Machine Learning-Assisted Method for Efficient and Accurate Antenna Modelling

Ju Tan

Department of Electronic and Electrical Engineering

The University of Sheffield

Supervisors: Prof. Jie Zhang, Mr. Eddie Ball

This thesis is submitted for the approval of the

Doctor of Philosophy

March 2023

Acknowledgements

Firstly, I would like to thank my family, my parents Mr. Maoye Tan and Ms. Cuie Huang, without your love and support, I would not be the person I am today.

I would like to thank my supervisor Prof. Jie Zhang for his patient, encouragement and guidance. Without his continuous guidance, I would not find the right way in my research and finish my research.

I would like to thank Dr. Yu Shao and Dr. Jiliang Zhang, for their patient and guidance, I have learned much in research work. I appreciate the working time we have spent together.

I would like to thank my girlfriend Miss. Huimin Liu, she is always companying me and support me.

I would like to thank all the friend and people who offered me support.

The last, I would always memorise my beloved grandfather, R. I. P. (20th/Mar/2023).

Abstract

Antenna modelling is an important tool for engineers and researchers in the field of telecommunications, as it allows for the design and optimisation of antennas in different scenarios and for a variety of applications. However, conventional methods of antenna modelling can be computationally expensive and time-consuming, which can limit the exploration of design space and lead to the inaccurate or even failed in antenna design and optimisation.

With the rapid development of wireless communication technology, antenna design has attracted extensive attention. As device for transmitting and receiving electromagnetic (EM) signals, antenna has a significant impact on the performance of wireless communication systems. Over the past decade, various new antenna and analysis methods have been proposed. Generally, the modelling and analysis of antenna are carried out in EM simulation software such as Computer Simulation Technology (CST) Microwave Studio, High-Frequency Structure Simulator (HFSS), which can be used to model and simulate various kinds of antennas, and the corresponding performance such as reflection coefficient, gain, radiation pattern and impedance of antenna can be directly obtained through simulation. Unfortunately, modern antenna design is more complicated because of the increasing number of design variables, complex structures, and environmental factors. Parametric sweep is an important function of EM simulation software that allows designers to get the information of an antenna under different conditions, the time cost to run an EM simulation for individual candidate solution varies from seconds to minutes, or even several hours. An antenna with complex structure may require thousands of EM simulation to model, the cost in time and computational resources are impractical and unacceptable for most designers and researchers.

To address these challenges, machine learning (ML) methods have been developed and applied to improve the efficiency and accuracy of antenna modelling. These methods involve using ML algorithms to train models on data, which can then be used to predict the performance of antennas for a given set of design variables. This thesis employs and combines different ML-assisted antenna modelling methods to reduce time, cost, and computational intensity in antenna design and accelerate the design process without compromising accuracy. First, quick estimation can be performed using the linear regression (LR) method based on limited data and computational resources to obtain guidance and check the feasibility of an antenna design. Then one of the ANN-based methods can be selected for antenna modelling and optimisation according to the antenna design complexity. These methods can be combined into a systematic antenna design process for modern antenna design. This set of processes can model and optimise antenna for different applications and scenarios with broad ranges of design variables. Compared to EM simulation-based and conventional ML-based antenna design methods. This process can perform accurate antenna modelling using significantly reduced time and computational resources and eliminate unnecessary costs in optimisation, fabrication and testing.

In the first part, a concrete embedded antenna is proposed to mitigate the space occupation and aesthetic problems of indoor dense small cell deployment. The LR method is employed to fast estimate the relationship between antenna performance (radiation efficiency, gain, and input impedance) and embedding ambient (embedding depth and concrete dielectric constant) since the EM simulation-based antenna modelling is time-consuming. The complex mutual coupling between the antenna and the concrete leads to a limited amount of simulated data, and LR can model and predict the performance parameters of the antenna with limited data and a few computing resources. LR can also use limited resources to evaluate the feasibility of antenna design before implementation and fabrication, which can reduce unnecessary overhead and identify potential issues in the antenna. The findings of this study are beneficial to antenna designers for indoor communication concrete embedding antenna design and deployment, as well as communication-friendly building materials.

In the second part, a heuristic algorithm-enhanced artificial neural network (ANN) is proposed to model concrete embedded antenna. The utilisation of ANN can handle the complex and non-linear relationship between inputs and outputs, and it can also make a prediction on antenna performance when new design points are given. A global optimisation algorithm is used to enhance ANN to eliminate local minima issues, and Bayesian regularisation (BR) is employed to improve the network prediction accuracy at new design points. The network accuracy and efficiency are higher than the conventional back-propagation ANN.

The third part proposes a multi-fidelity neural network for antenna modelling and optimisation. Two sources of simulated data are involved and combined to perform antenna modelling with a large amount of cheap and inaccurate models and a small amount of expensive and accurate models. The correlation between two sources of data can be learned adaptively by decomposing the correlation into linear and non-linear components. The feasibility of the approach is validated by three antenna structures, the results show that this method can make prediction for broad ranges of input parameters with satisfactory accuracy; then the surrogate model is directly applied in the optimisation algorithm framework to replace EM simulation to accelerate antenna optimisation procedure.

List of Publications

Published

J. Tan, Y. Shao, J. Zhang and J. Zhang, “Empirical formulas for performance prediction of concrete embedded atnenna," in *2020 Int. Sym. Antenna Propag. (ISAP)*, IEEE, pp. 201-202, 2021.

J. Tan, Y. Shao, J. Zhang and J. Zhang, “Artificial neural network application in prediction of concrete embedded antenna performance," in *2021 15th European Conf. Antennas Propag. (EuCAP)*, IEEE, pp. 1-5, 2021.

Submitted

J. Tan, Y. Shao, J. Zhang and J. Zhang, “Efficient antenna modelling and optimisation using multi-fidelity stacked neural network," *IEEE Trans. Antennas Propag.*.

Table of contents

List of figures

List of tables

Abbreviations

ANN Artificial neural network

BP Back propagation

BR Bayesian regularisation

BPNN Back propagation neural network

BRNN Bayesian regularisation neural network

CAD Computer-aided design

CEM Computational electromagnetics

CPW Circularly polarised wave

CST Computer Simulation Technology

DoE Design of Experiment

EM Electronmagnetic

FDTD Finite-difference time-domain

GA Genetic algorithm

GD Gradient descent

GPR Gaussian process regression

HLHS Heuristic Latin hypercube sampling

HF High-fidelity

HFNN High-fidelity neural network

HFSS High Frequency Simulation Software

LB Lower bound

LDW Linear decline weight

LF Low-fidelity

LR Linear regression

LFNN Low-fidelity neural network

LHS Latin hypercube sampling

LM Levenburg-Marquardt

MAPE Mean absolute percentage error

MF Multi-fidelity

MFSNN Multi-fidelity stacked neural network

MCS Monte Carlo sampling

MoM Method of moment

ML Machine learning

NLML Negative logarithm maximum likelihood

NMSE Normalised mean square error

MSE Mean square error

NN Neural network

NS Non-surrogate

PDE Patial differential equation

PEC Perfect electric conductor

PSO Particle swarm optimisation

PSO-BRNN Particle swarm optimisation and Bayeisan regularisation enhanced neural network

RBF Radial basis function

SE Squared exponential

SA Simulated annealing

SGD Stochastic gradient descent

SLT Statistical learning theory

SVM Support vector machine

SVR Support vector regression

UB Upper bound

UWB Ultra-wideband

Chapter 1

Introduction

1.1 Background

In 1864, British scientist James Maxwell predicted the existence of electromagnetic (EM) waves and created Maxwell's equations; thus a complete EM theory system was established. Then the German physicist Hertz confirmed the existence of EM waves via physical experiments. EM waves are a type of energy that travels through space in the form of oscillating electric and magnetic field. These waves are one of the fundamental elements of wireless communications, as EM waves carry information and are capable to be transmitted wirelessly over long distances. According to the frequency or wavelength, EM waves can be classified as radio waves, microwaves, infrared light, ultraviolet, visible light, X-rays, and gamma rays. With the continuous development of science and technology, human understanding of EM is also changing, and the application of EM waves is gradually proficient. The wireless communication system is an essential application of EM waves. Wireless communication refers to the transmission of data, voice, and video over a wireless network without physical connections between transmitters and receivers. As the carrier in a wireless communication system, and the EM waves fetch information between terminals, the convenience, mobility,

and flexibility of EM waves make wireless communication increasingly popular. Currently, wireless communication systems include a variety of technologies, such as cellular networks, Wi-Fi, Bluetooth, satellite communication, etc., and these technologies rely on different operation frequencies and modulation methods to fetch data efficiently and securely. Antennas are essential components in modern communication systems, as they are the interface between electronic circuitry and EM waves used to transmit and receive signals. Typically, an antenna system consists of a transceiver, a transmission line, and an antenna. The transmitter couples the alternating current (AC) signal containing information to the transmission line and radiates the alternative EM field confined in the transmission line into the free space through the transmitting antenna. The EM waves propagate in all directions as travelling waves in free space. At the receiving antenna position, the plane travelling wave will be received and converted into an AC signal, and finally processed at the receiving terminals. In the whole process, the antenna is responsible for signal conversion. The signals on the transmission line are converted into EM waves that can propagate freely, then the received travelling wave signals are converted into AC signal for back-end processing at the receiving antennas.

The essence of antenna design is that it can change the form of EM signals (AC on the transmission line, EM waves in space) through a specific structure to radiate or receive EM waves. In other words, a slight change in antenna geometries could interfere with the pattern of EM signal propagation and radiation. For example, in wireless communication, the receiving antennas are usually far away from the transmitted antennas and distributed in different spatial orientations. However, the omnidirectional antenna energy radiation manner does not always meet the actual wireless communication application in practice, antennas should be carefully designed for specific applications.

Antenna modelling is a process of creating a mathematical representation of an antenna's behaviour. The constructed model can be used to simulate antenna performance under various conditions, such as different frequencies, polarisations, and environmental situations. Therefore, antenna modelling is the most crucial step for antenna design and development, and it facilitates antenna optimisation to meet the required specifications.

One of the advantages of antenna modelling is to predict antenna performance. Antenna performance can be influenced by various factors, including the antenna's physical characteristics and other conditions (polarisation, frequency, and environment); using a mathematical model that can represent antenna behaviour and performance at a specific frequency range, the engineer can determine how well the antenna will perform and make any necessary adjustments to optimise antenna performance. Secondly, antenna modelling can improve antenna efficiency. The constructed models can be used to analyse the performance of an existing antenna, using the simulation tools to model the antenna's behaviour, engineers can identify potential issues with the antenna (e.g., poor radiation pattern, impedance mismatch, etc.). With the help of the modelling and analysis results, engineers can discover the root cause of the issue and develop solutions to improve the antenna's performance. Third, antenna modelling can be used to analyse the antenna behaviour at different frequencies range, the engineers can thus find out the suitable operation frequency for the antenna, then ensure the wireless communication system meets the required specifications. Also, the antenna modelling can give an inspection of an antenna's behaviour and performance with different fabricated materials, which facilitates the selection of materials as well as balances the difficulties in fabrication. Last but not least, antenna modelling significantly reduces the cost of antenna design and optimisation. The modelling results reveal the antenna behaviours and performance under different conditions and also provide guidance for engineers to identify areas where cost savings (time, money) can be made, such as size reduction without compromising the antenna performance.

EM simulation is essential and reliable for antenna modelling since the ability to accurately calculate the antenna performance under different conditions, the EM properties such as radiation pattern, gain, and impedance can be obtained through EM simulation. In addition, the EM simulation tools can provide visualised results that demonstrate the relationship between the antenna performance and input factors (e.g., antenna shape, size, fabricated materials). EM simulation tools use mathematical models and numerical methods to compute and simulate the desired performance of an antenna. These tools enable engineers to modify antenna shape, size, fabricated materials and environmental situations, and analyse

antenna performance under this wide range of conditions without the need for physical testing. With the help of EM simulation tools, the numerical results of an antenna under different conditions can be obtained for constructing mathematical models.

In summary, antenna modelling is an essential process for designing and optimising antennas for a variety of applications. EM simulation tools enable engineers to analyse the performance of an antenna under different conditions in a virtual setting, allowing engineers to modify antenna dimensions and identify mistakes in antenna design without conducting physical fabrication and testing.

1.2 Problem and Motivation

EM simulation tools such as Computer Simulation Technology (CST) Microwave Studio, High-Frequency Simulation Software (HFSS) and FEKO, are widely used methods for antenna modelling that allow designers to obtain antenna performance parameters (e.g., reflection coefficient, gain, radiation pattern, and impedance) without expensive and time-consuming prototype fabrication and testing. EM simulations are based on solving Maxwell's equations, which describe the behaviour of EM waves and the interactions between antennas and surroundings. One of the main advantages of EM simulations is the ability to provide physical insight into antenna performance, the visualised and numerical simulated results such as radiation pattern, gain, impedance and bandwidth, etc. With the guidance of these results, designers can efficiently design and optimise antenna performance for specific applications. In addition, EM simulations can also evaluate the antenna performance in a different environment and operating conditions, providing analysis into the impact of factors such as temperature, humidity, conductivity, and nearby objects. These evaluations can be crucial for antenna robustness and reliability in real-world environment operation. However, EM simulation requires a lot of computing resources, time, storage space and advanced EM knowledge. In addition, conventional EM field analysis methods such as the method of moments (MoM) and the finite element method (FEM) consume much time due to a large number of formula calculations.

With the continuous progress of wireless communication technology, antenna design has attracted widespread attention in radar systems, MIMO technology and other wireless devices. As a component for sending and receiving EM signals, the antenna can significantly affect the signal strength and quality. Over the past few years, various new antennas and emerging technologies has emerged, such as implantable antenna [1]-[3], liquid antenna [4] [5], dielectric embedded antenna [6] [7] and material integration antenna [8] [9]. The increasing number of considered design variable and more complex antenna structure can lead to a high computational overdue. Moreover, the modern antenna simulation not only evaluates the radiator itself but the surroundings, such as environmental factors, connectors and installation fixtures, or system-level tasks. Thus, the EM simulation could be a time-consuming process.

ML has become increasingly popular in antenna modelling and design in the last decade since it can learn from antenna simulated and measured data to accelerate the design and optimisation process. A well-trained ML model can accurately predict antenna performance in seconds, enabling designers evaluate antenna performance under a wide range of design space using a relatively short time.

1.3 Thesis Contribution

In this thesis, the ML based antenna modelling methods are investigated and combined as a systematic process to accelerate antenna design for different scenarios and different complexity without compromising accuracy. Considering the complexity and fault tolerance of modern antenna design, a quick estimation that can provide qualitative guidance and check feasibility of antenna design using small amount of data and computational resources is needed in the beginning. Then the antenna is modelled using ANN-based methods. Two different methods based on enhanced ANN are proposed to design antennas accurately with reduced time cost, different methods are selected according to the complexity of an antenna.

In the first part (Chapter 3), a concrete embedded antenna for indoor communication is proposed to mitigate the space occupation and aesthetic problems of indoor dense cell

deployment for the first time. A microstrip antenna with multiple layer configuration operating at 3.5 GHz is embedded in a concrete wall for indoor communication. The impact of embedding depth and concrete dielectric constant on antenna radiation efficiency, gain, and input impedance are investigated. Simple empirical formulas are obtained using LR method based on EM simulation. The LR model can give a prediction on new design point in a second, which saves more than 50 hours to evaluate antenna performance variation in new embedded ambient. The LR model can quickly estimate of antenna input characteristics with relatively small data and low computational resources, and this method is very suitable for testing the feasibility of antenna design and optimisation. The results of this work provide good guidance to antenna designers and architects for concrete embedded antenna deployment and communication-friendly building material selection.

In the second part (Chapter 4), a heuristic algorithm enhanced ANN is proposed to model the concrete embedded antenna. The ANN model takes antenna embedded depth and concrete dielectric constant as inputs and gives antenna radiation efficiency, gain and input impedance as outputs. The particle swarm optimisation (PSO) is employed to search the global optimal weights and bias for ANN, then Bayesian regularisation (BR) is used to train the ANN to overcome the over-fitting issue. The modelling and generalisation errors of enhanced ANN are 0.002% and 26.63%, respectively, which are better than the conventional back-propagation neural network (BPNN) (0.13% and 63.83%) and LR method (45.62% and 222.07%). In addition, the model training time cost of the enhanced network is much lower than BPNN, which is 848 iterations to 3084 iterations, and the time is 172 s to 583 s.

In Chapter 5, a multi-fidelity stacked neural network (MFSNN) is proposed to construct a surrogate model for antenna modelling and optimisation. Multi-fidelity (MF) modelling method is combined with ANN for the first time, which can significantly reduce the cost of gathering initial training data for surrogate model construction, up to 90% time saving is achieved in training data using EM simulation. Furthermore, the correlation between LF and HF models can be learned adaptively and accurately by decomposing the correlation into linear component and non-linear component. Three antenna examples are employed to test the proposed method. The results indicate that the MFSNN can accurately perform

antenna modelling with a reduced amount of HF training data points. The maximum training and testing errors of MFSNN in antenna modelling using reduced amount are 0.034% and 1.64% for antenna 1 (2 design variables, 10% HF data), 0.95% and 8.84% for antenna 2 (8 design variables, 30% HF data), 0.78% and 4.18% for antenna 3 (4 design variables, 10% HF data). The accuracy of MFSNN in antenna modelling is comparable to conventional ANN trained using 100% HF, and the MFSNN can save 38%, 66.74% and 56.3% time cost than EM-simulation based antenna optimisation for Antenna 1, Antenna 2 and Antenna 3; and save 79.26%, 56.3% and 78.34% time cost than conventional ANN based optimisation for Antenna 1, Antenna 2 and Antenna 3, without compromising model accuracy.

1.4 Thesis Organisation

The content of this thesis are arranged as follow: Chapter 2 is the literature review. The numerical method in antenna modelling and analysis is first introduced. Then the reviews of ML methods, the fundamental conception of three popular ML algorithms - ANN, support vector machine (SVM), and Gaussian process regression (GPR) are introduced, their application on antenna modelling, analysis and optimisation are reviewed. In the last part, the techniques that can save time and cost are introduced, and the MF framework and design of experiment (DoE) sampling methods are reviewed. Chapter 3 proposes a concrete embedded antenna for indoor communication and the LR method is used to model the proposed antenna. The antenna performance is investigated using full-wave EM simulation. The relationship between embedding ambient and antenna performance parameters is modelled. Simple empirical formulas are obtained that can fast predict the antenna's performance for any given embedding ambient. In Chapter 4, a heuristic algorithm enhanced ANN is proposed to model the concrete embedded antenna. The ANN model takes antenna embedded depth and concrete dielectric constant as inputs and gives antenna radiation efficiency, gain and input impedance as outputs. In Chapter 5, a MFSNN is proposed to construct surrogate model for antenna modelling and optimisation. Three antenna examples are employed to test the

effectiveness of MFSNN, the modelling and optimisation results demonstrate the efficiency and accuracy of MFSNN. Chapter 6 presents conclusion and future work.

Chapter 2

Literature Review

Overview

In this chapter, the literature reviews related to the ML method and machine learning-assisted antenna modelling method to support the current work and future work. In the first part, a brief review of numerical antenna analysis is presented. In the second part, an in-depth review of ML is given. Its history of development and application scenarios are introduced, then the three most popular ML methods - ANN, SVM and GPR are introduced, and the principles of these methods are presented. After that, the ML-assisted antenna modelling method is introduced. The reason why ML is popular and how ML is combined with antenna research are illustrated. The application of ML methods including ANN, SVM and GPR on antenna design and analysis are reviewed, and associated research works are given to demonstrate the contribution of ML in antenna research. The last part introduces MF modelling approach and DoE methods that can reduce the cost of antenna design.

2.1 Antenna Numerical Analysis

The antenna is the an essential component in wireless communication systems. Therefore antenna analysis and design are crucial factors in improving the performance of wireless communication systems. Antenna modelling aims to characterise the interaction of the EM field with the antenna via solving Maxwell's equation [10]. The introduction of computation

electromagnetics (CEM) makes the practical EM problems analytically calculable, such as radiation, scattering, and coupling. The CEM was employed in the antenna research field and greatly affected the development of antenna analysis. With the help of CEM, antennas with irregular shapes, complex structures, and complex configurations can be designed and analysed. Various numerical methods of CEM can accurately resolve the antenna problem, including MoM [11], Finite-difference time-domain (FDTD) [12] - [14] method and FEM [15] [16]. From the 1980s, commercial CEM software for antenna design, analysis and simulations emerged, such as HFSS, CST and FEKO, which can be directly used to analyse and optimise antenna using these EM solvers.

These numerical methods have been successfully applied to antenna design and analysis. In [18], a lossy circular microstrip antenna was analysed using MoM. In [19], a rectangular monopole antenna with parasitic loop elements was analysed by MoM; its input impedance, reflection coefficient, superficial current density distribution and radiation pattern were calculated, and the results were compared with EM solver simulation results. MoM is employed to investigate an indoor MIMO system performance for the first time in [20], and the relationship between array spacing and MIMO channel capacity was demonstrated. A body-worn UHF antenna system was designed using the FDTD method in [21], and the complex coupling between antenna and the human body is modelled. In [22], the FDTD method was employed to demonstrate that using the discontinuity-radiation concept can make the long dielectric rod antenna efficient; the calculated results agreed well with the experimental results. The FEM-based EM solver HFSS was employed to analyse a small UWB microstrip-fed tap monopole antenna in [15]. Similarly, in [16], a print dipole antenna integrated with an artificial magnetic conductor block was analysed, its radiation characteristics were simulated, and the direct optimisation was implemented using HFSS EM solver.

The CEM-based numerical methods and EM simulation provide physical insight into the antenna, which is regarded as a rigorous way to design and evaluate antenna. However, with the rapid development of modern wireless communications, antennas are getting more complex, and the requirements and constraints for designing and optimising are the most

concerns. The EM simulation will be computationally-intensive once more parameters and design specifications are considered. A large amount of EM simulations are necessary to obtain the variation of antenna characteristics with different design parameters so that the parameter sweeping can continuously evaluate the antenna performance for different geometries. As a result, the computational cost is the main challenge for antenna design and optimisation, and the method that can efficiently design and optimise antennas is required to address the issue.

2.2 Machine Learning

2.2.1 The History of Machine Learning

ML is a multi-disciplinary science covering probability theory, statistics and optimisation theory, etc. Its fundamental principle is to learn from data and previous experience, then automatically improve the performance of algorithms to achieve optimal performance [26]. Algorithms, experience and performance are the three essential elements of ML [27]. ML acquires experience through data, uses algorithms to build up models, then trains and evaluates models based on data. When the performance of the model meets the design requirements, the model can be used for testing, and can generate the corresponding output on the new input set. If the model does not meet the requirements, the model will continue to evaluate the error and adjust hyper-parameters until the convergence condition is met. The development process of ML can be roughly divided into three stages [28] [37], the "Deduction" stage (1950s - 1960s), the "Knowledge" stage (1970s - 1980s) and the Machine Learning period (1980s - now).

The 1950s to 1960s period was the initial stage of ML. The theoretical basis of this period was the neural network (NN) model that appeared in the 1940s. At that time, the purpose of ML was to enable the machine to deduce logically and make the machine intelligent. Some artificial intelligence (AI) programs can solve and prove some famous mathematical theorems. The most representative event during 'Deduction' period was the halma playing

program by A. Samuel [29]. However, in this period, due to the lack of stored data and insufficient understanding of ML, the ML system was far away from the criteria of AI.

Then between the 1970s and mid-1980s, in this period, there was a rapid development in ML. As a result, the ML system at this stage has acquired a large amount of knowledge and data, and the learning methods and strategies have been continuously improved at the same time. As a result, an "expert system" has been established based on a large-scale knowledge reserve [30]. However, faced with the knowledge reserve, the "expert system" is still unable to generalise the knowledge. Based on that, the "autonomous learning" conception was proposed.

From the 1980s, ML was booming. In 1986, the famous back-propagation (BP) algorithm was proposed by D. E. Rumelhart's team [58], which significantly made the NN algorithm popular again. Thus, the BP algorithm became one of the most widely used and frequently used ML algorithms. In the mid-1990s, statistical learning [31] became mainstream, and the representative algorithms were SVM [32] and kernel methods [33]. At the beginning of the 21st century, due to the increase in knowledge reserve and computer configuration, the ANN became mainstream again and deep learning [34] was proposed and applied, the ML systems performance in this stage has been dramatically improved. In addition, ML's basic theory and mathematical interpretability have been further strengthened and developed, so ML has received widespread attention from many fields. Some well-known ML techniques, such as pattern recognition [35], deep learning, biological information processing [36], etc., emerged during this phase.

With the rapid development of computer technology, computers at present have the ability to store and process a massive volume of data, which makes ML systems more accurate and intelligent. On the other hand, the scope of ML applications is expanding to different areas, such as autopilot [38], natural language processing [39], face recognition [40], and speech recognition [41].

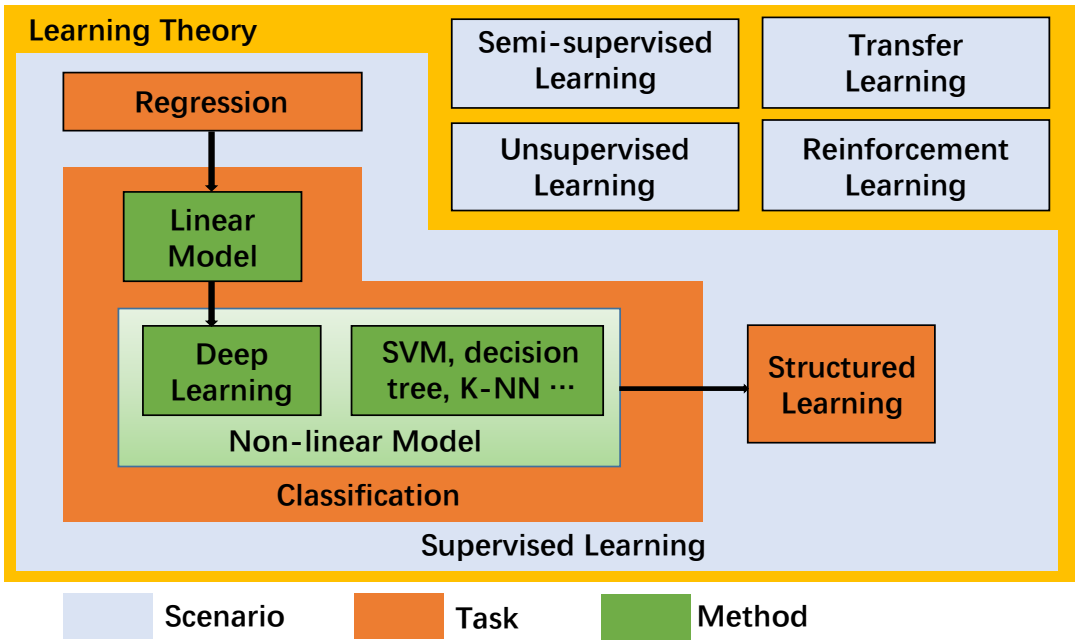

Fig. 2.1 Learning map [25].

2.2.2 The Categories of Machine Learning

The category map of machine learning methods is presented in Fig. 2.1. The ML methods can be briefly classified into four categories, which are supervised learning [42], semi-supervised learning [43], unsupervised learning [44] and reinforcement learning [45].

In supervised learning, the input data is called "training data", and each set of training data has its corresponding identification or results, such as "spam" or "non-spam" in an anti-spam system, handwritten digit recognition. While building a model, supervised learning establishes a learning process - the prediction result is compared with actual result in training data, and continuously adjusts the model parameters until the accuracy of model reaches the expectation. The common application scenarios of supervised learning are classification problems and regression problems, and the typical algorithms are Logistic Regression [47] and back-propagation NN [57].

In unsupervised learning, all the data is not specifically identified, and the model learns to infer the intrinsic structure of the data. The typical application for unsupervised learning is clustering, the Apriori algorithm [48] and K-means algorithm [49] are usually used.

In semi-supervised learning, the input data is partially labelled and partially unlabelled. The semi-supervised model can be used to make predictions. The model first needs to learn the internal structure of the data in order to organise the data correctly so that it can make accurate predictions. The application scenarios include classification and regression. The algorithms for semi-supervised learning are derivative of commonly used supervised learning algorithms, such as Graph Inference [50] and Laplacian SVM [51].

In reinforcement learning mode, the input data is used as model feedback. Unlike supervised learning, the input data is only used as the criteria to check the correction of the model. In the reinforcement learning procedure, the input data is directly fed back to the model, which will adjust and update according to the input data. Typical applications of reinforcement learning contain dynamic systems and robot control. For example, Q-Learning [52] and Temporal difference learning [53] are extensively used in reinforcement learning.

