Artificial Intelligence: New Frontiers in Real-Time Inverse Scattering and Electromagnetic Imaging

Marco Salucci, *Senior Member, IEEE*, Manuel Arrebola, *Senior Member, IEEE*, Tao Shan, *Member, IEEE*, and Maokun Li, *Senior Member, IEEE*

***Abstract*—In recent years, artificial intelligence (AI) techniques have been developed rapidly. With the help of big data, massive parallel computing, and optimization algorithms, machine learning (ML) and (more recently) deep learning (DL) strategies have been equipped with enhanced learning and generalization capabilities. Besides becoming an essential framework in image and speech signal processing, AI has been also widely applied to solve several electromagnetic (EM) problems with unprecedented computational efficiency, including inverse scattering and EM imaging. In this paper, a review of the most recent progresses in the application of ML and DL for such problems is given. We humbly hope a brief summary could help us to better understand the pros and cons of this research topic and foster future research in using AI to address paramount challenges in the field of EM vision.**

***Keywords*—Inverse scattering, electromagnetic imaging, artificial intelligence, machine learning, deep learning, learning by examples**

I. INTRODUCTION

During the last decades, Artificial Intelligence (AI) has been an important research topic in signal processing, computing science, and mathematical areas. Recently, the application of such techniques is rapidly spreading to many other fields, including electromagnetics (EM) [1]. Although pioneer works on the application of Machine Learning (ML) techniques to different problems in antennas, arrays, EM scattering or imaging are dated some years ago, they were mainly proof-of-concepts or they were dealing with size limited scenarios/applications. Currently, the real interest in the application of AI to EM problems comes from the potential of the more sophisticated Deep Learning (DL) paradigm, which is emerging as a powerful framework enabling unprecedented time and accuracy performance for solving a wide variety of complex EM problems [2]–[13].

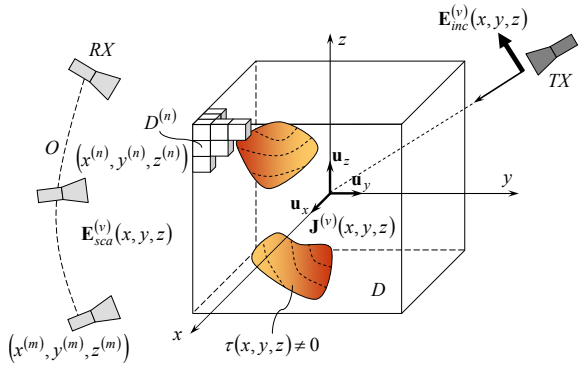


Fig. 1. Geometry of the *3D ISP*.

The number of AI-related publications in EM problems has been exponentially growing during the last years. However, most of them are works presented in conferences, which show partial or preliminary results but not a mature technique. Actually, the papers discussing more complete studies are published in journals and they are almost restricted to the last three years. When publications in ML or DL in the field of EM imaging or scattering are searched in IEEE Xplore database, most of journal papers have been published in 2020 or 2021, demonstrating the novelty of the topic as well as its current great interest because of the potential benefit that can be obtained from its development.

Although the development or improvement of EM techniques based on AI is at the beginning, many research works have been recently carried out in a number of applications such as EM scattering [2], [3], radar and remote sensing [4], target classification [5], shape reconstruction [6], inverse design [7], [8] and fingerprinting [9] of EM devices, and medical imaging [10]. Moreover, an ever-growing number of papers have been published on localization [11], human behavior monitoring [12], and EM compatibility (EMC) [13], among others.

Within this context, the aim of this paper is to provide a comprehensive review the most recent applications of ML

This work benefited from the networking activities carried out within the Project "CYBER-PHYSICAL ELECTROMAGNETIC VISION: Context-Aware Electromagnetic Sensing and Smart Reaction (EMvisioning)" (Grant no. 2017HZJXSZ) funded by the Italian Ministry of Education, University, and Research within the PRIN2017 Program (CUP: E64I19002530001) and the Project "SMARTOUR - Piattaforma Intelligente per il Turismo" (Grant no. SCN_00166) funded by the Italian Ministry of Education, University, and Research within the Program "Smart cities and communities and Social Innovation" (CUP: E44G14000040008).

The work was also supported in part by the Spanish Ministry of Science and Innovation and the Spanish Agency for Research within project ENHANCE-5G (PID2020-114172RB-C21 / AEI / 10.13039/501100011033) and the Government of Principality of Asturias within project AYUD/2021/51706.

It was also supported in part by the National Natural Science Foundation of China (61971263), the National Key R&D Program of China (2018YFC0603604), Institute for Precision Medicine, Tsinghua University, Beijing, China, and THE XPLORER PRIZE.

Marco Salucci is with the ELEDIA Research Center (ELEDIA@UniTN - University of Trento), DICAM - Department of Civil, Environmental, and Mechanical Engineering, Via Mesiano 77, 38123 Trento - Italy, e-mail: *marco.salucci@unitn.it*.

Manuel Arrebola is with Universidad de Oviedo, Department of Electrical Engineering, Signal Theory and Communications Group, Calle Luis Ortiz Berrocal, Campus Universitario, E33203 Gijon, Asturias - Spain, e-mail: *arrebola@uniovi.es*.

Maokun Li and Tao Shan are with Beijing National Research Center for Information Science and Technology, Microwave and Antenna Institute, Department of Electronic Engineering, Tsinghua University, 100084 Beijing - China, e-mail: *maokunli@tsinghua.edu.cn, shantao@tsinghua.edu.cn*.

and DL techniques to inverse scattering (IS) and EM imaging problems. To this end, Section II provides the mathematical formulation of the problem at hand. Section III and Section IV describe the learning-by-examples (LBE) framework and the most attracting/advanced DL methodologies for its solution, respectively. Then, Section V reviews the most recent applications of such techniques to several inverse scattering problems (ISPs) and imaging problems. Finally, some conclusions, remarks, trends, and possible roadmaps are discussed in Sect. VI, identifying challenges and opportunities in this widely unexplored research field, as well.

II. MATHEMATICAL FORMULATION

Without loss of generality[1], let us consider a three-dimensional (*3D*) volume D embedded within a homogeneous, isotropic, non-magnetic ($\mu = \mu_0 = 4\pi \times 10^{-7}$ [H/m]), and lossless ($\sigma = \sigma_0 = 0$ [S/m]) background medium of permittivity $\varepsilon = \varepsilon_0 = 8.85 \times 10^{-12}$ [F/m] [Fig. 1]. Under the hypothesis of monochromatic time-harmonic EM fields, the Maxwell's equations governing at frequency f all scattering phenomena inside/outside D can be rewritten in terms of the electric field phasor $\mathbf{E}(x,y,z) = E_x(x,y,z)\,\mathbf{u}_x + E_y(x,y,z)\,\mathbf{u}_y + E_z(x,y,z)\,\mathbf{u}_z$ [$E_p(x,y,z) \in \mathbb{C}$ and $\mathbf{u}_p$ being the unit vector along the p-th Cartesian coordinate, $p = \{x,y,z\}$ - Fig. 1], corresponding to its time-domain representation, $\mathbf{E}(x,y,z,t) = \Re\{\mathbf{E}(x,y,z)\exp(-j\omega t)\}$, where $\Re\{\,.\,\}$ is the real part, $j = \sqrt{-1}$, and $\omega = 2\pi f$. In order to non-invasively retrieve a guess of its *EM* characteristics, D is illuminated by means of V incident waves with associated electric fields $\mathbf{E}_{inc}^{(v)}(x,y,z)$, $v = 1,...,V$ [Fig. 1]. Defining the scattered field for $\forall\,(x,y,z) \in \mathbb{R}^3$ as

$$\mathbf{E}_{sca}^{(v)}(x,y,z) = \mathbf{E}_{tot}^{(v)}(x,y,z) - \mathbf{E}_{inc}^{(v)}(x,y,z)\,;\;\; v = 1,...,V \tag{1}$$

where $\mathbf{E}_{tot}^{(v)}(x,y,z)$ is the total electric field measured in presence of unknown scatterers, it turns out that, for $v = 1,...,V$,

$$\mathbf{E}_{sca}^{(v)}(x,y,z) = \omega^2 \varepsilon_0 \mu_0 \int\int\int_D \mathbf{G}(x,y,z,x',y',z') \cdot \mathbf{J}^{(v)}(x',y',z')\,dx'dy'dz'. \tag{2}$$

In (2) $\mathbf{G}(\,.\,)$ is the free-space dyadic Green's function [14]. Moreover,

$$\mathbf{J}^{(v)}(x,y,z) = \tau(x,y,z)\,\mathbf{E}_{tot}^{(v)}(x,y,z)\,;\; v = 1,...,V \tag{3}$$

is the induced equivalent current in D arising from the unknown contrast distribution

$$\tau(x,y,z) = [\varepsilon_r(x,y,z) - 1] + j\frac{\sigma(x,y,z)}{\omega\varepsilon_0} \tag{4}$$

where $\varepsilon_r(x,y,z) \geq 1$ and $\sigma(x,y,z) \geq 0$ are the relative permittivity and conductivity distributions, respectively [$\Rightarrow \tau(x,y,z) \geq 0$ for $(x,y,z) \in D$ - Fig. 1]. The scattered field radiated by the V equivalent sources is computed from the measured incident/total fields according to (1) at M probing locations $\left(x^{(m)}, y^{(m)}, z^{(m)}\right)$, $m = 1,...,M$ forming the so-called observation domain $O \notin D$ [Fig. 1].

In order to numerically solve the ISP at hand, a voxel-wise representation of the *EM* properties of D is adopted by employing N *3D* pulse basis functions

$$\mathcal{B}^{(n)}(x,y,z) = \begin{cases} 1 & \text{if} \quad (x,y,z) \in D^{(n)} \\ 0 & \text{otherwise} \end{cases} \qquad n = 1,...,N \tag{5}$$

$D^{(n)}$ being the n-th voxel of barycenter $\left(x^{(n)}, y^{(n)}, z^{(n)}\right)$ [$D = \bigcup_{n=1}^{N} D^{(n)}$ - Fig. 1]. Accordingly, the unknown equivalent currents for each v-th $(v = 1,...,V)$ illuminating direction and contrast distributions can be expressed as

$$\mathbf{J}^{(v)}(x,y,z) = \sum_{p=\{x,y,z\}} \sum_{n=1}^{N} J_p^{(v,n)} \mathcal{B}^{(n)}(x,y,z)\,\mathbf{u}_p; \tag{6}$$

and

$$\tau(x,y,z) = \sum_{n=1}^{N} \tau^{(n)} \mathcal{B}^{(n)}(x,y,z) \tag{7}$$

respectively, where $J_p^{(v,n)} = J_p^{(v)}\left(x^{(n)}, y^{(n)}, z^{(n)}\right)$ is the p-th ($p = \{x,y,z\}$) scalar component of the induced current, and $\tau^{(n)} = \tau\left(x^{(n)}, y^{(n)}, z^{(n)}\right)$, $n = 1,...,N$. Therefore, it is possible to rewrite (2) in compact matrix form as

$$\left[\begin{array}{ccc} \underline{E}_{sca,x}^{(v)} & \underline{E}_{sca,y}^{(v)} & \underline{E}_{sca,z}^{(v)} \end{array}\right]^T = \underline{\underline{G}}_{3D} \left[\begin{array}{ccc} \underline{J}_x^{(v)} & \underline{J}_y^{(v)} & \underline{J}_z^{(v)} \end{array}\right]^T \tag{8}$$

where $.^T$ is the transpose operator, $\underline{E}_{sca,p}^{(v)} = \left[E_{sca,p}^{(v,m)};\, m = 1,...,M\right]^T$, $E_{sca,p}^{(v,m)} = E_{sca,p}^{(v)}\left(x^{(m)}, y^{(m)}, z^{(m)}\right)$, and $\underline{J}_p^{(v)} = \left[J_p^{(v,n)};\, n = 1,...,N\right]^T$, $p = \{x,y,z\}$. Moreover, $\underline{\underline{G}}_{3D} \in \mathbb{C}^{3M \times 3N}$ is the external Green's matrix for the *3D* scenario at hand [14].

In many practical scenarios a *2D* reconstruction is sought instead of a *3D* one. Towards this end, Equation (2) is simplified to a scalar one under the assumption that the electric field is transverse-magnetic (*TM*)-polarized [i.e., $\mathbf{E}(x,y,z) = E_z(x,y,z)\,\mathbf{u}_z$] and that the unknown scatterers are infinitely-extended cylinders with invariant properties along the z-axis. Under such hypotheses,

$$E_{sca,z}^{(v)}(x,y) = \omega^2 \varepsilon_0 \mu_0 \int\int_D g_{2D}(x,y,x',y')\,J_z^{(v)}(x',y')\,dx'dy';\; v = 1,...,V \tag{9}$$

where $J_z^{(v)}(x',y') = \tau(x',y')\,E_{tot,z}^{(v)}(x',y')$ and $g_{2D}(x,y,x',y')$ is the *2D* Green's function [14]. Finally, after discretizing D into N pixels centered at $\left(x^{(n)}, y^{(n)}\right)$, $n = 1,...,N$, the following matrix expression is yielded

$$\underline{E}_{sca,z}^{(v)} = \underline{\underline{G}}_{2D} \underline{J}_z^{(v)}; \quad v = 1,...,V \tag{10}$$

where $\underline{\underline{G}}_{2D} \in \mathbb{C}^{M \times N}$ is the discrete Green's operator for the *2D* scenario [14]. In the following, the *AI*-driven computationally-efficient solution of the *3D*/*2D ISP* described

[1]For the sake of notation simplicity, free-space imaging is considered in the following [14] [15]. The extension to more complex scenarios (e.g., subsurface [16], biomedical [17] [18], through-the-wall [19] [20], *NDT/NDE* [21] [22]) only requires proper reformulations of the involved Green's operators and primary/induced sources.

