A Two-Scale Multi-Physics Deep Learning Model for Smart MEMS Sensors

José Pablo Quesada-Molina[1,2], Stefano Mariani[1]

[1]Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy
[2]Department of Mechanical Engineering, Universidad de Costa Rica, Rodrigo Facio Brenes Campus, San José, Costa Rica
Email: stefano.mariani@polimi.it

Abstract

Smart materials and structures, especially those bio-inspired, are often characterized by a hierarchy of length- and time-scales. Smart Micro Electro-Mechanical Systems (MEMS) are also characterized by different physical phenomena affecting their properties at different scales. Data-driven formulations can then be helpful to deal with the complexity of the multi-physics governing their response to the external stimuli, and optimize their performances. As an example, Lorentz force micro-magnetometers working principle rests on the interaction of a magnetic field with a current flowing inside a semiconducting, micro-structured medium. If an alternating current with a properly set frequency is let to flow longitudinally in a slender beam, the system is driven into resonance and the sensitivity to the magnetic field may result largely enhanced. In our former activity, a reduced-order physical model of the movable structure of a single-axis Lorentz force MEMS magnetometer was developed, to feed a multi-objective topology optimization procedure. That model-based approach did not account for stochastic effects, which lead to the scattering in the experimental data at the micrometric length-scale. The formulation is here improved to allow for stochastic effects through a two-scale deep learning model designed as follows: at the material scale, a neural network is adopted to learn the scattering in the mechanical properties of polysilicon induced by its polycrystalline morphology; at the device scale, a further neural network is adopted to learn the most important geometric features of the movable parts that affect the overall performance of the magnetometer. Some preliminary results are discussed, and an extension to allow for size effects is finally foreseen.

Keywords

Quantification, Lorentz Force Magnetometer

1. Introduction

Over the last years, we have witnessed an accelerated development in the technology of computer systems, with a huge increase in the available computation power. This evolution has been importantly boosted by demanding consumer applications and a growth in the market of high-performance computing. The development of affordable and highly specialized hardware, such as graphics processing units designed to optimize large data computations via parallel processing [1] [2], and the ease to access a vast amount of data, constitute key factors that contributed to circumvent the former bottlenecks limiting the use of data-driven algorithms like machine learning (ML). The paradigm proposed by such data-driven approaches is revolutionizing the research activity in numerous fields, materials science among them [3] [4] [5]. An appealing aspect of ML approaches lies in their potential to enable, accelerate and even simplify the identification of the underlying mappings governing the chain process-structure-property-performance. As shown in [5], diverse ML approaches have been successfully implemented to perform descriptive, predictive and prescriptive tasks in a straightforward data-driven manner.

Within this context, a very popular type of ML algorithms is represented by the artificial neural networks (ANNs). In their simplest form, feedforward neural networks (FFNNs) [6] [7] are obtained by assembling a number of layers of interconnected perceptrons; this architecture is typically referred to as a multilayer perceptron (MLP). By stacking a large enough number of layers, we enter into the realm of deep learning (DL), a subfield of ML that relies upon several levels of non-linear information processing and abstraction to produce complex feature learning tasks from unstructured input information [8]. A popular subtype of ANNs is given by the convolutional neural networks (CNNs). These architectures are formulated as a variant of FFNNs, to be well suited for data featuring a spatial correlation [9]. This feature makes CNNs capable to perform image recognition, time series forecasting and speech recognition, as CNNs are able to learn position and scale invariant structures in the data. Due to the compatibility of this feature with the typical data profiles to handle in materials science, a significant number of applications can be found, including material texture recognition [10] [11] and structure-property mapping [12] [13] [14].

In this work, we propose a two-scale ML approach based on an assembly of ANNs (CNNs and MLPs) at different length-scales, to provide an accurate property-performance mapping for polysilicon MEMS devices. We specifically focus on a single-axis Lorentz force magnetometer introduced in [15] [16]. Material- and geometry-related uncertainty sources, whose effects have been observed to get enhanced as the device footprint is reduces [17]-[22], are addressed

independently at the two scales: first, at the material (micro) scale, the effects of the morphology of the film constituting the movable structure are learned; next, at the device (meso) scale, the effects of the microfabrication process and of the geometry of the device are accounted for and learned on their own. A surrogate model is accordingly proposed, with the capability of an automatic detection of the effects of the uncertainty causes encoded in the output of the MEMS device, finally leading to an accurate one-to-one input-output mapping. As a further technical aspect, the ground-truth data used for training, validation and testing of the proposed ANN-based model have been obtained via a reduced-order model of the entire device discussed in [15] [16].