The three most popular supervised learning algorithms in antenna design and analysis - ANN, SVM, and GPR, will be reviewed in the following, their basic conception and principle are introduced.

2.2.3 Artificial Neural Network

Conception and Principle

Artificial neural network (ANN) was proposed in the 1940s by Warren McCulloch and Walter Pitts [78], and became one of the most popular ML algorithms. The ANN was invented based on the principle of brain neuroscience. It mimics the information processing and storing way of the human brain NN, which means this algorithm has the ability of learning and generalization. ANN is composed of a large number of neurons, which processes information by adjusting the weights and biases of the interconnection between neurons (as shown in Fig. 2.2), where x_1 to x_n are the signals generated by the signal source or other neurons, w_i donates the ith weight, θ is the threshold (also known as bias), thus the relationship between

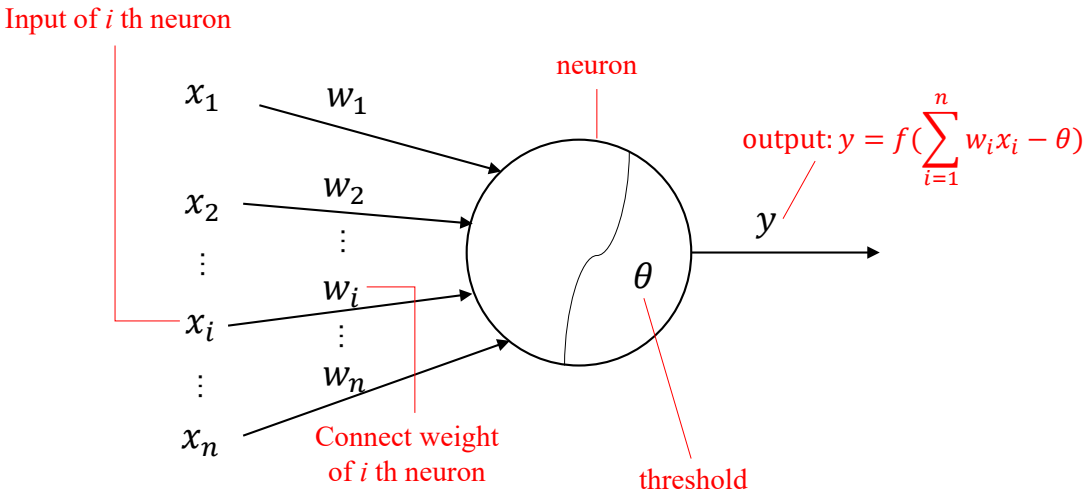

Fig. 2.2 Diagram of simple NN with single neuron [78].

input x and output y could be expressed as:

$$y = f(\sum_{i=1}^{n} w_i x_i - \theta), \tag{2.1}$$

where y stands for the output of neuron, f is the activation function, w_i is ith weight of neural network, x_i denotes ith point in input vector, and θ is the threshold of the neuron. The activation function is used to introduce non-linear factors to the network and compute the current neuron. Thus, the network could solve more complex problems properly. The selection of activation functions is essential when ANN is constructed since the ANN convergence speed and accuracy can be different when using different activation functions. Here are some commonly used activation function [54] are presented in Fig. 2.3 to Fig. 2.5, respectively:

• Sigmoid/ logistic function [55]:

$$g(x) = \frac{1}{1+e^{-x}}, \tag{2.2}$$

• Hyperbolic tangent function [56]:

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}, \tag{2.3}$$

• Rectified liner unit function [56]:

$$ReLU(x) = \begin{cases} 0, & x < 0 \\ x, & otherwise \end{cases} \tag{2.4}$$

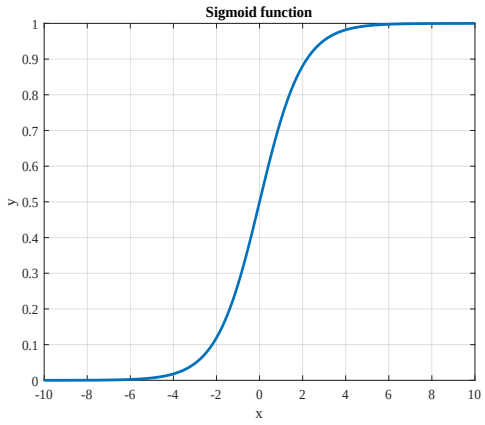

Fig. 2.3 Sigmoid activation function [55].

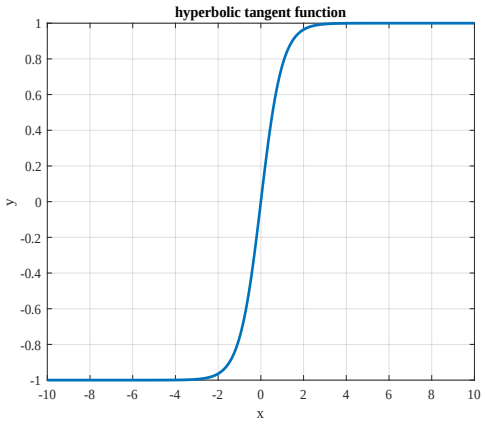

Fig. 2.4 Hyperbolic tangent activation function [56].

ANN Training

In order to train the ANN, a feasible algorithm is required. The BP algorithm [57] [58] is an extensively used gradient-based algorithm for ANN training. The main feature of BP NN is forward propagation and error backpropagation. The input signal is processed layer by layer from the input layer through the hidden layer to the output layer in the forwarding

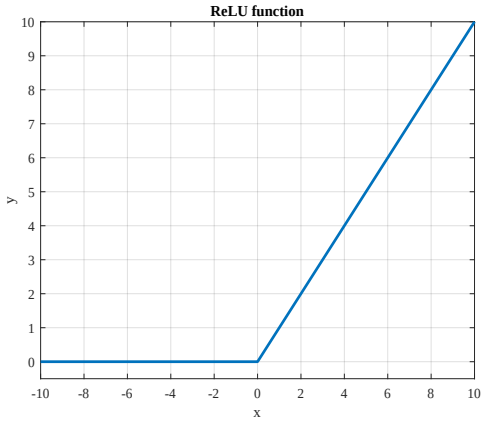

Fig. 2.5 ReLU activation function [56].

transmission of the signal. The neurons in each layer are only related to the output of the previous layer. When the results of the output layer cannot meet the requirement, the predicted results are backpropagated, and the weights and biases are adjusted according to the gradient computation until the predicted output of ANN approximates the expected output. Before using the BPNN for prediction or classification, the network needs to be trained so that the network can map the relationship between input and output. The training process of BPNN (shown in Fig. 2.6) can be summarised as the following steps:

(1) Network initialisation

Determine the topology of the network according to the complexity of the problem – the input neuron and output neuron of the network. It includes the number of input layer nodes n, the number of hidden layer nodes l, and output layer nodes m. The weights between the input layer and hidden layer w_{ij}, hidden layer and output layer w_{jk}, and corresponding biases a and b are initialised. Meanwhile, the relevant network parameters, such as activation functions and learning rate, are determined. The weights and biases are generally initialised to random numbers in the range of 0 to 1.

(2) Hidden layer output computation

Use the weight w_{ij} and the bias a between the input layer and the hidden layer to calculate the output H of the hidden layer:

$$H_i = f(\sum_{i=1}^{n} w_{ij}x_i - a_j), j = 1,2,...,l. \tag{2.5}$$

(3) Network output computation

Use the output H of the hidden layer in step 2, plus with the weights, w_{jk} and bias b between the hidden layer and the output layer, the network output can be calculated as:

$$O_k = \sum_{j=1}^{l} (H_j w_{jk}) - b_k, k = 1,2,...,m. \tag{2.6}$$

(4) Error back propagation

The difference between the original data Y and the network output O is referred as the prediction error of the network:

$$e_k = Y_k - O_k, k = 1,2,...,m. \tag{2.7}$$

(5) Update weights and biases

According to the error e calculated in the previous step, update the weights and biases in the network.

$$w_{ij} = w_{ij} + \eta H_j(1-H_j)x_i \sum_{k=1}^{m} w_{jk}e_k, i = 1,2,...,n; j = 1,2,...,l \tag{2.8}$$

$$w_{jk} = w_{jk} + \eta H_j e_k, j = 1,2,...,l; k = 1,2,...,m. \tag{2.9}$$

$$a_j = a_j + \eta H_j(1-H_j)x_i \sum_{k=1}^{m} w_{jk}e_k, j = 1,2,...,l, \tag{2.10}$$

$$b_k = b_k + e_k, k = 1,2,...,m, \tag{2.11}$$

where the η is the learning rate of neural network, the weights and biases will stop updating until the termination conditions (e.g. target error, number of epoch) are met.

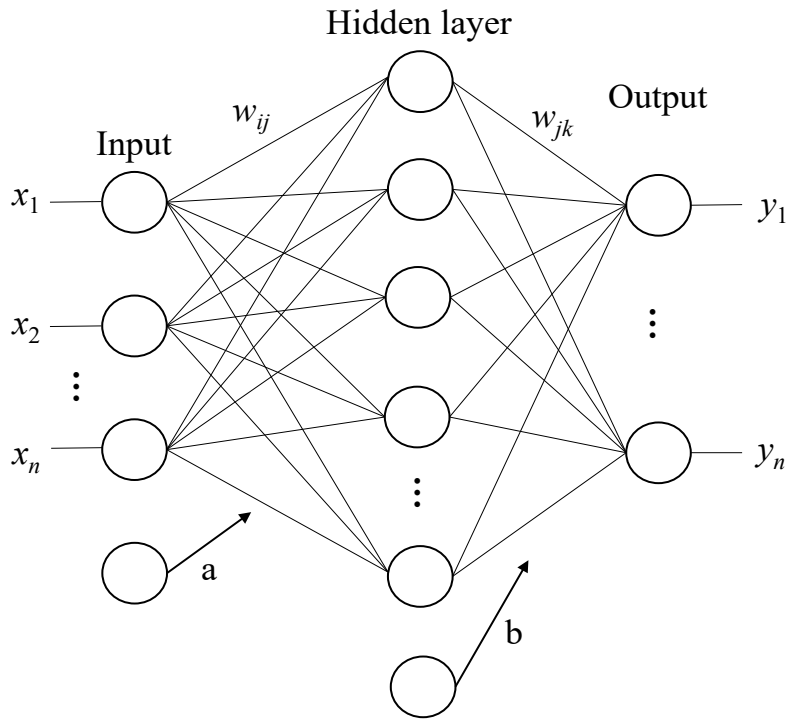

Fig. 2.6 Back-propagation neural network example.

Strategies for Improving ANN Performance

Even though the ANN can approximate any continuous function with arbitrary complexity with arbitrary accuracy [60], the local optimum and over-fitting issues are the main concerns of the gradient-based algorithms [64].

Gradient descent (GD) search is the most widely used parameter optimisation method. In this method, some initial values are selected as the initial point and the optimal parameters are iteratively sought. In each iteration, the gradient of the objective function in the current point is computed, and then the search direction is determined based on the gradient. For example, since the negative gradient direction is the fastest descent direction, the algorithm would search for the optimum along the negative gradient direction. Once the gradient of the objective function at the current point is zero, the algorithm determines that the local minimum has been reached, and the update amount will be zero, which indicates that the iterative update of the parameters is stopped. It is can be observed that if the objective function has only one local minimum, then the found local minimum is the global minimum.

However, if there are multiple local minimums for the objective function, the algorithm cannot guarantee that the found optimal solution is the global minimum. In practice, some strategies are employed to escape the local minimum to reach the global minimum. The optimisation such as PSO [64] [65], genetic algorithm (GA), stochastic gradient descent (SGD) and simulated annealing algorithm (SA) are combined with ANN for escaping the local minimum and the ANN performance can be significantly improved.

The over-fitting issue is another challenge, it is characterised by decreasing training errors and increasing testing errors. There are some ways to prevent over-fitting: early stopping [67], regularisation and dropout. Early stopping is employed to truncate the iteration epoch to prevent over-fitting, stopping iteration before the model converges on the training data. In order to obtain a well-performing NN, the algorithm will traverse the entire dataset over many epochs. If the number of epochs is inadequate, the network may be under-fitting, and vice versa, the too large number of epochs can lead to over-fitting. Early stopping aims to solve the problem that the number of epochs needs to be manually set. In the early stopping strategy, the data set is divided into training and testing sets. The training set is used to compute gradients, update weights and biases, and the testing is used to estimate network errors and generalisation capabilities. Once the training error decreases while the testing error keeps increasing, the training is stopped, and the weights and bias of the smallest testing error checkpoint are returned.

Regularisation is another commonly used solution to mitigate the over-fitting issue, It adds a penalty term into the objective function to describe the complexity of the network, such as the sum of the square of network weights and bias. The commonly used regularisation methods are L1 regularisation [68] and L2 regularisation [69], the objective functions under different regularisations are:

$$E_{L1} = \lambda \frac{1}{m} \sum_{k=1}^{m} e_k + (1-\lambda) \sum_{i} \|w_i\|_1, \tag{2.12}$$

where λ is the regularisation ratio, e_k denotes to kth error, m is the total number of calculated error, w_i denotes to the ith weight in network. The L1 normalisation of network

weights is considered in L1 regularisation, and the L1 regularisation is also known as Lasso regularisation.

$$E_{L2} = \lambda \frac{1}{m} \sum_{k=1}^{m} e_k + (1-\lambda) \sum_{i} \|w_i\|_2^2, \tag{2.13}$$

where λ is the regularisation ratio, e_k denotes to kth error, m is the total number of calculated error, w_i denotes to the ith weight in network. Different from L1 regularisation, L2 regularisation takes the square of Euclidean normalisation of network weights in regularisation, and the L2 regularisation is also known as Ridge regularisation.

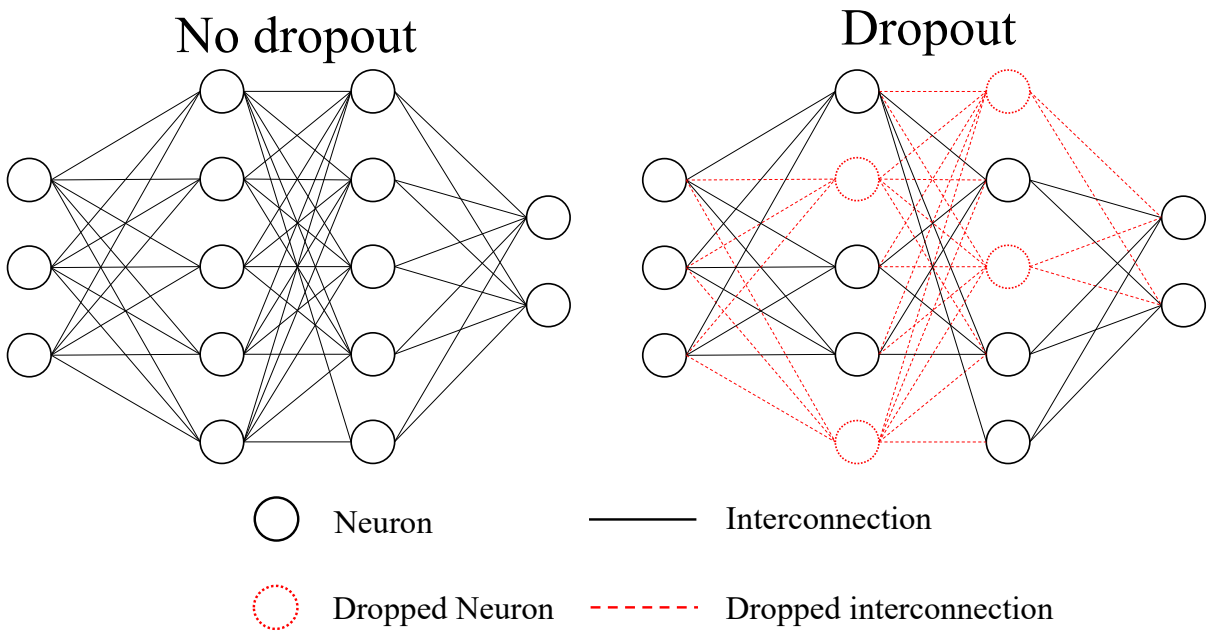

Fig. 2.7 Diagram of dropout [70].

Dropout is usually used during the neural network training process. It prevents overfitting via adjusting the construction of NN [70]. In a dropout operation, a random number of neurons in the hidden layer are dropped during the training with a certain probability, but the dropping is not the actual deletion. Instead, the activation function of these dropped neurons is set to be zero so that the output of these selected neurons is zero, or these neurons are disabled from being involved in the computation. By using dropout, at each iteration of training, the topology of NN would be different. Thus the correlation among networks is reduced.

2.2.4 Support Vector Machine

SVM was proposed based on statistical learning theory (SLT) by Vapnik [61] at the 1990s, which is widely popular in addressing classification problems [74]. Moreover, due to its excellent generalization capability, the SVM has been extensively applied to many fields (e.g. pattern recognition, data mining). The mathematical framework of SVM will be briefly presented as follows.

Linearly separable problem

SVM was first applied to the classification problem. As shown in Fig. 2.8, the solid red line can classify two different types of samples, and the two dashed lines pass through the points that are closet to the samples and dashed lines are paralleled to each other, the distance between two dashed lines is called the classification interval γ. The red line in Fig. 2.8 aims to separate the two types of samples correctly, and the classification interval γ needs to be maximised. Assuming that the normal vector of the three lines is w, and the expression for two dashed lines are $w^T x + b = 1$ and $w^T x + b = -1$, for upper side and downside, respectively. Thus, the equation of the red line can be known as $w^T x + b = 0$. Furthermore, the classification interval could be obtained as $\gamma = \frac{2}{||w||}$. In order to maintain the separable state of the sample, it is necessary to consider constraints according to the specific problem ($y_i(wx_i + b) - 1 \geq 0$ for the case in Fig. 2.8). The optimisation problem is to find out a hyperplane with a maximum margin for classification. In other words, the aim is to find the w and b that satisfy equation (2.14), which makes γ max, which is expressed in equation (2.15).

$$\begin{cases} w^{\mathrm{T}} x_i + b \geq 1, & y_i = 1; \\ w^{\mathrm{T}} x_i + b \leq -1, & y_i = -1. \end{cases} \tag{2.14}$$

$$\max \frac{2}{||w||}, s.t. y_i(wx_i + b) \geq 1, (i = 1, 2, ..., l). \tag{2.15}$$

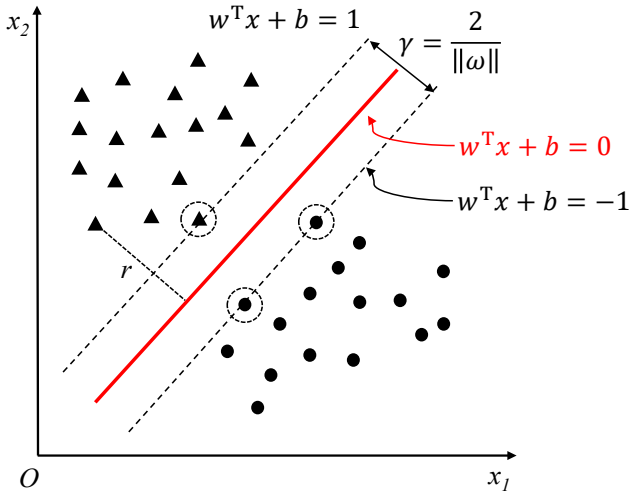

Fig. 2.8 Support vector and margin [73].

Equation (2.15) can be rewritten as:

$$\max \frac{1}{2}||w||^2, s.t. y_i(wx_i + b) \geq 1, (i = 1, 2, ..., l), \tag{2.16}$$

Equation (2.16) could be eventually deducted using equation (2.14) and (2.15), and it is the basic form of SVM.

Linearly inseparable problem

In practice, there may exist the problem that the samples are inseparable. In this case, it is necessary to map the input vector to a high-dimensional feature space, and the optimal classification hyperplane will be searched in this feature space. In this step, an appropriate kernel function is selected to transfer the inseparable problem from the initial space to high-dimensional feature space so that the problem is converted into a linearly separable problem. The usage of kernel function substitutes the inner product operation in high-dimensional space, which is $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. The kernel function is responsible for mapping the input vector between spaces, and its usage can avoid the dimension curse. Fig. 2.9 shows the kernel mapping between input space and high-dimensional space.

In some cases, the error may occur in sample classification for linearly inseparable problems. Therefore, the error tolerance should be considered to prevent the problem from

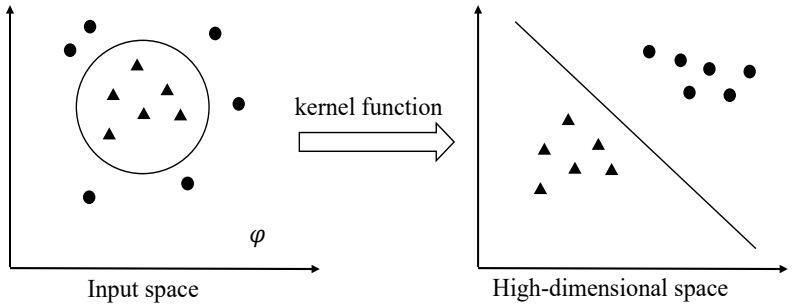

Fig. 2.9 Mapping between input space and high-dimensional space [73].

becoming unsolvable, and the slack variable ξ_i is added. In addition, for the misclassified sample, a penalty term C is added. Thus, the optimisation problem is formulated as follows:

$$\min_{w,b} \frac{||w||^2}{2} + C\sum_{i=1}^{l} \xi_i, s.t. y_i[wx_i + b] \geq 1 - \xi_i, \xi_i \geq (i = 1,2,...,l). \tag{2.17}$$

Support Vector Regression (SVR)

Given training data set $D = [(x_1,y_1),(x_2,y_2),...,(x_m,y_m)], x_i, y_i \in R^n, i = 1,2,...,l$, where x_i is the input data and y_i the output data. A real-valued function $f(x) = w^{\mathrm{T}} \cdot \phi(x) + b$ is being searched in R^n according to the given training dataset, where the $\phi(x)$ is the vector responsible for mapping the input vector into the high-dimensional feature space [74]. The weight vector and bias are w and b, respectively. In this way, for any given input vector $\boldsymbol{x}$, the corresponding output can be inferred from the function $f(x)$.

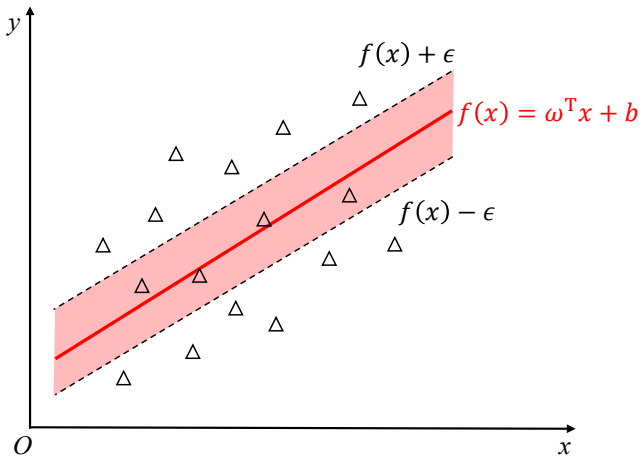

Fig. 2.10 Support vector regression diagram [73]

In the regression problem, the error is calculated as the difference between the model output $f(x)$ and actual value y. When $f(x) = y$, the error equals 0. In SVR, the maximum error tolerance between $f(x)$ and y is assumed to be ε, as shown in Fig. 2.10. In other words, an interval band with a width of 2ε is constructed using $f(x)$ as the centre. The training points located in the interval are not penalised, while the points that fall outside the boundary are penalised. Hence, the regression problem can be formulated as follows:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{m} l_\varepsilon((w^{\mathrm{T}} \cdot \phi(x) + b) - y_i), \tag{2.18}$$

where C is regularisation term, $l_\varepsilon(\cdot)$ is ε-insensitive loss function which expressed as:

$$l_\varepsilon(z) = \begin{cases} 0, & |z| \leq \varepsilon; \\ |z| - \varepsilon, & \text{otherwise.} \end{cases} \tag{2.19}$$

In equation (2.18), the $\phi(x_i)$ is defined by using kernel function K, the Gaussian kernels, or known as radial basis function (RBF), is commonly used in mapping as [76]:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\frac{(x_k - x_k')^2}{2\sigma^2}). \tag{2.20}$$

The model is trained to find optimal w and b by minimising the equation (2.18). It is noted that the error tolerance ε can significantly affect the model accuracy and sparsity. A larger ε can lead to sparser solution [75].

2.2.5 Gaussian Process Regression

Gaussian process [114] (GP) is a collection of random variables, the number of variables in this collection is finite, and all the variables are jointly Gaussian distributed. As a Gaussian distribution can be specified by a mean vector and a covariance matrix, a GP can be uniquely characterised using a mean function and a covariance function as:

$$f \sim \mathscr{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')), \tag{2.21}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are the samples in the design space $\mathbb{R}^N$, N is the number of samples. $m(\boldsymbol{x})$ and $k(\boldsymbol{x},\boldsymbol{x}')$ are the mean function and the co-variance function respectively. The co-variance matrix stores the correlation between two samples, which is calculated using covariance functions, some typical covariance functions are presented as following:

• squared exponential (SE) covariance function [71]:

$$k(\boldsymbol{x},\boldsymbol{x}') = \sigma_f^2 \exp(-\frac{(x_k - x_k')^2}{2l^2}), \tag{2.22}$$

where σ_f and l are the hyper-parameters of the co-variance function. σ_f is the scale amplitude parameter, while l is the length-scale parameter.

• Periodic kernel covariance function:

$$k(\boldsymbol{x},\boldsymbol{x}') = \sigma_f^2 \exp(-\frac{2\sin^2(\pi\frac{(x_k - x_k')}{p})}{2l^2}), \tag{2.23}$$

where p is the period term, other terms have the same meaning as SE function.

• Local periodic (LP) covariance function [72]:

$$k_{LP}(\boldsymbol{x},\boldsymbol{x}') = k_{SE}(\boldsymbol{x},\boldsymbol{x}') \times k_{PER}(\boldsymbol{x},\boldsymbol{x}'). \tag{2.24}$$

In order to carry out the predictions at the new sample points $\boldsymbol{X}^*$, the prior of zero mean is assumed over the training samples and new samples, and the prior distribution could be notated as:

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}^* \end{bmatrix} \sim \mathcal{N}(\boldsymbol{0}, \begin{bmatrix} K(\boldsymbol{X},\boldsymbol{X}) & K(\boldsymbol{X},\boldsymbol{X}^*) \\ K(\boldsymbol{X}^*,\boldsymbol{X}) & K(\boldsymbol{X}^*,\boldsymbol{X}^*) \end{bmatrix}). \tag{2.25}$$

With the appropriated estimated hyper-parameters, we can calculated the posterior mean and the variances of new sample points. The posterior mean is the prediction of the new sample points and the posterior variance indicate the uncertainty of prediction. Then the predicted mean $m(\boldsymbol{X}^*)$ and variance $\sigma^2(\boldsymbol{X}^*)$ are given as:

$$m(\boldsymbol{X}^*) = K(\boldsymbol{X}^*,\boldsymbol{X})[K(\boldsymbol{X},\boldsymbol{X})]^{-1}\boldsymbol{y}, \tag{2.26}$$

and

$$\sigma^2(\boldsymbol{X}^*) = K(\boldsymbol{X}^*,\boldsymbol{X}^*) - K(\boldsymbol{X}^*,\boldsymbol{X})[K(\boldsymbol{X},\boldsymbol{X})]^{-1}K(\boldsymbol{X},\boldsymbol{X}^*), \tag{2.27}$$

where $K(\boldsymbol{X},\boldsymbol{X})$ is the co-variance matrix that evaluates the correlation between the points in training samples, $K(\boldsymbol{X},\boldsymbol{X}^*)$ is the co-variance matrix between the training point and new point of interest. $\boldsymbol{X}$ donates the training inputs and $\boldsymbol{y}$ is the training outputs. For computing the prediction via (3) and (4), the hyper-parameters are estimated by minimising the negative logarithm maximum likelihood (NLML) as:

$$\log p(\boldsymbol{y}|\boldsymbol{X},\sigma_f,l) = -\frac{1}{2}(n\log 2\pi + \log|\boldsymbol{K}| + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{y}), \tag{2.28}$$

where the $\boldsymbol{X}$ is the training inputs and the $\boldsymbol{y}$ is the target output. The GP model is characterised with optimal mean function and covariance function after training. For any given input data, the GP model can give corresponding prediction along with the uncertainty of the predicted result.