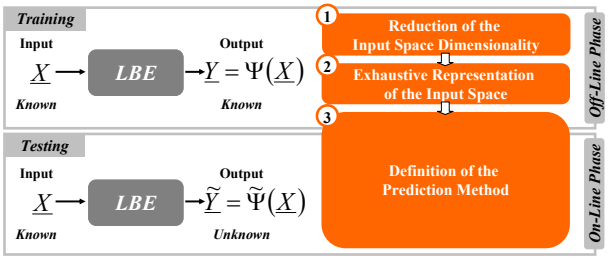


Fig. 2. (*LBE Solution Strategies*) - Equivalent definitions of *LBE* strategies as two-steps or three-steps approaches.

by equations (8)/(10) is formulated within the Learning-by-Examples (*LBE* - Sect. III) and Deep Learning (*DL* - Sect. IV) frameworks, respectively.

III. LEARNING-BY-EXAMPLES SOLUTION STRATEGIES

The term Learning-by-Examples (*LBE*) refers to a specific class of ML algorithms also known as "supervised learning" techniques. Their goal is to predict the unknown Q-dimensional output/response,

$$\underline{\widetilde{\mathcal{Y}}} = \left\{ \widetilde{\mathcal{Y}}_q;\ q = 1, ..., Q \right\} \tag{11}$$

associated to a known input sample of K variables/features

$$\underline{\mathcal{X}} = \{ \mathcal{X}_k;\ k = 1, ..., K \} \tag{12}$$

by means of a computationally-efficient but accurate surrogate model (*SM*)/digital twin (*DT*) $\widetilde{\Psi}(\,.\,)$ of the actual (but time-costly) input/output (*I/O*) function $\Psi(\,.\,) : \underline{\mathcal{X}} \rightarrow \underline{\mathcal{Y}}$. Mathematically, the *LBE* goal is formulated as follows

$$\begin{aligned} &\underline{\widetilde{\mathcal{Y}}} = \widetilde{\Psi}\left(\underline{\mathcal{X}}\right) \\ &\text{subject to } \left\| \underline{\widetilde{\mathcal{Y}}} - \underline{\mathcal{Y}} \right\| \rightarrow 0 \text{ and } \widetilde{\Delta t} \ll \Delta t \end{aligned} \tag{13}$$

where $\|\,.\,\|$ is the ℓ_2-norm, $\underline{\mathcal{Y}} = \Psi\left(\underline{\mathcal{X}}\right) = \{\mathcal{Y}_q;\ q = 1, ..., Q\}$ is the actual response of $\underline{\mathcal{X}}$, while $\widetilde{\Delta t}$ and Δt are the time required to make a prediction or an exact evaluation (through simulations or experiments) of the *I/O* relationship, respectively.

Depending on the nature of $\underline{\mathcal{Y}}$, two main families of *LBE* approaches can be identified, namely (*a*) classification and (*b*) regression techniques. Methods belonging to group (*a*) are aimed at retrieving a discrete-valued/integer label $\mathcal{Y} = \Psi\left(\underline{\mathcal{X}}\right) = \mathcal{L}\left(\underline{\mathcal{X}}\right) \in \mathbb{Z}$ ($Q = 1$) among a predefined set of $\mathcal{C}$ options/classes (i.e., $\mathcal{L}\left(\underline{\mathcal{X}}\right) \in \left\{ \mathcal{L}^{(c)};\ c = 1, ..., \mathcal{C} \right\}$). Otherwise, group (*b*) strategies predict a continuous-valued output $\underline{\mathcal{Y}} = \Psi\left(\underline{\mathcal{X}}\right) \in \mathbb{R}^Q$. The term "supervised" comes from the fact that a training set of S known *I/O* examples

$$\mathbb{T} = \left\{ \left[\underline{\mathcal{X}}^{(s)}, \underline{\mathcal{Y}}^{(s)} = \Psi\left(\underline{\mathcal{X}}^{(s)} \right) \right];\quad s = 1, ..., S \right\} \tag{14}$$

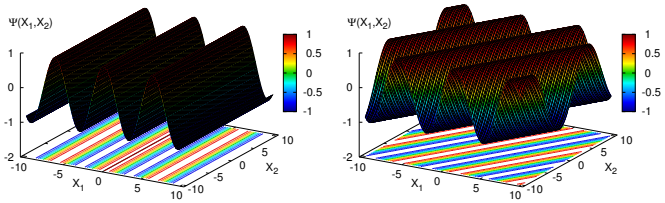
is exploited to build the *SM* $\widetilde{\Psi}(\,.\,)$ according to a "learning with a teacher" process [2].

[2]For completeness, "unsupervised learning" techniques such as clustering and association are aimed at automatically analyzing the structure of the input data to determine underlying similarities/relationships (i.e., "learning without a teacher").

There are many possible strategies to solve the *3D/2D ISP* (8)/(10) with high computational efficiency (often in real-time) within the *LBE* framework. They can be grouped in two main classes: (*i*) inverse learning (*IL*) and (*ii*) forward learning (*FL*) techniques. *IL* approaches directly solve the *ISP* by inverting the unknown relationship between data and unknown sources. The most *naive IL* implementation based on regression is to retrieve the unknown contrast function (i.e., $\underline{\mathcal{Y}} \leftarrow \underline{\tau}$, being $\underline{\tau} = \left\{ \tau^{(n)};\ n = 1, ..., N \right\}$) or the induced currents (i.e., $\underline{\mathcal{Y}} \leftarrow \underline{J}$, being $\underline{J} = \left\{ \underline{J}_p^{(v)};\ v = 1, ..., V;\ p = x, y, z \right\}$) starting from the scattered field (i.e., $\underline{\mathcal{X}} \leftarrow \underline{E}$, being $\underline{E} = \left\{ \underline{E}_{sca,p}^{(v)};\ v = 1, ..., V;\ p = x, y, z \right\}$) or equivalent/transformed data representations (e.g., the scattering matrix measured by a vector network analyzer [23]). However, owing to the very high complexity of predicting pixel-wise distributions because of the very large output space dimensionality (e.g., $Q \geq N$), "parametric" *IL* strategies are often a preferred choice [24] [25]. In this case, the *DT* is trained to recover a set of Q parameters/descriptors of the domain under test, i.e., $\underline{\mathcal{Y}} \leftarrow \underline{\mathcal{P}} = \{\mathcal{P}_q;\ q = 1, ..., Q\}$. For instance, the *ISP* can be reformulated to retrieve the position (i.e., $\underline{\mathcal{P}} = \left\{ x^{obj}, y^{obj}, z^{obj} \right\}$ [25] [26]) of an unknown object within D. Shaping/qualitative imaging is possible, as well, by letting the *SM* predicting geometrical descriptors of the target according to *a-priori* chosen parametric models (e.g., $\underline{\mathcal{P}} = \mathcal{P}_1 = \rho^{obj}$ for circular/spherical objects of radius ρ^{obj} [26], or $\underline{\mathcal{P}} = \left\{ l_x^{obj}, l_y^{obj}, l_z^{obj} \right\}$ for parallelepipeds of length l_x^{obj}, width l_y^{obj}, and depth l_z^{obj}, respectively [25]). It is worth pointing out that parametric approaches are typically used when the target can be modeled with a limited number of parameters. These latter (along with the model itself) must be carefully selected exploiting the underlying assumptions about the classes of shapes to which the scatterer can belong [27]. Finally, quantitative reconstructions can be performed by assuming that homogeneous scatterers are at hand (i.e., $\varepsilon_r\left(x, y, z\right) = \varepsilon_r^{obj}$ and $\sigma\left(x, y, z\right) = \sigma^{obj}$ for $\forall \left(x, y, z\right) \in \Omega$, Ω being the target support $\Rightarrow \underline{\mathcal{P}} = \left\{ \varepsilon_r^{obj}, \sigma^{obj} \right\}$ [24] [26]).

As for *IL* techniques implemented via classification strategies, the predicted label often corresponds to a specific "status" of D. Accordingly, multi-step approaches can be implemented to progressively infer information on the unknown targets starting from available data in O. As an example, in brain stroke imaging the *ISP* can be decomposed in sub-tasks triggered in a cascaded fashion and devoted at (*i*) detecting the presence of a stroke (i.e., binary classification problem, $\mathcal{C} = 2$: stroke present, $\mathcal{L}^{(1)} = 1$, or stroke absent, $\mathcal{L}^{(2)} = -1$), (*ii*) identifying its "nature" (i.e., $\mathcal{C} = 2$: ischaemic stroke, $\mathcal{L}^{(1)} = 1$, or hemorrhagic stroke, $\mathcal{L}^{(2)} = -1$), and finally (*iii*) localizing it (i.e., multi-class problem, $\mathcal{C} > 2$: $\mathcal{L}\left(\underline{\mathcal{X}}\right) \in \left\{ \mathcal{L}^{(c)};\ c = 1, ..., \mathcal{C} \right\}$, $\mathcal{L}^{(c)}$ identifying one among $\mathcal{C}$ sub-regions of D) [23].

On the other hand, *FL* strategies are based on the prediction of the forward scattering operator/phenomena (e.g., $\underline{\mathcal{Y}} \leftarrow \underline{E}$) starting from the pixel-wise (e.g., $\underline{\mathcal{X}} \leftarrow \underline{\tau}$) or parametric (i.e., $\underline{\mathcal{X}} \leftarrow \underline{\mathcal{P}}$) description of the unknown targets. Accordingly, the *DT* is regarded as a computationally-fast replacement of accurate but time-consuming forward scattering solvers when



(a) (b)

Fig. 3. (*LBE Solution Strategies*) - Plot of the $K = 2$-dimensional benchmark functions (*a*) $\Psi(\underline{\mathcal{X}}) = \cos(\mathcal{X}_1 + 0\mathcal{X}_2)$ and (*b*) $\Psi(\underline{\mathcal{X}}) = \cos\left(\mathcal{X}_1 \cos\left(\frac{\pi}{6}\right) - \mathcal{X}_2 \sin\left(\frac{\pi}{6}\right)\right)$.

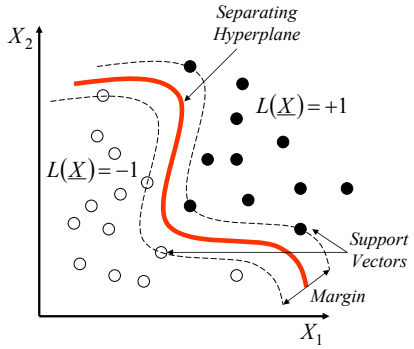


Fig. 4. (*LBE Solution Strategies*) - Pictorial sketch of the binary *SVM* classifier.

solving the *ISP* through iterative optimization approaches. In this framework, an effective recipe to mitigate the complexity of the underlying learning task is to exploit a regressor to directly estimate the data mismatch function to minimize, rather than the data itself. More precisely, a single scalar ($Q = 1$) is predicted quantifying the difference between measured, $\left\{\underline{E}_{sca,p}^{(v)};\, v = 1, ..., V;\, p = x, y, z\right\}$, and retrieved, $\left\{\underline{\widetilde{E}}_{sca,p}^{(v)}(\underline{\mathcal{X}});\, v = 1, ..., V;\, p = x, y, z\right\}$, data associated to a given trial guess $\underline{\mathcal{X}}$

$$\mathcal{Y}(\underline{\mathcal{X}}) = \frac{\sum_{p=\{x,y,z\}} \sum_{v=1}^{V} \sum_{m=1}^{M} \left| E_{sca,p}^{(v,m)} - \widetilde{E}_{sca,p}^{(v,m)}(\underline{\mathcal{X}}) \right|^2}{\sum_{p=\{x,y,z\}} \sum_{v=1}^{V} \sum_{m=1}^{M} \left| E_{sca,p}^{(v,m)} \right|^2}. \tag{15}$$

IL and FL strategies are commonly formulated as two-phases approaches as follows

1) *"Off-Line" Phase* - Build the training set $\mathbb{T}$ by generating S input samples within physically admissible bounds, $\underline{\mathcal{X}}^{(\min)} \leq \underline{\mathcal{X}}^{(s)} \leq \underline{\mathcal{X}}^{(\max)}$, $s = 1, ..., S$, then merging them with the corresponding simulated/measured responses $\underline{\mathcal{Y}}^{(s)}$, $s = 1, ..., S$. Finally, use $\mathbb{T}$ to train the *LBE* strategy and let it "learn" the *I/O* relationship.
2) "*On-Line*" *Phase* - Input to the trained *LBE* a previously-unseen test sample $\underline{\mathcal{X}}$ and let it predict the corresponding output as $\underline{\widetilde{\mathcal{Y}}} = \Psi(\underline{\mathcal{X}})$.

More in general, the *LBE* framework can be regarded as a three-step paradigm (Fig. 2), as described in the following Sections.

A. *Step 1:* Reduction of the Input Space Dimensionality

Because of the so-called "curse of dimensionality" [28] the number S of *I/O* training samples required to build a high-fidelity *DT* exponentially grows with the input space dimensionality, K. Therefore, when K is large [3], it is convenient to reduce the input space dimensionality by deriving a reduced set of $K' \ll K$ highly-informative features $\underline{\mathcal{X}}' = \{\mathcal{X}_k';\, k = 1, ..., K'\}$ so that both the learning complexity and the cost of the training phase are mitigated and, most importantly, the number of required samples S is kept as low as possible. Towards this goal, function independent/dependent feature selection/extraction techniques can be exploited to derive $\underline{\mathcal{X}}'$ from $\underline{\mathcal{X}}$ and build a reduced training set $\mathbb{T}' = \left\{\left[\underline{\mathcal{X}}'^{(s)}, \underline{\mathcal{Y}}^{(s)} = \Psi\left(\underline{\mathcal{X}}^{(s)}\right)\right];\;\; s = 1, ..., S\right\}$ to be exploited instead of $\mathbb{T}$.