The remainder of this paper is organized as follows. In Section 2, the model of the Lorentz force MEMS magnetometer is discussed, along with the relevant sources of uncertainty. Section 3 provides an illustration of the implementation of proposed data-driven model. Results are discussed in Section 4, and some concluding remarks and prospects for future activities are offered in Section 5.

2. Model-based Approach and Uncertainty Sources

Among the various geometries proposed in the literature for the resonating structure of Lorentz force MEMS magnetometers, we focus here on the one depicted in **Figure 1**. The movable part of the device is given by a slender beam connected to the substrate at the extreme cross-sections, in a clamped-clamped configuration. For sensing purposes, a couple of plates are connected to the beam at the mid-span. When a current flows in the longitudinal direction of the beam, the device can sense a magnetic field aligned with the out-of-plane direction due to in-plane motion of the beam (along axis y in the figure).

The beam features a length L, and a rectangular cross section of area $A = bh$ and moment of inertia $I = bh^3/12$, h being the in-plane width and b the out-of-plane thickness. g represents the gap between the surfaces of each parallel-plate capacitor. When the beam is driven into resonance, the amplitude v_{max} of the oscillations at the mid-span is a solution of the following equation, which governs the weakly coupled thermo-electro-magneto-mechanical problem at hand, see [15] [16]:

$$\left(\frac{F_0}{K_1}\right) = \left(2\left(1 - \frac{\omega}{\omega_1}\right)v_{max} + \frac{3}{4}\frac{K_3}{K_1}v_{max}^3\right)^2 + \left(\frac{d}{m\omega_1}v_{max}\right)^2 \qquad (1)$$

Here, in terms of effective values: m is the mass of the beam and of the attached plates; d is the damping coefficient; K_1 and K_3 are, respectively, the

Figure 1. SEM picture of the movable structure of the resonant MEMS magnetometer [16].

linear and cubic stiffnesses of the beam; F_0 is the amplitude of an external force, sinusoidally varying in time with a circular frequency ω; $\omega_1 = \sqrt{K_1/m}$ is the natural frequency of the beam in the linear regime. The above mentioned system features depend, on top of all, on Young's modulus, coefficient of thermal expansion, resistivity and thermal conductivity of the polysilicon constituting the beam. The mentioned stiffness terms show softening effects linked to the temperature raise caused by Joule effect, and to the electrostatic forces at the parallel plate capacitors. By increasing the driving voltage, v_{max} and the sensitivity to imperfections get enhanced; buckling and pull-in must be avoided, and so the operational conditions have to be properly designed, see [15] for a thorough discussion in this regard.

As the solution of Equation (1) depends on the flexural (EI) and axial (EA) rigidities of the beam, where E is the so-called apparent stiffness of the polysilicon under the assumed deformation, uncertainties come into play. Since geometry- and material-related uncertainties are ultimately governed by process variables that are stochastic in nature, probability distributions need to be adopted for the variables involved. In a sort of bottom-up multiscale approach to the problem of uncertainty quantification at these small scales, in our previous works we first considered E to be scattered due to the morphology of the polycrystalline film. As the scattering actually depends on the geometry of the beam, to lead to the well-known size effects, the probability distribution of E has been set with reference to the target geometry governed by the width h. The scattering induced by microfabrication, as measured by the overetch or etch-loss O, was next plugged in independently. The joint effects of scattering in these two main sources were therefore neglected; we aim here to overcome this issue and provide a general platform to deal with all the possible sources at the same time. This is of paramount importance since predicting the reliability of MEMS under the specific conditions to be encountered for the intended applications is necessary to release products to the market.