2.3 Machine-Learning-Assisted Antenna Surrogate Modelling Methods

With the increasing requirements for antenna performance in practical engineering, the antenna structure becomes more and more complex. In the antenna design, for the antenna with regular shapes, the expected antenna geometry can be obtained through the formula calculation (e.g., transmission line theory formula). However, for the antenna with irregular shapes and complex structures, it is necessary to build models often within the design space through full-wave EM simulation. Although the parameter sweep is running to obtain the optimal design, this process requires hundreds or even thousands of calculations, which is computationally expensive. On the other hand, as the complexity of the antenna increases, the relationship between antenna geometry and performance parameters is difficult to characterise as analytic formulas. Since there are inherent complex non-linear relationships, the numerical formulas can hardly be characterised.

Introducing ML methods can significantly reduce the enormous computation overhead and provide a feasible antenna analysis and design solution. In the past few decades, the machine-learning-assisted method has been widely used in CEM, antenna, passive device and circuit design, and many research achievements have been made. Since the 1980s, ML methods have attempted to solve the problem of EM signal propagation. The combination of ML methods and antenna design was proposed, and the efficiency and accuracy of ML methods became the research focus. ML-assisted antenna design is to use full-wave EM simulation to acquire the dataset, then train ML models with this dataset to capture the non-linear mapping relationship between antenna performance parameters (e.g., reflection coefficient, gain and axial ratio, etc.) and design geometries (e.g., patch length, width, substrate thickness, etc.). Once the ML model is successfully trained, the model can be used to evaluate the specific performance of a specific antenna. Compared with the full-wave EM simulation evaluation, the evaluation time cost of ML model (within a second) is much lower than full-wave EM simulation (usually takes minutes or even hours), and the accuracy of the ML model is comparable to the full-wave EM simulation. ML methods have been widely used in antenna research, including antenna modelling, analysis, synthesis, sensitivity analysis and optimisation.

Three most popular ML methods contains ANN [79] - [98], SVM [99] - [108] and GPR [114]-[123], the applications of these methods in antenna research are reviewed in the following.

2.3.1 Artificial Neural Network for Antenna Modelling

ANN is one of the most well-known ML methods and it has been introduced to the EM field and microwave engineering since the 1990s. In the late 1990s, ANN was introduced to microstrip analysis [79] and design [80] [81], and the feasibility and efficiency of ANN in antenna research are approved. In [79], the ANN was attempted to analyse the microstrip antenna, and it found that, compared to the numerical methods, the ANN can generate an analysis solution in a very short time. [80] used ANN in circular antenna design, the ANN was adjusted after testing with different sizes of training data and testing data, and the network

output agreed well with experiment results (the average error is less than 0.74%). A square-patch antenna problem is addressed using ANN in [81], the patch length is computed using ANN under the given substrate dielectric constant, thickness and antenna dominant-mode resonant frequency and the network parameters were optimised by trial and error way.

The ANN capability in addressing the non-linear and high-dimensional problem was preliminarily validated. The ANN started being used systematically in antenna synthesis and analysis (as shown in Fig. 2.9). [82] and [83] used ANN to calculate antenna input impedance, and the network predictions were comparable to the measured data. The antenna synthesis model was demonstrated in detail in [84], the antenna synthesis and analysis problem formulation was given, and a rectangular microstrip antenna was used to verify the model; different types of neural networks were used, and each synthesis ANN accuracy was reported above 90%. In [85], synthesis and analysis ANN models are trained with proper training algorithms, the accuracies were generally improved. In [86], the ANN synthesis model was first introduced in analysing single-feed circularly polarized microstrip antennas. A 3-layer ANN synthesis model was built, and different ANN topologies were tested and compared. The synthesis model's training and testing errors were reported to be less than 5%, comparable to the EM simulation and measurement results. A more general work has been produced in [89], ANN-based synthesis and analysis models for rectangular, circular and equilateral triangular microstrip patch antennas. The return-loss characteristics of different microstrip patch antennas have been reported, the average errors were limited within 2%.

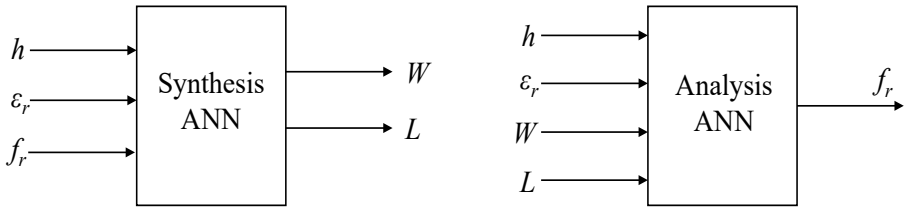

Fig. 2.11 Synthesis and analysis ANN model for antenna design [89]

The training strategies and algorithm can also significantly affect the accuracy and efficiency. The main drawbacks of the gradient-based algorithm are the slow convergence ratio and local minima issues. In addition, the over-fitting issue and generalisation capability

are the main concerns for ANN. For improving the ANN performance, the optimisation algorithms such as PSO [64] and GA [66] to address the local minima issue.

The utilisation of algorithms gradually becomes an essential aspect of antenna research. In [87], the Gaussian parametric model cooperated with ANN in the broadband antenna design problem. The Gaussian parametric model was used in the first step to obtaining the antenna impedance characteristics. Then these characteristics were taken as the inputs of ANN. The ANN was used to approximate the non-linear correlation between Gaussian parameters and antenna parameters, the results showed that this training method is efficient, accurate and robust. Up to seven training algorithms have been used in [88], the ANNs were trained to estimate the dimension of rectangular microstrip antennas, the RBF network was demonstrated to be optimal among other algorithms and the accuracy agreed well with the simulation results. [96] proposed a hybrid ML model that can map antenna performance characteristics onto the antenna design geometries. 10 different ML methods were combined to improve the generalisation capability of the ML model for analysing pin-fed patch antenna, the feasibility and efficiency were demonstrated, and the hybrid method can efficiently cross-validate the results. An ANN-based synthesis model for C-shaped antenna analysis using BR was presented in [98], and high efficiency and accuracy were achieved.

Besides antenna element analysis and design, ANN was also applied to reflectarray antenna analysis and modelling [90] - [93]. The utilisation of ANN accelerates the design and analysis of the reflectarray antenna, and high efficiency and accuracy were obtained. Recently, an ANN-based multi-parameter model [95] was developed to facilitate the procedure of antenna modelling, and it was shown that the results predicted by ANN agree well with the EM simulation. A consensus deep neural network [96] was proposed to reduce the uncertainty of single ANN and accelerate the antenna optimisation. In [97] an efficient knowledge-based ANN was proposed to design a circularly polarised lens antenna with multi-objectives.

2.3.2 Support Vector Machine for Antenna Modelling

The excellent generalisation capability of SVM makes it popular in antenna design. Furthermore, the amount of training data for SVM is small, thus the time cost in the training set generation by full-wave EM simulation is shrinking.

SVM is quite popular in antenna elements modelling. In [99], different types of millimetre wave (mmWave) transition structures, the modelling efficiency and accuracy were excellent compared to the EM simulation results. A SVR-based synthesis and analysis model was built in [100] in order to efficiently design microstrip lines, the model accuracy and efficiency were demonstrated to be comparable to ANN. In [101], SVR is introduced to model the characterisations of rectangular microstrip antenna, the performance parameters, including resonant frequency, operation bandwidth and input impedance, were modelled, and the results of SVR were compared with ANN's. [102] used SVM to accelerate the design procedure of rectangular patch antenna element and rectangular patch antenna array, the relationship between antenna geometries and performance values was explored, and the predicted results agreed with the simulation data. In addition, attempts have been made to model antenna elements with complex structures. A Bayesian SVR is employed to efficiently and accurately model planar antenna input characteristics in [103]. In [104], slotted microstrip antennas with irregular ground plane structures were characterised using SVM. The model can compute the radiation pattern at a specific resonant frequency. In [105], a broadband SIW cavity-backed slot antenna that operates at mmWave frequency band was modelled efficiently.

The SVM is also popular in modelling reflectarray unit cells. [106] used the SVM framework to characterise an array antenna cell unit, the radiation pattern was efficiently characterised and the result was compared with the MoM computed result, efficiency and accuracy were both demonstrated. In [107], the SVM was used to compute the reflection coefficient matrix of a dual-polarised reflectarray unit cell. The SVM models were built for different incident angles, for given design geometries, the models could accurately predict the real part, imaginary part and magnitude of the reflection coefficient. Three different types of space communication reflectarray antennas were accurately modelled in [108], and the cross-polar optimisation was implemented with SVM to significantly reduce the computation

time without compromising accuracy. A multi-frequency optimisation method based on SVM was proposed in [109], a very large shaped-beam reflectarray was fast modelled with high accuracy, then the antenna was optimised based on different frequency points. In [110], SVR is used to design a triple-band antenna, the model achieved an comparable accuracy to EM simulation using significantly reduced time. In [111], surrogate model based on SVR are constructed to investigate the impact of use of angles of incidence on reflectarray design; this method provides faster analysis manner without compromising accuracy. A graphene reconfigurable reflectarray is modelled using SVR in [112], the SVR accuracy is demonstrated to be better than RBF network and comparable to EM simulation. An SVR-enabled optimisation strategy is proposed in [113] to improve the cross-polarisation performance of reflectarray, and the over 9dB reduction in the maximum cross-polarisation level is achieved.

2.3.3 Gaussian Process Regression for Antenna Modelling

From the past decade, the GPR has received extensive attention in antenna research. Unlike ANN and SVM, the GPR can provide the uncertainty of the predicted results when given new inputs. GPR requires fewer training data patterns to train and has excellent generalisation capability, and these characteristics make it popular in the antenna research field. GPR was introduced to model the input characteristics of circularly polarised wave (CPW) fed slot antenna [114], planar dual-band microstrip patch antenna [115] [118], ultra-wideband (UWB) antenna [116] and compact microstrip antenna [122], the results showed the high accuracy that was comparable to EM simulation and numerically computed results. [119] investigate the relationship between substrate ground size and gain pattern of microstrip antennas using GPR. In [117], a two-stage method based on GPR to efficiently and accurately modelling and optimise antenna was proposed. Similarly, a multistage training strategy based on GPR was proposed in [121] and applied to antenna modelling and optimisation. GPR was used to construct an accurate surrogate model that can substitute EM simulation computation in [120] and [123], the time cost in optimising antenna structure was significantly reduced. Recently, a multi-branch method based on GPR is proposed in [124] to balance exploitation and

exploration in antenna design. A two-stage modelling with domain confinement is proposed in [125], and significantly reduced CPU cost is obtained compared to the conventional GPR method. An expedited variable-resolution method based on GPR is proposed in [126], and up to 76% time saving is obtained compared to conventional ML methods.

2.3.4 Summary

ML-assisted antenna modelling is a rapidly growing field with the potential to revolutionise the way antennas are designed and optimised. This approach has several advantages over traditional methods such as MoM, FTDT and FEM, including the ability to handle complex and non-linear relationships, optimise multiple performance metrics simultaneously, and make a prediction with limited data. However, despite these advantages, several challenges and limitations should be addressed to fully realised the potential of ML in antenna modelling.

One of the critical challenges is the shortage in ML algorithm interpretability. The numerical methods provide detailed information about the behaviour of an antenna and parameterise the specific performance using equations. While the ML surrogate model only provides the prediction of antenna performance of an antenna corresponding to the specific geometry, it is challenging to understand why ML surrogate models give certain predictions and to make necessary adjustments to the design. In addition, there also be a lack of transparency in the ML algorithm used in antenna modelling. The utilisation of black-box techniques makes it difficult to understand the underlying processes and parameters used by the algorithm. Thus, the accuracy of ML surrogate model is difficult to verify and validate.

The most essential limitation to using ML in antenna modelling is the large amount of data required to train an accurate ML surrogate. Antenna simulation and experimental data are scarce due to the computational intensity of EM simulation, so considerable dataset acquisition can be time-consuming and expensive, which can lead to inaccurate surrogate models, as the models that constructed using limited data cannot accurately represent the behaviour of antennas in all scenarios.

2.4 Efficient Model Construction Techniques

2.4.1 Multi-fidelity Modelling Method

Multi-fidelity (MF) modelling method can significantly reduce the cost of collecting initial training data for surrogate model construction by leveraging both low-fidelity (LF) and high-fidelity (HF) data [127] [128]. LF data is inaccurate and easy to obtain, its quantity is sufficient. HF data is accurate and expensive to obtain, and the amount is small. The MF method was first mentioned in [127] as "multi-level", the conception of autoregressive was proposed. In [128], the surrogate modelling and optimisation in the MF framework were demonstrated in detail, the methodology based on GPR was explained and its effectiveness was tested. [127] and [128] provided solid theoretical foundations for its application. The MF method can address both low-dimensional and high-dimensional problems with linear or non-linear correlations, and it was later applied to various fields, such as aerospace [129] [130], electromagnetism [131]-[133] and fluid mechanics [134] [135], etc., and its effectiveness and performance were extensively recognised. In some other work, the MF method was researched from a mathematical aspect [136]-[138].

The MF modelling approach is popular in antenna design since it can significantly accelerate data generation speed using EM simulation. Although the conventional antenna design depends on the data generated via HF EM simulation, the time consumption is high, so the dataset size is limited. The MF modelling framework generates the EM simulation data using different EM models.

In antenna research, MF modelling is conceived by reducing the mesh density of the antenna model in the EM solver platform. In addition, other options that could create an LF model are possibly including:

- Modelling metals with the perfect electric conductor (PEC);
- Ignoring dielectric dispersion and losses;
- Neglecting the metallisation thickness of patches and strips.
- Using discrete port instead of waveguide port.

The possible simplifications accelerate the LF model's simulation time, resulting in a 10 - 30 times faster simulation period than the HF model. As a result, the accuracy of the LF model is inadequate. However, it should be able to generally characterise the features, such as the trend and shape of antenna performance that the HF model presents.

The combination of GPR and multi-fidelity modelling is popular in antenna design. The MF modelling method is introduced and combined with GPR in [139], the MF approach is demonstrated to be efficient in antenna design without compromising accuracy. In [140], the correlation between LF and HF was exploited, and co-Kriging was used to learn this correlation to efficiently and accurately model the antenna. In [117], a two-stage modelling method based on the MF framework and GPR was proposed. In the first stage, the LF data is used to build the GP model, and its output is then used as an auxiliary dataset. The auxiliary and HF data were used together to train the GP model at the second stage. The final model can give accurate predictions for antenna performance and can be used in place of EM simulation for antenna optimisation. [141] and [142] used MF framework in multi-objective antenna optimisation, the efficiency was promoted without the loss of accuracy. Recently, a concept of variable-resolution approach has been proposed based on the MF approach to efficiently build a surrogate model and accelerate the antenna design process [126] [143], remarkable CPU resources are saved compared to the situation while using conventional modelling methods.

One advantage of MF-based ML antenna modelling method is that it allows for combining data from different simulation models and experimental data to create a more comprehensive and accurate dataset. As a result, the MF framework can improve the accuracy of the ML models by reducing the required data and improving the generalisation of ML models. In addition, MF based ML models facilitate a faster optimisation of antenna designs. With a more efficient model, it becomes possible to explore a larger design space in reducing time, leading to better designs in a relatively shorter time. However, the complexity of creating an effective MF model is high since the construction of the MF model requires a thorough understanding of the sources of data and the interdependencies between them. In addition, the data selection, weight and combination can affect the modelling accuracy and

efficiency. Another limitation of the MF modelling method is the risk of error propagation when using LF data. Although the accuracy of the ML model depends on the accuracy of training data, when LF data is used to approximate HF data, the errors between LF and HF data can propagate and impact the model accuracy, which may lead to unreliable and inaccurate models.

2.4.2 Design of Experiment Sampling Methods

The quality of sampling candidates can significantly affect the surrogate accuracy and generalisation capability. The conventional DoE method, such as Fractional factorial design, central composite design, and Monte Carlo sampling (MCS), can only arrange the sample at a specific corner or even at the repeated place [146], thus the sampling could not traverse the parameter space, the surrogate model cannot accurately model the target antenna. Latin Hypercube Sampling (LHS) is a commonly used sampling strategy in antenna modelling to explore the parameter space of the antenna design efficiently. The utilisation of LHS reduces the number of EM simulations required to achieve accuracy, and it helps to identify the most critical design parameters that have the most significant impact on antenna performance, enabling better optimisation manner of antenna design.

To address this issue, a sampling method that can globally fill the design space [147] is needed to promote the quality of the model.Latin hypercube sampling (LHS) is a popular sampling strategy for computationally demanding modelling because it can extract a relatively small sample from a large amount of information that contains variate uncertainty, characters and sensitivity [148]. LHS tends to sample in the whole parameter space, the whole space is stratified with equal probability and the data is randomly sampled in each interval without replacement [149]. LHS is extensively used to further reduce the number of sample points for model construction [117] [140] [141] [142] [144] [145], a multi-zone LHS is proposed in [152], which can effectively allocate samples with uniformity.

In [150] and [151], the heuristic Latin hypercube sampling (HLHS) is proposed, which aims to address the issue that the optimal variables of practical antenna design have a low probability located at the extreme edge of parameter space. It can further reduce the number

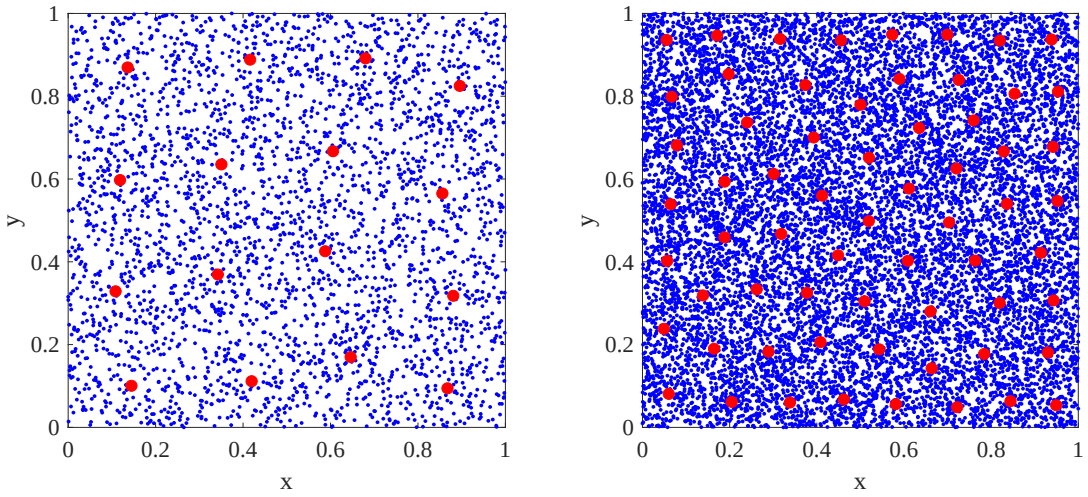

Fig. 2.12 Heuristic Latin hypercube sampling [150]

of sampled points. In this method, K-means clustering [49] is employed to improve the uniformity of sampling and further reduce the sampled points for EM simulation. A two-step sample is considered, LHS is firstly used to fill up the design space, then K-means clustering is applied to the previously sampled points to get a smaller amount of sample points, as shown in Fig. 2.12, where blue dots are generated by LHS and red dots are final samples. However, the sample boundary will shrink as the number of dimensions increases, as shown in Fig. 2.13, indicating poor uniformity and randomness of samples, leading to inaccurate surrogate construction.

2.5 Summary

ML-based antenna modelling methods have gained popularity in the last decade due to their ability to analyse complex datasets and generate accurate predictions. ML is a data-driven approach that can handle complex and large datasets, and ML can help to identify the correlation between different design parameters that conventional methods would miss. An essential advantage is that the ML method can reduce the time cost and computational resource requirement, accelerating the process of antenna design. However, ML model requires a large amount of data to train to ensure the model's accuracy and generalisation capability, and training with large dataset requires a significant amount of computational

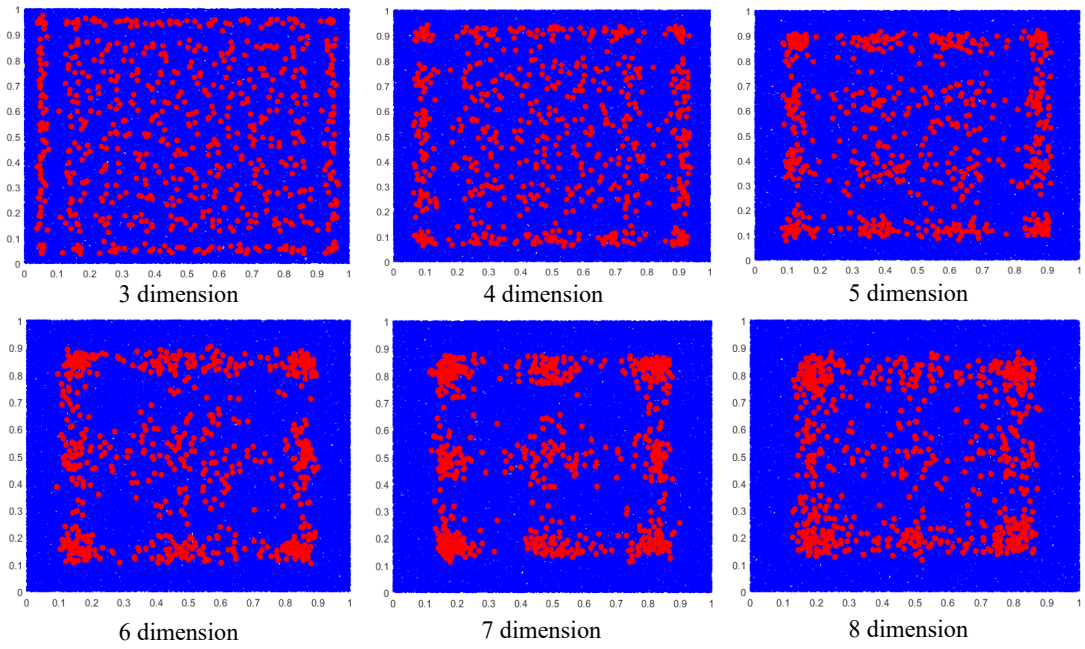

Fig. 2.13 Heuristic Latin hypercube sampling at high dimension

resources. Furthermore, the model may not accurately predict the antenna performance once the training data does not represent the actual operating conditions or with low uniformity. In addition, the antenna design with a large number of considered design parameters can be computationally expensive, and the data generation and model training can also be time-consuming, particularly for complex antenna designs. Currently, combinations of various algorithms are considered in most current research, including ML, evolutionary, and differential algorithms, to achieve efficiency and accuracy in antenna design.

DoE is a statistical approach that can be used to efficiently and effectively explore the design space in antenna design. An appropriate sampling method reduces the number of simulations required to achieve accurate results, and it helps identify the most critical design parameters that impact antenna performance. In addition, DoE can reveal the correlations between design parameters, and the candidates' distribution reflects the sampling quality. However, some limitations arise that a priori knowledge of the design space may not always be available or definable. Another potential issue is the number of dimensions in the design space, too few sample points can result in a sparse representation in the design space, which is the culprit of inaccurate model construction. While too much sample can lead to a high

computational budget. Last but not least, DoE cannot guarantee that the sampled points are optimal for model construction, the sampling depends on the initial point and the algorithm used to optimise the design.

The MF approach uses data from varying levels of fidelity to improve the overall performance of models or systems. As a result, MF allows faster and more efficient exploration of the design space and vastly reduces the time cost and computational intensity in simulation and data generation. In this case, sufficient data can be obtained using the MF approach. However, the insufficient accuracy of LF models can lead to inaccurate models and suboptimal designs. The transformation from HF to LF should keep consistent to ensure data quality. Another limitation is the risk of error propagation caused by LF data usage, which impacts model accuracy and system design.

Chapter 3

Empirical Formulas for Performance Prediction of Concrete Embedded Antenna

Overview

Ultra-dense small cell deployment is regarded as the most promising way to meet the traffic demand, the deploying small cells densely in buildings is anticipated to improve throughput in the fifth-generation (5G) wireless networks [153]. Particularly the communication capacity crunch problem is manifest in the indoor environment since a large amount of business takes place indoors. Network densification is a promising technology to address the above problem, and deploying small cells densely in buildings is anticipated to improve the throughput of 5G cellular communication [154]. However, deploying small cells with a number of antennas or antenna arrays will occupy large space which affects the usage of a building. Moreover, indoor small cells may cause aesthetic problems to house appearance. A feasible solution to mitigate these negative effects is to integrate antennas with the building materials, for example, embedding antennas into concrete walls.

So far, most research on concrete embedded antennas are focused on data and power transmission in wireless sensor network embedded in concrete to perform structural health monitoring [155] [156]. A compact and robust patch antenna for wireless monitoring

radiation efficiency, gain, input resistance and input reactance are difficult to compute due to the complex coupling between the antenna and concrete.

In Fig. 3.1, the considered antenna was designed in [163] and optimised to operate at 3.5 GHz, which is the indoor communication frequency for 5G networks. The antenna is completely embedded into the concrete slab. For the antenna unit, a microstrip patch is etched on a Rogers RT5880, with relative permittivity of 2.2 and $\tan\delta = 0.0009$. Two protective facesheets of Roger 3003 ($\varepsilon_r = 3$ and $\tan\delta = 0.001$) are employed to the top and bottom of the antenna structure. A porous honeycomb structure is added above the antenna, by doing this, the direct contact between metallisation and concrete can be avoided. Thus the antenna could preserve its electrical and mechanical robustness in the embedding ambient. The electrical properties and thickness for each layer of the antenna are shown in Table 3.1, and the dimensional parameter of the patch is provided in Table 3.2. The total size of the antenna system results in $60 \times 60 \times 11.5$ mm^3, and assuming that the concrete slab with nominal relative permittivity of $\varepsilon_r = 4$, $\tan\delta = 0.03$ and permeability $\mu = 1$. The comparisons of the antenna reflection coefficient before and after embedding are plotted in Fig. 3.3, which shows that the existence of concrete reduces the amplitude of the antenna reflection coefficient. Considering the concrete dielectric constant, the antenna reflection coefficients with different given concrete dielectric constants are presented in Fig. 3.4. It is observed that the larger dielectric constant results in reduced S_{11} amplitude and shifting centre frequency of the antenna.

Table 3.1 Electrical properties and thickness for each layer of antenna

Layer	Material	Dielectric constant	$\tan\delta$	Thickness (mm)
Upper facesheet	Roger 3003	3	0.001	$d_1 = 0.25$
Honeycomb	Air	1	0	$d_2 = 10$
Substrate	Roger 5880	2.2	0.0009	$d_3 = 1$
Lower facesheet	Roger 3003	3	0.001	$d_4 = 0.25$

Table 3.2 Dimensional parameters of the proposed antenna

Parameter	A	W	L	x	y	w	l
Value(mm)	60	33.85	28.39	8	2.69	3.12	23.805

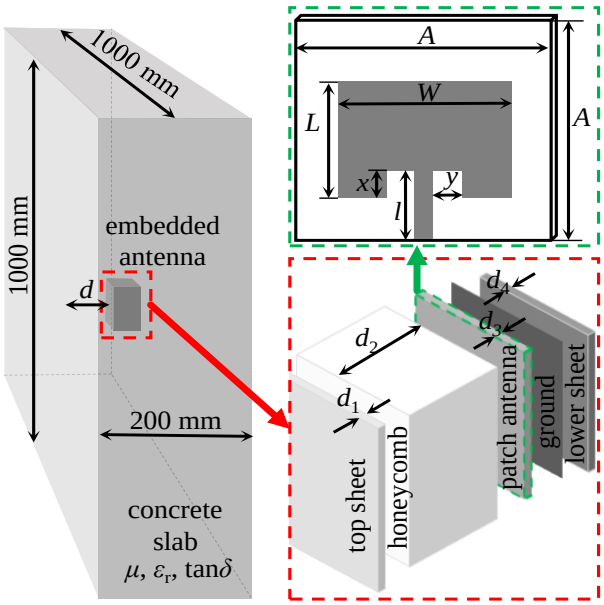

Fig. 3.1 Overview of system model geometry.

3.2 Empirical Analysis for Performance Parameters

In this section, the performance of the concrete embedded antenna is fitted using the LR method. The impact of the embedding ambient on antenna multi-parameters has been fitted based on the full-wave EM simulation results of the antenna. The embedding depth and the concrete dielectric constant on antenna radiation efficiency, gain, input resistance and input reactance are investigated. Some of the simple formulas are individually obtained using LR. The results in this section provide fundamental knowledge and guidance to antenna designers and architects for concrete embedding deployment and communication-friendly building materials selection. In order to develop mathematical expression to estimate the antenna performance as a function of embedding depth and dielectric constant, the continuous parameters (embedding depth and dielectric constant) are considered, and dimensions are fixed.