Indicating with $\underline{\underline{\mathcal{X}}} = \left\{\underline{\mathcal{X}}^{(s)};\, s = 1, ..., S\right\}$ and with $\underline{\underline{\mathcal{Y}}} = \left\{\underline{\mathcal{Y}}^{(s)};\, s = 1, ..., S\right\}$ the collected *I/O* samples in $\mathbb{T}$, respectively, "function independent" techniques derive $\underline{\underline{\mathcal{X}}}' = \left\{\underline{\mathcal{X}}'^{(s)};\, s = 1, ..., S\right\}$ by analyzing the data distribution within the input space (i.e., $\underline{\underline{\mathcal{X}}}' = \aleph\left(\underline{\underline{\mathcal{X}}}\right)$). Otherwise, "function dependent" strategies rely on the "observed" relation between $\underline{\underline{\mathcal{X}}}$ and $\underline{\underline{\mathcal{Y}}}$ (i.e., $\underline{\underline{\mathcal{X}}}' = \aleph\left(\underline{\underline{\mathcal{X}}}, \underline{\underline{\mathcal{Y}}}\right)$). Moreover, "feature selection" refers to the identification of a subset of K' features from the original K ones which are carrying the largest amount of information on the *I/O* function to predict. Differently, "feature extraction" is a process aimed at deriving a new set of K' features from the original ones. To better understand the difference between feature selection and extraction, let us consider in the following two rather simple and intuitive examples. In the first example, a regression model must be created to predict the function $\Psi(\underline{\mathcal{X}}) = \cos(\mathcal{X}_1 + 0\mathcal{X}_2)$ [$K = 2$, $Q = 1$ - Fig. 3(*a*)]. Clearly, $\Psi(\underline{\mathcal{X}})$ depends only on variable $\mathcal{X}_1$. Therefore, $\mathcal{X}_2$ can be simply discarded because it has no impact on the output, by letting $\mathcal{X}_1' = \mathcal{X}_1$ and $K' = 1$ (i.e., $\mathcal{X}_1$ is the selected feature). On the other hand, if $\Psi(\underline{\mathcal{X}}) = \cos\left(\mathcal{X}_1 \cos\left(\frac{\pi}{6}\right) - \mathcal{X}_2 \sin\left(\frac{\pi}{6}\right)\right)$ [Fig. 3(*b*)] then both $K = 2$ features have an equal impact on the output. However, it is possible to extract a new single ($K' = 1$) feature $\mathcal{X}' = \mathcal{X}_1 \cos\left(\frac{\pi}{6}\right) - \mathcal{X}_2 \sin\left(\frac{\pi}{6}\right)$ which completely captures the actual behavior of the *I/O* relation. Finally, as for the strengths and limitations of each technique, function dependent strategies are generally more appropriate than function independent ones in regression problems (e.g., for NDT-NDE [25]). On the other hand, it should be pointed out that the main advantage of feature extraction strategies over feature selection ones is that they do not require cumbersome sensitivity analyses to understand which features should be kept and which could be discarded [29].

In the following, let us briefly recall the basics of two feature extraction techniques for *ISP*s, namely the Principal

[3] Although such a claim is rather vague, it is not possible to rigorously *a-priori* indicate whether a given value of K should be considered as "large", because such a definition depends on the *ISP* at hand as well as on the adopted *LBE* strategy to solve it. However, to provide the reader with a general-purpose "rule-of-thumb" it is possible to consider $K > 50$ as an indicative threshold for defining a large-dimensionality *LBE* problem.

Component Analysis (*PCA*) [30] and the Partial Least Squares (*PLS*) [29].

Principal Component Analysis (*PCA*): The *PCA* is a function independent feature extraction technique. Given S K-dimensional input samples $\underline{\underline{\mathcal{X}}}$, its goal is to find the $K' \ll K$ "principal components", i.e., the directions along which the largest variance of data is observed. More in detail, the *PCA* applies the following linear transformation [30]

$$\underline{\underline{\mathcal{X}}}' = \underline{\underline{\mathcal{X}}} \times \underline{\underline{\mathcal{W}}} \tag{16}$$

where the columns of the $(K \times K')$ weight matrix $\underline{\underline{\mathcal{W}}}$, $\{\underline{\mathcal{W}}_k;\ k = 1, ..., K'\}$ are the eigenvectors of the covariance matrix of $\underline{\underline{\mathcal{X}}}$.

Partial Least Squares (*PLS*): Differently from the *PCA*, the *PLS* is a function dependent feature extraction technique. It seeks a new reduced set of features $\underline{\underline{\mathcal{X}}}'$ such that the *I/O* relationship is linearized as much as possible [29]. Towards this end, it expresses the output as follows

$$\underline{\underline{\mathcal{Y}}} = \left(\underline{\underline{\mathcal{X}}}' \times \underline{\underline{\mathcal{Q}}}^T\right) + \underline{\underline{\mathcal{E}}} \tag{17}$$

where $\underline{\underline{\mathcal{Q}}}$ is a $(Q \times K')$ matrix of weights, $\left(\underline{\underline{\mathcal{X}}}' \times \underline{\underline{\mathcal{Q}}}^T\right)$ is a linear approximation of $\underline{\underline{\mathcal{Y}}}$, and $\underline{\underline{\mathcal{E}}} = \underline{\underline{\mathcal{Y}}} - \left(\underline{\underline{\mathcal{X}}}' \times \underline{\underline{\mathcal{Q}}}^T\right)$ expresses the approximation error. There are in the literature several iterative approaches for deriving the optimal *PLS* transformation matrix $\underline{\underline{\mathcal{W}}}$ (16) such that $\underline{\underline{\mathcal{E}}}$ is minimized (e.g., the *SIMPLS* [25]). It should be pointed out that hypothesis (17) may be inaccurate for highly non-linear *I/O* relationships. For such cases, non-linear *PLS* algorithms have been proposed such as the orthogonal kernel *PLS* (*OKPLS*) [31], which exploits the so-called "kernel trick" [32] to reformulate the linear approximation (17) inside a higher-dimensionality space.

B. Step 2: Exhaustive Representation of the Input Space

Once the *ISP-DoF*s have been defined, the purpose of this step is to select the S *I/O* pairs to build $\mathbb{T}$ [4]. One-shot or iterative/adaptive sampling techniques can be exploited towards this goal. The former class of strategies performs a non-iterative sampling of the input space in order to comply with the "input space filling" (*ISF*) property, that is obtaining the most uniform possible spreading of training samples within the physically admissible bounds. The uniform full-factorial sampling is the most common strategy to achieve *ISF*. However, it becomes rapidly unfeasible since the number of generated training samples exponentially grows with the number K of uniformly-quantized variables. Orthogonal Arrays (*OA*s) [33] and Latin Hypercube Sampling (*LHS*) [34] are effective and widespread alternatives mitigating such an issue.

On the other hand, adaptive sampling techniques are based on the iterative selection of new training samples to yield a suitable balance between "exploration" (i.e., sampling regions with a low density of samples) and "exploitation" (i.e., adding samples where a large non-linearity of the underlying function has been observed) [35] [36] [37]. LOLA-Voronoi [35] and Output Space Filling (*OSF*) [24] [25] [31] [38] are two techniques belonging to this group.

In the following, two common sampling strategies in *ISP*s (i.e., the *LHS* and the *OSF*) are briefly described.

Latin Hypercube Sampling (*LHS*): *LHS* is a single-shot strategy with *ISF* properties [34]. Although it is based on a pseudo-random exploration of the input space, it mitigates - differently from the uniform random sampling - undesired phenomena of "under-sampling" and "over-sampling" for a user-defined number of training samples S. It works as follows

1) Divide the range of each variable into S uniform intervals;
2) For each dimension ($k = 1, ..., K$), randomly choose one point inside each s-th interval;
3) For each point of variable $k = 1$, randomly select one point of variable $k = 2$, ..., $k = K$ to form a K-dimensional sample;
4) Repeat Step 3 until S K-dimensional samples have been generated, each time excluding already selected points.

Output Space Filling (*OSF*): The *OSF* is an adaptive sampling strategy for solving *ISP*s within the *LBE* framework [24] [25] [31] [38]. In [24] [25] [31] it is exploited for "parametric" *IL* inversion to adaptively sample the parameters space generating S configurations of the unknown scatterers $\left\{\underline{\mathcal{P}}^{(s)};\ s = 1, ..., S\right\}$ such that a uniform exploration of the extracted features space, $\left\{\underline{\mathcal{X}}'^{(s)};\ s = 1, ..., S\right\}$, is yielded. It works as follows:

- *Step 1 - Initialization*
  1) Use the *LHS* to generate $S = S_0$ scatterer configurations, $\underline{\underline{\mathcal{Y}}}_0 = \left\{\underline{\mathcal{P}}^{(s)};\ s = 1, ..., S_0\right\}$, within physically-admissible bounds, then compute the scattered data associated to each sample, $\underline{\underline{\mathcal{X}}}_0 = \left\{\underline{E}^{(s)};\ s = 1, ..., S_0\right\}$;
  2) Apply a feature extraction technique (e.g., the *PLS*) to $\underline{\underline{\mathcal{X}}}_0$ and $\underline{\underline{\mathcal{Y}}}_0$ to derive the transformation rule $\underline{\underline{\mathcal{X}}}'_0 = \aleph\left(\underline{\underline{\mathcal{X}}}_0, \underline{\underline{\mathcal{Y}}}_0\right)$, then form the initial training set of S_0 samples as $\mathbb{T}' = \left(\underline{\underline{\mathcal{X}}}'_0, \underline{\underline{\mathcal{Y}}}_0\right)$;
- *Step 2 - Adaptive Sampling*
  1) Use the *LHS* to select Z "candidate" scatterer configurations $\left\{\underline{\mathcal{P}}^{(z)};\ z = 1, ..., Z\right\}$;
  2) Use the information in $\mathbb{T}'$ to quickly predict (via multi-dimensional linear interpolation [25]) the extracted features for each candidate, $\left\{\underline{\widetilde{\mathcal{X}}}'^{(z)};\ z = 1, ..., Z\right\}$;
  3) Select the candidate maximizing the minimum distance in the extracted features space from all samples in $\mathbb{T}'$,

$$\underline{\mathcal{P}}^{(*)} = \arg\left\{\max_{z=1,...,Z}\left[\min_{s=1,...,S}\left\|\underline{\widetilde{\mathcal{X}}}'^{(z)} - \underline{\mathcal{X}}'^{(s)}\right\|\right]\right\} \tag{18}$$

  then compute the data associated to $\underline{\mathcal{P}}^{(*)}$, $\underline{\mathcal{X}}^{(*)}$, and map it to the reduced features space as $\underline{\mathcal{X}}'^{(*)} = \aleph\left(\underline{\mathcal{X}}^{(*)}, \underline{\mathcal{P}}^{(*)}\right)$;

[4] In the following, for the sake of notation let us refer to the input space as the original K-dimensional one. All discussed concepts and theories are applicable to the K'-dimensional reduced features, as well.

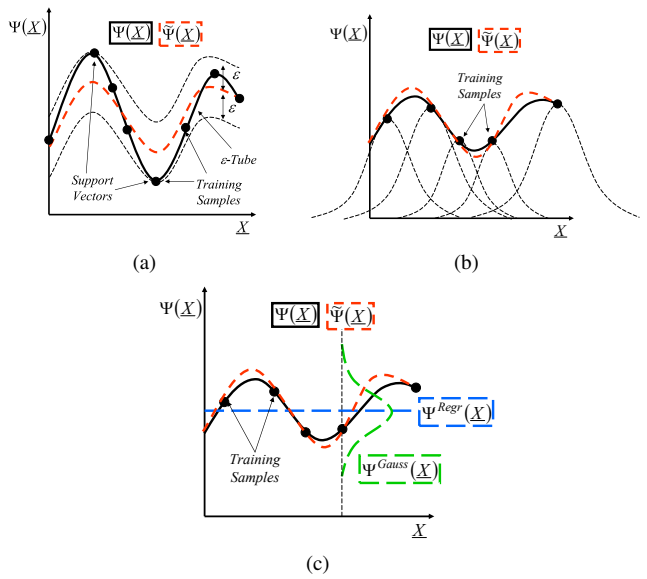


Fig. 5. (*LBE Solution Strategies*) - Pictorial sketch of the (*a*) *SVR*, (*b*) *RBF*, and (*c*) *OK* regressors.

4) Update the training set as $\mathbb{T}' = \mathbb{T}' \cup \left(\underline{\mathcal{X}}'^{(*)}, \underline{\mathcal{P}}^{(*)}\right)$ and let $S \leftarrow (S+1)$. Finally, go to Step 2.1 and iterate until the desired number of training samples has been generated.

C. *Step 3:* Definition of the Prediction Method

This last step is aimed at exploiting the information inside $\mathbb{T}/\mathbb{T}'$ defining the proper *DT* model $\widetilde{\Psi}(\,.\,)$. In the following, some of the most common prediction stategies for classification and regression in *ISP*s are described.

Support Vector Machines (*SVM*s): *SVM*s rapidly emerged in the *EM* community as a powerful alternative to well-established *LBE* classification techniques such as neural networks and decision trees [39] [40]. A binary ($\mathcal{C} = 2$) *SVM* is aimed at recovering the optimal separating hyper-plane (defined by means of a subset of training samples called "support vectors" - Fig. 4) between two classes by solving the following optimization problem

$$\begin{array}{ll} \min_{\underline{\alpha}} & \frac{1}{2}\underline{\alpha}^T \underline{\underline{A}}\,\underline{\alpha} - \underline{1}^T\underline{\alpha} \\ \text{subject to} & \underline{\mathcal{L}}^T\underline{\alpha} = 0;\ \ 0 \le \alpha_s \le C;\ \ s = 1, ..., S \end{array} \tag{19}$$

where $\underline{\alpha} = \{\alpha_s;\, s = 1, ..., S\}^T$ are Lagrange's multipliers, $\underline{1}$ is a column vector of all ones, $\underline{\mathcal{L}} = \left\{\mathcal{L}\left(\underline{\mathcal{X}}^{(s)}\right);\, s = 1, ..., S\right\}^T$, and $C > 0$ is the *SVM* regularization parameter [40]. Moreover, $\underline{\underline{A}}$ is a $(S \times S)$ matrix whose (a, b)-th entry $(a, b = 1, ..., S)$ is $\underline{\underline{A}}\big|_{a,b} = \mathcal{L}\left(\underline{\mathcal{X}}^{(a)}\right)\mathcal{L}\left(\underline{\mathcal{X}}^{(b)}\right)\mathcal{K}\left(\underline{\mathcal{X}}^{(a)}, \underline{\mathcal{X}}^{(b)}\right)$, $\mathcal{K}(\,.\,)$ being the so-called "kernel function" expressing the scalar product between input samples $\left(\underline{\mathcal{X}}^{(a)}, \underline{\mathcal{X}}^{(b)}\right)$ in a higher dimensionality space enabling the linear separation of classes even if $\Psi(\,.\,)$ is a non-linear function of $\underline{\mathcal{X}}$ [32] [40] (Fig. 4). One of the most common definitions of $\mathcal{K}(\,.\,)$ is the following

$$\mathcal{K}\left(\underline{\mathcal{X}}^{(a)}, \underline{\mathcal{X}}^{(b)}\right) = \exp\left(-\gamma\left\|\underline{\mathcal{X}}^{(a)} - \underline{\mathcal{X}}^{(b)}\right\|^2\right) \tag{20}$$

$\gamma > 0$ being a real weight. The solution of (19) yields the following binary decision function

$$\widetilde{\Psi}\left(\underline{\mathcal{X}}\right) = \text{sgn}\left\{\sum_{s=1}^{S}\mathcal{L}\left(\underline{\mathcal{X}}^{(s)}\right)\alpha_s\mathcal{K}\left(\underline{\mathcal{X}}, \underline{\mathcal{X}}^{(s)}\right) + o\right\} \tag{21}$$

where o is a bias term. The extension to the multi-class case (i.e., $\mathcal{C} > 2$) is straightforward and it is generally leveraging on the "one-against-one" strategy [41]. More in detail, a set of $\mathcal{C} \times \frac{(\mathcal{C}-1)}{2}$ binary *SVM* classifiers (21) is built using the training data corresponding to each pair of classes. Therefore, the predicted label for a given input test sample $\underline{\mathcal{X}}$ corresponds to the class collecting the maximum number of "votes" among the different trained *SVM*s.

