



# Analysis of three-dimensional potential problems in non-homogeneous media with physics-informed deep collocation method using material transfer learning and sensitivity analysis

Hongwei Guo<sup>2,3,4</sup> · Xiaoying Zhuang<sup>2,3,4</sup> · Pengwan Chen<sup>5</sup> · Naif Alajlan<sup>1</sup> · Timon Rabczuk<sup>1</sup>

Received: 2 August 2021 / Accepted: 12 February 2022  
© The Author(s) 2022

## Abstract

In this work, we present a deep collocation method (DCM) for three-dimensional potential problems in non-homogeneous media. This approach utilizes a physics-informed neural network with material transfer learning reducing the solution of the non-homogeneous partial differential equations to an optimization problem. We tested different configurations of the physics-informed neural network including smooth activation functions, sampling methods for collocation points generation and combined optimizers. A material transfer learning technique is utilized for non-homogeneous media with different material gradations and parameters, which enhance the generality and robustness of the proposed method. In order to identify the most influential parameters of the network configuration, we carried out a global sensitivity analysis. Finally, we provide a convergence proof of our DCM. The approach is validated through several benchmark problems, also testing different material variations.

**Keywords** Deep learning · Collocation method · Potential problem · PDEs · Sampling method · Activation function · Non-homogeneous · Transfer learning · Sensitivity analysis · Physics-informed

## Nomenclature

$k(\mathbf{x})$	Position-oriented material function
$\phi$	Potential function
$q$	Flux of potential field
$\mathbf{n}$	Unit normal vector to a surface
$w_{jk}^l$	Weight between neuron $k$ in hidden layer $l - 1$ and neuron $j$ in hidden layer $l$
$b_j^l$	Bias of neuron $j$ in layer $l$
$\sigma$	Activation function

$\theta$	Hyperparameters including all weights and biases
Loss( $\theta$ )	Loss function for training
$\mathbf{x}_\Omega$	Collocation points to discretize the physical domain
$\mathbf{x}_\Gamma$	Collocation points to discretize the boundaries
MSE	Mean square error loss form
$\phi^h(\mathbf{x};\theta)$	Potential function approximated by Neural networks

✉ Timon Rabczuk  
timon.rabczuk@uni-weimar.de

Hongwei Guo  
hongwei.guo@iop.uni-hannover.de

Xiaoying Zhuang  
zhuang@iop.uni-hannover.de

Pengwan Chen  
pwchen@bit.edu.cn

Naif Alajlan  
najlan@ksu.edu.sa

<sup>1</sup> ALISR Laboratory, College of Computer and Information Sciences, King Saud University, PO Box 51178, Riyadh 11543, Saudi Arabia

<sup>2</sup> Computational Science and Simulation Technology, Institute of Photonics, Leibniz Universität Hannover, Appelstr. 11, 30167 Hannover, Germany

<sup>3</sup> Department of Geotechnical Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, People's Republic of China

<sup>4</sup> Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education, Tongji University, 1239 Siping Road, Shanghai 200092, People's Republic of China

<sup>5</sup> State Key Laboratory of Explosion Science and Technology, Beijing Institute of Technology, No. 5, South Street, Zhongguancun, Haidian District, Beijing 100081, People's Republic of China

$G(x)$	Governing equation
$\tilde{\phi}, \tilde{q}$	Potential field and flux prescribed at boundaries
$\eta_i$	Learning rate
$EE_i$	Elementary effect for each input factor
$\mu_i^*$	Mean of the distribution of the elementary effects of each input
$\sigma_i$	Standard deviation of the distribution of the elementary effects of each input
$\Lambda_j$	Spectral curve of the Fourier progression
$S_i^{\text{FAST}}$	First-order FAST sensitivity indices
$S_{T_i}$	Total order FAST sensitivity indices
$e$	Relative error to measure the model accuracy
$E_a$	Analytical solution
$\hat{E}_a$	Predicted solution
$\ \cdot\ $	$L_2$ -norm

## 1 Introduction

Recent years have witnessed the vast growing application of neural networks in physics, this is partly due to the fact that by training the neural network, high-dimensional raw data can be converted to low-dimensional codes [1], and thus the high-dimensional PDEs can be directly solved using a ‘meshfree’ deep learning algorithm, which improves computing efficiency and reduces the complexity of problems. The deep learning method deploys a deep neural network architecture with nonlinear activation functions that introduces the nonlinearity that the system as a whole needs for learning nonlinear patterns. This lends some credence to the application of a physics informed machine learning method in discovering the physics behind the potential problems in non-homogeneous media, which is a wide range of problems in physics and engineering.