Radiation efficiency is one of the most important parameters of antenna performance. The scale of antenna loss characterises it, thus it is able to indicate the feasibility of the antenna system in the concrete structure and reflect the amount of power that the antenna system can radiate in the concrete structure. After being buried or embedded, the antenna's

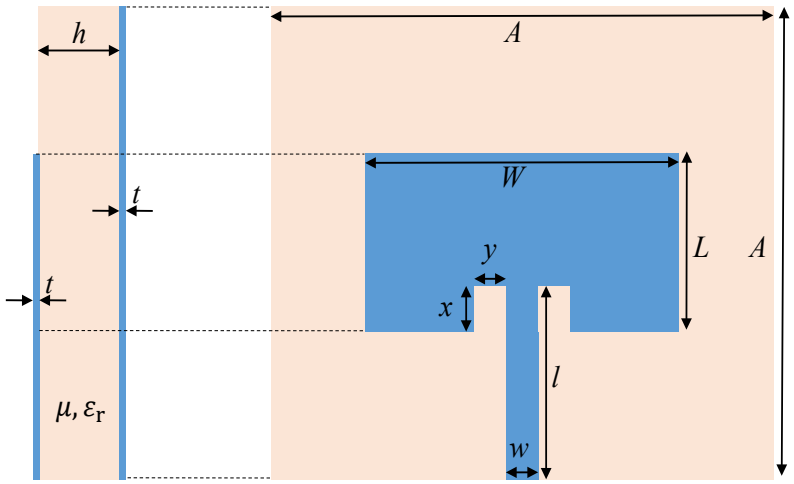

Fig. 3.2 Schematic of antenna unit.

efficiency is particularly affected due to the interaction with concrete, which can lead to the variation of gain. Gain is an essential concern of antenna performance and it characterises the strength of an antenna to transmit and receive the signal in a specific direction, which directly determines the quality of wireless communication. Inappropriate embedding can severely degrade an antenna's radiation efficiency and gain. Fig. 3.5 and Fig. 3.6 show the simulation results of radiation efficiency η_{rad} and gain G vary against embedding depth d and concrete dielectric constant ε_{r}, respectively. It is noted that the increase of d makes η_{rad} and G decrease in a fluctuation, while the increasing ε_{r} leads to the decline of η_{rad} and G.

The input impedance of an antenna is normally determined by antenna configuration, operation frequency and ambient environment. Here we consider the effects of embedding depth and dielectric constant of concrete on the input impedance of the antenna. The input impedance Z_{in} consists of input resistance R_{in} and input reactance X_{in}, which can be written as:

$$Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}}. \tag{3.1}$$

The variations of R_{in} and X_{in} are shown in Fig. 3.7 and Fig. 3.8, respectively. Both R_{in} and X_{in} appear to be a damped oscillation with the increase of embedded depth. In addition, the increasing dielectric constant of concrete ε_{r} leads to decreasing amplitude in the input resistance R_{in} but increasing X_{in}.

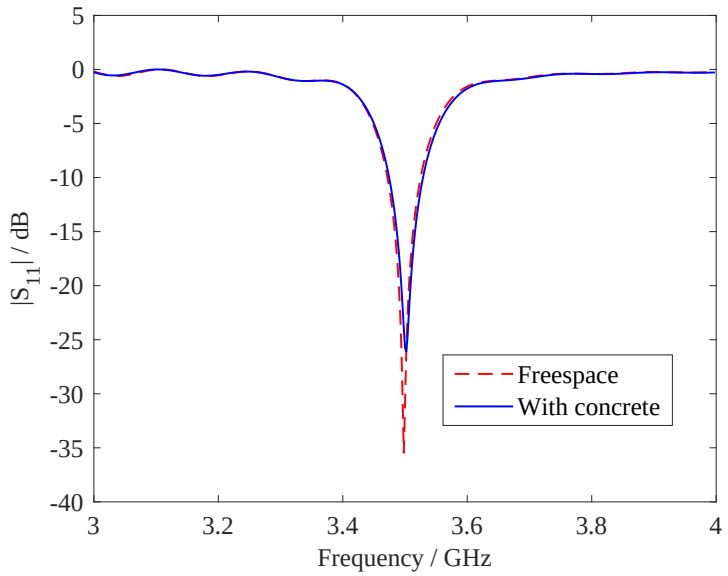

Fig. 3.3 Simulated $|S_{11}|$ of the antenna in freespace and inside concrete slab of 1000 $\times$ 1000 $\times$ 200 mm^3, with nominal $\varepsilon_r = 4$ and $\tan\delta = 0.03$.

In order to obtain a fast and straightforward prediction of the performances of concrete embedded antenna, empirical formulas are fitted out from the above-simulated results, where the embedding depth d is measured in meters, and the dielectric constant of concrete ε_r is a dimensionless parameter. Since the radiation efficiency η_{rad} decreases in a fluctuation with the increase of the embedding depth, the relationship between η_{rad} and d is formulated as a linear decreasing function superposed with a damping sinusoidal function, and η_{rad} is measure in percentage (%):

$$\eta_{rad} = A_1(\varepsilon_r)\exp^{B_1(\varepsilon_r)d} + \sin[C_1(\varepsilon_r)d] + D_1(\varepsilon_r)[d + E_1(\varepsilon_r)]. \tag{3.2}$$

Also, gain G could be formulated in the same form as η_{rad} because of the linear relationship $G = D \times \eta_{rad}$, and G is measured in dBi and given as:

$$G = A_2(\varepsilon_r)\exp^{B_2(\varepsilon_r)d} + \sin[C_2(\varepsilon_r)d] + D_2(\varepsilon_r)[d + E_2(\varepsilon_r)]. \tag{3.3}$$

In formula (3.2), A_1, B_1, C_1, D_1 and E_1 are undetermined coefficients which are functions of ε_r. By LR, formulas of η_{rad} for the different dielectric constants of concrete are listed in

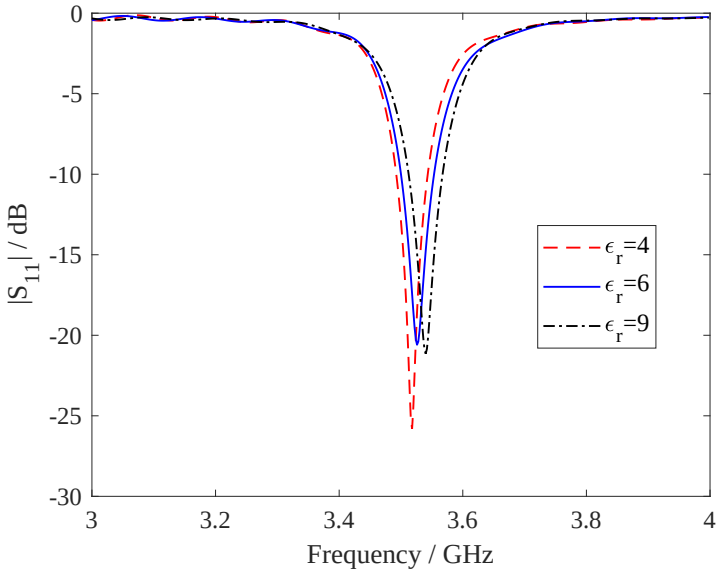

Fig. 3.4 Simulated $|S_{11}|$ with different concrete permittivity (ε_r = 4, 6, 9) for the antenna in the frequency band of interest.

Table 3.3 Formulas of radiation efficiency for different dielectric constants

ε_r	Regression formulation
4	$\eta_{\text{rad}} = 11.79e^{-27.52d}\sin(297.70d) - 152.00\cdot(d-0.30)$
5	$\eta_{\text{rad}} = 12.97e^{-26.33d}\sin(338.10d) - 149.00\cdot(d-0.27)$
6	$\eta_{\text{rad}} = 14.59e^{-25.98d}\sin(372.60d) - 146.20\cdot(d-0.26)$
7	$\eta_{\text{rad}} = 16.06e^{-24.30d}\sin(404.10d) - 145.20\cdot(d-0.25)$
8	$\eta_{\text{rad}} = 17.35e^{-23.79d}\sin(433.20d) - 143.50\cdot(d-0.24)$
9	$\eta_{\text{rad}} = 18.15e^{-22.37d}\sin(461.10d) - 142.40\cdot(d-0.23)$

Table 3.3. After a careful inspection of the results in Table 3.3, the coefficient in formula (3.2) can be formulated as a linear function of the dielectric constant of concrete. The LR for the five coefficients in (3.2) are shown in Fig. 3.9, which demonstrate the high accuracy of the linear fitting. The fitted formulas fort the five coefficients in (3.2) are listed in Table 3.4. Substituting the results in Table 3.4 into (3.2), the empirical formula of η_{rad} as a function of the embedding depth d and the dielectric constant of concrete ε_r can be written as:

$$\begin{aligned}\eta_{\text{rad}} = (1.33\varepsilon_r + 6.52)\exp^{(\varepsilon_r - 31.56)d} + \sin[(32.39\varepsilon_r + 173.90)d] \\ + (1.87\varepsilon_r - 158.50)[d + (1.30\times 10^{-2}\varepsilon_r - 0.34)](\%).\end{aligned} \tag{3.4}$$

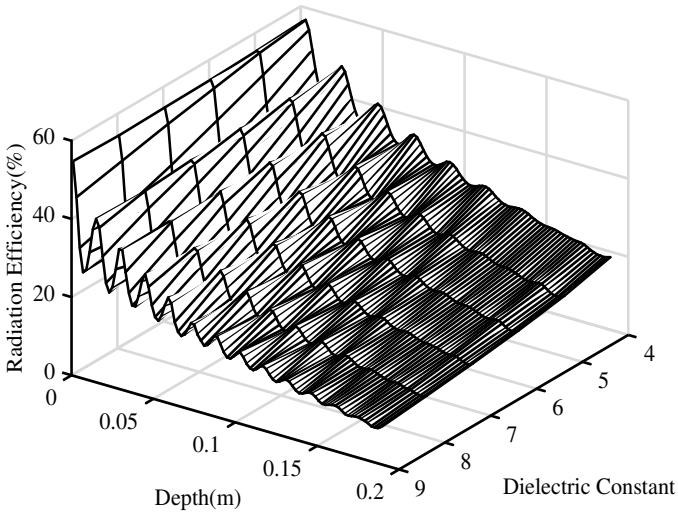

Fig. 3.5 Simulated radiation efficiency η_{rad} against embedded depth and concrete dielectric constant.

Table 3.4 Linear fitting formulas for coefficients in equation (3.2)

ε	Formulation
A_1	$A_1 = 1.33\varepsilon_{\text{r}} + 6.52$
B_1	$B_1 = \varepsilon_{\text{r}} - 31.56$
C_1	$C_1 = 32.39\varepsilon_{\text{r}} + 173.90$
D_1	$D_1 = 1.87\varepsilon_{\text{r}} - 158.50$
E_1	$E_1 = 1.30 \times 10^{-2}\varepsilon_{\text{r}} - 0.34$

Using the same procedure, the formula of antenna gain G as a function the embedding depth d and the dielectric constant of concrete ε_{r} is written as (measured in dBi):

$$\begin{aligned} G = {} & (0.533\varepsilon_{\text{r}} - 0.89)\exp^{(1.47\varepsilon_{\text{r}} - 32.75)d} + \sin[(22.09\varepsilon_{\text{r}} + 250.80)d] \\ & + (-1.43\varepsilon_{\text{r}} - 9.98)[d + (0.05\varepsilon_{\text{r}} - 0.54)](\text{dBi}). \end{aligned} \tag{3.5}$$

As the input resistance and the input reactance oscillated damply with the embedding depth, they can be modelled by a sinusoidal function with attenuated amplitude and both are measured in Ohm (Ω):

$$R_{\text{in}} = A_3(\varepsilon_{\text{r}}) + B_3(\varepsilon_{\text{r}})\exp^{C_3(\varepsilon_{\text{r}})d}\cos[D_3(\varepsilon_{\text{r}})d + E_3(\varepsilon_{\text{r}})], \tag{3.6}$$

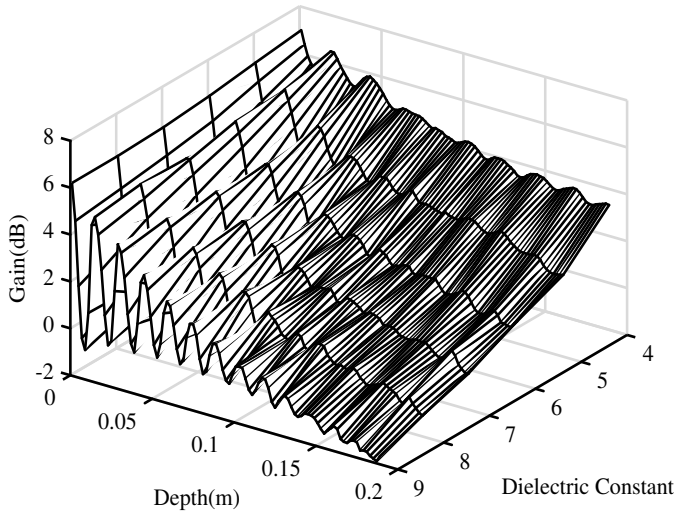

Fig. 3.6 Simulated gain G against embedded depth and concrete dielectric constant.

$$X_{\text{in}} = A_4(\varepsilon_{\text{r}}) + B_4(\varepsilon_{\text{r}}) \exp^{C_4(\varepsilon_{\text{r}})d} \cos[D_4(\varepsilon_{\text{r}})d + E_4(\varepsilon_{\text{r}})]. \tag{3.7}$$

Following the same fitting procedure, the formulas of the input resistance R_{in} and the X_{in} for different dielectric constants are listed in Table 3.5 and Table 3.6. Coefficients in formulas (3.8) and (3.9) can be determined by linear fitting from results in Table 3.5 and Table 3.6, via the same procedure as for radiation efficiency. The final empirical formulas of R_{in} and X_{in} as function of the embedding depth d and the dielectric constant of concrete ε_{r} are:

$$\begin{aligned} R_{\text{in}} = {} & (-2.23\varepsilon_{\text{r}} + 40.13) + (0.42\varepsilon_{\text{r}} + 10.60) \exp^{(-0.20\varepsilon_{\text{r}} - 18.58)d} \cos[(29.67\varepsilon_{\text{r}} + 174.90)d \\ & + (-0.04\varepsilon_{\text{r}} + 6.56)], \end{aligned} \tag{3.8}$$

$$\begin{aligned} X_{\text{in}} = {} & (0.76\varepsilon_{\text{r}} - 6.16) + (0.65\varepsilon_{\text{r}} + 10.52) \exp^{(-0.36\varepsilon_{\text{r}} - 17.94)d} \cos[(29.35\varepsilon_{\text{r}} + 175.50)d \\ & + (-0.02\varepsilon_{\text{r}} + 8.13)]. \end{aligned} \tag{3.9}$$

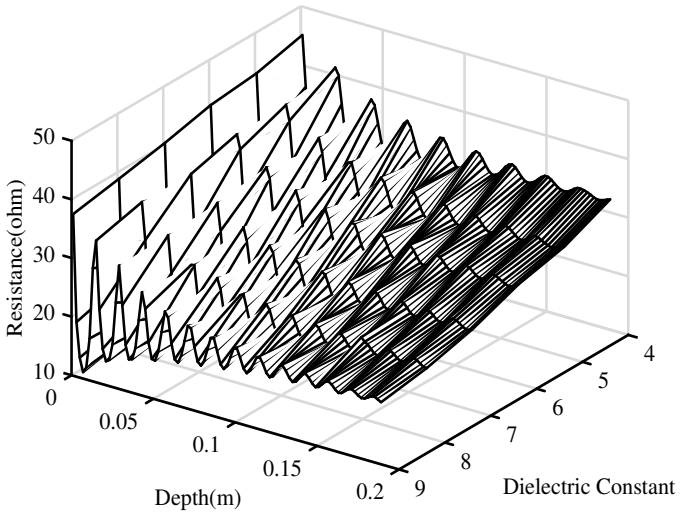

Fig. 3.7 Simulated input resistance R_{in} against embedded depth and concrete dielectric constant.

Table 3.5 Formulas of resistance for different dielectric constants

ε	Regression formulation
4	$R_{\text{in}} = 32.01 + 12.14e^{-19.34d}\cos(289.4d + 6.417)$
5	$R_{\text{in}} = 28.43 + 12.68e^{-19.68d}\cos(324.5d + 6.343)$
6	$R_{\text{in}} = 26.64 + 13.21e^{-19.77d}\cos(356.3d + 6.327)$
7	$R_{\text{in}} = 23.98 + 16.74e^{-20.05d}\cos(385.5d + 6.274)$
8	$R_{\text{in}} = 22.22 + 14.05e^{-20.18d}\cos(412.8d + 6.243)$
9	$R_{\text{in}} = 20.68 + 14.15e^{-20.42d}\cos(438.3d + 6.215)$

The fitted result of antenna gain G and input reactance X_{in} are presented in Fig. 3.10 and Fig. 3.11, respectively. The gain of the proposed antenna decrease in a fluctuating manner with the increasing of embedding depth d, since a larger d introduces more absorption loss by the concrete. The fluctuation is caused by the interference of multiple reflection inside the concrete slab. The local maximum of gain occurs once the multiple reflection is in phase. It is worth noting that the fluctuation period is approximately equal to half-wavelength in the concrete. The X_{in} oscillates with gradually decaying amplitudes as d increases, as it is shown in Fig. 3.8. The concrete slab can be modelled as a lossy transmission line, so the variation of X_{in} against d is similar to the reactance fluctuation along a transmission line.

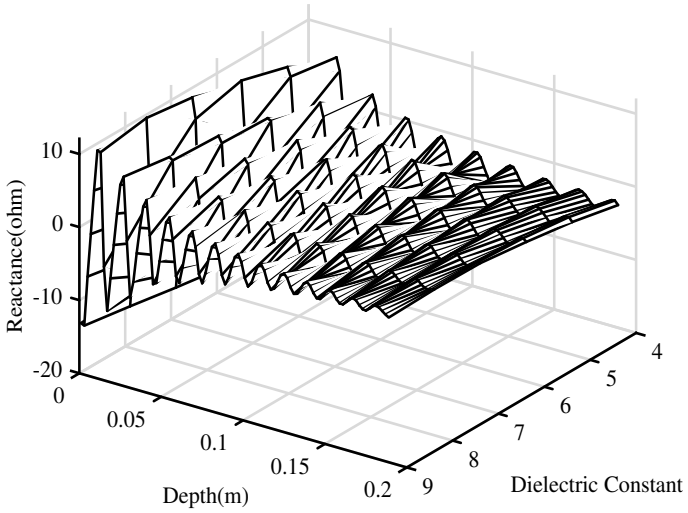

Fig. 3.8 Simulated input reactance X_{in} against embedded depth and concrete dielectric constant.

Table 3.6 Formulas of reactance for different dielectric constants

ε	Regression formulation
4	$X_{\text{in}} = -2.97 + 13.04e^{-1.92d}\cos(28.86d + 8.04)$
5	$X_{\text{in}} = -1.99 + 13.71e^{-1.98d}\cos(32.35d + 8.02)$
6	$X_{\text{in}} = -1.12 + 14.48e^{-2.02d}\cos(35.51d + 8.01)$
7	$X_{\text{in}} = -0.43 + 15.18e^{-2.07d}\cos(38.42d + 7.98)$
8	$X_{\text{in}} = 0.06 + 15.65e^{-2.09d}\cos(41.06d + 7.95)$
9	$X_{\text{in}} = 0.37 + 15.89e^{-2.11d}\cos(43.62d + 7.93)$

3.3 Discussion

It can be observed from the fitted results that the radiation efficiency and gain of the proposed concrete embedded antenna decrease in a fluctuating manner with increasing embedded depth since deeper embedded depth introduces more absorption loss by the concrete. The interference of multiple reflections inside the concrete slab causes the fluctuation. The local maximum of gain occurs when the multiple reflections are in phase. It is worth noting that the fluctuation period is approximately equal to half-wavelength in the concrete. The wavelength

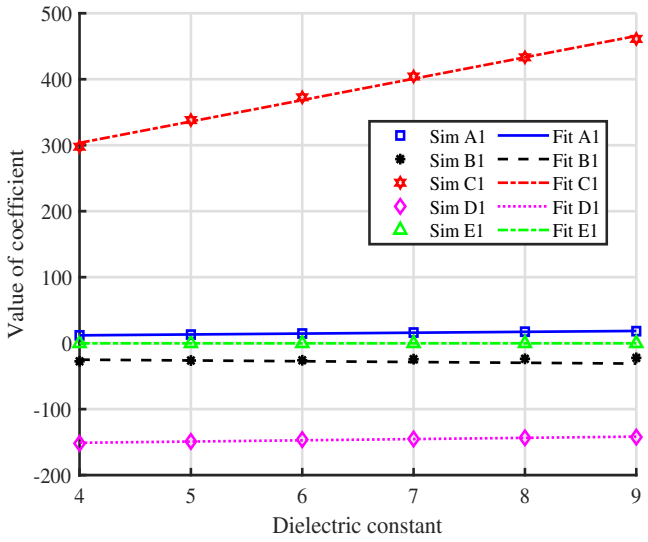

Fig. 3.9 Linear regression for the five coefficient in formula (3.2).

in concrete is calculated by equation (3.10):

$$\lambda = \frac{\lambda_0}{\sqrt{\varepsilon_r}}, \tag{3.10}$$

where λ_0 is the wavelength in free space. In addition, the radiation efficiency and gain decrease with increasing dielectric constant of concrete, which indicates concrete with lower dielectric constant is beneficial to radiation. The fitted result of gain is shown in Fig. 3.10, which is helpful for selecting optimal embedded depth and dielectric constant to achieve peak gain.

The input impedance oscillates with gradually decaying amplitude as embedded depth increases, and the fitted result of reactance is shown in Fig. 3.11. The concrete slab can be modelled as a lossy transmission line, so the variation of input impedance against embedded depth is similar to the impedance fluctuation along transmission line. Reflection coefficient of the antenna is determined by input impedance through the following equation:

$$\Gamma = \frac{Z_{in} - Z_0}{Z_{in} + Z_0}, \tag{3.11}$$

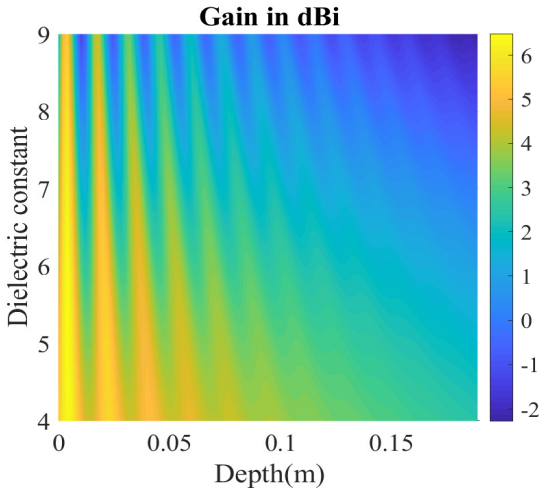

Fig. 3.10 Fitted gain G against embedding depth and concrete dielectric constant.

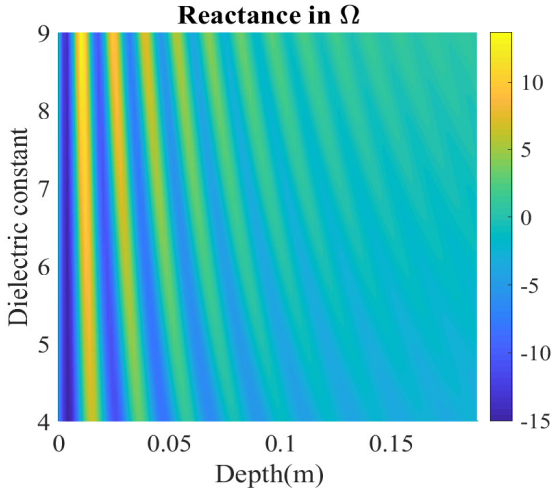

Fig. 3.11 Fitted input reactance X_{in} against embedding depth and concrete dielectric constant

where Z_0 is the characteristic impedance of feed line. With the empirical formulas fitted out in last section, reflection coefficient can be easily and fast predicted.

Comparisons between the fitted and simulated results are shown in Fig. 3.12, in which the fitted results agree with simulated results well. The fitted formulas can characterise the antenna's radiation efficiency, gain, input resistance and input reactance using given embedding depth and concrete dielectric constant. In order to validate the formulas, a new dataset is used. The considered embedding depth d is sampled with step width of 0.004

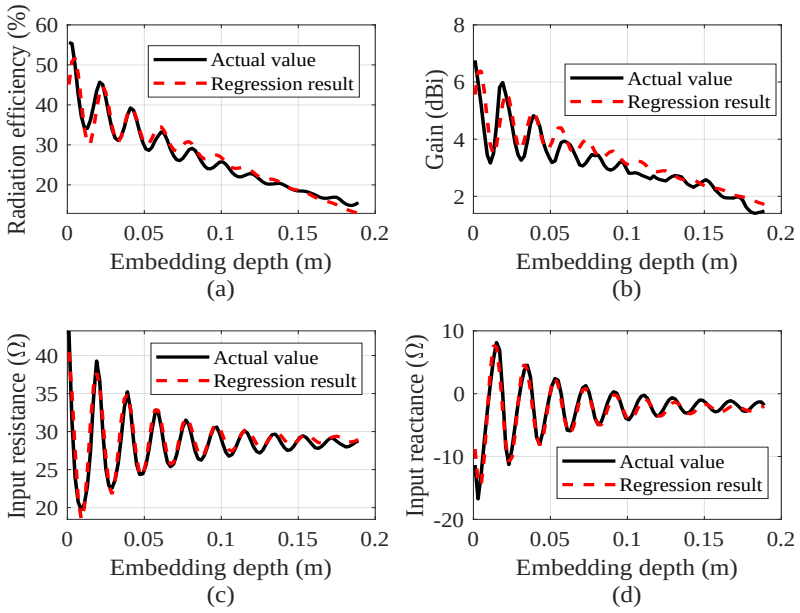

Fig. 3.12 Modelling results comparison between regression results and CST response.

meters in the range from 0.001 meter to 0.189 meter, and the concrete dielectric constant is set to be 4.5. The generalisation result is shown in Fig. 3.13.

Table 3.7 Regression model errors[1] for different antenna performance parameter

Objectives	η_{rad}	G	R_{in}	X_{in}
Modelling	78.69%	49.32%	29.11%	25.34%
Generalisation	414.6%	124.45%	314.69%	34.52%

[1] Measured in NMSE

In addition, the LR model saves considerable time in estimating variations in antenna performance parameters. The EM simulation-based method takes more than an hour to run an individual simulation, so evaluating new design points can be time-consuming. In the generalisation capability test stage, the LR method saves more than 50 hours to predict the variations of desired performance parameters under different embedding ambient, even though the generalisation accuracy is not good.

LR is easy to understand and interpret, the coefficients of the equation provide information about the variation of a specific antenna parameter as considered design variable changing. The LR method requires a small amount of data and low computational resources to train. Fitted formulas can quickly estimate the relationship between desired antenna performance

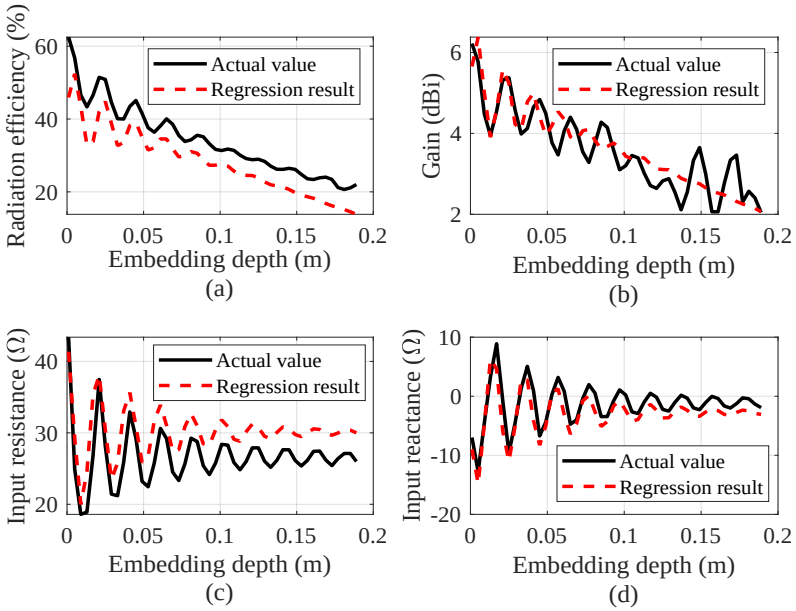

Fig. 3.13 Generalisation results comparison between regression results and CST response.

parameters and design variables using a limited dataset and small computational resources. However, LR's modeling accuracy needs to be improved, and the formulas show limited generalisation capability when the new design point is given. In addition, the LR model has limited power in handling multiple input variables, as well as multiple outputs (antenna performance parameters).