To predict the scattering in E, a statistical investigation of the effects of the polycrystal morphology due to the topology of grain boundaries and to the lattice orientation for each single grain in the considered aggregate, a stochastic homogenization analysis has been conducted in [17]. By allowing for a beam geometry featuring $h = 2$ μm, the mean and standard deviation of the probability distribution of E have been obtained as 150.1 GPa and 5.5 GPa, respectively. Though not reported here for brevity, the mean value has resulted to be only slightly affected by h; the standard deviation has shown instead a significant dependence on the size, with a smaller scattering for larger volumes of polysilicon.

Regarding the geometry-related uncertainties, the overetch O has been sampled out of a microfabrication-tailored normal distribution featuring zero mean (since the target is the designed geometry) and a standard deviation of 50 nm, see [18] [20]. Assuming O to vary stochastically, the in-plane film width given by h–$2O$ varies too, so that the bending and axial rigidities turn out to be affected,

the former nonlinearly. Also the gap at capacitors is modified as $g + 2O$, if the etch defects are assumed to be homogenous and isotropic features of the micro-fabrication process.

3. Novel Two-Scale Artificial Neural Network-Based Model

To account for the inherent stochastic effects associated to the microstructure of the slender beam, we already proposed to consider the film as a concatenation or assemblage of subdomains, each one digitally-generated through a regularized Voronoi tessellation [17], see **Figure 2**. Such subdomains are termed statistical volume elements (SVEs), as they bring features of the actual statistical properties of the film morphology.

In the images of the SVEs, the pixels ranging within the interval [0, 255] represent the in-plane lattice orientation angle. Due to the symmetry displayed by the elastic properties of single-crystalline silicon, see [23], the Young's modulus of each grain is invariant under a series of affine transformations. This feature provides the ground for a data augmentation procedure wherein seven new images could be generated for each SVE, all linked to the same ground- truth data characterizing the original one. The considered affine transformations are: three counter clockwise 90° rotations, and four mirror transformations (horizontal, vertical and about the two diagonals). Each beam is then assumed to be composed of an SVE and a collection of its data augmented instances, whose concatenation order and frequency are randomly assigned.

Moving to the proposed ANN-based approach, **Figure 3** provides a sketch of the global topology of the regression model. Two stages can be distinguished: in the first one, a 128×128 pixels SVE image is fed into a ResNet-based model, developed in [24], which is leveraged to provide the estimation of the effective Young's modulus $\hat{E}$ of the SVE; in the second one, the same SVE image along with the just predicted value $\hat{E}$ and the overetch O are taken as input. Regarding the mentioned value of O, its value is sampled from the microfabrication-governed probability distribution. This second stage handles multiple and mixed data inputs, *i.e.* images and numerical data, and provides as output a normalized oscillation amplitude v_{max}/v_{max}^{*}, the normalization term v_{max}^{*} corresponding to the oscillation amplitude associated to the reference value
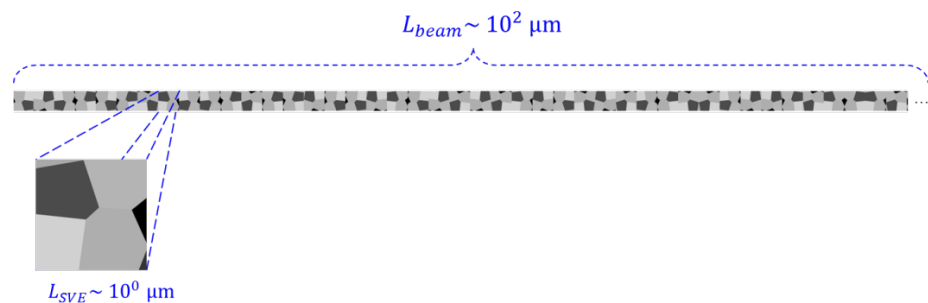

Figure 2. Considered slender beam regarded as a random concatenation of one individual SVE and its corresponding instances.

$E^* = 149.3$ GPa of the Young's modulus and to null overetch. This normalization factor has been introduced in order to focus on the trend in the results, rather than on the amplitude itself.