Support Vector Regression (*SVR*): *SVR* defines a function $\widetilde{\Psi}\left(\underline{\mathcal{X}}\right)$ jointly exhibiting (*i*) the flattest possible behavior and (*ii*) a deviation not larger than a given threshold ϵ from the S training samples [i.e., $\left\|\widetilde{\Psi}\left(\underline{\mathcal{X}}^{(s)}\right) - \Psi\left(\underline{\mathcal{X}}^{(s)}\right)\right\| \le \epsilon,\ s = 1, ..., S$ - Fig. 5(*a*)] [42]. For $Q = 1$ [5], the following problem is solved during the off-line phase

$$\begin{array}{ll} \min_{\underline{\alpha},\,\underline{\alpha}^*} & \frac{1}{2}\left(\underline{\alpha} - \underline{\alpha}^*\right)^T\underline{\underline{B}}\left(\underline{\alpha} - \underline{\alpha}^*\right) + \epsilon\sum_{s=1}^{S}\left(\alpha_s - \alpha_s^*\right) \\ & + \sum_{s=1}^{S}\Psi\left(\underline{\mathcal{X}}^{(s)}\right)\left(\alpha_s - \alpha_s^*\right) \\ \text{subject to} & \underline{1}^T\left(\underline{\alpha} - \underline{\alpha}^*\right) = 0; \\ & 0 \le \alpha_s,\, \alpha_s^* \le C;\ \ s = 1, ..., S \end{array} \tag{22}$$

to derive the following regression *SM*

$$\widetilde{\Psi}\left(\underline{\mathcal{X}}\right) = \sum_{s=1}^{S}\left(-\alpha_s + \alpha_s^*\right)\mathcal{K}\left(\underline{\mathcal{X}}, \underline{\mathcal{X}}^{(s)}\right) + o. \tag{23}$$

In (22) $\underline{\alpha} = \{\alpha_s;\, s = 1, ..., S\}^T$ and $\underline{\alpha}^* = \{\alpha_s^*;\, s = 1, ..., S\}^T$ are Lagrange's multipliers, while $\underline{\underline{B}}$ is a $(S \times S)$ matrix with entries $\underline{\underline{B}}\big|_{a,b} = \mathcal{K}\left(\underline{\mathcal{X}}^{(a)}, \underline{\mathcal{X}}^{(b)}\right)$ $(a, b = 1, ..., S)$. As for the *SVR* constant C, it determines the trade-off between the flatness of $\widetilde{\Psi}\left(\underline{\mathcal{X}}\right)$ and the "penalty" given to training deviations larger than ϵ. Therefore, it must be properly calibrated via suitable cross-validation strategies [43] so that accurate predictions can be achieved without suffering from over-fitting issues (i.e., the incapability of generalization, yielding large regression errors for previously-unseen inputs).

Radial Basis Functions (*RBFs*): The *RBF* a particular artificial neural network whose prediction $\widetilde{\Psi}\left(\underline{\mathcal{X}}\right)$ is expressed in terms of a linear combination of kernel functions depending on the radial distance between $\underline{\mathcal{X}}$ and $\underline{\mathcal{X}}^{(s)}$, $s = 1, ..., S$ [44] [Fig. 5(*b*)]. Mathematically,

$$\widetilde{\Psi}\left(\underline{\mathcal{X}}\right) = \sum_{s=1}^{S}\lambda_s\mathcal{K}\left(\underline{\mathcal{X}}, \underline{\mathcal{X}}^{(s)}\right) \tag{24}$$

where λ_s, $s = 1, ..., S$ are real weights determined during the training phase. Equation (24) complies with

$$\widetilde{\Psi}\left(\underline{\mathcal{X}}^{(i)}\right) = \sum_{s=1}^{S}\lambda_s\mathcal{K}\left(\underline{\mathcal{X}}^{(i)}, \underline{\mathcal{X}}^{(s)}\right) = \Psi\left(\underline{\mathcal{X}}^{(i)}\right);\quad i = 1, ..., S \tag{25}$$

[5] Without loss of generality, in the following let us refer to the scalar regression case, the extension to multiple outputs ($Q > 1$) being straightforward [24].

meaning that the *RBF* makes no error on training locations, differently from the *SVR* [Fig. 5(*b*) vs. Fig. 5(*a*)]. Therefore, it is particularly suitable for those scenarios where the *I/O* relationship is deterministic (i.e., not affected by noise, as it happens in *FL-LBE* strategies).

Kriging: Kriging is a regression technique based on the Bayesian theory [43] [39]. It treats the *I/O* relationship as the realization of a stochastic process [Fig. 5(*c*)]

$$\Psi\left(\underline{\mathcal{X}}\right) = \Psi^{Regr}\left(\underline{\mathcal{X}}\right) + \Psi^{Gauss}\left(\underline{\mathcal{X}}\right) \tag{26}$$

where $\Psi^{Regr}\left(\,.\,\right)$ is a regression function capturing the general trend of $\Psi\left(\,.\,\right)$ and $\Psi^{Gauss}\left(\,.\,\right)$ is a Gaussian process statistically modeling the errors (or "residuals") made by $\Psi^{Regr}\left(\,.\,\right)$ with respect to $\Psi\left(\,.\,\right)$. Although the definition of the former term leads to different Kriging models, its most common implementation is $\Psi^{Regr}\left(\underline{\mathcal{X}}\right) = \Psi_0$ [i.e., constant regression, ordinary Kriging (*OK*) - Fig. 5(*c*)]. On the other hand, $\Psi^{Gauss}\left(\,.\,\right)$ is assumed to have zero mean and covariance proportional to the weighted distance of input samples [43] The *OK* prediction at any point $\underline{\mathcal{X}}$ is computed as

$$\widetilde{\Psi}\left(\underline{\mathcal{X}}\right) = \Psi_0 + \sum_{s=1}^{S} w_s r_s\left(\underline{\mathcal{X}}\right) \tag{27}$$

where $\Psi_0 = \left(\underline{1}^T \underline{\underline{\mathcal{R}}}^{-1} \underline{\Psi}\right) / \left(\underline{1}^T \underline{\underline{\mathcal{R}}}^{-1} \underline{1}\right)$,

$$\underline{r}\left(\underline{\mathcal{X}}\right) = \left\{ r_s\left(\underline{\mathcal{X}}\right) = \prod_{k=1}^{K} \exp\left(-\theta_k \left|\mathcal{X}_k - \mathcal{X}_k^{(s)}\right|^{\beta_k}\right);\ s = 1, ..., S \right\}^T, \tag{28}$$

where $\{(\theta_k,\, \beta_k)\,;\ k = 1, ..., K\}$ are $2K$ hyper-parameters estimated during the off-line phase, and the weighting coefficients $\underline{w} = \{w_s;\ s = 1, ..., S\}^T$ are computed as

$$\underline{w} = \underline{\underline{\mathcal{R}}}^{-1}\left(\underline{\Psi} - \underline{1}\Psi_0\right). \tag{29}$$

In the previous expressions, $\underline{\Psi} = \left\{ \Psi\left(\underline{\mathcal{X}}^{(s)}\right);\ s = 1, ..., S \right\}^T$, while $\underline{\underline{\mathcal{R}}}$ is a $(S \times S)$ matrix whose $(a,\, b)$-th entry is $\underline{\underline{\mathcal{R}}}\big|_{a,b} = \prod_{k=1}^{K} \exp\left(-\theta_k \left|\mathcal{X}_k^{(a)} - \mathcal{X}_k^{(b)}\right|^{\beta_k}\right)$. One paramount advantage of the Kriging over the *SVR* and *RBF* is its capability of providing an estimation of the "uncertainty" associated to any prediction, $\mathcal{U}\left(\underline{\mathcal{X}}\right)$, being $\mathcal{U}\left(\underline{\mathcal{X}}^{(s)}\right) = 0$ and $\widetilde{\Psi}\left(\underline{\mathcal{X}}^{(s)}\right) = \Psi\left(\underline{\mathcal{X}}^{(s)}\right)$ for $s = 1, ..., S$ (i.e., exact interpolation of training samples with null uncertainty) [43].

IV. DEEP LEARNING SOLUTION STRATEGIES

In the following, the basics of the DL paradigm are presented (Sect. IV A). Moreover, the fully data-driven (Sect. IV B), the knowledge-assisted (Sect. IV C), and the physics-embedded (Sect. IV D) learning approaches are described.

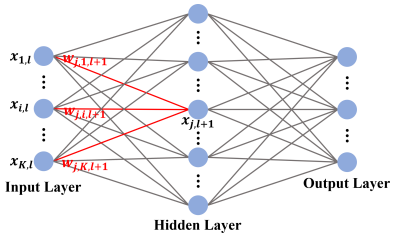


Fig. 6. Structure of FCN with one hidden layer. The connectivity between l-th and $l+1$-th layer is also denoted in the picture.

A. Basics of Deep Learning Approach

Fully Connected Network: Fully connected network (FCN), also known as multi-layer perceptron (MLP), is an important type of artificial neural networks (ANNs) [45]. FCN consists of input layer, hidden layer and output layer, as shown in Figure 6. Let $\mathcal{X}_{i,l}$ denote the output of the i-th neuron in the l-th layer, the output of the j-th neuron in the *l+1*-th layer with $\mathcal{X}_{i,l}$ as input can be expressed as [46]:

$$\mathcal{X}_{j,l+1} = \mathcal{F}(\sum_{i=1}^{K} \mathcal{W}_{j,i,l+1} \cdot \mathcal{X}_{i,l} + b_j)\,, \tag{30}$$

where $\mathcal{F}$, $\mathcal{W}_{j,i,l+1}$, b_j, K represent nonlinear activation function, weight vector between i-th and j-th neurons, bias of j-th neuron, and number of neurons in the l-th layer respectively.

Convolutional Neural Network: Convolutional neural network (CNN) has been widely used in image classification [47], target recognition [48], image analysis [49], inverse problems [50], etc. Unlike FCN, CNN applies convolution operation instead of the matrix vector multiplication, as shown in Figure 7 [51]. Convolution operation brings the sparse interactions and parameter sharing by using convolutional kernels to connect a subset of neurons in the previous layer. The typical relationship of neurons between the l-th and *l-1*-th layer of CNN can be expressed as [52]:

$$\mathcal{X}_l = \mathcal{F}(\mathbb{P}(\mathcal{X}_{l-1} \otimes \mathcal{K}_l + b_l)), l = 1 \cdots L\,, \tag{31}$$

where $\mathcal{F}$, $\mathbb{P}$ and $\otimes$ represent the nonlinear activation function, pooling operation and convolution operation respectively, b_l denotes the bias of neurons in the l-th layer, L denotes the number of layers. Assuming $\mathcal{X}_{l-1}$ is two-dimensional, the convolution operation can be defined as [51]:

$$(\mathcal{X}_{l-1} \otimes \mathcal{K}_l)_{i,j} = \sum_{m=1}^{M} \sum_{n=1}^{N} \mathcal{X}_{l-1}(i-m, j-n)\mathcal{K}_l(m,n) \tag{32}$$

where m, n and M, N denote the indices and size of the 2-D convolutional kernel. The pooling operation can be regarded as down sampling. It can reduce the size of input feature map and the number of CNN parameters. The commonly used pooling operations include max pooling, average pooling, stochastic pooling [53], spatial pyramid pooling [54], and etc. The nonlinear activation function can introduce nonlinearity into CNN, including Sigmoid function, Tanh function, ReLU, PReLU, Leaky ReLU and RReLU [55], etc. Many important CNNs show good learning capability and provide insights into

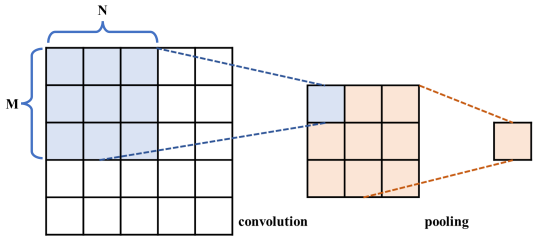


Fig. 7. Schematic of convolution and pooling operation in the CNN.

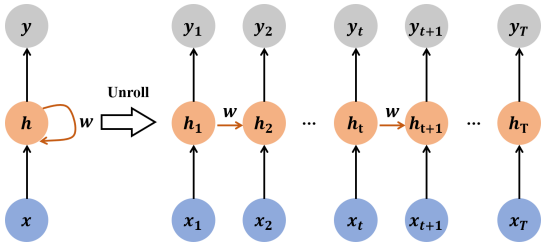


Fig. 8. The schematic of a standard RNN and the unrolled structure.

the design of CNNs, such as AlexNet [56], GoogleNet [57], U-Net [58], ResNet [59], VGGNet [60], etc.