The current wave of deep learning started around 2006, when Hinton et al. [2, 3] introduced deep belief nets and unsupervised learning procedures that could create layers of feature detectors without needs of labelled data. Equipped with deep learning model, information can be extracted from complicated raw input data with multiple levels of abstraction through a layer-by-layer process [4]. Various variants, such as multilayer perceptron (MLP), convolutional neural networks (CNN) and recurrent/recursive neural networks (RNN) [5], have been developed and applied to, e.g. image processing [6, 7], object detection [8, 9], speech recognition [10, 11], biology [12, 13] and even finance [14, 15]. Over the past decade, it has been widely used in applications due to high performance demonstrated. Deep learning can learn features from data automatically, and the features can be used to get the approximation of solutions to differential equations [16], which cast light on the possibility of using deep learning as functional approximators.

Artificial neural networks (ANN) stand at the center of the deep learning revolution, it can be traced back to the 1940s [17], but they became especially popular in the past few decades due to the vast development in computational power and sophisticated machine learning algorithms, such as backpropagation technique and advances in deep neural networks. Due to the simplicity and feasibility of ANNs to deal with nonlinear and multi-dimensional problems, they were applied in inference and identification by data scientists [18]. They were also adopted to solve partial differential equations (PDEs) [19–21], but shallow ANNs are unable to learn the complex nonlinear patterns effectively. With improved theories incorporating unsupervised pre-training, stacks of auto-encoder variants, and deep belief nets, deep learning with enhanced learning abilities can also serve as an interesting alternative to classical methods such as FEM.

According to the universal approximation theorem [22, 23], any continuous function can be approximated by a feedforward neural network with one single hidden layer. However, the number of neurons of the hidden layer tends to increase exponentially with increasing complexity and non-linearity of a model. Recent studies show that DNNs render better approximations for nonlinear functions [24]. Some researchers employed deep learning for the solution of PDEs. E et al. developed a deep learning-based numerical method for high-dimensional parabolic PDEs and back-forward stochastic differential equations [25, 26]. Raissi et al. [27] introduced physics-informed neural networks for supervised learning of nonlinear partial differential equations. Beck et al. [28] employed deep learning to solve nonlinear stochastic differential equations and Kolmogorov equations. Sirignano and Spiliopoulos [29] provided a theoretical proof for deep neural networks as PDE approximators, and concluded that they converge as the number of hidden layers tend to infinity. Karniadakis et al. presented physics-informed neural networks for various applications including fluid mechanics [30]. For problems in solid mechanics, we presented a Deep Collocation Method (DCM) in [31, 32], which has been the basis for a stochastic deep collocation method with neural architecture search strategy for stochastic flow analysis in heterogeneous media. We found that physics-informed deep learning model can account for stochastic disturbance/uncertainties efficiently and stably in [33]. An alternative to physics informed neural networks based on the strong form, such as the DCM, the Deep Energy Method (DEM) [34–37] takes advantage of the total potential energy in the loss instead of a BVP.

The problems of potential represent a category of physical and engineering problems. For some physical parameters in potential problems, for example, heat conductivity, permeability, permittivity, resistivity, magnetic permeability, tends to have a spatial distribution, and they can vary with respect

to one or more coordinates. To deal with these problems, the non-homogeneous problems are translated into homogeneous problems with some classes of material variations. The steady-state heat conduction analysis of FGMs analysis is a representative of potential problems. Due to the inherent mathematical difficulties, closed-form solutions exist in a few simple cases. Some traditional powerful methods, such as the finite element method (FEM), the boundary element method (BEM), and the method of fundamental solutions (MFS) and the dual reciprocity method (DRM) were used to solve the potential problems [38, 39]. The ‘meshfree’ physics-informed neural networks offered a novel and robust approach in discovering the nonlinear patterns behind the potential patterns, especially for higher dimensions.

The learning ability of deep neural networks strongly relies on the optimization algorithm and the neural network configurations, such as the activation function, number of neurons and layers, weight initialization methods, number of iterations, and so on. In this paper, we therefore compare different parameters to offer suggestions on the choice of a favourable configuration for the physics-informed neural network. Moreover, to increase the generality and robustness of the physics-informed deep learning based collocation method, the material transfer learning technique is integrated in the model, which will reduce the computation costs for different material variation types and help to improve the numerical results. Further, to unveil those influencing parameters for the proposed model, a global sensitive analysis is supplemented in the paper, which will be instructive for setting up physics-informed neural networks.

