi.com/2504-5377/7/1/15.

Jung, D., Junek, G.V., Park, J.S., Kumashi, S.R., Wang, A., Li, S., Grijalva, S.I., Fernandez, N., Cho, H.C., Wang, H., 2021. A CMOS 21 952-pixel multi-modal cell-based biosensor with four-point impedance sensing for holistic cellular characterization. IEEE J. Solid-State Circuits 56, http://dx.doi.org/10.1109/JSSC.2021.3085571.

Khan, A., Vibhute, A.D., Mali, S., Patil, C., 2022. A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. Ecol. Inform. 69, 101678. http://dx.doi.org/10.1016/j.ecoinf.2022.101678, URL: https://www.sciencedirect.com/science/article/pii/S1574954122001285.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Laborde, C., Pittino, F., Verhoeven, H.A., Lemay, S.G., Selmi, L., Jongsma, M.A., Widdershoven, F.P., 2015. Real-time imaging of microparticles and living cells with CMOS nanocapacitor arrays. Nature Nanotechnol. 10 (9), http://dx.doi.org/10.1038/nnano.2015.163.

Lai, L.H., Lin, W.Y., Lu, Y.W., Lui, H.Y., Yoshida, S., Chiou, S.H., Lee, C.Y., 2023a. A 460 800 pixels CMOS capacitive sensor array with programmable fusion pixels and noise canceling for life science applications. IEEE Trans. Circuits Syst. II 70 (5), 1734–1738. http://dx.doi.org/10.1109/TCSII.2023.3261445.

Lai, P.H., Tseng, L.S., Yang, C.M., Lu, M.S., 2023b. Design and characterization of a 16 × 16 CMOS capacitive DNA sensor array. IEEE Sens. J. http://dx.doi.org/10.1109/JSEN.2023.3253123.

LeCun, Y., Bengio, Y., et al., 1995. Convolutional networks for images, speech, and time series. In: The Handbook of Brain Theory and Neural Networks, Vol. 3361, No. 10. Citeseer, p. 1995.

Lee, J., Chae, K.J., 2021. A systematic protocol of microplastics analysis from their identification to quantification in water environment: A comprehensive review. J. Hazard. Mater. 403, http://dx.doi.org/10.1016/j.jhazmat.2020.124049.

Lombardo, F., Pittino, F., Goldoni, D., Selmi, L., 2023. Nanoparticles simulated and measured datasets. URL: https://github.com/DanieleGoldoni/nanoparticle_dataset_sizing_LombardoPittnoGoldoniSelmi_EAAI.

Lorenzo-Navarro, J., Castrillón-Santana, M., Santesarti, E., De Marsico, M., Martínez, I., Raymond, E., Gómez, M., Herrera, A., 2020. SMACC: A system for microplastics automatic counting and classification. IEEE Access 8, 25249–25261. http://dx.doi.org/10.1109/ACCESS.2020.2970498.

Massarelli, C., Campanale, C., Uricchio, V.F., 2021. A handy open-source application based on computer vision and machine learning algorithms to count and classify microplastics. Water 13 (15), http://dx.doi.org/10.3390/w13152104, URL: https://www.mdpi.com/2073-4441/13/15/2104.

Meiler, V., Pfeiffer, J., Bifano, L., Kandlbinder-Paret, C., Fischerauer, G., 2023. Approaches to detect microplastics in water using electrical impedance measurements and support vector machines. IEEE Sens. J. 23 (5), 4863–4872. http://dx.doi.org/10.1109/JSEN.2023.3236375.

Nix, D.A., Weigend, A.S., 1994. Estimating the mean and variance of the target probability distribution. In: Proceedings of 1994 Ieee International Conference on Neural Networks, Vol. 1. ICNN'94, IEEE, pp. 55–60.