Recurrent Neural Network: Recurrent neural network (RNN) has wide applications in processing sequential data, such as speech recognition [61], time series processing [62], machine translation [63], etc. A standard RNN can output a sequence $\mathbf{Y} = \mathcal{Y}_1, \cdots, \mathcal{Y}_T$ with the input of a sequence $\mathbf{X} = \mathcal{X}_1, \cdots, \mathcal{X}_T$. The structure of a standard RNN is illustrated in Figure 8. In RNN, a sequence of hidden state $\mathbf{h} = h_1, \cdots, h_T$ is introduced to represent the information of all previous time step. The relationship between the input and the hidden state can be expressed as [61]:

$$h_t = \mathcal{F}(\mathcal{W}_x \mathcal{X}_t + \mathcal{W}_h h_{t-1} + b_h) \,, \tag{33}$$

where $\mathcal{W}_x, \mathcal{W}_h$ denote the weight matrix of the input and the hidden state, t ranges from 1 to T, b_h and $\mathcal{F}$ denote the bias of the hidden state and the nonlinear activation function. Then, the output can be calculated based on the hidden state [61]:

$$\mathcal{Y}_t = \mathcal{W}_y h_t + b_y \,, \tag{34}$$

where $\mathcal{W}_y$ and b_y denote the weight matrix and bias of the output. Long short-term memory (LSTM) [64], bidirectional recurrent neural networks (BRNN) [65] and gated recurrent neural networks(GRNN) [66] are three important RNN architectures. LSTM introduces a memory cell to replace the hidden layer of the standard RNN in order to avoid the gradient vanishing [67]. In BRNN, information can propagate towards both forward and backward directions [65]. GRNN applies the reset gate into the vanilla LSTM [6] to reduce the number of parameters [68].

Generative Adversarial Network: Generative adversarial network (GAN) is a powerful method in image processing, such as image synthesis [69], image-to-image translation [70], etc. The structure of GAN consists of a generator and a discriminator. GAN aims to make the generator capture high-dimensional distribution of training data via an adversarial process [71]. The generator and discriminator are trained and optimized simultaneously. The generator generates samples satisfying the distribution of training data, and the discriminator determines whether the samples are from the distribution of training data or not. The global optimality of GAN is mathematically guaranteed [71]. Furthermore, GAN can learn deep and hieratical representations of data and construct the corresponding structured latent space by carefully designing the discriminator [72]. The learned deep and structured data representations are vital to signal processing or image processing [72]. Various architectures of GANs are reported such as convolutional GAN [73], conditional GAN [74], bidirectional GAN [75], etc.

[6]The vanilla LSTM is the commonly used LSTM model that usually consists of a single hidden layer of LSTM units.

B. Fully Data-driven Learning Approach

Fully data-driven learning approach aims to approximate the inverse operators by learning the mappings between the properties of scatterers and the measured scattering fields via the training process:

$$\Psi_\Theta : \mathbb{Y} \mapsto \mathbb{X} \,, \tag{35}$$

where Ψ represents the learned inverse operator, Θ denotes the corresponding parameters, $\mathbb{Y}$ and $\mathbb{X}$ denote the scattering fields and properties of scatterers respectively. In this approach, DL is regarded as a black box to approximate the highly nonlinear input/output (I/O) function Eq. (35). The objective function of direct learning approach evaluates the mismatch of the scatterers' properties:

$$obj = \min_{\Theta} \Upsilon(\Psi_\Theta(\mathbb{Y}), \mathbb{X}) \,, \tag{36}$$

where Υ denotes the metric function measuring the mismatch of scatterers' properties.

Direct Learning Scheme: Direct learning scheme builds a single DNN to regress the relationship between the scattering fields and scatterers. ANNs are applied into the inversion of multiple plasma parameters [76], the reconstruction of randomly-shaped profiles [77]. The CNN is built to realize the super-resolution dielectric imaging of micro-structures [78], qualitatively reconstruct the coordinates and radius of circular scatterers [79], inversion based on phaseless data [80]. More complicated architectures of DNNs are designed for practical applications, such as the blends of CNN and RNN for non-destructive testing of micro-structure [81], a multi-branch CNN for reconstructing 3D moisture distributions of stored grain [82].

Two-step scheme: Generally, two-step schemes build two different DNNs of which one reconstructs the initial guesses of scatterers from the scattering fields and another refines the preliminary inversion for better resolutions. A complex-valued CNN and a residual CNN can be applied for the initial guesses and the refinement respectively [83]. In [84], the extreme learning machine is used to generate the preliminary inversions of scatterers. Another method of two-step scheme first applies an autoencoder to construct the latent space by encoding the high-resolution dielectric images into representation vectors.

Then the CNN is trained to learn the mappings between the measured scattering fields and the latent space. The final model concatenates the CNN and the decoder part of autoencoder [85].

C. Knowledge-assisted Learning Approach

Knowledge-assisted learning approach approximates the inverse operator with the combination of data and knowledge. Instead of directly mapping the measured data to property of scatterers, this approach learns the mappings between the reconstructed model and the intermediate physical quantities that are transformed based on the knowledge of ISP and associated numerical methods. In the knowledge-assisted learning approach, the inverse operator can be expressed as:

$$\Psi_\Theta : \mathbb{Y} \mapsto \psi_{\mathbb{X}}(\mathbb{X}), \tag{37}$$

where $\psi_{\mathbb{X}}$ is the transform function of the scatterers properties. The Ψ_Θ can be defined as:

$$\Psi_\Theta = \psi^1 \circ \cdots \circ \psi^k \circ \cdots \circ \psi^K, \tag{38}$$

where $\psi^1 \cdots \psi^K$ are different parts of the learned inverse operator and $\circ$ denotes the combination of operators. Some of them are data-driven and others are the mathematical transforms. The objective function of knowledge-assisted learning approach can be described as:

$$obj = \min_{\Theta} \Upsilon(\Psi_\Theta(\mathbb{Y}), \psi_{\mathbb{X}}(\mathbb{X})) \ . \tag{39}$$

Knowledge-assisted learning approach integrates DNNs with the traditional ISP algorithms. It combines the advantages of both DNNs and traditional ISP algorithms. DNNs are applied as the data-driven nonlinear function approximators. Traditional ISP algorithms help transform the ISP with physics prior and DNNs only need to learn mappings between intermediate physical quantities.

Knowledge-data Sequential Scheme: Knowledge-data sequential scheme first generates the initial guesses of inversions by the non-iterative inversion method, and then DNNs are trained to enhance or refine the initial inversion for better performance. This scheme maintains more physics prior and simplifies the learning task of DNNs, which can achieve better results than fully data-driven approach. The non-iterative inversion methods in this scheme are usually simple to achieve good computational efficiency.

Back projection method is combined with U-Net to reconstruct the contrast distributions [86]. DeepNIS is proposed based the back propagation method and the complex-valued residual CNN [87]. The performance and dynamic evolution behavior of DeepNIS are discussed in [88]. The backpropagation scheme is based on the similar idea of DeepNIS but the U-Net is employed instead of the complex-valued CNN [3], [14]. The backpropagation scheme is further extended in [89] by replacing the U-Net with a GAN. The discriminator of the GAN can enforce the generator to capture more linearities of ISPs, which is more suitable for reconstruction of high contrast target. Born approximation and Monte Carlo method are combined to generate the preliminary reconstructions that is input to a 3-D U-Net to solve 3D ISPs [90]. Born approximation is adaptable for a wider frequency and more suitable for the low-frequency ISPs compared to the backpropagation method [91].

The dominant current scheme is proposed in [3] to improve the performance of the backpropagation scheme. The input of U-Net is the contrast derived by the dominant component of the induced current. The generalization capabilities of the backpropagation and dominant current scheme on the unseen data samples during the training process are further investigated in [92]. In order to evaluate the confidence of the backpropagation and dominant current scheme, a Bayesian CNN is applied to reconstruct the contrast distribution and predict the corresponding uncertainty of reconstructions at the same time [93]. Inspired by the new integral equation method [94], the modified contrast scheme is further proposed by taking the modified contrast [94] as input [95]. The modified contrast scheme shows better generalization ability and learning capability in ISPs with higher contrast. The knowledge-assisted learning approach is applied to solve ISPs with phaseless data and the input of CNN is generated by the Levenberg–Marquardt methods and CSI methods [80]. The artifacts of breast imaging are prevented by applying U-Net to enhance the multimodal microwave-ultrasound CSI results [96]. In [97], U-Net is applied in the 3D breast imaging with the 3D CSI reconstruction as input. A GAN is built for ISPs with an inhomogeneous background, and the input of the GAN is generated by the distorted Born method and the backpropagation method [98]. U-Net is trained for classifications with uncertainty quantification of breast imaging based on the microwave and ultrasound properties generated by the Gauss-Newton Inversion algorithm [99]. Instead of generating initial guesses of inversions via training process, supervised descent method is trained to learn and store the descent directions, and the pre-learned descent directions can help update the inversion models in the microwave imaging [100].

Data-knowledge Sequential Scheme: Data-knowledge Sequential Scheme uses results of DNNs to improve the performance of traditional ISP algorithms. DNNs are applied to generate the initial guesses of inversion that are the initial value of the traditional ISP algorithms. Good initial values can improve the final performance of inversion in many ISP algorithms, especially iterative/deterministic methods. A contrast source network (CS-Net) is designed to be embedded into the subspace optimization method [101]. The input of CS-Net is the signal subspace of the contrast source and the output is the predicted total contrast source. The predicted total contrast source is then refined in the iterative procedure of the subspace optimization method. The CNN is trained to generate the initial images in multimodality microwave imaging [102]. The input of the CNN are the multimodal images, such as MR, CT, or ultrasound images and the output are the preliminary dielectric images that are used as the initial guess of model-based microwave imaging method. The CNN incorporates prior information from other imaging modalities, which further reduces the ill-posedness and non-linearity of microwave imaging. A CNN is built to denoise the contrast in the iterative process of the linear model method and the denoising CNN can be regarded as a regularization item

[103]. U-net is modified to improve the full-wave inversion by predicting the absent low-frequency scattered field based on the measured high-frequency data, and the predicted low-frequency scattered field is inverted as an initial guess for the high-frequency data inversion [104].

D. Physics-embedded Learning Approach

Physics-embedded learning approach incorporates the physics prior into DNN to improve the generalization ability and learning efficiency. The design of DNNs is usually inspired by the mathematical model or the physics prior. Physics-embedded learning approach is an important method to solve partial differential equations (PDEs) by investigating the mathematical connections between deep learning, PDEs and associated numerical methods, such as physics-informed neural network [105], DeepXDE [106], PDE-Net [107], [108], etc. Various works have also been reported to apply physics-embedded learning approach to accelerate EM forward modeling, which is of great inspiration for applications in ISPs. The solving process of matrix equations formulated in the method of moments is transformed as the optimization process of the DNN parameters where the Adam [109] and stochastic gradient descent [110], [111] are employed [112], [113]. Finite difference time domain method is implemented based on the recurrent convolutional neural network by interpreting the finite difference operator and time marching scheme as the convolutional operator and the recurrent architecture [114]. Physics embedded deep neural network is designed to solve 2D volume integral equations by unfolding the conjugate gradient method as an iterative DNN with Green's function embedded [115]. Physics-informed supervised residual learning is proposed as a general framework for forward modeling by incorporating the fixed-point iterative method into the deep residual neural network [59], [116].

The induced current learning method (ICLM) is proposed in [117] by designing a cascaded end-to-end CNN to predict the induced currents. The cascaded end-to-end CNN is inspired by the basis-expansion strategy and multi-labels are derived to guide the training of the CNN. In [118], the iterative Born approximation is interpreted as an ANN and the error backpropagation algorithm of DL is applied to reconstruct the properties of scatterers. It is indicated in [119] that the error backpropagation algorithm of DNN could be applied in the beam propagation method. FBPConvNet is proposed by investigating the link between CNN and the unrolled iterative method with the convolutional operators [120]. The input of the CNN is generated by filtered back projection [121]. The deep convolutional framelets are reported for perfect reconstruction by extending the convolution framelets [122] of low-rank Hankel matrix in inverse problems [123]. Inspired by the low rank structure of ISPs, SwitchNet is proposed in [124] by introducing sparsely connected switch layers. SwitchNet wires the connectivity of the network in nonlocal fashion, which is in line with the global impact between scatterers and electric fields. By interpreting the linearized forward map as the one-dimensional convolutions, the DNN is built to solve far field pattern and seismic imaging problems [125].

V. APPLICATIONS

This section is aimed at providing an overview of the recent applications of AI-based techniques for solving inverse scattering (Sect. V-A) and EM imaging (Sect. V-B, with particular focus on biomedical imaging) problems. Radar applications are discussed, as well, given their high relevance within the EM visioning framework.

A. EM Inverse Scattering

ISPs are aimed at retrieving information from the field scattered by a scenario when it is illuminated with one or several sources. They can be classified into different groups, depending on the information to be retrieved and/or the field of application. Although the final goal of the ISP is to retrieve qualitative (i.e., shape) and quantitative (i.e., material composition) information on unknown targets from non-invasive measurements of the scattered field [14], many works found in the literature are focused in overcoming the inherent limitations to inverse scattering, which is a non-linear problem. The first attempts to solve ISPs with shallow NNs have been concerned with the parametric inversion of the scatterers [126], [127]. However, most of the works on AI applied to ISPs are much more recent and they are mainly focused on DL techniques, and the topic is increasing rapidly the attention of the scientific community.

In the framework of application of DL to general ISP problems, three CNN-based approaches have been proposed in [3]. On the other hand, in [80] it is proposed a learning-based inversion approach in the frame of the U-net CNN to quantitatively image unknown scatterers located in homogeneous background from the amplitude-only measured total field. Terming the contrast source network, that learns the noise space components of the radiation operator is also achieved by the use of an alternative CNN architecture [101]. An alternative to real valued CNN is proposed in [128] and [129], using complex-valued CNN to solve the ISPs and overcome the classic limitations. In [129] a FFNN model with complex-valued (CV) data stream and the corresponding CV backpropagation training algorithm are combined to realize CV-CNNs. In other to bridge the gap between physical knowledge and learning approaches, an induced current learning method (ICLM) can be used, to incorporate merits in traditional iterative algorithms into the architecture of convolutional neural network (CNN) [117]. Other works as [3] also proposes the use of CNNs for ISPs to alleviate the computational cost of classic algorithms. For classification problems, Bayesian CNN (BCNN) can be used to quantify the uncertainties in solving ISPs [93]. With Monte Carlo dropout, the proposed BCNN is able to directly predict the pixel-based uncertainties of the widely used DLSs in ISPs.