Regarding the first stage, the ResNet-based regression model exploits residual learning: local and translational invariant low-level features such as colors, edges and shapes are extracted in the initial convolutional layers. These features are then combined through further convolution operations in deeper layers, to achieve complex levels of abstraction and extract high-level features. A flatten operation is then used to pass the output feature maps to a fully connected layer featuring 100 nodes. Finally, a linear activation function is applied to the output neuron, to get the output in terms of the estimated Young's modulus $\hat{E}$. More details about training, validation and testing of this model can be found in [25].

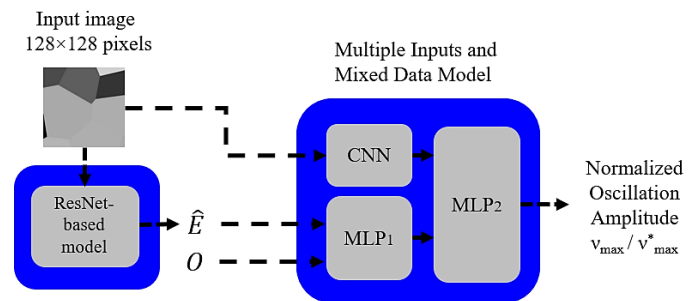Regarding the second stage, **Figure 4** shows the corresponding sub-architectures

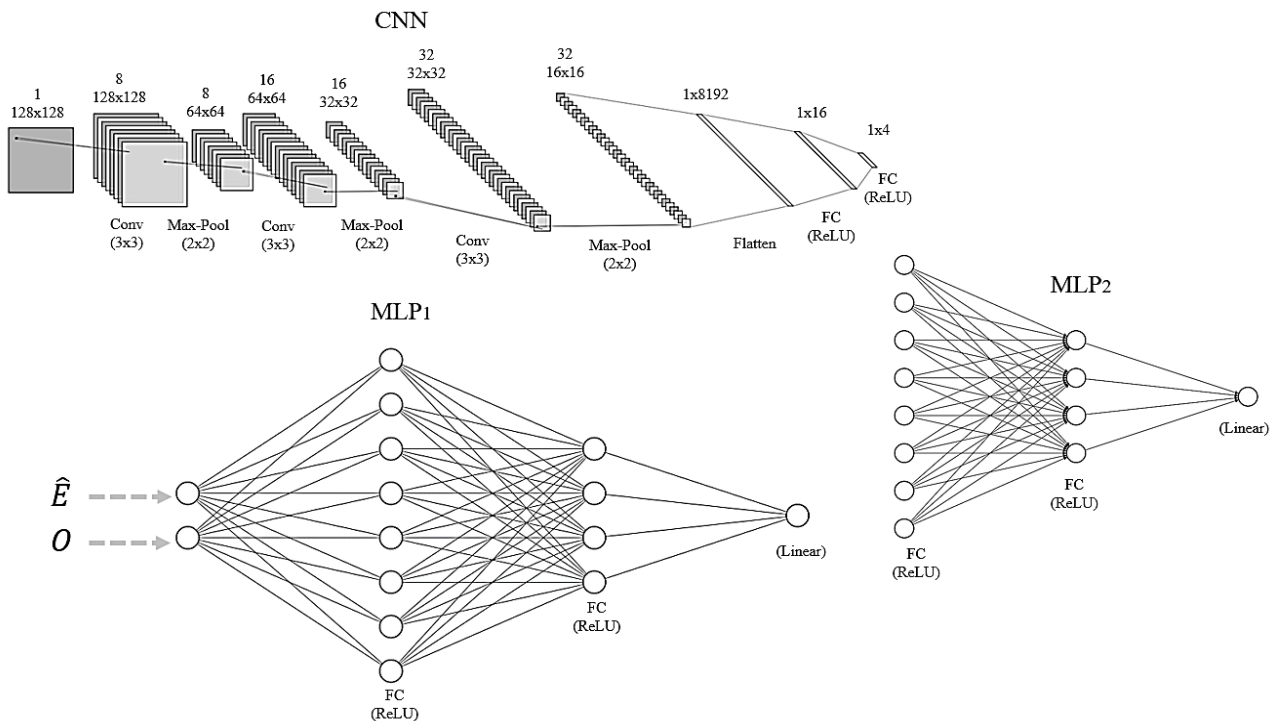**Figure 3.** Topology for the proposed data-driven model.