3.4 Summary

A sandwiched microstrip antenna is embedded into a concrete wall for indoor wireless communication. Antenna performances are investigated with varying embedded depth and dielectric constant of concrete. It is found that the radiation efficiency and gain decrease in fluctuation with increasing embedded depth, and the input impedance displays a damped periodical oscillation with embedded depth. Larger concrete dielectric constant leads to smaller radiation efficiency and gain. Both embedded depth and dielectric constant have a significant effect on antenna performance. LR method is used to fit out simple empirical formulas for fast performance prediction of concrete embedded antenna. The LR models give predicted curves that can reveal the variation of antenna specific performance parameters

under different embedding ambients. Considerable time saving is achieved when the estimation of antenna performance on new design points is needed. Even though the modelling and generalisation accuracy could be much better, the results in this work are helpful for concrete embedded antenna deployment and communication-friendly building materials selection. In addition, this method can also helpful to check the feasibility of modern antenna design for other applications and scenarios.

Chapter 4

Heuristic Algorithm Enhanced Artificial Neural Network Based Antenna Performance Modelling

Overview

ANN has already been recognised as a feasible tool for microwave modelling and simulation in recent years [164], which can learn and solve complex and non-linear problems relatively quickly. To a certain extent, ANN could be used as a surrogate model that substitutes the computationally intensive EM simulation solver. By far, ANN has been successfully applied to various antenna applications, such as antenna optimisation [95] [121] and antenna analysis and synthesis [86]. In [165], the ANN-based models were presented to compute the resonant frequency of the antenna with lower error. Generally, the gradient-based training algorithm, such as the BP algorithm, is used in the ANN training process. However, the slow convergence ratio and local optimum issues are the main drawbacks of gradient-based algorithm [166]. As a result, the optimisation algorithm such as particle swarm optimisation (PSO) [167] and genetic algorithm (GA) [168] are combined with ANN and can significantly improve the performance of ANN. On the other hand, the over-fitting issue and generalisation capability are the main concerns for the ANN model. In order to

improve the accuracy of ANN, the early-stopping strategy and Bayesian Regularisation (BR) [169] are introduced to address these issues, which prevents training from occurring and effectively improves the generalisation capability of the network.

In the present work, ANN is utilised to compute and predict the performance of a concrete embedded antenna for indoor communications. A hybrid ANN model with PSO, BP and BR (PSO-BRNN) and a classic backpropagation NN (BPNN) are developed for the computation and prediction. The performance of PSO-BRNN in terms of modelling and generalisation errors is compared with the LR method and BPNN.

4.1 Antenna System Model

A structurally integrated antenna with a multi-layer configuration mentioned in Chapter 3 is selected because of its excellent mechanical and electrical performance. The antenna is fully embedded in a solid concrete slab, as shown in Fig. 4.1, and the concrete has a dimension of $1000\,\text{mm} \times 1000\,\text{mm} \times 200\,\text{mm}$. The embedding depth d of the antenna is measured as the distance between the top concrete-air interface and the top surface of the antenna. The effect embedding depth d and concrete dielectric constant ε_r on antenna performance will be investigated. Thus other electrical property, such as loss tangent, is fixed to 0.03 ($\tan\sigma = 0.03$). The proposed antenna is optimised to operate at 3.5 GHz, and its geometries are kept intact. The antenna unit is sandwiched among a lower facesheet (LF), a honeycomb structure, and an upper facesheet (UF) for better electrical and mechanical characteristics in the concrete wall. The complex relationship between antenna embedding ambient (embedding depth and concrete dielectric constant) and antenna performance parameters (radiation efficiency, gain, input resistance and reactance) using ANN.

4.2 Artificial Neural Network

ANN is developed in the last few decades and has been extensively applied in the engineering and science field due to its ability to learn the relationship between inputs and outputs in a

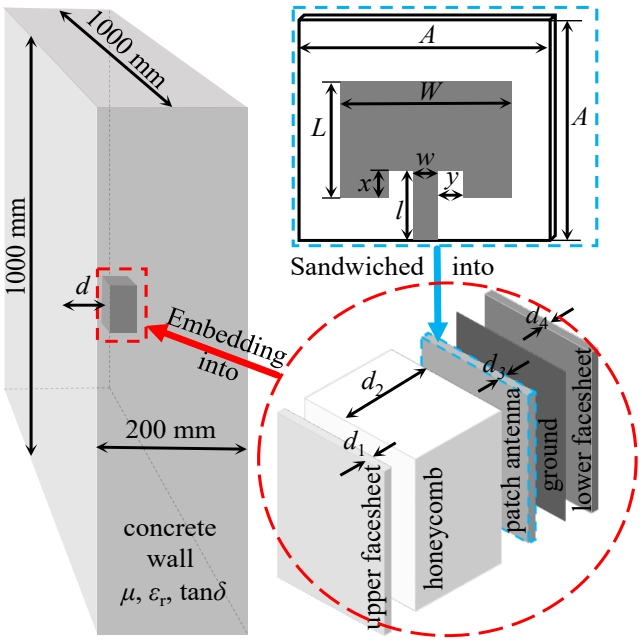

Fig. 4.1 Antenna model geometry. The parameters of antenna unit are A=60 mm, W=33.85 mm, L=28.39 mm, x=8 mm, y=2.69 mm, w=3.12 mm, and l=23.805 mm.

fast and flexible way. An ANN normally consists of three components, the input layer and output layer in the first and the last layer, and the rest is hidden layer(s). Each neuron j in the ith hidden layer sums up all inputs x_j by multiplying weights w_{ij}.

$$y_j = f(\sum_i w_{ij} x_j) + b_i, \tag{4.1}$$

where $f(.)$ is the activation function (e. g. sigmoid function, hyperbolic tangent, Relu and radial basis function etc.), which can introduce non-linear factor to solve problem that linear models cannot solve. w_{ij} stands for the jth weight in the ith layer. x_j denotes jth input sample in input vector, b_i is bias vector in ith layer.

4.2.1 ANN Architecture

In this work, ANN is utilised to compute and predict the concrete embedded antenna's performance proposed in Chapter 3. The PSO is employed to search the global optimal weights and biases for ANN for faster and more efficient computation and prediction. The BR

is used in the model training stage in order to overcome over-fitting issue. For comparison, a conventional BPNN based on a GD algorithm is developed as the reference, and its network parameters (topology, learning rate) are the same as PSO-BRNN for a reasonable comparison.

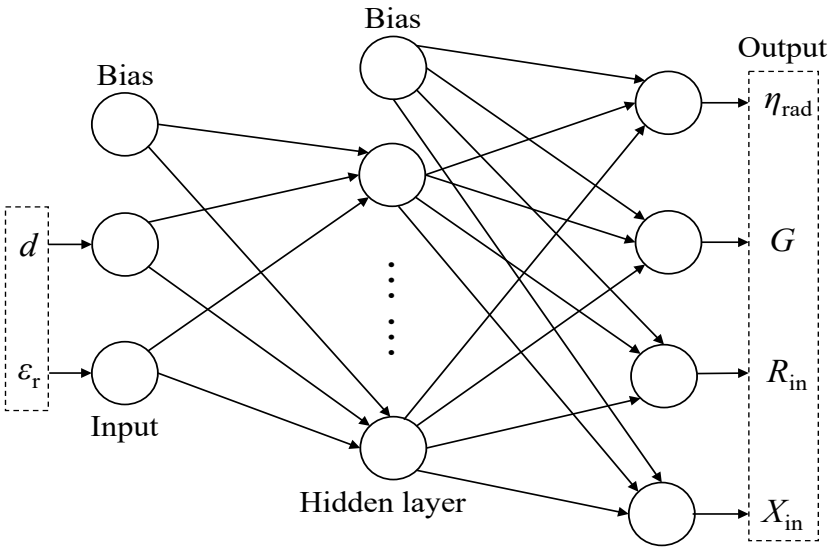

Fig. 4.2 Architecture of neural network model

For the concerned input variables, the antenna embedding depth d and concrete dielectric constant ε_{r} are selected, while the antenna's radiation efficiency η_{rad}, gain G, input resistance R_{in} and input reactance X_{in} are considered as outputs. The suggested network architecture consists of the hidden layer (the number of neurons and layer are adjusted), two input neurons (for d and ε_{r}) and four output neurons (for η_{rad}, G, R_{in} and X_{in}), the diagram of ANN is shown in Fig. 4.2. The number of hidden layers and neuron are adjusted, all the neurons are fully connected to the output layer that gives the desired values of antenna performance. The activation function used in the hidden layer is tangent sigmoid (equation (4.2)), while the simple linear function is used in the output layer.

$$y = \text{tansig}(\text{x}) = \frac{2}{1+\exp^{-2\text{x}}} - 1. \tag{4.2}$$

Given a data set $\boldsymbol{D} = [\boldsymbol{x}_j, \boldsymbol{y}_j]^T$, which consists of inputs vector $\boldsymbol{x}_j$ and outputs vector $\boldsymbol{y}_j$, a supervised non-linear regression task is going to be solved by ANN. The relationship

between the inputs vector and the outputs vector could be written as:

$$\boldsymbol{y}_j = f(\boldsymbol{x}_j), \tag{4.3}$$

where the corresponding inputs vector and outputs vector of ANN model are:

$$\boldsymbol{x}_j = [d, \varepsilon_{\mathrm{r}}]^{\mathrm{T}}, \tag{4.4}$$

$$\boldsymbol{y}_j = [\eta_{\mathrm{rad}}, G, R_{\mathrm{in}}, X_{\mathrm{in}}]^{\mathrm{T}}. \tag{4.5}$$

4.2.2 ANN Construction and Data Preparing

The range of inputs is 0.001 meters to 0.189 meters with a step width of 0.001 meters for the embedding depth d, and 4 to 9 with a step width of 1 for the concrete dielectric constant ε_{r}. The data set for ANN training is generated via the same CST model in Chapter 3 and is shown in Fig. 4.2, every single variable with a length of 1134. For balancing the scale of all variables and obtaining accurate prediction, all the obtained data have been normalised between 0 and 1 using equation (4.6), since it can avoid the error caused by a different order of magnitude.

$$\boldsymbol{X}_{\mathrm{norm}} = \frac{\boldsymbol{x}_j - \min(\boldsymbol{x}_j)}{\max(\boldsymbol{x}_j) - \min(\boldsymbol{x}_j)}. \tag{4.6}$$

The data set $\boldsymbol{D}$ is used for ANN training stage. Due to the utilisation of BR, there is no validation data involved in model training. All the ANN training and optimisations are performed suing MATLAB 2020b on an Intel Xeon W2135 3.70 GHz machine with 32 GB RAM.

4.3 ANN Training

4.3.1 PSO and ANN

The conventional ANN utilises the gradient-based method to train generally, and the convergence of ANN strongly depends on the initial guess of weights and bias. OThe BP algorithm is a well-known training method for neural networks, it is based on the GD algorithm. Hence, the initial point of weights and bias is essential for the BP training. If the weights and bias are not initialised properly, the results are likely to get stuck in a local optimum and consequently, the solution is not the best.

PSO is a random search algorithm based on group cooperation developed by emulating birds' foraging behavior. An effective evolutionary algorithm can find the global maximum or minimum of the objective function. In this work, the mean square error (MSE) of the neural network is considered as the evaluated fitness in PSO, which is calculated as:

$$E = \frac{1}{N}\sum_{1}^{N}\sum_{1}^{k}(\hat{y}_j - y_j)^2, \tag{4.7}$$

where $\hat{y}_j$ is the network outputs vector, y_j is the outputs vector of data set $\boldsymbol{D}$. N is the total number of data, k is the total number of output. In the present work, the N and k are 1134 and 4, respectively. Equation (4.13) is the objective function that needs to be optimised in PSO. Since the neural network learning process is mainly to update the weights and biases, thus the location of the particles in PSO corresponds to the values of all weights and biases in the network. E is taken as the fitness function of the PSO algorithm. All the weights and biases are randomly initialised in the range of 0 to 1, and then these values are adjusted by the PSO algorithm until the global minimum of the fitness function is found. In each iteration, the fitness function of each particle is calculated. The corresponding position P_i and velocity V_i are updated according to the calculated value of the fitness function, personal best p_{best} and global best g_{best}, the updated regulations for particles are:

$$V_i = \omega V_i + c_1\phi_1(p_{\text{best}} - P_i) + c_2\phi_2(g_{\text{best}} - P_i), \tag{4.8}$$

$$P_i = P_i + V_i, \tag{4.9}$$

where c_1 and c_2 are acceleration coefficients, ϕ_1 and ϕ_2 are random and positive number with uniform distribution ranged between 0 and 1, p_{best} is the personal best position of particle, and g_{best} is the global optimum position of particle. ω is the inertial weight, the linear decline weight (LDW) strategy is used to manipulate ω for the optimum solution search. The larger ω facilitates global searching, while the smaller ω is beneficial to precise local searching. The LDW strategy is expressed as:

$$\omega = \omega_{\max} - \frac{t \times (\omega_{\max} - \omega_{\min})}{t_{\max}}, \tag{4.10}$$

where $\omega_{\max}$ is the maximum inertial weight, $\omega_{\min}$ is the minimum inertial weights, t is current iteration, and $t_{\max}$ is the maximum iteration of PSO.

In the beginning, the ANN is built with a specific topology. Thus the dimension of a particle can be determined. For this work, the dimension of each particle equals the total number of weights and biases in the network. All the weights and biases in this network are going to be optimised by the PSO algorithm. 500 particles are employed and the computation iterates 700 times for searching the global optimum solution. Firstly, the number of the particle is selected. The particle positions and velocities are randomly initialised, and each particle i is characterised by its position vector $\boldsymbol{X}_i$ and velocity V_i. The position boundary $[-X_{\max}, X_{\max}]$, velocity boundary $[-V_{\max}, V_{\max}]$, inertial weight range $[w_{\min}, w_{\max}]$, acceleration coefficients c_1 and c_2, and the maximum iteration are defined, and these parameters are presented in Table 4.1.

In each iteration of PSO, the value of the fitness function of an individual particle is calculated, and the velocity and position of all particles are updated using the regulation in equation (4.8) and equation (4.9). Once the optimisation criteria meet, the PSO iteration terminates. g_{best} stores the global optimum solution for the network weights and bias, then the g_{best} is reshaped and assigned according to the topology of network which is prepared to be trained.

Table 4.1 PSO parameters

Parameter	Value
Number of particle	700
Position boundary	[-1,1]
Velocity boundary	[-0.8,0.8]
Inertial weight	[0.2,1]
Learning factor c_1	2
Learning factor c_2	2
Maximum iteration	1000

4.3.2 Bayesian Regularisation and ANN

BR is used to mitigate the potential over-fitting issue that may occur in the ANN training process. The over-fitting and over-training can lead to the loss of regression accuracy and generalisation of the network.To overcome the over-fitting problem, the BR adds regularisation term to the objective function as:

$$F = \beta E_D + \alpha E_w, \tag{4.11}$$

where the F is the objective function after introducing BR, E_D is the sum of squared errors of the network, $E_w = \frac{1}{m}\sum_{i=1}^{m} w_i^2$ is the sum squared errors of the weights in the network, m is the total number of weights. α and β are the hyper-parameters that need to be estimated and tuned in the training process. Network weights w are regarded as random variables and its density function is written as:

$$P(w|D,\alpha,\beta,M) = \frac{P(D|w,\beta,M)P(w|\alpha,M)}{P(D|\alpha,\beta,M)}, \tag{4.12}$$

where D represents the dataset, and M is the ANN topology information (dimension of inputs and outputs, number of hidden layers and neuron in hidden layer); $P(w|D,\alpha,\beta,M)$ is the posterior distribution of ANN weights, $P(D|w,\beta,M)$ is the likelihood function represents the training data occurrence probability with given weights, $P(w|\alpha,M)$ is the prior density of weights before data is fed. The BR algorithm is explained in detail in [169]. In general, all the noise in data is assumed to be Gaussian additive noise. With this assumption, the

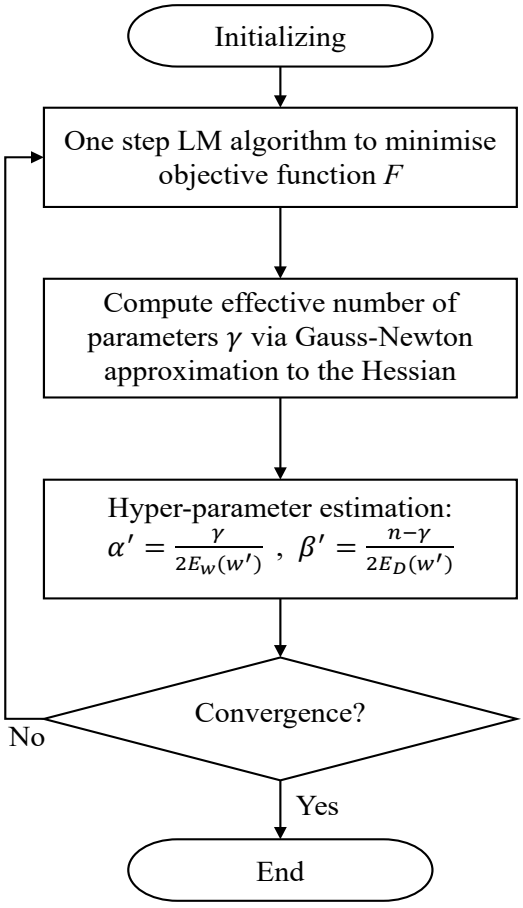

Fig. 4.3 The training process of PSO-BRNN.

probability density function of weights in equation (4.11) could be estimated. Then the hyper-parameters α and β are determined by solving the the Hessian matrix of F at the minimum point. Gauss-Newton approximation is used to solve Hessian matrix while the Levenburg-Marquardt (LM) training algorithm is employed to search the minimum point, the training process terminates once the training goal is met. The flow chart of BR is presented in Fig. 4.3, and the summary of the major step for particle swarm optimisation and Bayesian regularisation enhanced neural network (PSO-BRNN) training is shown in Fig. 4.4.

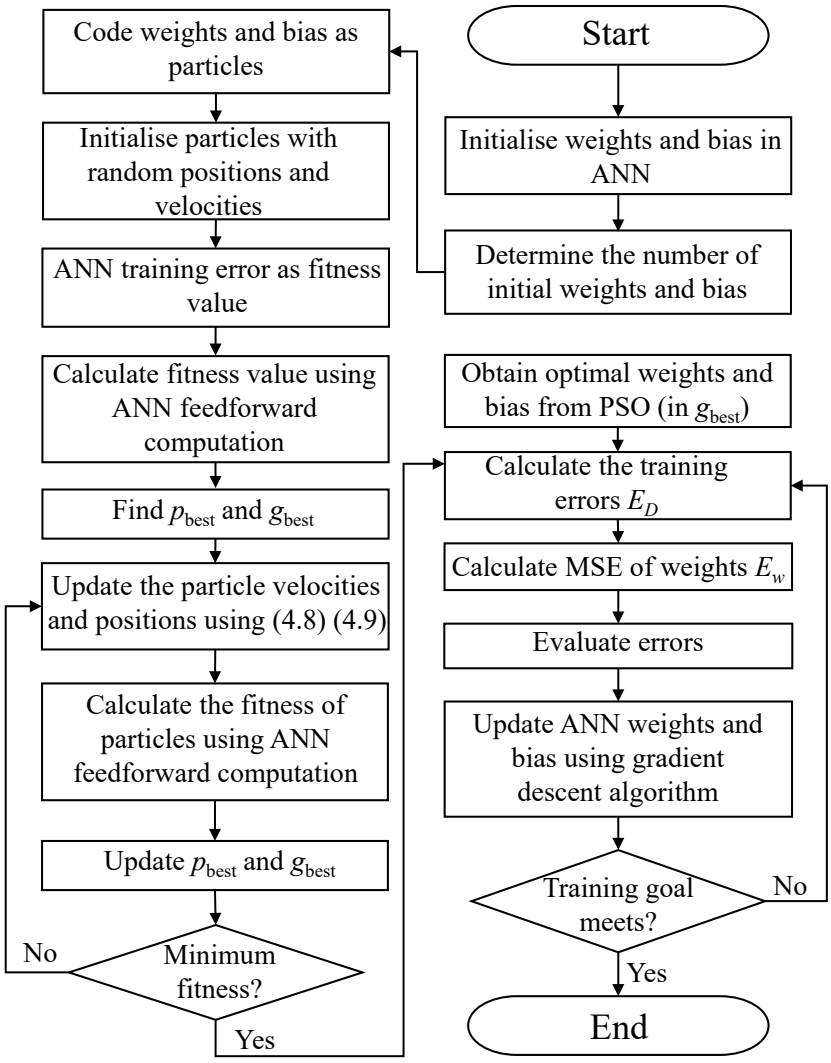

Fig. 4.4 The training process of PSO-BRNN.

4.4 Results

PSO is used for the ANN learning process, and weights and biases are adjusted to reach the minimum error between the ANN response and actual values. The training performance of PSO is compared with a BPNN, and the convergence and regression accuracy comparisons are exhibited. Fig. 4.5 compares PSO and BP algorithms in terms of convergence rate. It can be observed that PSO performs better than the BP algorithm. PSO converges faster than BP algorithm, and the iteration is terminated with a lower mean square error (MSE). The normalised mean square error (NMSE) is calculated in equation (4.13). The training error

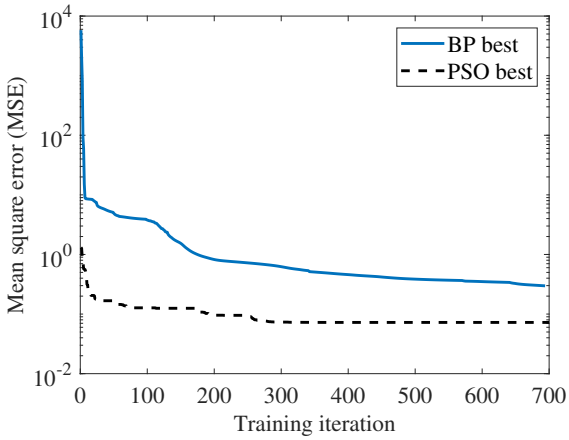

Fig. 4.5 Comparison of convergence between PSO and BP algorithm.

and time cost are illustrated in Table 4.2. It shows that the error of PSO-BRNN is much lower than BPNN, with 0.002% to 0.13%. In addition, by applying the PSO and BR, the iteration times of convergence are lower than BPNN, thus the result in the reduction of training time of PSO-BRNN (172 seconds) than BPNN (583 seconds).

$$NMSE = \frac{1}{N}\sum_{n=1}^{N_{\mathrm{LF}}} |\hat{y} - y|^2, \tag{4.13}$$

Fig. 4.6 presents the antenna performance prediction results of PSO-BRNN and BPNN with the actual values as reference. It can be observed that the learning accuracy of PSO-BRNN is better than the BPNN, the BPNN cannot map the fluctuation as detailed as PSO-BRNN while the embedding depth increases. This problem is caused by the local minima issue, once a network is trained with GD-based algorithm, the local minima is likely to be considered the best result by the network, so the training terminates. Therefore the error between network estimation and the actual value cannot be further minimised, then the weights and biases in the network stop adjusting and maintain in a plateau. While the PSO-BRNN is trained with optimum initialisation of weights and bias, it can map the slight fluctuation of the antenna performance, indicating the learning ability of PSO-BRNN is better.

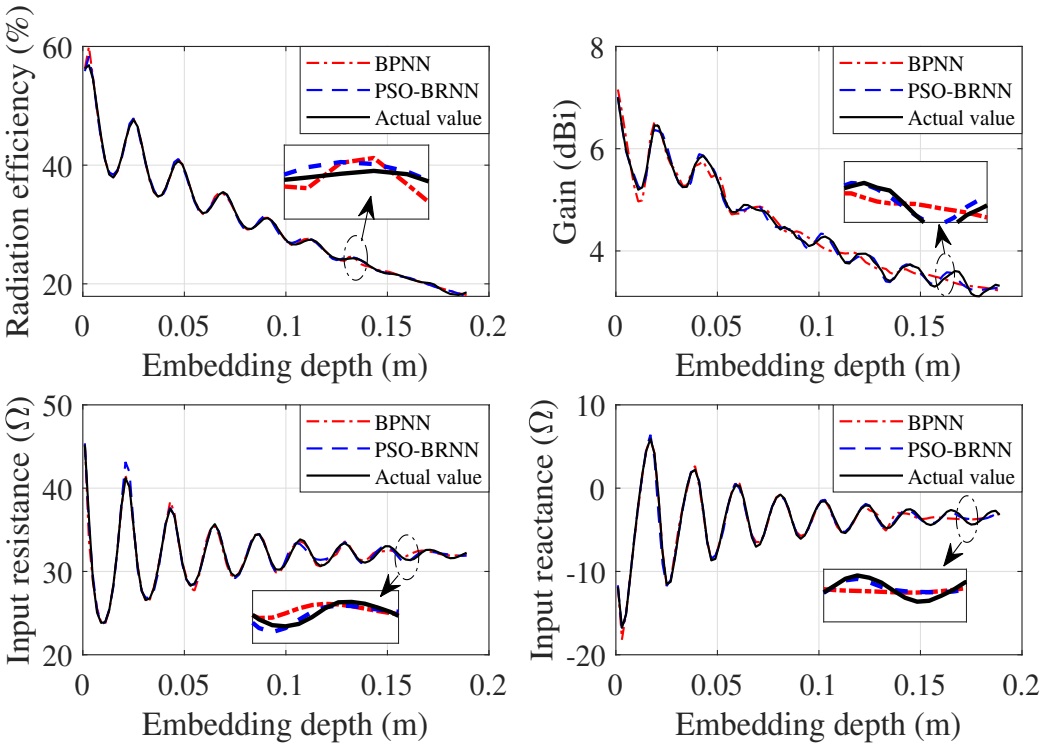

Fig. 4.6 Training performance comparison between BRNN and BPNN

The generalisation capability is essential for networks, and the network performance is mainly measured by its generalisation capability. For testing the generalisation capability of the trained ANN model, data other than that used in the training process is introduced, which is the testing dataset. The selected embedding depth d and concrete dielectric constant ε_r are exclusive from the dataset $\boldsymbol{D}$, the d is sampled with a step width of 0.004 meters and ranged from 0.001 meters to 0.189 meters and the ε_r is 4.5. The outputs of different networks are obtained and depicted in Fig. 4.7. The generalisation errors for PSO-BRNN, BPNN, and LR is presented in Table 4.3. Although, it can be observed that the PSO-BRNN performs better

Table 4.2 Training errors of antenna modelling with different methods

Algorithm	NMSE[1] [2] (%)	Iteraion	Time (s)
PSO-BRNN	0.002%	848	172
BPNN	0.13%	3084	583
LR	45.62%	N/A	N/A

[1] Measured in NMSE
[2] Average NMSE is considered

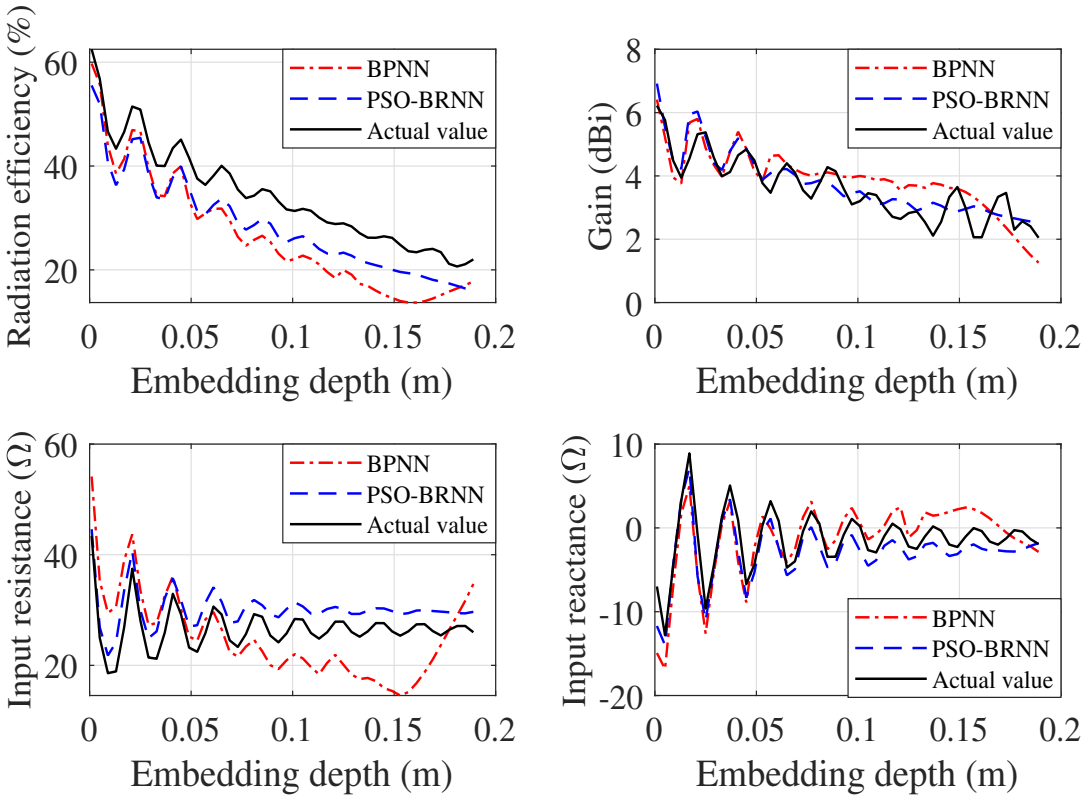

Fig. 4.7 Generalisation capability comparison between BRNN and BPNN

generalisation capability on new design points than BPNN and LR, most of the features of actual data can be captured by PSO-BRNN.