Alternatively, a significant reduction of the CPU time required by gradient-like deterministic retrieval techniques has been yielded by training a DNN to learn descent directions [100], [130]. In [104], a deep-learning-based low-frequency (LF) data prediction scheme is proposed to solve the highly nonlinear inverse scattering problem (ISP) with strong scatterers. The nonlinearity of ISP is alleviated by introducing the LF

components in full-wave inversion. In this scheme, a DNN is trained to predict the absent LF scattered field data from the measured high-frequency (HF) data. Then, the predicted LF data and measured HF data can be inverted by a frequency-hopping technique. In [115], DL and physical simulation are combined, providing a strategy for real-time imaging without loosing reliability and accuracy. It is performed by the design of an iterative DNN to solve full-wave ISP in 2D, embedding the Forward modelling NNs that predict the scattered field in an inversion neural network.

Apart from CNN and DNN, other ML techniques have also been proposed for solving different issues in ISPs. A method named as the modified contrast scheme (MCS), is proposed to tackle nonlinear ISPs [95]. A local-wave amplifier coefficient is used to form the modified contrast, which can alleviate the global nonlinearity in original ISPs without decreasing the accuracy of the problem formulation. [103] exploits a linear-model-based network (LMN) learning strategy, which benefits from both model complexity and data learning. Quantification of uncertainty is a major issue in ISPs and therefore in DL applied to ISP. In [131], a practical uncertainty estimation method framed in the Bayesian theory is introduced for DL inversion of EM data. This overcomes the classic approach based on deterministic prediction that does not provide uncertainty estimates. In order to reconstruct dielectric targets, a structural similarity (SSIM) loss function is introduced and combined with the more classic approach based on a pixel-wise mean squared error (MSE) between the reconstructed image and its reference ones [132]. As an alternative, a dual-module machine learning scheme is proposed to reconstruct inhomogeneous scatterers with high contrasts and large electrical dimensions [84]. The first nonlinear mapping module (NMM) is an extreme learning machine (ELM), which is used to convert the measured scattered fields at the receiver arrays into the preliminary images of the scatterers. The second image-enhancing module (IEM) is a convolutional neural network (CNN), which is used to refine further the images from the NMM to obtain high-accuracy pixel-based model parameter distribution in the inversion domain. Compared with the traditional approximate methods such as backpropagation, the NMM-IEM machine learning can produce the preliminary image with a much higher accuracy but the unknown weight matrices of the ELM are only solved once during training. Another important issue in practical applications is the identification and management of damaged data. In ISPs, EM measurements often contain damaged data due to malfunctioning receivers, which can severely influence the inversion performance. Thus, in [133], a new receiver approximation machine learning (RAML) method is proposed to repair the data from the damaged receivers and the repaired data are the input to the dual-module nonlinear mapping module–image enhancing module machine learning scheme for the 2-D ISP. Finally, the real-time retrieval of the characteristics of a defect with eddy current testing in a non destructive testing and evaluation framework has also been addressed [25]. A statistical learning approach is developed to deal with the inversion problem at hand, taking into account the computational efficiency. More in detail, a feature extraction technique based on partial least squares (PLS) is combined with an output space filling (OSF) adaptive sampling scheme for generating optimal training databases, while accurate and robust reconstructions are performed with a support vector regression (SVR) algorithm.

As previously mentioned, qualitative ISPs are aimed at retrieving the location and shape of one or multiple unknown scatterers. Within this context, many works have proposed the use of AI techniques ISPs to improve the overall performances [6], [134], [135]. For example, in [6] the authors show that given a properly trained neural network, single frequency reconstructions can be very competitive with multifrequency techniques that do not use neural networks. On the other hand, a ISP is applied to salt delineation in [134]. In this work the authors investigate the mapping of subsurface electrical resistivity distributions from EM data with CNNs and demonstrate they are able to reconstruct arbitrary shape more efficiently than with classic methods. Another application linked to shape reconstruction is the detection of fault signals that can be particularized in different schemes of fault detection [135]. In this case, resonant grounding distribution systems are studied and method of faulty feeder detection based on the continuous wavelet transform (CWT) and convolutional neural network (CNN) is proposed.

In some of the applications of ISPs the goal is not to obtain a high definition image or profile of the scatter but to get enough information to classify it among a set of possible solutions, with low error. In classic solutions a full ISP is solved and then image processing techniques are applied. However, sometimes a large amount of data is required in order to get high-resolution images to ensure the performance of the processing technique. However, this is one of the main applications of AI techniques and therefore they can be applied to classification problems in ISPs. Among other applications, these techniques can be applied in biometrics [5] for personal identification. Actually, microwave biometric scans have recently gained attraction as a non-contact technique due to their robustness to environmental lighting and unobtrusiveness. In [5], the microwave signature of the human forearm is exploited as a biometric modality. The system, among several evaluations, is tested by collecting microwave samples from human volunteers' forearms and classifying the data using Support Vector Machines and Naive Bayesian classifiers. In other different application, the electric discharge states in gases can be detected based on the information on visual images [136], and different states of corona discharge in plasma can be identified by applying four kinds of machine learning algorithms to extract color, brightness, and shape information characteristics of visible images. The four different machine learning algorithms are support vector machine (SVM), K-nearest neighbor regression (KNN), single layer perceptron (SLP), and decision tree algorithms. Finally, unsupervised learning is applicable to classification that does not know the number of specific categories in advance, and sparse auto-encoders (SAE) are widely used for feature extraction of unsupervised learning. Therefore, [137] proposes an electromagnetic signal classification system based on SAE which is combined with the NN clustering algorithm.

B. EM Imaging

Although Imaging can be seen as a particular case of ISPs, the number and relevance of works in this topic makes it worth of a separate discussion and in particular the case of application to medical diagnostic. As a general baseline, the research has been mainly focused on the application of DL and NN. This is the case of [138], where a general microwave image reconstruction is presented. It is based on the conversion of a 24 x 24 samples acquisition into a 128 x 128 image throughout a NN. The main contribution regards the training process, which is divided into a two-stage training method in order to reduce the complexity of this task. DL is also applied in [98], but in this case the authors try to solve the inhomogeneous background ISP. To alleviate the burden of nonlinearity and ill-posedness of the ISP, the distorted-Born backpropagation scheme is introduced to quantitatively reconstruct a rough image of the unknown object. Then, this is the input of a generative adversarial network (GAN), which outputs the fine reconstructed image of the relative permittivity. An alternative to most common solutions is the use of Greedy Pursuit Algorithms (GPAs) to reconstruct sparse signals [139].

Medical Imaging (MI): Electromagnetic Imaging is a non-invasive technique and therefore it is very interesting for brain diagnosis and early detection of some kinds of tumours, among other diseases. The application of DL techniques to this discipline is of great interest as in general imaging since it can speed up the process without losing accuracy. Although medical imaging based on DL is at the beginning of the development approaches have already been discussed in the literature [10], [99], [140]–[148]. A focus body area in medical imaging is the head and in particular the brain, since this is a part of the body with difficult exploration with other methods. In [140] a brain anomaly localization algorithm in an unsupervised ML framework is presented for EM brain imaging. The method is based on expected value estimation and takes the advantage of the highly symmetrical human brain. The algorithm processes signals collected from pairs of antennas that are positioned symmetrically around the head, discretizes the imaging domain into pixels, and computes the statistical fields between the antennas on the left and right sides of the head. Then, it concatenates their intensities along the axis normal to the imaging domain to compute the expected value for every pixel. The imaging results demonstrate the capability of the proposed algorithm to localize bleeding and estimate its size with less than 10% error in less than a minute, which makes it suitable for real-time use in emergency stroke scenarios. It is worth pointing out that DNNs are known for being data hungry machines, and in many practical cases, such as electromagnetic medical imaging, there is not enough training data to feed them. However, a deep domain adaptation technique can be customized for matching distributions of complex-valued electromagnetic data [141], showing improved performance over regular ones. On the other side, in [142], SVR is applied to functional magnetic resonance imaging (fMRI) in the framework of brain networks, where parts of the brain are segregated based on functionality and they are then connected with many interactions (functional integration). The SVR models are applied to each functional part of the brain and they are then connected in a more complex network. The "Deep D-Bar" approach [143] , [10] and a DCS-based technique [144] have been proposed for the realtime (e.g., $\Delta t \leq 8ms$) electrical impedance tomography of the chest. According to authors' vision, DL could provide a way to incorporate more versatile prior information to mitigate the ill-posedness. DNNs have also been proposed for suppressing off-axis scattering in ultrasound channel data, in the frequency domain via the short-time Fourier transform [145]. The authors provide results of simulations as well as experimental results based on physical phantom and in vivo measurements, showing a relevant improvement in the contrast ratio (CR) compared with classic imaging. CNNs has been used to improve the identification and classification of human breast tissues through ultrasound [99] or combined ultrasound-microwave [146] imaging. Quantitative tomographic reconstructions of the dielectric properties (complex-valued permittivity) and the ultrasonic properties (compressibility and attenuation) as well as their combination, with the corresponding actual tissue-type classification constitute the training set. Finally, two reconstruction algorithms based on an autoencoder and a fully connected NN respectively [147], obtained better performance than traditional algorithms. DNNs are common tools within the image processing community to perform classification and they have been applied in the detection of melanoma and lymph node [148].

Radar: Scattering problem is an inherent issue in radar application, and it becomes more challenging in synthetic aperture radar (SAR) imagery, usual in onboard flying radars. In order to obtain a robust and confident system, the techniques for signal processing and IS should be reliable and real-time requirements are usually demanded. In this framework, DNNs are natural candidates for performing accurate automatic target recognitions and improve radar performance. The radar problem can be either a classification one or an imaging one. In [4], a CNN has been implemented for high-accuracy image classification to avoid overfitting when small training databases are at hand. Similarly, a generative DNN is applied in SAR [149] to learn a hierarchical representation of the features of the targets. On the other hand, polarimetric SAR image classification has been addressed with a deep CNN incorporating expert knowledge on the interpretation of the scattering mechanisms and polarimetric feature mining [150]. The problem of rough surface estimation has been addressed in [151], using a novel CNN inversion method by letting the CNN learn the nonlinear relationship between inverted images and predicted surface descriptors. Moreover, DNNs have been demonstrated for microwave remote sensing of vegetated areas in [152] where a ML scheme has been used to predict the polarimetric bistatic scattering cross section of a finite dielectric cylinder modelling a corn canopy in C-band. The "radar-Siamese" has been proposed to automatically extract robust features for an accurate material classification, based on the radar signature of the media [153]. Moreover, complex-valued CNNs (CV-CNN) are proposed specifically for synthetic aperture radar (SAR) image interpretation [154].

These CNNs utilize both amplitude and phase information of complex SAR imagery. Moreover, a complex backpropagation algorithm based on stochastic gradient descent is derived for CV-CNN training. In other way, CNNs can be combined with compressive sensing (CS) to get high resolution with reduced number of antenna elements and measurements [155]. The use of CNNs alleviates the inherent limitations of CS: high computational complexity and requirement of parameter tuning to ensure good image reconstruction under different noise, sparsity and undersampling levels. Another case of scatter classifier is presented in [156]. The traditional 3-D variational Born iterative method (VBIM) is combined with the unsupervised machine-learning expectation maximization algorithm (EMA). DL has also greatly emerged as a powerful approach in Ground Penetrating Radar (GPR). In this kind of systems, AI techniques can be used for different purposes as signal processing and imaging [2], [157] detecting and classification of buried objects [158] , [159], material identification [160], or landmine detection [161], among others.

Finally, CNNs can also be applied to improve the real-time performance of radars in safety critical systems in Advanced Driver Assistance Systems (ADAS) [162]. They are proposed to be used as part of the radar system to detect and classify objects. SVMs have been employed in radar electronic reconnaissance in the framework of cognitive radio to recognize modulated signals in complex EM environment [163]. In particular they are proposed for the recognition of radar signal modulation under low signal-to-noise ratio, which can be seen as other application of classifiers based on ML techniques.

Table I demonstrates a categorical summary of various deep learning approaches and AI-based applications of inverse scattering and electromagnetic imaging problems. Despite the great flexibility of the categorization, Table I can still provide a practical overview of how artificial intelligence can be applied into the inverse scattering and electromagnetic imaging.

TABLE I
CATEGORICAL SUMMARY OF DEEP LEARNING APPROACHES AND AI-BASED APPLICATIONS OF INVERSE SCATTERING AND ELECTROMAGNETIC IMAGING PROBLEMS.

Fully Data-driven Learning Approach	
Direct Learning Scheme	[76]–[82]
Two-step scheme	[83]–[85]
Knowledge-assisted Learning Approach	
Knowledge-data Sequential Scheme	[3], [80], [86]–[90] [92], [93], [95]–[100]
Data-knowledge Sequential Scheme	[101]–[104]
Physics-embedded Learning Approach	
[117]–[121] [123]–[125]	,
Applications in Electromagnetic Inverse Scattering	
[3], [5], [6], [25], [80], [84], [93], [95], [100], [101], [103], [104], [115], [117], [126]–[137]	
Applications in Electromagnetic Imaging	
General Imaging:	[98], [139], [147]
Medical Imaging:	[10], [99], [140]–[148]
Radar:	[2], [4], [149]–[163]

VI. CONCLUSIONS

In this paper, we briefly review recent research in the application of AI techniques, especially those based on DL, to solve ISPs and imaging problems. This is a rapidly developing area, as we can see from the reference list. However, using ML in electromagnetic engineering is not a new concept and many researchers have made contribution since 1990s. The idea of building a surrogated model from data to substitute the complex solution process based on differential equations and to relieve the computational bottleneck is tantalizing. With the help of big data, massively parallel computing, and optimization algorithms, DNNs with millions of parameters can be trained, allowing a successful leap in their learning and generalization capabilities. This improvement also helps ISPs and imaging. These problems are inherently nonlinear, ill-posed, and more importantly, dealing with measured data, thus very suitable for the application of ML. With DNNs, it has been demonstrated in many studies that the accuracy and efficiency can be improved in solving such problems. Their success is demonstrated by the large amount of publications appeared in the scientific literature during the last few years, as well as by the contributions within this Special Issue of the IEEE Transactions on Antennas and Propagation.