Figure 4. Sub-architectures of the multiple inputs and mixed data stage of the proposed model.

adopted. Although not explicitly shown in the figure, the four-neuron output layer of the CNN and the single-neuron output layer of the first multilayer perceptron (MLP$_1$), are fully connected to the eight-neuron input layer of the second multilayer perceptron (MLP$_2$). As far as the CNN architecture is concerned, it features a consecutive application of Convolution, ReLU activation, Batch Normalization and Max-Pool; these operations are applied as many times as the number of filters defined for the convolution. The dimensionality of the output space after the convolution is controlled by the output filters, which have been 8, 16 and 32 in the present scheme. The height and width of the 2D convolution window, referred to as the kernel size, have been set to (3, 3) pixels. A case-insensitive padding has been used to guarantee the size of the output feature-maps to be the same of the input feature-maps. The application of the Max-Pool layers enabled a down-sampling of the input height and width. After the output of the last Max-Pool layer, a flatten layer has been adopted to pass the information onto an encoder composed of a standard MLP equipped with a 16-node hidden layer and a 4-node output layer. To prevent overfitting, a dropout layer (not shown in the sketch) has been applied to the 16-node fully connected layer, randomly setting its input units to 0 with a dropout rate of 0.5 measuring the fraction of the input units to drop during training. MLP$_1$ is simply composed of a sequence of regular fully-connected layers. The specific sequence was made of an 8-node, a 4-node and a 1-node fully connected layers; only the output layer was activated linearly, while the two hidden layers feature a ReLU function. An identical configuration was chosen for MLP$_2$. **Table 1** summarizes the most relevant hyperparameters set during the training.

The implementation of the proposed approach was done making use of Keras, a Python interface for TensorFlow, a free and open-source software library for machine learning. Regarding the hardware, a GeForce GTX 1050 Ti GPU was exploited for the training of the model.

4. Results

A validation is first reported to assess the capability of the ResNet-based regression model to yield accurate predictions of the morphology-governed polysilicon Young's modulus. To carry out this investigation, a dataset composed of 989

Table 1. Hyperparameters adopted for training of the multiple inputs and mixed data model.

Hyperparameter	Value
Total number of epochs	400
Patience (early stopping)	100
Mini-Batch Size (B.S)	10
Learning rate α	$1 \cdot 10^{-3}$
Optimizer	Adam
Loss function	MSE

parent SVEs has been considered and not used during the training; a second dataset composed of 6923 instances has been then generated though data augmentation. Once the datasets have been assembled, they have been independently fed into the ResNet model to generate the predictions of the effective Young's modulus $\hat{E}$. The parity plot in **Figure 5** shows the correlation between the ground-truth data, generated in this case via numerical homogenization, and the predicted data, produced by the ResNet-based NN architecture. The identity map is included in the plot to clearly identify the ideal behavior of the regression model. Each blue dot in the graph represents the mean prediction over the seven images corresponding to instances associated to the same parent SVE; it can be seen that there is no significant difference between the performance of the model regarding parent SVEs or through the data augmentation procedure.