Oh, S., Hossen, I., Luglio, J., Justin, G., Richie, J.E., Medeiros, H., Lee, C.H., 2021. On-site/in situ continuous detecting ppb-level metal ions in drinking water using block loop-gap resonators and machine learning. IEEE Trans. Instrum. Meas. 70, 1–9. http://dx.doi.org/10.1109/TIM.2021.3115215.

Otter, D.W., Medina, J.R., Kalita, J.K., 2020. A survey of the usages of deep learning for natural language processing. IEEE Trans. Neural Netw. Learn. Syst. 32 (2), 604–624.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8024–8035.

Pennacchio, A., Giampaolo, F., Piccialli, F., Cuomo, S., Notomista, E., Spinelli, M., Amoresano, A., Piscitelli, A., Giardina, P., 2022. A machine learning-enhanced biosensor for mercury detection based on an hydrophobin chimera. Biosens. Bioelectron. 196, 113696. http://dx.doi.org/10.1016/j.bios.2021.113696, URL: https://www.sciencedirect.com/science/article/pii/S0956566321007338.

Pittino, F., Scarbolo, P., Widdershoven, F., Selmi, L., 2015. Derivation and numerical verification of a compact analytical model for the AC admittance response of nanoelectrodes, suitable for the analysis and optimization of impedance biosensors. IEEE Trans. Nanotechnol. 14 (4), http://dx.doi.org/10.1109/TNANO.2015.2434106.

Pittino, F., Selmi, L., 2014. Use and comparative assessment of the CVFEM method for Poisson–Boltzmann and Poisson–Nernst–Planck three dimensional simulations of impedimetric nano-biosensors operated in the DC and AC small signal regimes. Comput. Methods Appl. Mech. Engrg. 278, 902–923.

Renault, C., Laborde, C., Cossettini, A., Selmi, L., Widdershoven, F., Lemay, S.G., 2021. Electrochemical characterization of individual oil micro-droplets by high-frequency nanocapacitor array imaging. Faraday Discuss. 233, http://dx.doi.org/10.1039/d1fd00044f.

Sangkham, S., Faikhaw, O., Munkong, N., Sakunkoo, P., Arunlertaree, C., Chavali, M., Mousazadeh, M., Tiwari, A., 2022. A review on microplastics and nanoplastics in the environment: Their occurrence, exposure routes, toxic studies, and potential effects on human health. In: Marine Pollution Bulletin, Vol. 181. Elsevier Ltd, http://dx.doi.org/10.1016/j.marpolbul.2022.113832.

Senevirathna, B., Lu, S., Dandin, M., Basile, J., Smela, E., Abshire, P., 2019. High resolution monitoring of chemotherapeutic agent potency in cancer cells using a CMOS capacitance biosensor. Biosens. Bioelectron. 142, http://dx.doi.org/10.1016/j.bios.2019.111501.

Stadlbauer, B., Cossettini, A., Morales E., J.A., Pasterk, D., Scarbolo, P., Taghizadeh, L., Heitzinger, C., Selmi, L., 2019. Bayesian estimation of physical and geometrical parameters for nanocapacitor array biosensors. J. Comput. Phys. 397, http://dx.doi.org/10.1016/j.jcp.2019.108874.

Widdershoven, F., Cossettini, A., Laborde, C., Bandiziol, A., Van Swinderen, P.P., Lemay, S.G., Selmi, L., 2018. A CMOS pixelated nanocapacitor biosensor platform for high-frequency impedance spectroscopy and imaging. IEEE Trans. Biomed. Circuits Syst. 12 (6), http://dx.doi.org/10.1109/TBCAS.2018.2861558.

Widdershoven, F., Van Steenwinckel, D., Überfeld, J., Merelle, T., Suy, H., Jedema, F., Hoofman, R., Tak, C., Sedzin, A., Cobelens, B., Sterckx, E., Van Der Werf, R., Verheyden, K., Kengen, M., Swartjes, F., Frederix, F., 2010. CMOS biosensor platform. In: Technical Digest - International Electron Devices Meeting. IEDM, http://dx.doi.org/10.1109/IEDM.2010.5703484.