However, there are two sides to every coin. We cannot overlook the limitations of DL. Its learning and generalization ability is not unbounded and is limited by the network structure, the available dataset, and the computing power. Unlike Maxwell's equations, transferring from one scenario to another usually requires new design of the networks and training from the beginning. Moreover, datasets in electromagnetics measured at different occasions are different in both contents and formats, unlike images and speech recordings. All these challenges the applications of machine learning techniques in electromagnetic engineering. Therefore, it may be important to clearly describe the problem and specify the boundary of application when machine learning is used.

EM theory provides a powerful tool for research and engineering. It clearly discovers the physics and has a good generalization ability. However, EM modeling of real-world phenomena can never be exact because of the gap between reality and theory. On the other hand, measured data are close to the real world, but they often contain various information and are usually noisy. A hybridization of these two may allow us more chance in solving nonlinear and ill-posed problems with better accuracy and efficiency. It may also extend the applicability of electromagnetic inverse problems in real world. Recent research in inverse scattering and imaging has proven its feasibility. In the future, we may benefit from the development of machine learning platforms, take advantages of data and parallel computing, and develop new algorithms for electromagnetic sensing and imaging.

REFERENCES

[1] A. Massa *et al.*, "DNNs as applied to electromagnetics, antennas, and propagation—A review," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2225–2229, 2019.

[2] I. Giannakis *et al.*, "A machine learning-based fast-forward solver for ground penetrating radar with application to full-waveform inversion," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4417–4426, 2019.

[3] Z. Wei and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 1849–1860, 2019.

[4] S. Chen *et al.*, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans Geosci Remote Sens*, vol. 54, no. 8, pp. 4806–4817, 2016.

[5] A.-A. Nabulsi *et al.*, "Machine learning classification of s-band microwave scattering measurements from the forearm as a novel biometric technique," *IEEE Open J. Antennas Propag*, vol. 1, pp. 118–125, 2020.

[6] Y. Kalepu *et al.*, "Reconstructing dispersive scatterers with minimal frequency data," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 62–66, 2020.

[7] P. Naseri and S. V. Hum, "A generative machine learning-based approach for inverse design of multilayer metasurfaces," *IEEE Trans. Antennas Propag.*, 2021.

[8] T. Shan *et al.*, "Coding programmable metasurfaces based on deep learning techniques," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 10, no. 1, pp. 114–125, 2020.

[9] A. Stern *et al.*, "EMFORCED: EM-based fingerprinting framework for remarked and cloned counterfeit IC detection using machine learning classification," *IEEE Trans. VLSI Syst.*, vol. 28, no. 2, pp. 363–375, 2019.

[10] M. Capps and J. L. Mueller, "Reconstruction of organ boundaries with deep learning in the D-bar method for electrical impedance tomography," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 3, pp. 826–833, 2020.

[11] X. Wang *et al.*, "BiLoc: Bi-modal deep learning for indoor localization with commodity 5GHz WiFi," *IEEE access*, vol. 5, pp. 4209–4220, 2017.

[12] Y. Kim and Y. Li, "Human activity classification with transmission and reflection coefficients of on-body antennas through deep convolutional neural networks," *IEEE Trans. Antennas Propag.*, vol. 65, no. 5, pp. 2764–2768, 2017.

[13] D. Shi *et al.*, "A novel method for intelligent EMC management using a "knowledge base"," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 6, pp. 1621–1626, 2018.

[14] X. Chen, *Computational methods for electromagnetic inverse scattering*. Wiley Online Library, 2018.

[15] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*. CRC press, 2020.

[16] R. Persico, *Introduction to ground penetrating radar: inverse scattering and data processing*. John Wiley & Sons, 2014.

[17] A. Fhager *et al.*, "Microwave diagnostics ahead: Saving time and the lives of trauma and stroke patients," *IEEE Microw. Mag.*, vol. 19, no. 3, pp. 78–90, 2018.

[18] N. K. Nikolova, "Microwave imaging for breast cancer," *IEEE Microw. Mag.*, vol. 12, no. 7, pp. 78–94, 2011.

[19] W. Shao and T. McCollough, "Advances in microwave near-field imaging: Prototypes, systems, and applications," *IEEE Microw. Mag.*, vol. 21, no. 5, pp. 94–119, 2020.

[20] K. Xu *et al.*, "A fast integral equation-based method for solving electromagnetic inverse scattering problems with inhomogeneous background," *IEEE Trans. Antennas Propag.*, vol. 66, no. 8, pp. 4228–4239, 2018.

[21] Z. Liu *et al.*, "Electromagnetic imaging of damages in fibered layered laminates via equivalence theory," *IEEE Trans. Comput. Imaging*, vol. 4, no. 2, pp. 219–227, 2018.

[22] R. Zoughi, *Microwave non-destructive testing and evaluation*. Amsterdam, The Netherlands: Kluwer, 2000.

[23] M. Salucci *et al.*, "Multi-step learning-by-examples strategy for real-time brain stroke microwave scattering data inversion," *Electronics*, vol. 10, no. 1, p. 95, 2021.

[24] M. Salucci *et al.*, "Real-time electrical impedance tomography of the human chest by means of a learning-by-examples method," *IEEE J. Electromagn. RF Microw. Med. Biol.*, vol. 3, no. 2, pp. 88–96, 2019.

[25] M. Salucci *et al.*, "Real-time NDT-NDE through an innovative adaptive partial least squares SVR inversion approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6818–6832, 2016.

[26] E. Bermani *et al.*, "An innovative real-time technique for buried object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 927–931, 2003.

[27] M. Salucci, L. Poli, P. Rocca, and A. Massa, "Learned global optimization for inverse scattering problems-matching global search with computational efficiency," *IEEE Trans. Antennas Propag.*, 2022.

[28] J. Friedman *et al.*, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1, no. 10.

[29] S. Wold *et al.*, "PLS-regression: a basic tool of chemometrics," *Chemometrics and intelligent laboratory systems*, vol. 58, no. 2, pp. 109–130, 2001.

[30] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.

[31] M. Salucci *et al.*, "A nonlinear Kernel-based adaptive learning-by-examples method for robust NDT/NDE of conductive tubes," *J Electromagnet Wave*, vol. 33, no. 6, pp. 669–696, 2019.

[32] J. Shawe-Taylor *et al.*, *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[33] A. S. Hedayat *et al.*, *Orthogonal arrays: theory and applications*. Springer Science & Business Media, 2012.

[34] M. D. McKay *et al.*, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 42, no. 1, pp. 55–61, 2000.

[35] K. Crombecq *et al.*, "A novel hybrid sequential design strategy for global surrogate modeling of computer experiments," *SIAM J. Sci. Comput.*, vol. 33, no. 4, pp. 1948–1974, 2011.

[36] S. S. Garud *et al.*, "Design of computer experiments: A review," *Comput. Chem. Eng.*, vol. 106, pp. 71–95, 2017.

[37] H. Liu *et al.*, "A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design," *Struct. Multidiscip. Optim.*, vol. 57, no. 1, pp. 393–416, 2018.

[38] S. Bilicz *et al.*, "Kriging-based generation of optimal databases as forward and inverse surrogate models," *Inverse Probl*, vol. 26, no. 7, p. 074012, 2010.

[39] A. Massa *et al.*, "Learning-by-examples techniques as applied to electromagnetics," *J Electromagnet Wave*, vol. 32, no. 4, pp. 516–541, 2018.

[40] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[41] A. Mathur and G. M. Foody, "Multiclass and binary SVM classification: Implications for training and classification users," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 241–245, 2008.

[42] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput*, vol. 14, no. 3, pp. 199–222, 2004.

[43] A. Forrester *et al.*, *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.

[44] Y.-K. Sui *et al.*, "An efficient global optimization algorithm based on augmented radial basis function," *Int. J. Simul. Multi. Des. Optim.*, vol. 2, no. 1, pp. 49–55, 2008.

[45] K. Hornik *et al.*, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks," *Neural Netw*, vol. 3, no. 5, pp. 551–560, 1990.

[46] M. H. Hassoun *et al.*, *Fundamentals of artificial neural networks*. MIT press, 1995.

[47] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, 2017.

[48] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Lect Notes Artif Int*, vol. 9, no. 2, pp. 85–112, 2020.

[49] S. M. Anwar *et al.*, "Medical image analysis using convolutional neural networks: a review," *J. Med. Syst.*, vol. 42, no. 11, pp. 1–13, 2018.

[50] M. T. McCann *et al.*, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 85–95, 2017.

[51] I. Goodfellow *et al.*, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[52] J. Bouvrie, "Notes on convolutional neural networks," 2006.

[53] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," *arXiv preprint arXiv:1301.3557*, 2013.

[54] K. He *et al.*, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015.

[55] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378*, 2018.

[56] A. Krizhevsky *et al.*, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.

[57] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[58] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[59] K. He *et al.*, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[60] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[61] A. Graves *et al.*, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee, 2013, pp. 6645–6649.

[62] Y. Qin *et al.*, "A dual-stage attention-based recurrent neural network for time series prediction," *arXiv preprint arXiv:1704.02971*, 2017.

[63] S. Liu *et al.*, "A recursive recurrent neural network for statistical machine translation," 2014.

[64] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[65] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process*, vol. 45, no. 11, pp. 2673–2681, 1997.

[66] J. Chung *et al.*, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[67] Z. C. Lipton *et al.*, "A critical review of recurrent neural networks for sequence learning," *arXiv preprint arXiv:1506.00019*, 2015.

[68] Y. Yu *et al.*, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, 2019.

[69] S. Reed *et al.*, "Generative adversarial text to image synthesis," in *International Conference on Machine Learning*. PMLR, 2016, pp. 1060–1069.

[70] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134.

[71] I. J. Goodfellow *et al.*, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[72] A. Creswell *et al.*, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, 2018.

[73] A. Radford *et al.*, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[74] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[75] J. Donahue *et al.*, "Adversarial feature learning," *arXiv preprint arXiv:1605.09782*, 2016.

[76] Y. Zhang *et al.*, "Neural network model for parameter inversion in electromagnetic wave and plasma interaction systems," *IEEE Trans. Plasma Sci.*, vol. 48, no. 6, pp. 2143–2152, 2020.

[77] M. Ambrosanio *et al.*, "Machine learning for microwave imaging," in *2020 14th European Conference on Antennas and Propagation (EuCAP)*, 2020, pp. 1–4.

[78] P. Ran *et al.*, "Subwavelength micro-structure probing by binary-specialized methods: contrast source and convolutional neural networks," *IEEE Trans. Antennas Propag.*, 2020.

[79] H. Yang and J. Liu, "A qualitative deep learning method for inverse scattering problems." *Appl Comput Electrom*, vol. 35, no. 2, 2020.

[80] K. Xu *et al.*, "Deep learning-based inversion methods for solving inverse scattering problems with phaseless data," *IEEE Trans. Antennas Propag.*, vol. 68, no. 11, pp. 7457–7470, 2020.

[81] P. Ran *et al.*, "Electromagnetic micro-structure non-destructive testing: Sparsity-constrained and combined convolutional recurrent neural network methods," *Electronics*, vol. 9, no. 11, p. 1750, 2020.

[82] V. Khoshdel *et al.*, "A multi-branch deep convolutional fusion architecture for 3D microwave inverse scattering: stored grain application," *Neural Comput. Appl.*, pp. 1–13, 2021.

[83] H. M. Yao *et al.*, "Two-step enhanced deep learning approach for electromagnetic inverse scattering problems," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2254–2258, 2019.

[84] L.-Y. Xiao *et al.*, "Dual-module NMM-IEM machine learning for fast electromagnetic inversion of inhomogeneous scatterers with high contrasts and large electrical dimensions," *IEEE Trans. Antennas Propag.*, vol. 68, no. 8, pp. 6245–6255, 2020.

[85] W. Shao and Y. Du, "Microwave imaging by deep learning network: Feasibility and training method," *IEEE Trans. Antennas Propag.*, vol. 68, no. 7, pp. 5626–5635, 2020.

[86] Y. Sun *et al.*, "Efficient and accurate inversion of multiple scattering with deep learning," *Opt. Express*, vol. 26, no. 11, pp. 14 678–14 688, 2018.

[87] L. Li *et al.*, "DeepNIS: Deep neural network for nonlinear electromagnetic inverse scattering," *IEEE Trans. Antennas Propag.*, vol. 67, no. 3, pp. 1819–1825, 2018.

[88] L. Li *et al.*, "Performance analysis and dynamic evolution of deep convolutional neural network for electromagnetic inverse scattering," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2259–2263, 2019.

[89] L. Guo *et al.*, "Complex-valued pix2pix—deep neural network for nonlinear electromagnetic inverse scattering," *Electronics*, vol. 10, no. 6, p. 752, 2021.

[90] J. Xiao *et al.*, "Fast electromagnetic inversion of inhomogeneous scatterers embedded in layered media by born approximation and 3-D U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1677–1681, 2019.

[91] W. C. Chew, *Waves and fields in inhomogeneous media*. IEEE press, 1995.

[92] Z. Wei and X. Chen, "Generalization capabilities of deep learning schemes in solving inverse scattering problems," in *2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*. IEEE, 2019, pp. 215–216.

[93] Wei, Zhun and Chen, Xudong, "Uncertainty quantification in inverse scattering problems with bayesian convolutional neural networks," *IEEE Trans. Antennas Propag.*, vol. 69, no. 6, pp. 3409–3418, 2020.

[94] Y. Zhong *et al.*, "A new integral equation method to solve highly nonlinear inverse scattering problems," *IEEE Trans. Antennas Propag.*, vol. 64, no. 5, pp. 1788–1799, 2016.

[95] Y. Zhou *et al.*, "An Improved Deep Learning Scheme for Solving 2D and 3D Inverse Scattering Problems," *IEEE Trans. Antennas Propag.*, 2020.

[96] V. Khoshdel *et al.*, "Enhancement of multimodal microwave-ultrasound breast imaging using a deep-learning technique," *Sens.*, vol. 19, no. 18, p. 4050, 2019.

[97] V. Khoshdel *et al.*, "Full 3D microwave breast imaging using a deep-learning technique," *J Imaging*, vol. 6, no. 8, p. 80, 2020.