The investigation has been then moved at the device level. To carry out the training of the model at this stage, a dataset composed of 1839 samples has been gathered, each input being characterized by the specific SVE image and relevant $\hat{E}$ and O values. The output, in terms of the normalized oscillation amplitude, has been obtained from the reduced-order analytical model briefly described in Section 2. The output values of v_{max}/v_{max}^{*}, being $v_{max}^{*} = 28$ nm, have turned out to be characterized by: a mean value of 0.99 and a standard deviation of 0.15, if associated to the training set; a mean value of 0.98 and a standard deviation of 0.15, if associated to the validation set (consisting of 460 samples).
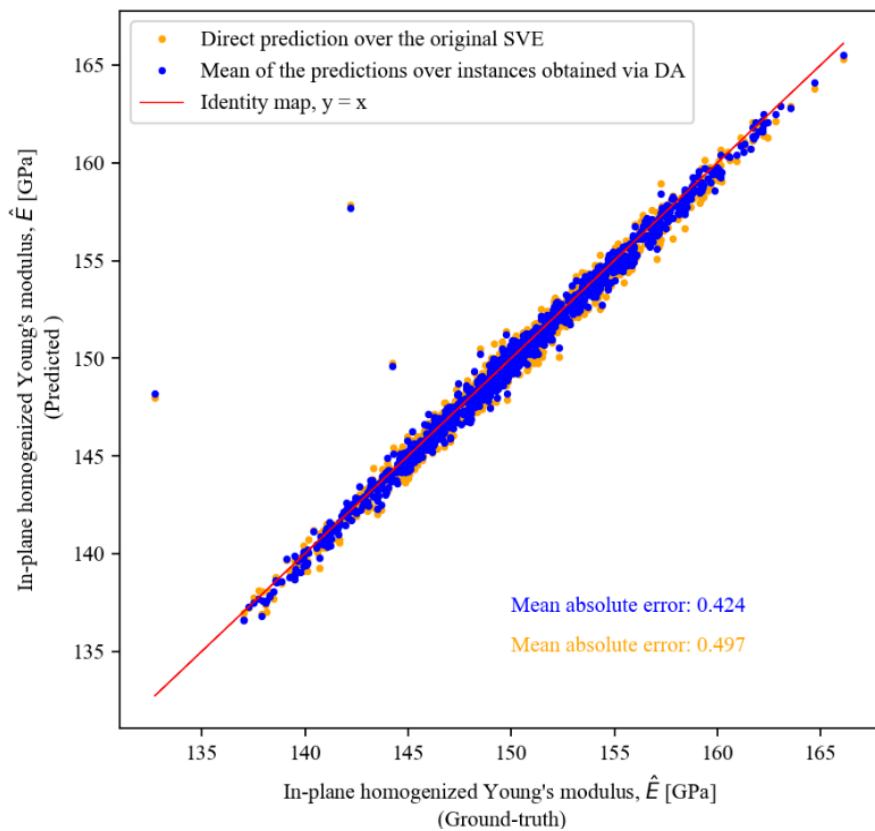

Figure 5. Performance of the ResNet-based regression model.

Although the multiple inputs and mixed data model is constituted by the interconnection of the three sub-architectures discussed in Section 3, the training stage occurs simultaneously over the second stage only. The minimum validation loss has been attained after 327 epochs; afterwards, no considerable improvement has been observed. The parity plot in **Figure 6** testifies the relevant performance, again in terms of the correlation between the predicted values and the corresponding ground-truth data: as expected from the evolution of the loss, not reported for brevity, a noteworthy good agreement between predicted and ground-truth data is observed. The coefficient of determination R^2 almost equal to 1 states that the sum of squares of residuals equals zero, and so the predicted device behaviour lies perfectly along the line of best fit. To finally assess the ability of the model to adapt to new, previously unseen data, drawn from the same probability distributions used to train and tune the model, a test set consisting of 119 samples has been generated. The resulting output values feature a mean value of 0.99 and a standard deviation of 0.14, with the predictions reproducing almost exactly the identity map.

5. Conclusions

A data-driven framework has been proposed and effectively implemented for modeling the response of a resonant Lorentz force MEMS magnetometer. Digitally-generated representative images of the microstructure, together with values of overetch sampled out of a microfabrication-tailored statistical distribution, have been exploited to train the model and provide accurate predictions of the normalized maximum oscillation amplitude of the movable part of the device.
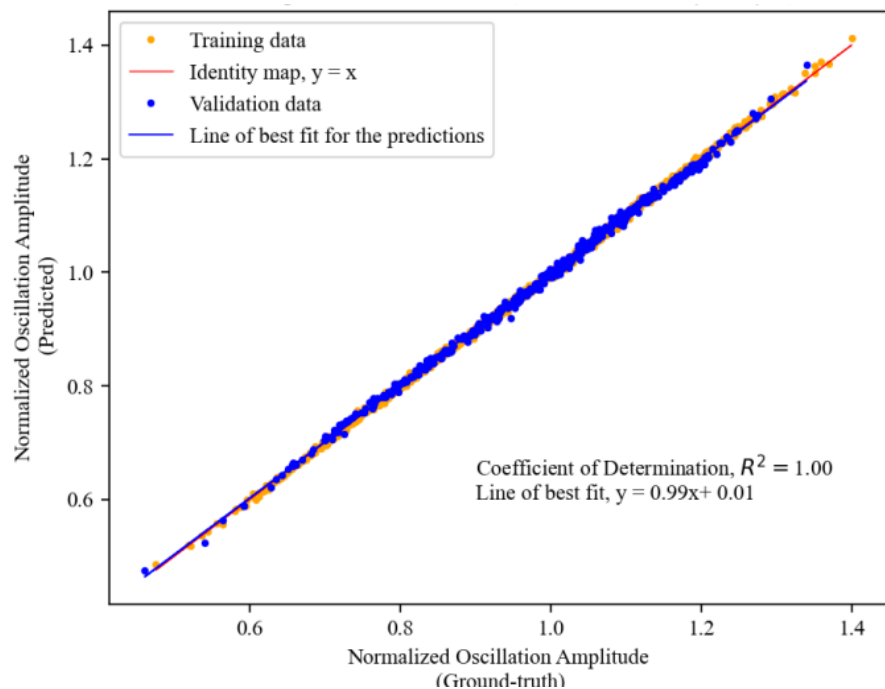