Table 4.3 Generalisation errors of antenna modelling with different methods

Algorithm	NMSE[1] [2] (%)
PSO-BRNN	26.63%
BPNN	63.83%
LR	222.07%

[1] Measured in NMSE
[2] Average NMSE is considered

4.5 Discussion

Compared to the LR-based method in Chapter 3, the non-linear relationships between inputs (embedded depth and concrete dielectric constant) and outputs (radiation efficiency, gain,

input resistance and input reactance) can be captured simultaneously using ANN rather than constructing a single model for individual performance parameter, which demonstrates that ANN-based model has good ability in handling antenna design with variate tasks. The combination of the PSO algorithm and BR method can improve the efficiency and accuracy in the training stage, even though the data amount is small. The utilisation of PSO prevents ANN trapping in local minima, and BR helps ANN adjust hyper-parameters to mitigate the over-fitting issue. However, compared to LR, the ANN-based model requires computational resources to generate data and training models. In addition, ANN-based model may show limited ability in modelling the antennas if training data is expensive to obtain. The results reflect an issue in data that a larger amount of data is needed to improve the overall performance, but the training data is expensive to be obtained due to the complex coupling between the antenna and concrete. In this case, a more efficient method to address the issue is needed to ensure the amount of training data and time cost.

4.6 Summary

In this work, the ANN-based method has been presented to predict the performance of a concrete embedded antenna. A hybrid ANN (PSO-BRNN) is trained to predict the performance of the concrete embedded antenna, and the training and generalisation errors are compared to the BPNN. The PSO algorithm is utilised to search for the global optimum weights and bias for ANN, and the BR algorithm is employed to overcome the over-fitting issue of ANN. Compared to LR and BPNN, PSO-BRNN exhibits an accurate manner in computation and prediction with a reduction in NMSE, and PSO-BRNN is more efficient than BPNN since it uses less iteration times to train. The generalisation capacity of different networks is tested with the new design points, the outputs of PSO-BRNN reveals an excellent generalisation capability, and its learning ability excels BPNN and LR. The results indicate that the PSO-BRNN is an effective method for the concrete embedded antenna performance prediction for indoor communication.

Chapter 5

Efficient Antenna Modelling Using Multi-Fidelity Stacked Neural Network

Overview

The computational EM-based simulation is an accurate and reliable way to evaluate antenna. However, with the rapid development of modern wireless communications, the structure of antennas is getting more complex, which leads to more requirements and constraints for design and optimisation. In this premise, the full-wave EM simulation solver is computationally-intensive once the more complex antenna structures and the increasing number of design variables are considered. The conventional antenna optimisation routine posts a challenge since a large number of evaluations of antennas are required to search for the optimal antenna design, especially the global optimisation using meta-heuristics algorithm (e.g., PSO [63], GA [66]) is extremely computationally expensive.

In Chapter 4, ANN was applied to a concrete embedded antenna modelling problem. The results show that the ML method could efficiently and accurately model an antenna. The trained model could substitute the EM simulation to calculate the antenna performance. However, a large amount of training data is required to ensure the accuracy of the ANN surrogate model. Inadequate training data can lead to the inaccurate surrogate and over-fitting issues. Since the training data is usually generated by HF EM simulation, the time cost of obtaining adequate training data is high. In this case, GPR becomes popular in antenna

design because it can give accurate predictions using fewer training patterns and provide uncertainties on new design points. However, once the antennas with a complex situation such as an increasing number of design variables, integration and environmental factors, the time cost of obtaining training data using EM simulation is still high.

MF modelling method can significantly reduce the cost of gathering initial training data for surrogate model construction by leveraging both LF and HF data [128]. In EM simulation, the fidelity is generally defined according to the mesh size: HF models are generated by the highly accurate and computationally expensive simulations with fine-discretisation; while the LF models obtained by simulations with coarse mesh size are less accurate, but their evaluations are very fast. By exploiting the correlation between LF and HF model, sufficient LF data can be utilised along with a small amount of HF data to accurately build up a surrogate model. The combination of GPR and MF modelling is popular in antenna design and improve antenna modelling efficiency without compromising accuracy [117] [121] [140] [150].

Compared to GPR, ANN can be flexibly interfered with (e.g., add and drop neurons or layers, train from checkpoints) and transferred. By applying techniques such as regularisation and dropout, ANN can obtain comparable efficiency and accuracy as GPR. The combination of a multi-fidelity approach and the deep neural network has been applied to fluid mechanics [170] and partial differential equation (PDE) solving [171]. The results indicate that the multi-fidelity-based neural network has excellent expressive capability and can mitigate the over-fitting issue while a small amount of training data is used. However, to the best of our knowledge, MF-based NN applied in antenna modelling and optimisation has not been reported yet. On the other hand, the MF-based GPR framework in current antenna modelling and optimisation works is based on a linear autoregressive scheme, which assumes the correlation between LF and HF data is linear. However, the correlation between different fidelities can go beyond linearity if the ranges of input parameters are large [172]. In this work, MFSNN is proposed to construct a surrogate model for antenna optimisation. The specific contributions of this work include:

• MF based stacked neural network is introduced in antenna modelling, which significantly reduces the HF training datasets required for ANN surrogate model construction.

• Both linear and non-linear correction between LF and HF models are considered in order to adaptively exploit the multi-fidelity information fusion in antenna modelling, which improves the accuracy of the proposed surrogate model.

• Accurate global surrogate models for broad ranges of input parameters are constructed by the proposed neural network efficiently, which can be applied in the global optimisation framework directly.

5.1 Overview of Multi-Fidelity Neural Network

A surrogate composed of three full-connected independent neural networks is presented in Fig. 5.1, which are a low-fidelity neural network (LFNN), a non-linear HF neural network (HFNN_1) and a linear HF neural network (HFNN_2). The MF modelling method is applied to the neural network construction, which aims to exploit the relationship between LF and HF data. The LFNN (x_L, x_L, θ) is used to approximate LF data, where (x_L, x_L) donates LF training dataset and θ is the corresponding hyper-parameters. The relationship between LF and HF data could be expressed as:

$$\hat{y}_H = F(x, \hat{y}_L), \tag{5.1}$$

where $F(.)$ is an unknown function that is used in reveal the correlation between different fidelities, $\hat{y}_L$ and $\hat{y}_H$ are predicted results of LFNN and HFNN respectively, x is the input of surrogate. Considered the potential correlation between LF and HF data, $F(.)$ is decomposed into non-linear component and linear component, which are denoted as HFNN_1 and HFNN_2, respectively, as shown in Fig. 5.1. The training dataset HFNN_1 and HFNN_2 is denoted as ($[x_\mathrm{H}, \hat{y}_L]$, y_H), in which $\hat{y}_L$ is the predicted LF response for the input of x_H. The hyper-parameters of HFNN_1 and HFNN_2 are φ_1 and φ_2 respectively. HFNN_1 and HFNN_2 are trained parallelly and the final output of the entire network is sum of the two HFNN's outputs, in this way the correlation between LF and HF data can be learned adaptively.

In the training process, the bias and weights of the neural network are adjusted by minimising the cost functions and optimising the hyper-parameters. In our stacked ANN framework, each network has individual cost function, which are shown as following:

$$f_{\text{total}} = f_{\text{LFNN}} + f_{\text{HFNN}} + \lambda \sum \omega_{i,j}^2, \tag{5.2}$$

where

$$f_{\text{LFNN}} = \frac{1}{N_{\text{LF}}} \sum_{n=1}^{N_{\text{LF}}} |\hat{y}_{\text{LF}} - y_{\text{L}}|^2, \tag{5.3}$$

$$f_{\text{HFNN}} = f_{\text{HFNN}_1} + f_{\text{HFNN}_2}, \tag{5.4}$$

$$f_{\text{HFNN}_t} = \frac{1}{N_{\text{HF}}} \sum_{n=1}^{N_{\text{HF}}} |\hat{y}_{\text{HFNN}_t} - y_{\text{H}}|^2, t = 1,2. \tag{5.5}$$

The total cost of the whole network is measured as the sum of individual costs and regularisation item cost. λ is the L_2 regularisation ratio for weights $w_{i,j}$ of HFNN_1. L_2 regularisation is an extensive strategy to prevent the potential over-fitting issue, especially when the amount of training data is small. The network is trained using the Adam algorithm, and the limited memory BFGS (L-BFGS) algorithm is employed for searching optimum locally. In order to start from good initial point, the Xavier's initialisation method is introduced [173] [174]. Xavier's initialisation balances the variance of inputs and outputs, which is expressed as:

$$w_{i,j}, b_{i,j} \sim U(-\sqrt{\frac{6}{n_{\text{in}} + n_{\text{out}}}}, \sqrt{\frac{6}{n_{\text{in}} + n_{\text{out}}}}), \tag{5.6}$$

where the $w_{i,j}$ and $b_{i,j}$ are the weights and bias of the whole network, n_{in} and n_{out} are the dimensionalities of input and output, respectively.

The method starts with obtaining the training data using EM simulation with different fidelities. The LFNN is firstly constructed using LF simulated results, then its prediction on the input of the HF dataset is stacked with HF data input x_{H}, plus with the HF output y_{H} to

construct and train the HFNN. The whole network is then utilised to approximate the HF data and make a prediction on new design points.

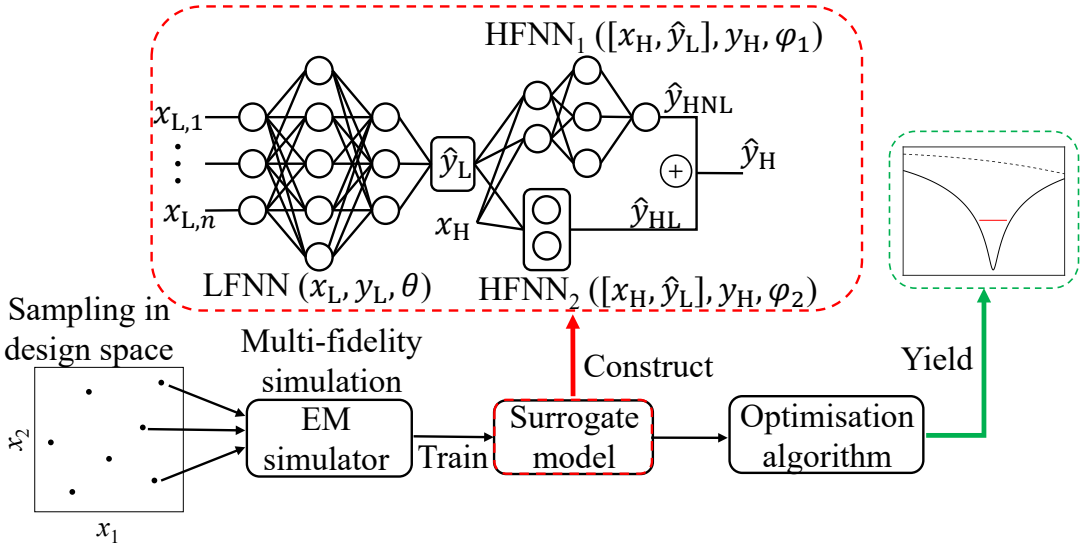

Fig. 5.1 Schematic of multi-fidelity based stacked neural network for antenna modelling.

In order to model the antenna, firstly, the design variables of the target antenna that can significantly affect the antenna performance are considered. Then the antenna is constructed using CST studio. The antenna should be carefully adjusted and could work appropriately in the frequency of interest. In addition, the LF antenna model is created using the simplification options mentioned in Section 2.4.1, and the trend and shape of full-wave EM simulation results should be generally similar to the results of the HF antenna model.

In the second stage, the design space of the antenna is defined by the lower bound (LB) and upper bound (UB) of individual antenna design variables, and the constraint conditions are possibly taken into consideration depending on the antenna's physical characteristics. In order to accurately model an antenna with a relatively small amount of samples, some DoE sampling techniques are used to arrange space-filling samples within the design space. Thus the sample could uniformly cover the design space. The testing data could be obtained using DoE as well.

In the data acquisition stage, import the antenna geometries that are sampled by DoE techniques into CST studio, the parameter sweep is implemented on both the HF model

and LF model, and the corresponding performance parameter of interest (e.g. reflection coefficient, gain and axis ratio etc.) could be obtained at each simulation iteration.

In the training stage, firstly, the LF data is used to build up the LFNN, and then the inputs of HF data are fed to get the response which can be treated as 'residual'. The model in step 1 is stacked with a composite network containing a linear and non-linear neural network. The HF inputs are stacked with residuals to train the HFNNs. Three networks are trained at the same time, and the weights and biases of the three networks are trained and updated simultaneously.

5.2 Surrogate Based Antenna Modelling

In this section, three antenna cases (a slot dipole antenna, a dual-frequency slot antenna and an ultra-wideband monopole antenna) are employed to examine the effectiveness of the MFSNN. The implementation environment of MFSNN is Python 3.7, TensorFlow r1.15 GPU version, CUDA 10.0 and cuDNN 7.4. The deep neural network is run on a Dell PowerEdge C4140 GPU node with Intel Xeon Gold 6138 2.00 GHz and 384 GB RAM, and an Nvidia Tesla V100 graphic unit (32 GB) is employed for GPU acceleration. All the EM models are modelled and simulated using CST Microwave Studio (2021 version) on the DELL workstation with Intel Xeon W2135 3.70 GHz and 32 GB RAM.

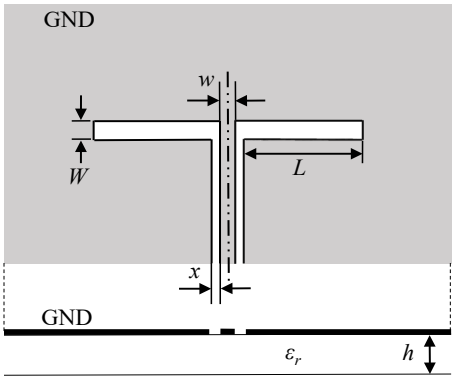

Fig. 5.2 Geometry of CPW-fed slot dipole antenna (Antenna 1).

5.2.1 CPW-fed Slot Dipole Antenna (Antenna 1)

A 50 Ω coplanar waveguide (CPW) line fed slot dipole antenna [117] with the infinite ground is shown in Fig. 5.2. The considered design variable for this antenna is $\boldsymbol{x} = [W, L]$, and the lower bound and upper bound are $[5, 28]$ mm and $[10, 50]$ mm, respectively. Other parameters are fixed, where the feed line width w = 0.5 mm, feed lines gap s = 4.0 mm, substrate thickness h = 1.6 mm and substrate dielectric constant ε_r = 4.4. The frequency band of interest is 2 - 2.7 GHz, and the modelling interest is the real part and imaginary part of the reflection coefficient (Re$\{S_{11}\}$ and Im$\{S_{11}\}$) within the frequency band of interest.

138 geometries are selected using LHS within the design space, along with 10 uniformly sampled frequency points to form the training data. The testing data contains 100 new geometries that are selected using LHS within the design space as well, and 71 equally-spaced points are sampled with each geometry.

The LF and HF antenna models are constructed and simulated using CST Microwave Studio. The HF model is obtained by simulation with fine mesh density in CST, while the LF model is simulated with coarse mesh (reducing the number of cells), relaxed steady-state criteria (reducing the simulation duration), simplified excitation port (e.g. discrete port instead of waveguide port), and simplified material (e.g. perfect conductor and dielectric without loss). The HF model for Antenna 1 with fine mesh density and high accuracy is constructed with 130,000 mesh cells, and the simulation time is 87 s. In contrast, the LF model is constructed with 5,000 mesh cells, and the simulation time is 7 s. The responses of the HF and LF model for one sample are presented in Fig. 5.3. It can be observed that the LF model has a similar trend as the HF model, even though the LF model is not accurate enough.

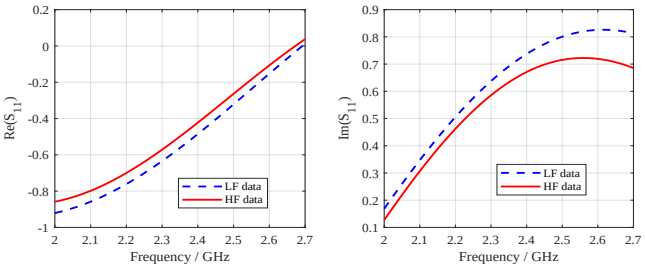

Fig. 5.3 Response of $Re(S_{11})$ and $Im(S_{11})$ against frequency for Antenna 1.

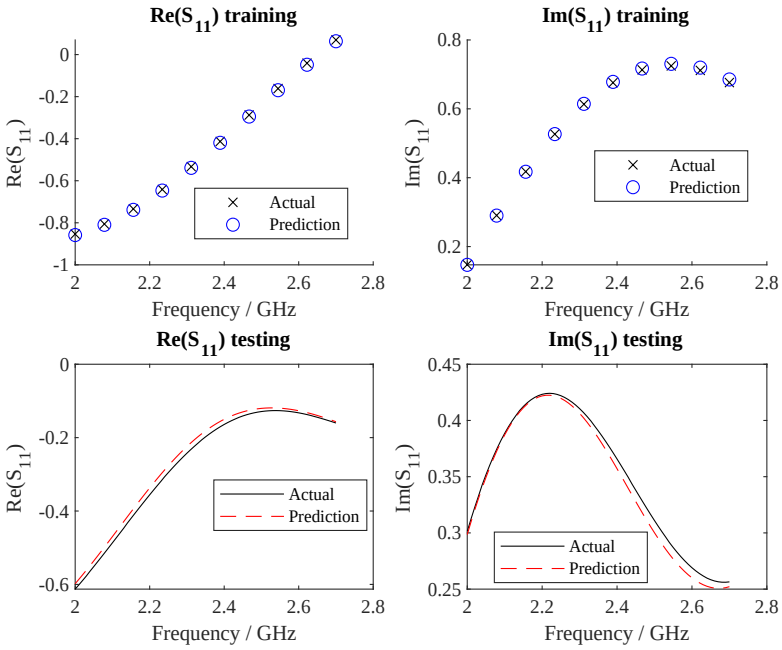

Fig. 5.4 Antenna 1 modelling results using 10% HF data.

The training data is simulated as both LF and HF models to construct the surrogate model, while the testing data is only simulated as the HF model for testing the generalisation ability of the proposed stacked neural network. The surrogate modelling is repeatedly implemented using different proportions of HF data, the proportion is defined as $\frac{n_{\mathrm{HF}}}{n} \times 100\%$, n is the number of total geometries used in the training data which is 138 in this example, and n_{HF} is the number of geometries that are simulated using HF simulation. The percentage of HF data in the training dataset varies from 70% to 10%. For example, if the percentage of HF data is 10%, we use 124 LF datasets to train LFNN, and 14 HF datasets to train HFNN_1 and HFNN_2. The training and testing errors are listed in Table 5.1, where the 100% proportion of HF data is set as a reference which means all the 200 training datasets are generated by HF simulation and are used to train a conventional ANN. The training and testing results of the surrogate model using 10% HF data are shown in Fig. 5.4. Since the simulation time of the LF model is only about one-twentieth of the HF model, a large amount of LF data is used to train LFNN to predict the variation trend of the reflection coefficient. After that, a small amount of HF data is used to train HFNN to compensate for the discrepancy between LF and HF models. Table 5.1 and Fig. 5.4 show that even if the amount of HF data is only 10% of

Table 5.1 Training and testing errors of Antenna 1 modelling with different HF data proportion

Proportion (%)	Tr[1][Re]	Tr[Im]	Te[1][Re]	Te[Im]
100	0.001%[2]	0.002%	0.22%	0.27%
70	0.004%	0.003%	0.75%	0.81%
60	0.008%	0.009%	0.74%	0.82%
50	0.016%	0.011%	0.78%	0.84%
40	0.013%	0.019%	0.75%	0.81%
30	0.019%	0.017%	0.75%	0.82%
20	0.016%	0.015%	0.75%	0.82%
10	0.034%	0.018%	1.64%	1.10%

[1] Tr stands for training, Te stands for testing.
[2] Normalised mean square error in percentage.

LF data, the prediction error on the testing dataset is about 1.5%. The proposed MFSNN has dramatically reduced the demand for HF training datasets, which can be served as a good global surrogate model since it exhibits high accuracy in a wide range of design variables.

5.2.2 Dual-Frequency Slot Antenna (Antenna 2)

A dual-frequency slot antenna [175] composed by an annular slot and a cross fed by microstrip line is presented in Fig. 5.5. The considered design vector is $\boldsymbol{X} = [L_1, L_2, R, d_p, W_p, L_s, L_t, W_t]^{\mathrm{T}}$, where L_1 and L_2 are the lengths of the two rectangular slots forming the cross slot, R is the radius of the annular slot, d_p and W_p are the depth and width of the pair of notches on the annular slot, L_s is the length of the stub of the microstrip line, L_t and W_t are the length and width of the impedance transformer. The LB and UB of the variation range of the design vector are $[22, 18, 12, 1, 1, -2, 10.5, 0.3]$ mm and $[34, 32, 18, 6, 6, 5, 22.8, 2.8]$ mm, respectively. Other parameters are substrate thickness h = 1.45 mm, dielectric constant of substrate $\varepsilon_{\mathrm{r}} = 4.2$, width of the feeding line W_f = 2.96 mm. The frequency band of interest is 1 - 7 GHz.

For training purposes, 700 geometries are sampled within the design space using LHS with 50 uniformly sampled frequency points per geometry. The testing dataset is composed of 100 new geometries obtained using LHS, with 121 equally-spaced frequency points per geometry. The HF model for Antenna 2 is constructed with 370,818 mesh cells in CST, and

its simulation costs 324s. While the LF model is constructed in 16,770 mesh cells whose simulation time is 12 s. The simplifications in the LF model are the same as those in Antenna 1, and we can see that the simulation time of the LF model is 27 times shorter than that of the HF model. The antenna responses of different fidelities are shown in Fig. 5.6.

The modelling process is similar to Antenna 1. The proportion of HF data in the training dataset varies from 70% to 30%, and the corresponding training and testing errors are listed in Table II. The training and testing results of the surrogate model on Antenna 2 using 30% HF data are shown in Fig. 5.7. In this example, the dimension of the design space is high and the variation range of each design variable is large. Table 5.2 and Fig. 5.7 indicate that when the percentage of HF data is 30%, the testing errors on new geometries are less than 9%. So the proposed MFSNN framework can construct an accurate global surrogate model for complex antenna structures with just a small amount of HF data.

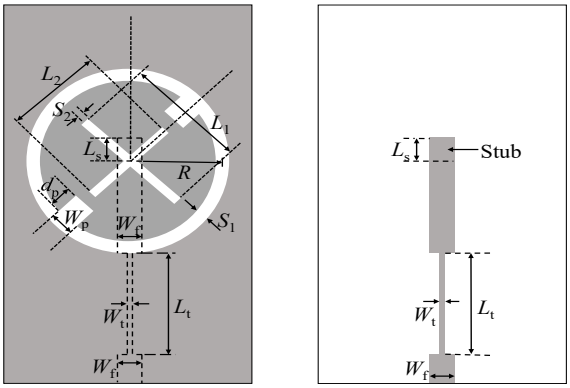

Fig. 5.5 Structure of dual-frequency slot antenna (Antenna 2).

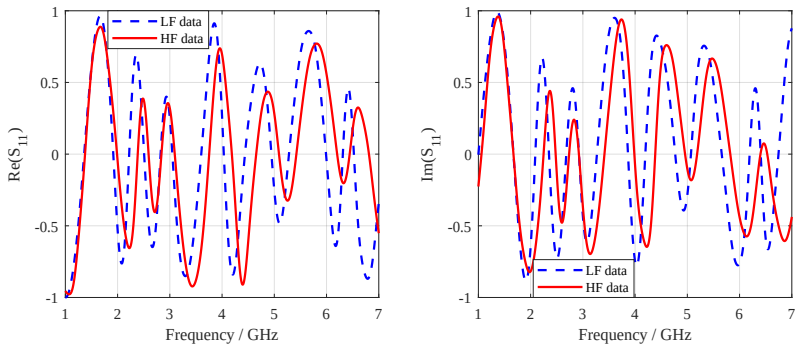

Fig. 5.6 Response of $Re(S_{11})$ and $Im(S_{11})$ against frequency for Antenna 2.

Table 5.2 Training and testing errors of Antenna 2 modelling with different HF data proportion

Proportion (%)	Tr[1][Re]	Tr[Im]	Te[1][Re]	Te[Im]
100	0.001%[1]	0.001%	1.33%	1.51%
70	0.019%	0.017%	2.13%	2.67%
60	0.030%	0.025%	2.93%	3.15%
50	0.040%	0.032%	3.55%	4.01%
40	0.069%	0.048%	6.81%	7.27%
30	0.095%	0.088%	8.84%	8.26%

[1] Tr stands for training, Te stands for testing.
[2] Normalised mean square error in percentage.

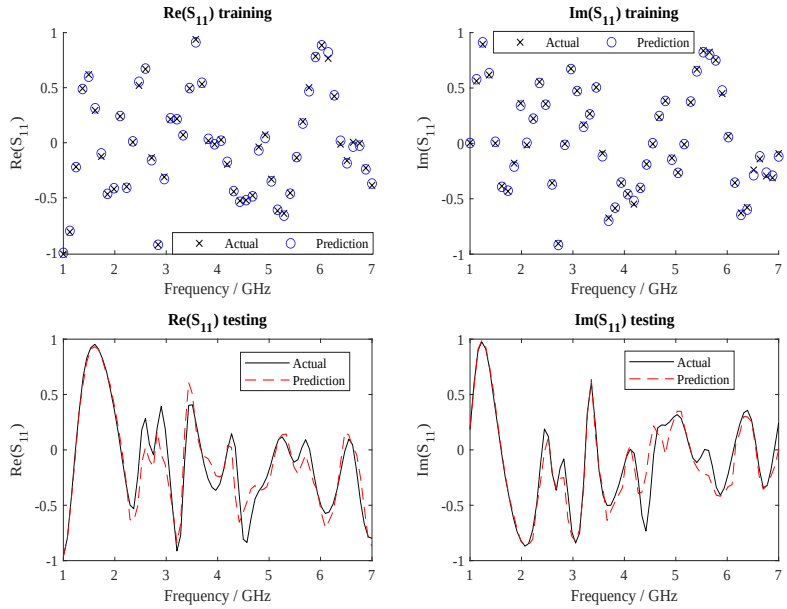

Fig. 5.7 Antenna 2 modelling results using 30% HF data.

5.2.3 Wideband Monopole Antenna (Antenna 3)

The antenna structure is shown in Fig. 5.8, in which an elliptical monopole is fed by a 50 Ω coplanar waveguide (CPW) line with the chambered ground. The design variable is $\boldsymbol{X} = [R_x, R_y, G_x, G_y]$, whose LB and UB are $[5,5,6,6]$ mm and $[15.5, 15.5, 11, 11]$ mm, respectively. Other dimensions are fixed (substrate dielectric constant $\varepsilon_r = 4.2$, substrate thickness h = 1.6 mm, H_1 = 7 mm, w_1 = 3 mm, w_0 = 1.6 mm, and the width W and length L of the dielectric substrate are 31 mm and 44.3 mm). The frequency band of interest is 3

- 10 GHz, and the modelling interests are the real part and imaginary part of the reflection coefficient (Re$\{S_{11}\}$ and Im$\{S_{11}\}$) within the frequency band of interest.