[98] X. Ye *et al.*, "An inhomogeneous background imaging method based on generative adversarial network," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 11, pp. 4684–4693, 2020.

[99] P. Mojabi *et al.*, "Tissue-type classification with uncertainty quantification of microwave and ultrasound breast imaging: A deep learning approach," *IEEE Access*, vol. 8, pp. 182 092–182 104, 2020.

[100] R. Guo *et al.*, "Supervised descent learning technique for 2-D microwave imaging," *IEEE Trans. Antennas Propag.*, vol. 67, no. 5, pp. 3550–3554, 2019.

[101] Y. Sanghvi *et al.*, "Embedding deep learning in inverse scattering problems," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 46–56, 2019.

[102] G. Chen *et al.*, "Learning-assisted multimodality dielectric imaging," *IEEE Trans. Antennas Propag.*, vol. 68, no. 3, pp. 2356–2369, 2020.

[103] H. Zhou *et al.*, "Linear-model-inspired neural network for electromagnetic inverse scattering," *IEEE Antennas Wireless Propag. Lett.*, vol. 19, no. 9, pp. 1536–1540, 2020.

[104] Z. Lin *et al.*, "Low-Frequency Data Prediction With Iterative Learning for Highly Nonlinear Inverse Scattering Problems," *IEEE Trans. Microw. Theory Techn.*, 2021.

[105] M. Raissi *et al.*, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.

[106] L. Lu *et al.*, "DeepXDE: A deep learning library for solving differential equations," *SIAM Rev.*, vol. 63, no. 1, pp. 208–228, 2021.

[107] Z. Long *et al.*, "PDE-Net: Learning pdes from data," in *International Conference on Machine Learning*. PMLR, 2018, pp. 3208–3216.

[108] Z. Long *et al.*, "PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network," *J. Comput. Phys.*, vol. 399, p. 108925, 2019.

[109] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[110] Y. LeCun *et al.*, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[111] Q. V. Le *et al.*, "On optimization methods for deep learning," in *ICML*, 2011.

[112] L. Guo *et al.*, "Investigation of Adam for low-frequency electromagnetic problems," in *2020 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO)*. IEEE, 2020, pp. 1–2.

[113] L. Guo *et al.*, "Application of stochastic gradient descent technique for method of moments," in *2020 IEEE International Conference on Computational Electromagnetics (ICCEM)*. IEEE, 2020, pp. 97–98.

[114] Guo, Liangshuai *et al.*, "Study on a recurrent convolutional neural network based FDTD method," in *2019 International Applied Computational Electromagnetics Society Symposium-China (ACES)*, vol. 1. IEEE, 2019, pp. 1–2.

[115] R. Guo *et al.*, "Physics embedded deep neural network for solving volume integral equation: 2D case," *IEEE Trans. Antennas Propag.*, 2021.

[116] T. Shan *et al.*, "Physics-informed supervised residual learning for 2D electromagnetic forward modeling," *arXiv preprint arXiv:2104.13231*, 2021.

[117] Z. Wei and X. Chen, "Physics-inspired convolutional neural network for solving full-wave inverse scattering problems," *IEEE Trans. Antennas Propag.*, vol. 67, no. 9, pp. 6138–6148, 2019.

[118] U. S. Kamilov *et al.*, "A recursive Born approach to nonlinear inverse scattering," *IEEE Signal Process. Lett.*, vol. 23, no. 8, pp. 1052–1056, 2016.

[119] U. S. Kamilov *et al.*, "Optical tomographic image reconstruction based on beam propagation and sparse regularization," *IEEE Trans. Comput. Imaging*, vol. 2, no. 1, pp. 59–70, 2016.

[120] K. H. Jin *et al.*, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.

[121] M. Beister *et al.*, "Iterative reconstruction methods in X-ray CT," *Physica medica*, vol. 28, no. 2, pp. 94–108, 2012.

[122] R. Yin *et al.*, "A tale of two bases: Local-nonlocal regularization on image patches with convolution framelets," *SIAM J. Imaging Sci.*, vol. 10, no. 2, pp. 711–750, 2017.

[123] J. C. Ye *et al.*, "Deep convolutional framelets: A general deep learning framework for inverse problems," *SIAM J. Imaging Sci.*, vol. 11, no. 2, pp. 991–1048, 2018.

[124] Y. Khoo and L. Ying, "SwitchNet: a neural network model for forward and inverse scattering problems," *SIAM J. Sci. Comput.*, vol. 41, no. 5, pp. A3182–A3201, 2019.

[125] Y. Fan and L. Ying, "Solving inverse wave scattering with deep learning," *arXiv preprint arXiv:1911.13202*, 2019.

[126] S. Caorsi and P. Gamba, "Electromagnetic detection of dielectric cylinders by a neural network approach," *IEEE Trans Geosci Remote Sens*, vol. 37, no. 2, pp. 820–827, 1999.

[127] I. T. Rekanos, "Neural-network-based inverse-scattering technique for online microwave medical imaging," *IEEE Trans. Magn.*, vol. 38, no. 2, pp. 1061–1064, 2002.

[128] H. M. Yao *et al.*, "Enhanced deep learning approach based on the deep convolutional encoder–decoder architecture for electromagnetic inverse scattering problems," *IEEE Antennas Wireless Propag. Lett.*, vol. 19, no. 7, pp. 1211–1215, 2020.

[129] X.-M. Pan *et al.*, "On phase information for deep neural networks to solve full-wave nonlinear inverse scattering problems," *IEEE Antennas Wireless Propag. Lett.*, 2021.

[130] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Probl*, vol. 33, no. 12, p. 124007, 2017.

[131] S. Oh and J. Byun, "Bayesian uncertainty estimation for deep learning inversion of electromagnetic data," *IEEE Geosci. Remote Sens. Lett.*, 2021.

[132] Y. Huang *et al.*, "Deep learning-based inverse scattering with structural similarity loss functions," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4900–4907, 2020.

[133] H.-J. Hu *et al.*, "Nonlinear electromagnetic inversion of damaged experimental data by a receiver approximation machine learning method," *IEEE Antennas Wireless Propag. Lett.*, 2021.

[134] S. Oh *et al.*, "Salt delineation from electromagnetic data using convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 519–523, 2018.

[135] M.-F. Guo *et al.*, "Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1291–1300, 2017.

[136] Q. Ye *et al.*, "A corona recognition method based on visible light color and machine learning," *IEEE Trans. Plasma Sci.*, vol. 48, no. 1, pp. 31–35, 2019.

[137] K. Ren *et al.*, "Pulses classification based on sparse auto-encoders neural networks," *IEEE Access*, vol. 7, pp. 92 651–92 660, 2019.

[138] W. Shao and Y. Du, "Microwave imaging by deep learning network: Feasibility and training method," *IEEE transactions on antennas and propagation*, vol. 68, no. 7, pp. 5626–5635, 2020.

[139] A. I. Sandhu *et al.*, "ANN-assisted CoSaMP Algorithm for Linear Electromagnetic Imaging of Spatially Sparse Domains," *arXiv preprint arXiv:1911.06514*, 2019.

[140] A. Brankovic *et al.*, "Unsupervised algorithm for brain anomalies localization in electromagnetic imaging," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 1595–1606, 2020.

[141] A. Al-Saffar *et al.*, "Closing the gap of simulation to reality in electromagnetic imaging of brain strokes via deep neural networks," *IEEE Trans. Comput. Imaging*, vol. 7, pp. 13–21, 2020.

[142] J. Richiardi *et al.*, "Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 58–70, 2013.

[143] S. J. Hamilton and A. Hauptmann, "Deep D-bar: Real-time electrical impedance tomography imaging with deep neural networks," *IEEE Trans. Med. Imaging*, vol. 37, no. 10, pp. 2367–2377, 2018.

[144] Z. Wei *et al.*, "Dominant-current deep learning scheme for electrical impedance tomography," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 9, pp. 2546–2555, 2019.

[145] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imaging*, vol. 37, no. 9, pp. 2010–2021, 2018.

[146] P. Mojabi *et al.*, "CNN for compressibility to permittivity mapping for combined ultrasound-microwave breast imaging," *IEEE J. Multiscale Multiphysics Comput. Tech.*, 2021.

[147] J. Xiao *et al.*, "Deep learning image reconstruction simulation for electromagnetic tomography," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3290–3298, 2018.

[148] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[149] Q. Song and F. Xu, "Zero-shot learning of SAR target feature space with deep generative neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2245–2249, 2017.

[150] S.-W. Chen and C.-S. Tao, "PolSAR image classification using polarimetric-feature-driven deep convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 627–631, 2018.

[151] T. Song *et al.*, "Inversion of rough surface parameters from SAR images using simulation-trained convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 7, pp. 1130–1134, 2018.

[152] H. Chen *et al.*, "Machine learning-assisted analysis of polarimetric scattering from cylindrical components of vegetation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 155–165, 2018.

[153] J. Weiß and A. Santra, "One-shot learning for robust material classification using millimeter-wave radar system," *IEEE Sens. Lett.*, vol. 2, no. 4, pp. 1–4, 2018.

[154] Z. Zhang *et al.*, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, 2017.

[155] Q. Cheng *et al.*, "Compressive sensing radar imaging with convolutional neural networks," *IEEE Access*, vol. 8, pp. 212 917–212 926, 2020.

[156] J. Li *et al.*, "3-D voxel-based reconstruction of multiple objects buried in layered media by VBIM hybridized with unsupervised machine learning," *IEEE Trans. Geosci. Remote Sens.*, 2021.

[157] X. L. Travassos *et al.*, "Artificial neural networks and machine learning techniques applied to ground penetrating radar: A review," *Appl. Comput. Inform.*, 2020.

[158] U. Ozkaya and L. Seyfi, "Deep dictionary learning application in GPR B-scan images," *Signal, Image and Video Processing*, vol. 12, no. 8, pp. 1567–1575, 2018.

[159] Y. Wan *et al.*, "Robust and efficient classification for underground metal target using dimensionality reduction and machine learning," *IEEE Access*, vol. 9, pp. 7384–7401, 2021.

[160] J. Sonoda and T. Kimoto, "Object identification form GPR images by deep learning," in *2018 Asia-Pacific Microwave Conference (APMC)*. IEEE, 2018, pp. 1298–1300.

[161] F. Picetti *et al.*, "Convolutional autoencoder for landmine detection on GPR scans," in *2018 41st Int. Conf. Telecommun. Signal Process.* IEEE, 2018, pp. 1–4.

[162] A. P. Sligar, "Machine learning-based radar perception for autonomous vehicles using full physics simulation," *IEEE Access*, vol. 8, pp. 51 470–51 476, 2020.

[163] L. Gao *et al.*, "Fusion image based radar signal feature extraction and modulation recognition," *IEEE Access*, vol. 7, pp. 13 135–13 148, 2019.

Marco Salucci received the M.S. degree in Telecommunication Engineering from the University of Trento, Italy, in 2011, and the Ph.D. degree from the International Doctoral School in Information and Communication Technology of Trento in 2014. He was a Post Doc Researcher at CentraleSupélec, in Paris, France, and a Post Doc Researcher at the Commissariat à l'énergie atomique et aux énergies alternatives (CEA), France. He is currently a Researcher at the University of Trento, Italy, and member of the ELEDIA Research Center. Dr. Salucci is a Member of the IEEE Antennas and Propagation Society and he was a Member of the COST Action TU1208 "Civil Engineering Applications of Ground Penetrating Radar". He is the Associate Editor for Communications and Memberships of the IEEE Transactions on Antennas and Propagation. Moreover, he serves as an Associate Editor of the IEEE Transactions on Antennas and Propagation and as a reviewer for different international journals including IEEE Transactions on Antennas and Propagation, IEEE Antennas and Wireless Propagation Letters, IEEE Journal on Multiscale and Multiphysics Computational Techniques, and IET Microwaves, Antennas & Propagation. His research activities are mainly concerned with inverse scattering, biomedical and GPR microwave imaging techniques, antenna synthesis, and computational electromagnetics with focus on System-by-Design methodologies integrating optimization techniques and Learning-by-Examples methods for real-world applications.

Manuel Arrebola (S'99, M'07, SM'17) was born in Lucena, Spain. He received the M.Sc. degree in telecommunication engineering from the University of Malaga, Málaga, Spain, in 2002, and the Ph.D. degree from the Technical University of Madrid (UPM), Madrid, Spain, in 2008.

From 2003 to 2007, he was with the Electromagnetism and Circuit Theory Department, UPM, as a Research Assistant. In 2005, he was with the Microwave Techniques Department, Universität Ulm, Ulm, Germany, as a Visiting Scholar. In 2007, he joined the Electrical Engineering Department, Universidad de Oviedo, Gijón, Spain, where he is currently an Associate Professor. In 2009, he enjoyed a research stay at the European Space Research and Technology Center, European Space Agency, Noordwijk, The Netherlands. In 2018, he was with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada, and in 2019 with the Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh, UK, both as Visiting Professor. His current research interests include the development of efficient analysis, design and optimization techniques of reflectarray and transmitarray antennas for applications in near and far fields, as well as novel technologies in antenna design.

Dr. Arrebola was a co-recipient of the 2007 S.A. Schelkunoff Transactions Prize Paper Award from the IEEE Antennas and Propagation Society.

Tao Shan received the B.S. degree in electronic engineering, Xidian University, Xi'an, China, in 2016, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2021. He is currently working as a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University. His current research interests include fast algorithms in computational electromagnetics, optimization algorithms for phased electromagnetic surface, deep learning and its applications in computational electromagnetics.

Maokun Li (M'04, SM'20) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from University of Illinois at Urbana-Champaign in 2004 and 2007, respectively. After graduation, he worked as a senior research scientist at Schlumberger-Doll Research, Cambridge, MA, USA. In 2014, he joined the department of electronic engineering, Tsinghua University, Beijing, China. His research interests include fast algorithms in computational electromagnetics and their applications in antenna modeling, electromagnetic compatibility analysis, geophysical exploration, biomedical imaging, etc. He also serves as the associate editor for IEEE Transactions on Antennas and Propagation, IEEE Transactions on Geoscience and Remote Sensing. He was also among the recipients of the 2017 IEEE Ulrich L. Rohde Innovative Conference Paper Award and the 2019 PIERS Young Scientist Award.