Figure 6. Performance of the multiple inputs and mixed data model.

While the training of the proposed ML-based model depends on ground-truth data obtained through a formerly developed reduced-order model of the device, a novelty and main advantage of the current approach rests on the capability to produce accurate one-to-one mappings on the basis of specific microstructural representative images. Some approximations have been introduced when dealing with the representation of the slender beam microstructure, considered as a random concatenation of one individual SVE and of its corresponding instances. This assumption has avoided a further homogenization procedure to upscale the Young's modulus from the SVE scale to the device one. Accordingly, the over-etch has been assumed homogeneous over the entire device.

In future activities, improvements are envisaged to extend the approach to the most general case by implementing an additional ANN to perform the second foreseen homogenization step. An additional ANN will be trained to lead to automatic defect detection. In this manner, the entire regression model will only require as input the representative image of the entire microstructure, which will ideally encode all the relevant information to account for the uncertainties.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Owens, J.D., Houston, M., Luebke, D., Green, S., Stone, J.E. and Phillips, J.C. (2008) GPU Computing. *Proc. IEEE.*, **96**, 879-899. https://doi.org/10.1109/JPROC.2008.917757

[2] Nickolls, J. and Dally, W.J. (2010) The GPU Computing Era. *IEEE Micro*, **30**, 56-69. https://doi.org/10.1109/MM.2010.41

[3] Schmidt, J., Marques, M.R.G., Botti, S. and Marques, M.A.L. (2019) Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *Npj Comput Mater.*, **5**, 83. https://doi.org/10.1038/s41524-019-0221-0

[4] Himanen, L., Geurts, A., Foster, A.S. and Rinke, P. (2019) Data-Driven Materials Science: Status, Challenges, and Perspectives. *Adv. Sci.*, **6**, Article ID: 1900808. https://doi.org/10.1002/advs.201900808

[5] Bock, F.E., Aydin, R.C., Cyron, C.J., Huber, N., Kalidindi, S.R. and Klusemann, B. (2019) A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics. *Front. Mater.,* **6**, 110. https://doi.org/10.3389/fmats.2019.00110

[6] Haykin, S. (1998) Neural Networks: A Comprehensive Foundation. Prentice Hall. Upper Saddle River, New Jersey.

[7] Russell, S.J. and Norvig, P. (2016) Artificial Intelligence: A Modern Approach.

Pearson Education Limited, Malaysia.

[8] Deng, L. (2014) Deep Learning: Methods and Applications. *FNT in Signal Processing*, **7**, 197-387. https://doi.org/10.1561/2000000039

[9] Schmidhuber, J. (2015) Deep Learning in Neural Networks: An Overview. *Neural Networks*, **61**, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003

[10] Cang, R. and Ren, M.Y. (2016) Deep Network-Based Feature Extraction and Reconstruction of Complex Material Microstructures. In *42nd Design Automation Conference*, American Society of Mechanical Engineers. https://doi.org/10.1115/DETC2016-59404

[11] Lubbers, N., Lookman, T. and Barros, K. (2017) Inferring Low-Dimensional Microstructure Representations Using Convolutional Neural Networks. *Phys. Rev. E.*, **96**, Article ID: 052111. https://doi.org/10.1103/PhysRevE.96.052111