The training datasets comprise 200 geometries selected using LHS within the design space, with 50 uniformly sampled frequency points per geometry. In addition, the testing data contains 100 new geometries obtained by LHS, with 121 equally-spaced frequency points for each geometry.

The LF and HF antenna models are constructed and simulated using CST Microwave Studio. The HF model is obtained by simulation with fine mesh density in CST, while the LF model is simulated with coarse mesh, the LF and HF model response of Re$\{S_{11}\}$ and Re$\{S_{11}\}$ are presented in Fig. 5.9. The HF model is constructed using 274,220 mesh cells and the simulation time is 213 s, while LF model is constructed in 12,354 mesh cells and taken 10 s to simulate.

For model training, 200 antenna geometries are sampled from the design space using LHS, with 50 frequency points with each antenna geometry. Thus the design vector is written as $\boldsymbol{x} = [R_x, R_y, G_x, G_y, f]^{\mathrm{T}}$. In order to test the accuracy of antenna modelling, 100 antenna geometries are sampled using LHS from design space and excluded from the training dataset. Each geometry is concatenated with 121 equally spaced frequency points, and the testing dataset is generated using HF model simulation. Similar to the above examples, the HF data is fed to MFNN in different proportions, from 70% to 10%, and the modelling results using 10% is shown in Fig. 5.10. The modelling performance for this example is presented in Table 5.3.

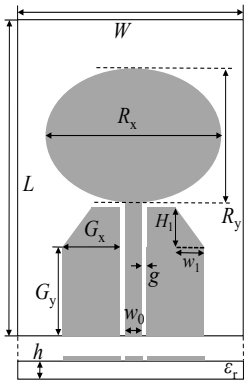

Fig. 5.8 Structure of the wideband CPW-fed monopole antenna (Antenna 3).

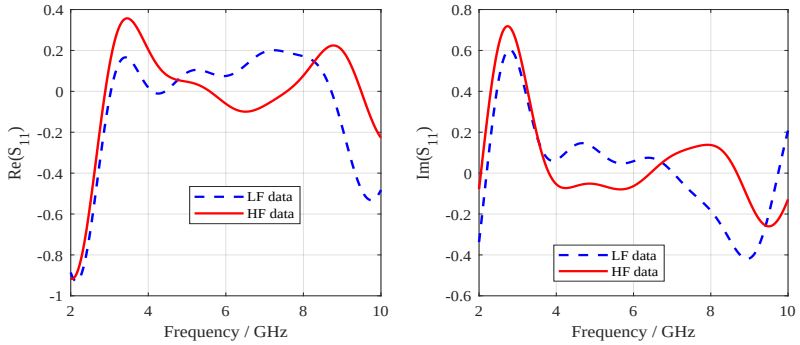

Fig. 5.9 Response of $Re(S_{11})$ and $Im(S_{11})$ against frequency for Antenna 3.

Table 5.3 Training and testing errors of Antenna 3 modelling with different HF data proportion

Proportion (%)	Tr[1][Re]	Tr[Im]	Te[1][Re]	Te[Im]
100	0.001%[2]	0.002%	0.11%	0.15%
70	0.06%	0.04%	0.61%	0.57%
60	0.13%	0.15%	0.92%	0.88%
50	0.22%	0.28%	1.32%	1.67%
40	0.33%	0.35%	1.84%	1.94%
30	0.41%	0.39%	2.34%	2.63%
20	0.52%	0.46%	2.73%	2.81%
10	0.78%	0.62%	3.39%	4.18%

[1] Tr stands for training, Te stands for testing.
[2] Normalised mean square error in percentage.

5.3 Discussion

To investigate the effect of decomposition in HF data approximation, the Antenna 2 modelling result using 30% HF data (210 geometries are evaluated as HF model) by different manners are presented in Fig. 5.11. Fig. 5.11 (a) shows the prediction results of a single conventional ANN trained by only 210 HF datasets. Due to insufficient training data, the conventional ANN cannot extract the features from the training data and the generalisation ability is limited, thus giving inaccurate prediction. For the result in Fig. 5.11 (b), both LFNN and HFNN are employed in modelling. However, only linear network ($HFNN_2$) is used in HFNN. The accuracy for this situation is poor, which indicates that the correlation between LF and

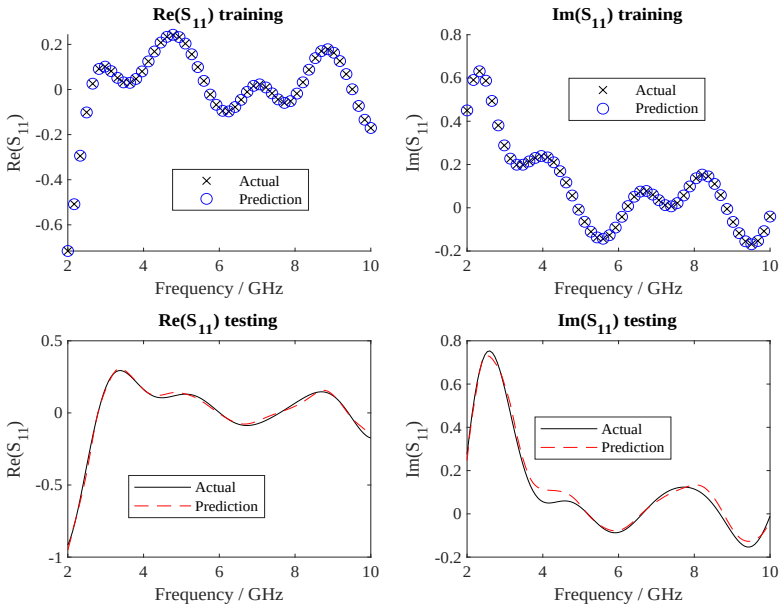

Fig. 5.10 Antenna 3 modelling results using 10% HF data.

HF models goes beyond linearity, especially for higher frequency band, since coarse mesh leads to large error at higher frequency where the discrepancy between LF and HF models is complex. Fig. 5.11 (c) gives the modelling results of LFNN connected with the non-linear network ($HFNN_1$). Compared to the results in Fig. 5.11 (b), the surrogate model in this case obtains higher accuracy except deteriorating in a few details, it appears this NN has learned most of the features of the data. The modelling results of the proposed MFSNN are shown in Fig. 5.11 (d), in which the prediction results agree well with the actual HF results. Compared to Fig. 5.11 (a), the MFSNN model yields accurate prediction with a small amount of HF data since it learns prior knowledge from sufficient LF data; compared to Fig. 5.11 (b) and (c), the MFSNN gains the highest prediction accuracy since it decomposes the correlation between LF and HF models into linear and non-linear components so that the correlation can be learned adaptively.

5.4 Application on antenna optimisations

In this section, the aforementioned MFSNN surrogate model is applied to antenna optimisation to validate the accuracy and efficiency of the proposed modelling method. The three antennas discussed above are employed to demonstrate the optimisation process. The

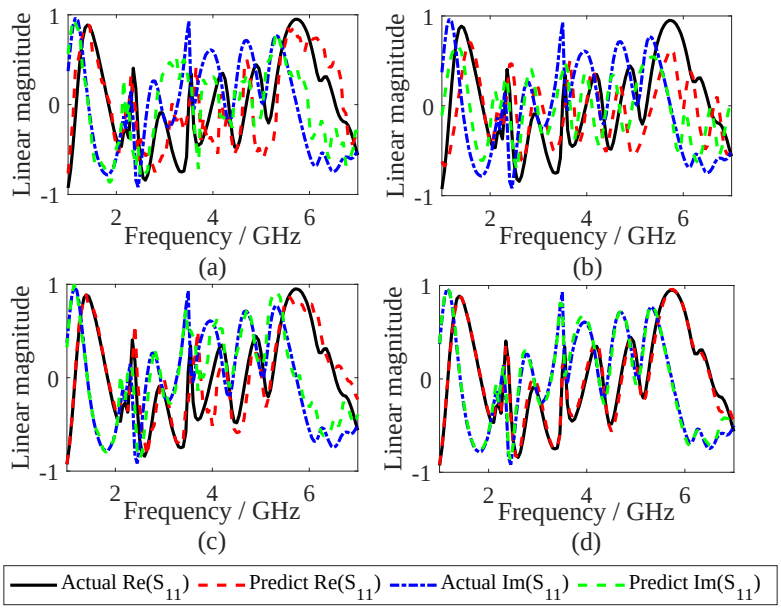

Fig. 5.11 Testing results for Antenna 2 modelling with 30% HF data using (a) conventional ANN; (b) LFNN + $HFNN_2$; (c) LFNN + $HFNN_1$; (d) LFNN + $HFNN_1$ + $HFNN_2$.

objective is to minimise the maximum S_{11} within the frequency band of interest (2-2.7 GHz for Antenna 1; 2.35-2.45 GHz and 5.75-5.85 GHz for Antenna 2; 3.1-10.6 GHz for Antenna 3). The PSO algorithm is utilised to search the optimum candidate in the design space. For comparison, the optimisation is implemented based on three methods. In method 1, the non-surrogate (NS) method is used, and each candidate in PSO is evaluated by HF full-wave simulation. In method 2, the surrogate model is constructed by the conventional ANN which is trained by sufficient HF data. In method 3, the surrogate model is constructed by the proposed MFSNN and trained by sufficient LF data along with a small amount of HF data. The proportion of HF data is 10% for Antenna 1, 30% for Antenna 2, and 10% for Antenna 3. In methods 2 and 3, the candidates in PSO are evaluated by the corresponding surrogate model.

The antenna optimisation results are shown in Fig. 5.12, in which the HF full-wave responses of the initial design and optimized designs for the three methods are provided. It can be seen that all the three methods can obtain the optimized design which fulfils the objective. Although the proposed MFSNN introduces some modelling errors due to the

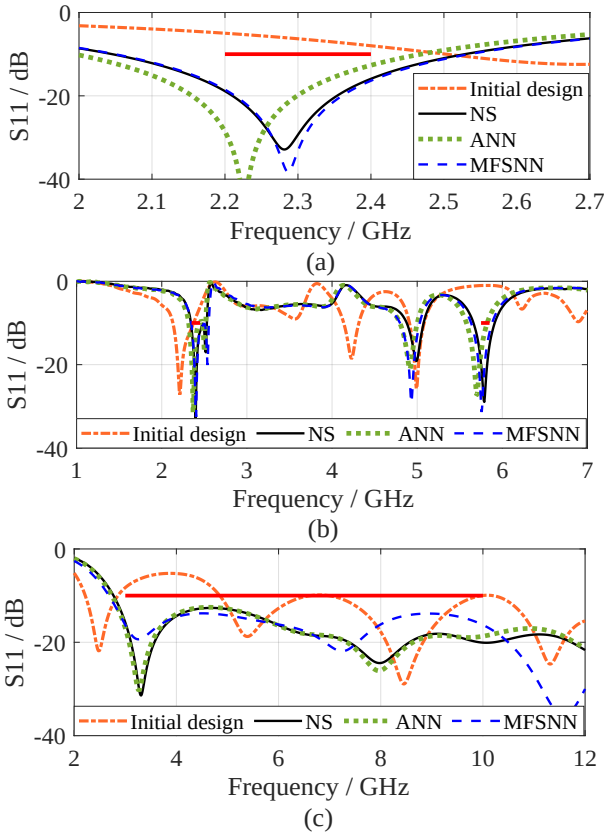

Fig. 5.12 Antenna optimisation results of the three methods along with the responses of initial design for (a) Antenna 1, (b) Antenna 2, and (c) Antenna 3. The objectives are drawn with red solid line.

limited HF training data, it is accurate enough to surrogate the full-wave simulation in a wide design space for optimisation application.

The comparisons of the cost of the three methods are shown in Table 5.4, in which the maximum S_{11} in the frequency band of interest, the number of iterations in PSO and the amount of full-wave simulations are given. The maximum S_{11} are below -10 dB and the required PSO iterations for meeting the design specification are similar for all three methods of each antenna. In the non-surrogate method, every candidate in each iteration has to call HF full-wave simulation to evaluate the objective function, so the computational cost is very high. In contrast, the cost of candidate evaluation in surrogate-based optimisation can be neglected since it is just a fast-forward pass of the NN in our model. However, the construction of conventional ANN requires a large number of HF training data to ensure its prediction

Table 5.4 Comparisons of the three optimisation methods

Antenna	Method	max$\|S_{11}\|^2$	Optimisation cost[3]	EM simulation [4]
1	NS[1]	-32.87 dB	5	48(H)
	ANN	-44.84 dB	4	138 (H)
	MFSNN	-38.28 dB	5	14 (H) + 124 (L)
2	NS[1]	-11.94 dB	10	1000(H)
	ANN	-10.16 dB	9	700 (H)
	MFSNN	-11.39 dB	11	210 (H) + 490 (L)
3	NS[1]	-12.64 dB	5	150(H)
	ANN	-12.50 dB	4	200 (H)
	MFSNN	-13.80 dB	7	20 (H) + 180 (L)

[1] NS stands for non-surrogate method.
[2] maximum value of $|S_{11}|$ at frequency band of interest: 2 to 2.7 GHz for Antenna 1, and 2.35 to 2.45 GHz and 5.75 to 5.85 GHz for Antenna 2, and 3.1 to 10.6 GHz for Antenna 3, optimisation objective is max $|S_{11}| \leq$ -10 dB.
[3] Number of iterations of PSO.
[4] It counts the total required amount of EM simulations in surrogate model construction stage and optimisation stage; where (H) refers to EM simulation of HF models, (L) is for LF models.

Table 5.5 Time cost of different method in antenna optimisation

Antenna	NS[2]	ANN	MFSNN
1	4,176 s	12,006 s (D^3) + 475 s (T^3)	2,086 s (D) + 503 s (T)
2	324,000 s	226,800 s (D) + 19,800 s (T)	73,920 s (D) + 33,840 s (T)
3	31,950 s	42,600 s (D) + 2,820 s (T)	6,060 s (D) + 3,780 s (T)

[1] NS stands for non-surrogate method.
[2] (D) refers to data generation using EM simulation, (T) refers to surrogate model training.

accuracy. Although the ANN method eliminates the full-wave simulation in optimisation stage, it heavily relies on HF EM simulation in the construction stage, so its full-wave simulation cost is comparable to or even larger (for Antenna 1) than the non-surrogate method. The proposed MFSNN has significantly reduced the requirement of HF EM simulations in model construction by learning the prior from LF data and exploiting the correlation between LF and HF models, and it can be applied in the global optimisation framework directly to replace the computationally expensive full-wave simulation and accelerate the optimisation procedure. The time cost details of three methods in example antennas optimisation are given in Table 5.5, it can be observed that the MFSNN saves significant time cost than NS and ANN methods. For Antenna 1, MFSNN saves 38% time cost than NS method and 79.26%

than ANN method; Antenna 2, MFSNN saves 66.74% time cost than NS method and 56.3% than ANN; 69.2% and 78.34% time cost saving are achieved rather than using NS and ANN for Antenna 3.

5.5 Summary

A stacked neural network based on the MF framework is introduced for antenna modelling and optimisation, which can construct an accurate surrogate model in a wide design space with just a small amount of HF training data. Furthermore, the correlation between LF and HF models can be learned adaptively and accurately by decomposing the correlation into linear component and non-linear component. Two antenna examples are used to demonstrate the model construction and the modelling errors are discussed in different cases. Furthermore, the application of antenna optimisation is presented and the performances of different optimisation methods are compared. Last, the MFSNN-based surrogate model presented in this paper can be efficiently applied in but not limited to optimisation, robust design and sensitivity analysis. More application scenarios are future research topics.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

The ML-assisted antenna modelling methods in the thesis were mainly investigated with three works. In the first part, the LR method is employed to model a concrete embedding antenna under indoor communication scenario. The simulated data is obtained using full-wave EM simulation, including antenna radiation efficiency, gain, input resistance and input reactance under different embedding depths and concrete dielectric constants. Four empirical formulas are fitted based on the simulated data, each performance parameter is characterised as a 2-dimensional equation with embedding depth and concrete dielectric constant as independent variables. The regression models can give an accurate prediction of the antenna performance, but the generalisation capability still needs to be improved. The linear regression model costs low computational resources in antenna modelling because of its simplicity, and it can provide a quick estimation of antenna behaviour. This method is relatively easy to apply to wide range of antennas. However, the LR model has limited capability in solving complex antenna modelling tasks, and only one antenna performance is considered in a model. Another potential limitation is that the model accuracy heavily depends on the quality and representativeness of the training data, once training data is not representative of the antenna's operating conditions or includes outliers or errors, the regression model can be inaccurate and unreliable. In conclusion, LR method is a useful tool for modelling antenna

behaviour, particularly for simple antenna design and where a quick performance estimate is needed. The training data should be carefully inspected and processed. Otherwise the model will be inaccurate and unreliable.

The second part introduces a heuristic algorithm-enhanced ANN method to efficiently model the concrete embedded antenna in the indoor communication scenario. Considering the limitation of linear regression in antenna modelling, the ANN algorithm is employed in modelling. Four considered antenna performance parameters (radiation efficiency, gain, input resistance and input reactance) can be simultaneously predicted based on given inputs (embedding depth and concrete dielectric constant), which accelerates the modelling process, and the ANN model is easily implemented. Since the complex coupling between antenna and concrete, the size of training data is limited, which can lead to the over-fitting issue. The utilisation of BR prevents this issue. In addition, the PSO algorithm is used to address the local minima issue. The PSO-BRNN performs better than conventional BPNN and linear regression in modelling accuracy (1.79% to 3.92% and 8.7%) and generalisation capability. However, the data size requirement is more considerable than LR, so more computational resources are needed in the data acquisition stage. The ANN-based model, which heavily relies on HF EM simulation, is suitable for the antenna design with an intermediate number of design variables and not particularly complex structures.

In the third part, MFSNN is introduced in antenna modelling and optimisation. Considering the computational and time cost of EM simulation in Chapter 3 and Chapter 4, an efficient modelling method is needed to address the data acquisition budget for ensuring the accuracy of the surrogate model. The utilisation of the MF approach significantly reduces the time cost in training data acquisition and also ensures the amount of training data surrogate model construction. Different types of antenna (single-band, dual-band, and ultra-wide band antenna) are employed to validate the effectiveness of the MFSNN surrogate model. The training and testing accuracy of MFSNN surrogate model are improved by correctly exploiting the correlation between LF and HF models. A large amount of LF data and a small amount of HF data are sufficient for training MFSNN. The network accuracy indicates that the MFSNN has higher efficiency in antenna modelling and optimisation than EM-based

simulation and conventional ANN method, up to 70% time cost is saved, and the accuracy is comparable to the conventional ANN with 100% HF data. The MFSNN performs well in antenna optimisation. It can optimise the antenna based on the design specification at a lower cost than EM simulation-based method and conventional ANN method. However, a sampling method with high uniformity is needed to address the high randomness and low uniformity issues in LHS, which can improve the representativeness and quality of training data. Then the surrogate model testing accuracy can be further improved.

6.2 Future Work

This thesis investigates te ML-assisted antenna modelling methods for accelerating the antenna design, analysis and optimisation process. However, there are some of the issues that can be addressed to improve future research.

The process of ML-assisted antenna modelling briefly includes:

1. Problem formulation: considered variables and desired performance parameters.

2. Define antenna design space and simple candidates within design space.

3. Collect data from running EM simulation (CST, HFSS, FEKO) of considered antennas.

4. Preprocessing dataset: (a)dataset examination (errors and outliers); (b)training set and testing set; (c) data tagging (inputs and outputs); (d) data normalisation.

5. Build an ML model and apply training algorithms, and evaluate the accuracy of model according to the error between model outputs and actual values based on specific evaluation criteria.

6. Surrogate model applications (antenna optimisation, sensitivity analysis, tolerance analysis, etc.).

According to the process, some improvements should be made in future research in order to achieve better performance:

• A sampling method with higher uniformity is needed to address the high randomness and low uniformity issue of LHS, the sampling quality can significantly affect the surrogate accuracy and generalisation capability. In addition, the aforementioned shrinking issue in

sampling candidates as the dimension grows should be addressed. Otherwise the surrogate can not model antenna behaviours in a wide range of design space.

• Network topology selection and optimisation. The current research on ANN-based antenna design are uses trial-and-error method to construct a network. However, an inappropriate network structure may lead to under-fitting, over-fitting, extended training times. Therefore, an algorithm is required to construct ANNs correctly to address the mentioned issues.

• Adaptively learning rate tuning during training. The error of ANN will decrease once the hyper-parameter and network structure are properly set. However, for the situation where the error remains unchanged, the learning rate should be adjusted to avoid the waste of computational resources and time. Meanwhile, a self-correction mechanism is needed to address the increasing error once the learning rate is adjusted.

• A multi-objective antenna optimisation should be considered, exploring different trade-offs between objectives and select the best design can satisfy all requirements. Practical antenna optimisation should consider multiple conflicting objectives that need to be optimised simultaneously.

• A transfer learning based on the MF modelling approach is considered in future work to improve the surrogate model's generalization capability. The constructed model is expected to model and optimise the identical type antennas (dual-band, triple-band, etc.) based on the prior knowledge from the pre-train stage. Meanwhile, the MF modelling approach can reduce the time and computational budget in large dataset generation.

References

[1] A. Z. Zaki, E. K. Hamad, T. G. Abouelnaga, H. A. Elsadek, S. A. Khaleel, A. J. A. Al-Gbur and Z. Zakaria, “Design and modeling of ultra-compact wideband implantable antenna for wireless ISM band," *Bioengineering*, vol. 10, no. 2, p. 216, 2023.

[2] S. Ahmad, B. Manzoor, K. N. Paracha, S. Haider, M. Liaqat, A. J. A. Al-Gburi, A. Ghaffar, M. Alibakhshikenari and M. Dalarsson, “A wideband bear-shaped compact size implantable antenna for in-body communications," *Appl. Sci.*, vol. 12, no. 6, p.2859, 2022.

[3] S. S. Singhwal, L. Matekovits, I. Peter and B. K. Kanaujia, “A study on application of dielectric resonator antenna in implantable medical devices," *IEEE Access*, no. 10, pp.11846-11857, 2022.

[4] K. Sharma, D. K. Upadhyay, H. Parthasarathy, and R. Gurjar, “Analysis and design of liquid antenna," *Int. J. RF. Microw. C. E.*, vol. 32, no. 5, p.e23101, 2022.

[5] J. O. Martínez, J. R. Rodríguez, Y. Shen, K. F. Tong, K. K. Wong and A.. G.Armada, “Toward liquid reconfigurable antenna arrays for wireless communications," *IEEE Commun. Mag.*, vol. 60, no. 12, pp.145-151, 2022.

[6] Y. Tawk and J. Costantine, “A polarization reconfigurable 3D printed dual integrated quadrifilar helix antenna array embedded within a cylindrical dielectric mesh," *IEEE Trans. Antennas Propag.*, vol. 71, no. 1, pp. 309-317, 2023.

[7] Z. X. Xia and K. W. Leung, “3-D-printed wideband circularly polarized dielectric resonator antenna with two printing materials," *IEEE Trans. Antennas Propag.*, vol.70, no. 7, pp. 5971-5976, 2022.

[8] T. Alam and M. Cheffena, “Integrated microwave antenna/sensor for sensing and communication applications," *IEEE Trans. Microw. Theory Tech.*, vol. 70, no. 11, pp. 5289-5300, 2022.

[9] S. Jayant, G. Srivastava and S. Kumar, “Quad-port UWB MIMO footwear antenna for wearable applications," *IEEE Trans. Antennas Propag.*, vol. 70, no. 9, pp. 7905-7913, 2022

[10] P. Sumithra and D. Thiripurasundari, “Review on computational electromagnetics," *Adv. Electronmagn.*, vol. 6, no. 1, pp. 42-55, 2017.

[11] R. F. Harrington, *Field computation by moment methods*, Wiley - IEEE Press, 1993.

[12] K. Yee, “Numerical solution of initial boundary value problems involving Maxwell’s equations in isotropic media," *IEEE Trans. Antennas Propag.*, vol. 14, pp. 302-307, 1966.

[13] P. A. Tirkas and C. A. Balanis, “Finite-different time-domain method for antenna radiation," *IEEE Trans. Antenna Propag.*, vol. 40, no. 3, pp. 334-340, 1992.

[14] N. Kaneda, B. Houshmand and T. Itoh, “FDTD analysis of dielectric resonators with curved surfaces," *IEEE Trans. Microw. Theory Tech.*, vol. 45, no. 9, pp. 1645-1649, 1997.

[15] J. L. Volakis, A. Chatterjee and L. C. Kempel, *Finite element method for electromagnetics: with applications to antennas, microwave circuits, and scattering*, IEEE Press, 1998.

[16] O. C. Zienkiewicz, R. L. Taylor and Z. Z. Jian, *The finite element methond: its basis and fundamentals*, Elsevier, 2005.

[17] R. W. P. King, *Transmission-line Theory*, New York: Dover, 1965.

[18] N. Y. Bliznyuk and A. I. Nosich, “Numerical analysis of lossy circular microstrip antenna," *Telecommun. Radio Eng.*, vol.55, no. 8, pp. 15-23, 2001.

[19] K. Q. da Costa, V. A. Dmitriev and L. D. F. de Carvalho, “Numerical analysis by method of moment (MoM) of a rectangular monopole antenna with parasitic loop elements," in *Proc. 11st URSI Commission F Open Symposium on Radio Wave Propagation and Remote Sensing, Rio de Janeiro*, 2007.

[20] L. Wang, Q. Chen, Q. Yuan and K. Sawaya, “Numerical analysis on MIMO performance of the modulated scattering antenna array in indoor environment," *IEICE Trans. Commun.*, vol. 94, no. 6, pp. 1752-1756, 2011.

[21] W. G. Scanlon and N. E. Evans, “Numerical analysis of bodyworn UHF antenna systems," *Electron. Commun. Eng. J.*, vol. 13, no. 2, pp. 53-64, 2001.

[22] T. Ando, J. Yamauchi and H. Nakano, “Numerical analysis of a dielectric rod antenna-demonstration of the discontinuity-radiation concept," *IEEE Trans. Antennas Propag.*, vol. 51, no. 8, pp. 2007-2013, 2003.

[23] A. Eldek, “Numerical analysis of a small ultra-wideband microstrip-fed tap monopole antenna," *Prog. Electromagn. Res.*, vol.65, pp. 59-69, 2006.

[24] A. Erentok, D. Lee and R. W. Ziolkowski, “Numerical analysis of a printed dipole antenna integrated with a 3-D AMC block," *IEEE Antennas Wirel. Propag. Lett.*, vol. 6, pp. 134-136, 2007.

[25] S. Sah, “Machine learning: a review of learning types," *Preprints*, 2020.

[26] E. I. N. Issam and M. J. Murphy, “What is machine learning?," *Machine Learning in Radiation Oncology*, pp. 3-11, 2015.

[27] X. Wang, Z. Lei, X. Zhang, B. Zhou and J. Peng, “Machine learning basics,"*Deep Learning*, pp. 98-164, 2016.

[28] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning*, Springer Science & Business Media, 2011.

[29] J. McCarthy and E. A. Fergenbaum, “In memoriam: Arthur Samuel: Pioneer in machine learning," *AI Magazine*, vol. 11, no. 3, p. 10, 1990.

[30] E. A. Fergenbaum, “Roots of knowledge systems and expert system application", *Artificial Intelligence in Perspective*, vol. 59, pp. 233-240.

[31] V. N. Vapnik, *Statistical learning theory*, Wiley, New York, 1998.

[32] S. Suthaharan, “Support vector machine," *Machine learning models and algorithms for big data classification*, pp. 207-235, 2016.

[33] T. Hofmann, B. Schölkopf and A. J. Smola, “Kernel methods in machine learning," *Ann. Stat.*, vol. 36, no. 3, pp. 1171-1220, 2008.

[34] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*, MIT Press, 2016.

[35] Y. Anzai, *Pattern recognition and machine learning*, Elsevier, 2012.

[36] N. Kriegeskorte, “Deep neural networks: a new framework for modelling biological vision and brain information processing," *Annu. Rev. Vis. Sci.*, vol. 1, pp. 417-446, 2015.

[37] A. L. Fradkov, “Early history of machine learning,” *IFAC-PaperOnLine*, vol. 53, no. 2, pp. 1385-1390, 2020.

[38] M. Daily, S. Medasani, R. Behringer and M. Trivedi, “Self-driving cars," *Computer*, vol. 50, no. 12, pp. 18-23, 2017.