[12] Cecen, A., Dai, H., Yabansu, Y.C., Kalidindi, S.R. and Song, L. (2018) Material Structure-Property Linkages Using Three-Dimensional Convolutional Neural Networks. *Acta Materialia*, **146**, 76-84. https://doi.org/10.1016/j.actamat.2017.11.053

[13] Cang, R., Li, H., Yao, H., Jiao, Y. and Ren, Y. (2018) Improving Direct Physical Properties Prediction of Heterogeneous Materials from Imaging Data via Convolutional Neural Network and a Morphology-Aware Generative Model. *Computational Materials Science*, **150**, 212-221. https://doi.org/10.1016/j.commatsci.2018.03.074

[14] Frankel, A.L., Jones, R.E., Alleman, C. and Templeton, J.A. (2019) Predicting the Mechanical Response of Oligocrystals with Deep Learning. *Computational Materials Science*, **169**, Article ID: 109099. https://doi.org/10.1016/j.commatsci.2019.109099

[15] Bagherinia, M., Bruggi, M., Corigliano, A., Mariani, S. and Lasalandra, E. (2014) Geometry Optimization of a Lorentz Force, Resonating MEMS Magnetometer. *Microelectronics Reliability*, **54**, 1192-1199. https://doi.org/10.1016/j.microrel.2014.02.020

[16] Bagherinia, M. and Mariani, S. (2019) Stochastic Effects on the Dynamics of the Resonant Structure of a Lorentz Force MEMS Magnetometer. *Actuators*, **8**, 36. https://doi.org/10.3390/act8020036

[17] Mariani, S., Martini, R., Ghisi, A., Corigliano, A. and Beghi, M. (2011) Overall Elastic Properties of Polysilicon Films: A Statistical Investigation of the Effects of Polycrystal Morphology. *Int J Mult Comp Eng*, **9**, 327-346. https://doi.org/10.1615/IntJMultCompEng.v9.i3.50

[18] Mirzazadeh, R. and Mariani, S. (2017) Uncertainty Quantification of Microstructure-Governed Properties of Polysilicon MEMS. *Micromachines*, **8**, 248. https://doi.org/10.3390/mi8080248

[19] Mariani, S., Ghisi, A., Mirzazadeh, R. and Azam, S.E. (2018) On-Chip Testing: A Miniaturized Lab to Assess Sub-Micron Uncertainties in Polysilicon MEMS. *Micro and Nanosystems*, **10**, 84-93. https://doi.org/10.2174/1876402911666181204122855

[20] Mirzazadeh, R., Eftekhar Azam, S. and Mariani, S. (2018) Mechanical Characterization of Polysilicon MEMS: A Hybrid TMCMC/POD-Kriging Approach. *Sensors*, **18**, 1243. https://doi.org/10.3390/s18041243

[21] Ghisi, A. and Mariani, S. (2019) Effect of Imperfections due to Material Heterogeneity on the Offset of Polysilicon MEMS Structures. *Sensors*, **19**, 3256. https://doi.org/10.3390/s19153256

[22] Dassi, L., Merola, M., Riva, E., Santalucia, A., Venturelli, A., Ghisi, A. And Mariani, S. (2021) A Stochastic Model to Describe the Scattering in the Response of Polysilicon MEMS. *Engineering Proceedings*, **2**, 95.

https://doi.org/10.3390/engproc2020002095

[23] Hopcroft, M.A., Nix, W.D. and Kenny, T.W. (2010) What Is the Young's Modulus of Silicon? *Journal of Microelectromechanical Systems*, **19**, 229-238. https://doi.org/10.1109/JMEMS.2009.2039697

[24] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *In* 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR). https://doi.org/10.1109/CVPR.2016.90

[25] Quesada-Molina, J.P., Rosafalco, L. and Mariani, S. (2020) Mechanical Characterization of Polysilicon MEMS Devices: a Stochastic, Deep Learning-Based Approach. *In* 21*st International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems* (EuroSimE), 1-8. https://doi.org/10.1109/EuroSimE48426.2020.9152690