[39] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi and E. Mayer, “Applying natural language processing and machine learning techniques to patient experience feedback: A systematic review," *BMJ Health Care Inform.*, vol. 28, no. 1, 2021.

[40] H. Wechsler, J. P. Phillips, V. Bruce, F. F. Soulie and T. S. Huang, *Face recognition: From theory to applications*, Springer Science & Business Media, 2012.

[41] A. Ganapathiraju, J. E. Hamaker, and J. Picone, “Applications of support vector machines to speech recognition," *IEEE Trans. Signal. Process.*, vol. 52, pp. 2348-2356, 2004.

[42] T. Jiang, J. Gradus and A. J . Rosellini, “Supervised machine learning: a brief primer," *Behav. Ther.*, vol. 51, no. 5, pp. 675-687, 2020.

[43] X. Zhu and A. B. Goldberg, “Introduction to semi-supervised learning," *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 3, no. 1, pp. 1-130, 2009.

[44] R. Gentleman and V. J. Carey, “Unsupervised machine learning," *Bioconductor Case Studies*, pp. 137-158, 2008.

[45] L. P. Kaelbling, M. L. Littman and A. W. Moore, “Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237-285, 1996.

[46] K. Weiss, T. M. Khoshgoftaar and D. Wang, “A survey of transfer learning," *J. Big Data*, vol. 4, no. 1, pp. 1-40, 2016.

[47] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein and M. Klein, *Logistic Regression*, New York: Springer-Verlag, 2002.

[48] M. Hegland, “The apriori algorithm - a tutorial," in *Mathematics Computation in Imaging Science and Information Processing*, pp. 209-262, 2007.

[49] D. Steinley, “K-means clustering: a half-century synthesis, " *Br. J. Math. Stat. Psychol.*, vol. 59, no. 1, pp. 1-34, 2006.

[50] A. Subramanya and P. P. Talukdar, “Graph-based semi-supervised learning," *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 8, no. 4, pp. 1-125, 2014.

[51] L. Gómez-Chova, G. Camps-Valls, J. Munoz-Mari and J. Calpe, “Semisupervised image classification with Laplacian support vector machines," *IEEE Geosci. Remote Sensing Lett.*, vol. 5, no. 3, pp. 336-340, 2008.

[52] C. J. Watkins and P. Dayan, “Q-learning," *Machine learing*, vol. 8, no. 3, pp. 279-292, 1992.

[53] G. Tesauro, “Practical issues in temporal difference learning," *Adv. Neural Inf. Process Syst.*, vol.4, pp. 259-266, 1991.

[54] P. Sibi, S. A. Jones and P. Siddarth, “Analysis of different activation functions using back propagation neural networks," *J. Theor. Appl. Inf. Technol.*, vol. 47, no. 3, pp. 1264-1268, 2013.

[55] A. A. Minai and R. D. Williams, “On the derivatives of the sigmoid," *Neural Networks*, vol. 6, no. 6, pp. 845-853, 1993.

[56] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, “Activation functions in deep learning: a comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92-108, 2022.

[57] M. Buscema, “Back propagation neural networks," *Subst. Use Misuse*, vol. 33, no. 2, pp. 233-270, 1998.

[58] Y. Chauvin, D. E. Rumelhart, *Backpropagation: Theory, Architectures, and Applications*, Psychology Press, 2013.

[59] F. J. Pineda, “Generalization of back-propagation to recurrent neural networks," *Physical Review Lett.*, vol. 59, no. 19, pp. 2229-2232, 1987.

[60] K. Hornik, M. Stinchcombe and H. White, “Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, 359-366, 1989.

[61] B. E. Boser, I. M. Guyon and V. N. Vapnik, “A training algorithm for optimal margin classifiers," In *Proc. ACM 5th Workshop Computat. Lern. Theory*, pp. 144-152, 1992.

[62] D. E. Goldberg, *Genetic algorithm in search, optimization and machine learning*, Addision-Wesley, Boston, MA, 1989.

[63] J. Kennedy and R. Eberhart, “Particle swarm optimization," in *Proc. 1995 Int. Conf. Neural Networks*, vol. 4, Perth, Australia, pp. 1942-1948, 1995.

[64] E. A. Grimaldi, F. Grimaccia, M. Mussetta and R. E. Zich, “Pso as an effective learning algorithm for neural network application," In *Proc. ICCEA 3rd Int. Conf. Computat. Electromagn. Appl.*, pp. 557-560, 2004.

[65] I. Vilović, N. Burum and M. Brailo, “Microstrip antenna design using neural networks optimized by PSO," In *21st ICECom Int. Conf.*, pp. 1-4, 2013.

[66] D. F. Cook, C. T. Ragsdale and R. L. Major, “Combining a neural network with a genetic algorithm for process parameter optimisation," *Eng. Appl. Artif. Intell.*, vol. 13, pp. 391-396, 2000.

[67] Y. Yao, L. Rosasco and A. Caponnetto, “On early stopping in gradient descent learning," *Constr. Approx.*, vol. 26, no. 2, pp. 289-315, 2007.

[68] M. Y. Park and T. Hastie, “L1-regularization path algorithm for generalized linear models," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 69, no. 4, pp. 659-677, 2007.

[69] T. V. Laarhoven, “L2 regularization versus batch and weight normalization," *arXiv preprint arXiv: 1706.05350*.

[70] P. Baldi and P. J. Sadowski, “Understanding dropout," *Adv. Neural Inf. Process syst.*, vol. 26, pp. 2814-2822, 2013.

[71] A. G. Wilson , *Covariance Kernels for Fast Automatic Pattern Discovery and Extrapolation with Gaussian Processes*, University of Cambridge Ph.D. thesis, 2014 .

[72] D. Duvenaud, J. R. Lloyd, R. Grosse, J. B. Tenenbaum and Z. Ghahramani, “Structure discovery in nonparametric regression through compositional kernel search," *arXiv preprint arXiv: 1302.4922*, 2013 .

[73] M. Mohri, A. Rostamizadeh and A. Talwalkar, *Foundations of Machine Learning*. The MIT Press, 2012.

[74] S. R. Gunn, “Support vector machines for classification and regression,"*SISI Tech Report*, p. 6459, 1998.

[75] A. J. Smola and Scholkopf, “A tutorial on support vector regression," *Statistics Comput.*, vol. 14, pp. 199-222, 2004.

[76] O. Chapelle, V. Vapnik and O. Bousquet, “Choosing kernel parameters for support vector machines," *Machine Learn*, vol. 46, pp. 131-161, 2002.

[77] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, The MIT Press, 2016.

[78] M. Warren and W. Pitts, “A logical calculus of ideas immanent in nervous activity," *Bull. Math. Biol.*, vol. 5, no. 4, pp. 115-133, 1943.

[79] L. Vegni and A. Toscano, “Analysis of microstrip antennas using neural networks," *IEEE Trans. Mag.*, vol. 33, no. 2, pp. 1414-1419, 1997.

[80] R. K. Mishra and A. Patnaik, “Design of circular microstrip antenna using neural networks," *IETE J. Res.*, vol. 44, no. 1-2, pp. 35-39, 1998.

[81] R. K. Mishra and A.Patnaik, “Neural network-based CAD models for the design of square-patch antennas," *IEEE Trans. Antennas Propag.*, vol. 46, no. 12, pp. 1890-1891, 1998.

[82] S. Devi, D. C. Panda and S. S. Pattnaik, “A novel method of using artificial neural networks to calculate input impedance of circular microstrip antenna," In *IEEE Antenans Propag. Soc. Int. Symp.*, vol. 3, pp. 462-465, 2002.

[83] K. Guney and N. Sankaya, “Artificial neural networks for calculating the input resistance of circular microstrip antennas," *Microw. Opt. Technol. Lett.*, vol. 37, no. 2, pp. 107-111, 2003.

[84] N. Türker, F. Güneş and T. Yildirim, “Artificial neural design of microstrip antennas," *Tur. J. Elec. Eng. & Comp. Sci.*, vol. 14, no. 3, pp. 445-453, 2006.

[85] J. L. Narayana, K. S. Krishna and L. P. Reddy, “Design of microstrip antennas using artificial neural networks," In *Int. Conf. Computat. Intell. Multimed. Appl.*, vol .1, pp. 332-334, 2007.

[86] Z. Wang, S. Fang, Q. Wang and H. Liu, “An ann-based synthesis model for the single-feed circularly-polarized square microstrip antenna with truncated corners," *IEEE Trans. Antennas Propag.*, vol. 60, no. 12, pp. 5989-5992, 2012.

[87] Y. Kim, S. Keely, J. Ghosh and H. Ling, “Application of artificial neural networks to broadband antenna design based on a parametric frequency model," *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 669-674, 2007.

[88] V. V. Thakare and P. Singhal, “Artificial intelligence in the estimation of patch dimensions of rectangular microstrip antennas," *Circuits Syst.*, vol. 2, no. 4, p. 330, 2011.

[89] P. P. Bhagat, D. Pujara and D. Adhyaru, “Analysis and synthesis of microstrip patch antenna using artificial neural networks," In *IEEE Asia-Pacific Conf. Antennas Propag.*, pp.120-121, 2012.

[90] P. Robustillo, J. Zapata, J. A. Encinar and M. Arrebola, “Design of a contoured-beam reflectarray for a Eutelsat European coverage using a stacked-patch element characterized by an artificial neural network," *IEEE Antennas Wirel. Propag. Lett.*, pp. 977-980, 2012.

[91] A. Freni, M. Mussetta and P. Pirinoli, “Neural network characterization of reflectarray antennas," *Int. J. Antennas Propag.*, 2012.

[92] P. Robustillo, J. Zapata, J. A. Encinar and J. Rubio, “ANN characterization of multi-layer reflectarray elements for contoured-beam space antennas in the Ku-band," *IEEE Trans. Antennas Propag.*, vol. 60, no. 7, pp. 3205-3214, 2012.

[93] F. Güne¸s, S. Nesil, and S. Demirel, “Design and analysis of Minkowski reflectarray antenna using 3-D CST Microwave Studio-based neural network model with particle swarm optimization,” *Int. J. RF Microw. C. E.*, vol. 23, no. 2, pp. 272-284, 2013.

[94] A. Kayabaşı and A. Akdağlı, “An application of ann model with Bayesian regularization learning algorithm for computing the operating frequency of C-shaped patch antennas," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 1, no. 5, pp. 1-5, 2016.

[95] L.-Y. Xiao, W. Shao, F.-L. Jin, and B.-Z. Wang, “Multiparameter modeling with ann for antenna design," *IEEE Trans. Antennas Propag.*, vol. 66, no. 7, pp. 3718-3723, 2018.

[96] Z. Ž. Stanković, D. I. Olćan, N. S. Dončov, and B. M. Kolundžija, “Consensus deep neural networks for antenna design and optimization," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 5015-5023, 2021.

[97] Y. F. Liu, L. Peng, and W. Shao, “An efficient knowledge-based artificial neural network for the design of circularly polarized 3-D-printed lens antenna," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 5007-5014, 2022.

[98] A. Kayabaşı and A. Akdağlı, “An application of ann model with Bayesian regularization learning algorithm for computing the operating frequency of C-shaped patch antennas," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 1, no. 5, pp. 1-5, 2016.

[99] J. Meng and L. Xia, “Support vector regression model for millimetre wave transitions," *Int. J. Infrared Millimeter Waves*, vol. 28, no. 5, pp. 413-421, 2007.

[100] F. Güneş, N. T. Tokan and F. Gürgen, “Support vector design of the microstrip lines," *Int. J. RF Microwave Comput-Aided. Eng.*, vol. 18, no. 4, pp. 326-336, 2008.

[101] N. T. Tokan and F. Gunes, “Support vector characterization of the microstrip antennas based on measurement," *Prog. Electromagn. Res. B*, vol. 5, pp. 49-61, 2008.

[102] Z. Zheng, X. Chen and K. Huang, “Application of support vector machines to the antenna design," *Int. J. RF Microw. Comput-Aided. Eng.*, vol. 21, no. 1, pp. 85-90, 2011.

[103] J. P. Jacobs, S. Koziel and S. Ogurtsov, “Computationally efficient multi-fidelity Bayesian support vector regression modeling of planar antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 61, no. 2, pp. 980-984, 2012.

[104] C. Roy, T. Khan and B. K. Kanaujia, “Performance parameters prediction of slotted microstrip antennas with modified ground plane using support vector machine.” *Int. J. Microw. Wirel. Technol.*, vol. 9, no. 5, pp. 1169-1177, 2017.

[105] Q. Wu, H. Wang and W. Hong, “Broadband millimeter-wave SIW cavity-backed slot antenna for 5g applications using machine-learning-assisted optimization method,” in *Proc. 2019 Int. Workshop Antenna Technol. (iWAT)*, Miami, FL, USA, pp. 9-12, 2019.

[106] R. G. Ayestaran, M. F. Campillo and F. Las-Heras, “Multiple support vector regression for antenna array characterization and synthesis," *IEEE Trans. Antennas Propag.*, vol. 55, no. 9, pp. 2495-2501, 2007.

[107] D. R. Prado, J. A. López-Fernández, G. Barquero, M. Arrebola, and F. Las-Heras, “Fast and accurate modeling of dual-polarized reflectarray unit cells using support vector machines,” *IEEE Trans. Antennas Propag.*, vol. 66, no. 3, pp. 1258–1270, 2018.

[108] D. R. Prado, J. A. López-Fernández, M. Arrebola and G. Goussetis, “Support vector regression to accelerate design and crosspolar optimization of shaped-beam reflectarray antennas for space applications," *IEEE Trans. Antennas Propag.*, vol. 67, no. 3, pp. 1659-1668, 2018.

[109] D. R. Prado, J. A. López-Fernández, M. Arrebola and M. R. Pino, “Wideband shaped-beam reflectarray design using support vector regression analysis," *IEEE Antennas Wirel. Lett.*, vol. 18, no. 11, pp. 2287-2291, 2019.

[110] M. E. Yiğit, G. Ö. Günel, M. E. Aydemir and t. Günel, “Soft computing approach to design a triple-band slotted microstrip patch antenna," *Appl. Sci.*, vol. 12, no. 23, p.11923, 2022.

[111] D. R. Prado, J. A. Lopez-Fernandez, M. Arrebola, and G. Goussetis, “On the use of the angle of incidence in support vector regression surrogate models for practical reflectarray design,” *IEEE Trans. Antennas Propag.*, vol. 69, no. 3, pp. 1787–1792, 2021.

[112] L. Shi, Q. Zhang, S. Zhang, G. Liu, and C. Yi, “Accurate characterization of graphene reconfigurable reflectarray antenna element by SVR,” *IEEE J. Multiscale Multiphys. Comput. Techn.*, vol. 6, pp. 50–55, 2021.

[113] D. R. Prado, P. Naseri, J. A. López-Fernández, S. V. Hum and M. Arrebola, “Support Vector Regression-Enabled Optimization Strategy of Dual Circularly-Polarized Shaped-Beam Reflectarray with Improved Cross-Polarization Performance," *IEEE Trans. Antennas Propag.*, vol. 71, no. 1, pp. 497-507, 2023.

[114] J. P. de Villiers and J. P. Jacobs, “Gaussian process modeling of CPW-fed slot antennas," *Prog. Electromagn. Res.*, vol. 98, pp. 233-249, 2009.

[115] J. P. Jacobs and J. P. de Villiers, “Modeling of planar dual-band microstrip patch antenna using Gaussian process regression," In *The 3rd European Wirel. Technol. Conf.*, pp. 253-256, 2010.

[116] J. P. Jacobs and J. P. de Villiers, “Gaussian-process-regression-based design of ultrawide-band and dual-band CPW-fed slot antennas," *J. Electromagn. Waves Appl.*, vol. 24, no. 13, pp. 1763-1772, 2010.

[117] J. P. Jacobs and S. Koziel, “Two-stage framework for efficient Gaussian process modeling of antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 62, no. 2, pp. 706-713, 2013.

[118] J. P. Jacobs, “Efficient resonant frequency modelling for dual-band microstrip antennas by Gaussian process regression," *IEEE Antennas Wirel. Propag. Lett.*, vol. 14, pp. 337-341, 2014.

[119] J. P. Jacobs, “Characterisation by Gaussian processes of finite substrate size effects on gain patterns of microstrip antennas," *IET Microw. Antennas Propag.*, vol. 10, no. 11, pp. 1189-1195, 2016.

[120] M. Papadrakakis, V. Papadopoulos, G. Stefanou and V. Plevris, “Surrogate modeling of antenna radiation characteristics by Gaussian processes,", *VII European Comput. Methods Appl. Sci. Eng.*, pp. 5-10, 2016.

[121] Q. Wu, H. Wang and W. Hong, “Multistage collaborative machine learning and its application to antenna modeling and optimization," *IEEE Trans. Antennas Propag.*, vol. 68, no. 5, pp. 3397-3409, 2020.

[122] K. Sharma and G. P. Pandey, “Efficient modelling of compact microstrip antenna using machine learning," *AEU-Int. J. Electron. Commun.*, vol. 135, p. 153739, 2021.

[123] D. Ustun and F. Tokatas, “Surrogate-based computational analysis and design for H-shaped microstrip antenna," *J. Electromagn. Waves Appl.*, vol. 35, no. 1, pp. 71-82, 2021.

[124] W. Chen, Q. Wu, C. Yu, H. Wang and W. Hong, “Multibranch machine learning-assisted optimization and its application to antenna design," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 4985-4996, 2022.

[125] A. Pietrenko-Dabrowska, S. Koziel and L. Golunski, “Two-stage variable-fidelity modeling of antennas with domain confinement," *Scientific Reports*, vol. 12, no. 1, p. 17275, 2022.

[126] S. Koziel and A. Pietrenko-Dabrowska, “Expedited variable-resolution surrogate modeling of miniaturized microwave passives in confined domains," *IEEE Trans. Microw. Theory Tech.*, vol. 70, no. 11, pp.4740-4750, 2022.

[127] M. Kennedy and A. O’Hagan, “Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, vol. 87, no. 1, pp. 1-13, 2000.

[128] A. I. Forrester, A. Sobester, and A. J. Keane, “Multi-fidelity optimization via surrogate modeling,” *Proc. Royal Society*, vol. 463, pp. 3251–3269, 2007.

[129] K. Geiselhart, L. Ozoroski, J. Fenbert, E. Shields and W. Li, “Integration of multifidelity multi- disciplinary computer codes for design and analysis of supersonic aircraft," in *49th AIAA Aerospace Science Meeting Including the New Horizons Forum and Aerospace Exposition*, p. 465, 2011.

[130] S. Alestra, E. Kapushev, M. Belyaev, E. Burnaev, M. Dormieux, A. Cavailles, D. Chaillot and E. Ferreira, “Building data fusion surrogate models for spacecraft aerodynamic problems with incomplete factorial design of experiments," *Adv. Mater. Res.*, vol. 1016, pp. 405-412, 2014.

[131] A.Zaytsev, “Reliable surrogate modeling of engineering data with more than two levels of fidelity," in *Proc. 7th Int. IEEE Conf. Mechanical Aerosp. Eng.*, pp. 341-345, 2016.

[132] S. Koziel, “Adaptively adjusted design specifications for efficient optimization of microwave structures," *Prog. Electromagn. Res. B*, vol. 21, pp. 219-234, 2010.

[133] Y. Song, Q. S. Cheng and S. Koziel, “Multi-fidelity local surrogate model for computationally efficient microwave component design optimization," *Sensors*, vol. 19, no. 13, p, 3023, 2019.

[134] H. Babaee, P. Perdikaris, C. Chryssostomidis and G. E. Karniadakis, “Multi-fidelity modelling of mixed convection based on experimental correlations and numerical simulations,” *J. Fluid Mech.*, vol. 809, pp. 895-917, 2016.

[135] P. Perdikaris, M. Raissi, A. Damianou, N. Lawrence and G. E. Karniadakis, “Nonlinear information fusion algorithm for data-efficient multi-fidelity modelling,” in *Proc. Math. Phys. Eng. Sci.*, vol. 473, no. 2198, p. 20160751, 2017.

[136] A. Zaytsev and E. Burnaev, “Large scale variable fidelity surrogate modelling," *Ann. Math. Artif. Intell.*, vol. 81, no. 1, pp. 167-186, 2017.

[137] C. Park, R. T. Haftka and N. H. Kim, “Remarks on multi-fidelity surrogates," *Struct. Multidiscipl. Optim.*, vol. 55, no. 3, pp. 1029-1050, 2017.

[138] J. Song, Y. Chen and Y. Yue, “A general framework for multi-fidelity Bayesian optimization with Gaussian processes," in *The 22nd Int. Conf. Artif. Intell. Statist.*, pp. 3158-3167, 2019.

[139] S. Koziel and S. Ogurtsov, “Model management for cost-efficient surrogate-based optimization of antennas using variable-fidelity electromagnetic simulations,” *IET Microw. Antennas Propag.*, vol. 6, no. 15, pp. 1643–1650, 2012.

[140] S. Koziel, S. Ogurtsov, I. Couckuyt and T. Dhaene, “Variable-fidelity electromagnetic simulations and co-kriging for accurate modeling of antennas,” *IEEE Trans. Antennas Propag.*, vol. 61, no. 3, pp. 1301–1308, 2013.

[141] S. Koziel, S. Ogurtsov, “Multi-objective design of antennas using variable-fidelity simulations and surrogate models,” *IEEE Trans. Antennas Propag.*, vol. 61, no. 12, pp. 5931–5939, 2013.

[142] S. Koziel, A. Bekasiewicz, I. Couckuyt, and T. Dhaene, “Efficient multi-objective simulation-driven antenna design using co-kriging,” *IEEE Trans. Antennas Propag.*, vol. 62, no. 11, pp. 5900–5905, 2014.

[143] S. Koziel, A. Pietrenko-Dabrowska and A. G. Raef, “Knowledge-based expedited parameter tuning of microwave passives by means of design requirement management and variable-resolution EM simulations," *Scientific Reports*, vol. 13, no. 1, p. 334, 2023.

[144] J. Laurenceau and P. Sagaut, “Building efficient response surfaces of aerodynamic functions with kriging and cokriging, " *AIAA J.*, vol. 46, no. 2, pp. 498-507, 2008.

[145] M. Tejaswini and M. Sivapragasam, “ Multi-objective design optimization of turbine blade leading edge for enhanced aerothermal performance," *Sadhana-Acad. P. Eng. S.*, vol. 46, no. 4, p. 190, 2021.

[146] J. C. Helton, J. D. Johnson, C. J. Sallaberry and C. B. Storlie, “Survey of sampling-based methods for uncertainty and sensitivity analysis," *SANDIA REPORT*, Sandia National Laboratories, 2006.

[147] R. Zhang and Z. Wang, “The theoretical and numerical analysis of computer experiments," *Chinese J. Appl. Probab. Statist.*, no. 4, pp. 420-435 (in Chinese), 1994.

[148] J. P. C. Kleijnen, “An overview of the design and analysis of simulation experiments for sensitivity analysis," *Eur. J. Oper. Res.*, vol. 164, no. 2, pp. 287-300, 2005.

[149] M. D. McKay, “Latin hypercube sampling as a tool in uncertainty analysis of computer models," In *Proc. 24th Conf. Winter Simul.*, pp. 557-564, 1992.

[150] Z. Zhang, H. Chen and Q. S. Cheng, “Surrogate-assisted quasi-newton enhanced global optimization of antennas based on heuristic hypersphere sampling," *IEEE Trans. Antennas Propag.*, vol. 69, no. 5, pp. 2993-2998, 2020.

[151] Z. Zhang, F. Jiang, Y. Jiao and Q. S. Cheng, “Low-cost surrogate modelling of antennas using two-level Gaussian process regression mothod," *Int. J. Numer. Model.*, p. e2886, 2020.

[152] C. Chen, J. Long, W. Chen, Z. Liu and J. Guo, “Modeling and prediction of spindle dynamic precision using the Kriging-based response surface method with a novel sampling strategy," *Nonlinear Dyn.*, vol, 111, no. 1. pp. 559-579, 2023.

[153] J. G. Andrews et al., “What will 5G be?” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[154] N. Bhushan et al., “Network densification: the dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82-89, 2014.

[155] K. M. Shams and M. Ali, “Wireless power transmission to a buried sensor in concrete," *IEEE Sensors J.*, vol. 7, no. 12, pp. 1573-1577, 2007.

[156] M. F. Rad and L. Shafai, “Embedded microstrip patch antenna for structural health monitoring applications," *IEEE Antennas Propag. Soc. Int. Symp. San Diego, CA*, pp. 1-4, 2008.

[157] G. S. Mauro, G. Castorina, A. F. Morabito, L. Di Donato and G. Sorbello, “Effects of lossy background and rebars on antennas embedded in concrete structures," *Microw. Opt. Technol. Lett.*, vol. 58, no. 11, pp. 2653-2656, 2016.

[158] G. Castorina, L. Di Donato, A. F. Morabito, T. Isemia and G. Sorbello, “Analysis and design of a concrete embedded antenna for wireless monitoring applications [antenna applications corner]," *IEEE Antennas Propag. Mag.*, vol. 58, no. 6, pp. 76-93, 2016.

[159] S. H. Jeong and H. W. Son, “UHF RFID tag antenna for embedded use in a concrete floor," *IEEE Antennas Wirel. Propag. Lett.*, no. 10, pp. 1158-1161, 2011.

[160] C. S. You, W. Hwang, H. C. Park, R. M. Lee and W. S. Park, “"Microstrip antenna for SAR application with composite sandwich construction: surface-antenna-structure demonstration," *J. Compos. Mater.*, vol. 37, no. 4, pp. 315-364, 2003.

[161] C. You, M. M. Tentzeris and W. Hwang, “Multilayer effects on microstrip antennas for their integration with mechanical structures," *IEEE Trans. Antennas Propag.*, vol. 55, no. 4, pp. 1051-1058, 2007.

[162] J. Zhou, J. Huang, L. Song, D. Zhang and Y. Ma, “Electromechanical co-design and experiment of structurally integrated antenna," *Smart Mater. Struct.*, vol. 24, no. 3, p. 037004, 2015.

[163] X. Jin and M. Ali, “Simple empirical formulas to estimate the dielectric constant and conductivity of concrete," *Microw. Opt. Technol. Lett.*, vol. 61, no. 2, pp. 386-390, 2019.

[164] Q. Wu, H. Wang and W. Hong, “Machine-learning-assisted optimization and its application to antenna designs: opportunities and challenges," *China Commun.*, vol. 17, no. 4, pp. 152-164, 2020.

[165] S. Sagiroglu and A. Kalinli, “Determining resonant frequencies of various microstrip antennas within a single neural model trained using parallel tabu search algorithm,” *Electromagnetics.*,, vol. 25, pp. 551-565, 2005.

[166] E. A. Grimaldi, F. Grimaccia, M. Mussetta and R. E. Zich, “PSO as an effective learning algorithm for neural network applications,” in *Proc. 3rd Int. Conf. Comput. Electromagn. Appl.*, pp. 557-560, 2004.

[167] H. Moayedi, A. Moatamediyan, H. Nguyen, X. Bui, D. T. Bui and A. S. A. Rashid, “Prediction of ultimate bearing capacity through various novel evolutionary and neural network models,” *Eng. Comput.*, vol. 36, no. 2, pp. 671-687, 2020.

[168] D. F. Cook, C. T. Ragsdale and R. L. Major, “Combining a neural network with a genetic algorithm for process parameter optimization,” *Eng. Appl Artif. Intell.*, vol. 13, pp. 391-396, 2000.

[169] D. J. MacKay, “A practical bayesian framework for backpropagation networks,” *Neural Computat.*, vol. 4, no. 3, pp. 448-472, 1992.

[170] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phy.*, no. 378, pp. 686-707, 2019.

[171] X. H. Meng and G .E. Karniadakis, “A composite neural network that learns from multi-fidelity data: application to function approximation and inverse PDE problems," *J. Comput. Phy.*, vol. 401, p. 109020, 2020.

[172] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis, “Nonlinear information fusion algorithm for data-efficient multi-fidelity modelling," In *Proc. R. Soc. A*, vol. 473, p. 20160751, 2017.

[173] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13rd Int. Conf. Artif. Intell. Statist.*, vol. 9, pp. 249-256, 2010.

[174] J. Sirignano and K. Spiliopoulos, “Scaling limit of neural networks with the Xavier initialization and convergence to a global minimum," 2019, *arXiv: 1907.04108*, [Online]. Available: http://arxiv.org/abs/1907.04108

[175] Y. Shao and Z. Chen, “A design of dual-frequency dual-sense circularly-polarized slot antenna,", *IEEE Trans. Antennas Propag.*, vol. 60, no. 11, pp. 4992-4997, 2012.