The paper is organised as follows: First, the three-dimensional potential problem with in-homogeneous media is presented. Then we introduce the physics-informed deep learning-based collocation method, which includes the neural network architecture, activation functions, sampling methods, a convergence proof, the material transfer learning and sensitivity analysis. Subsequently, a sufficient survey of numerical examples is presented, which investigated different neural network configurations, material transfer learning

and model sensitivity analysis. Finally, the effectiveness of the deep learning method is demonstrated for solving three-dimensional potential problems in non-homogeneous media.

## 2 The governing equation for 3D problems of potential

The general partial differential equation for potential function  $\phi$  defined on a region  $\Omega$  bounded by surface  $\tau$ , with an outward normal  $\mathbf{n}$ , can be written as:

$$(k(\mathbf{x})\phi_{,i})_{,i} = k(\mathbf{x})\phi_{,ii} + k_{,i}(\mathbf{x})\phi_{,i} = 0, \tag{1}$$

where  $k$  is a position-oriented material function. Equation (1) is the field equation for a wide range of problems in physics and engineering, such as heat transfer, incompressible flow, gravity field, shaft torsion, electrostatics and magnetostatics, some of which are shown in Table 1 [40].

The Dirichlet  $\tau_D$  and Neumann boundary  $\tau_N$  conditions are given as:

$$\begin{aligned} \phi(\mathbf{x}, t) &= \bar{\phi}, \mathbf{x} \in \tau_D, \\ -k(\mathbf{x})\frac{\partial\phi(\mathbf{x}, t)}{\partial\mathbf{n}} &= \bar{q}, \mathbf{x} \in \tau_N, \end{aligned} \tag{2}$$

where  $\mathbf{n}$  is the unit outward normal to  $\tau_N$ . The material properties of functionally graded materials (FGMs) vary gradually in space. Classical variations of  $k(\mathbf{x})$  take the form  $k(\mathbf{x}) = k_0f(\mathbf{x})$ ,  $k_0$  denoting a reference value and  $f(\mathbf{x})$  is the material property variation function. Among the most common variation functions are the quadratic, exponential and trigonometric:

$$\begin{aligned} \text{Parabolic} : f(\mathbf{x}) &= (a_1 + a_2\mathbf{x})^2 \\ \text{Exponential} : f(\mathbf{x}) &= (a_1e^{\beta\mathbf{x}} + a_2e^{-\beta\mathbf{x}})^2 \\ \text{Trigonometric} : f(\mathbf{x}) &= (a_1\cos\beta\mathbf{x} + a_2\sin\beta\mathbf{x})^2. \end{aligned} \tag{3}$$

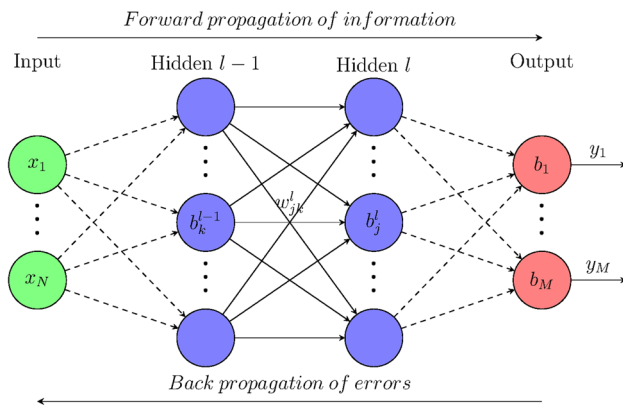
The governing equations for different material variations in the  $z$ -direction are summarized in Table 2:

**Table 1** Problems belong to the category of problems of potential

Problems	Scalar function $\phi$	$k(\mathbf{x})$	Boundary condition	
			Dirichlet	Neumann
Heat transfer	Temperature $T$	Thermal conductivity ( $k$ )	$T = \bar{T}$	Heat flow $q = -k\frac{\partial T}{\partial n}$
Ground water flow	Hydraulic head $H$	Permeability ( $k$ )	$H = \bar{H}$	Velocity flow $q = -k\frac{\partial H}{\partial n}$
Electrostatic	Electrostatic potential $V$	Permittivity ( $\epsilon$ )	$V = \bar{V}$	Electric flow $q = -k\frac{\partial V}{\partial n}$
Electric conduction	Electropotential $E$	Resistivity ( $k$ )	$E = \bar{E}$	Electric current $q = -k\frac{\partial E}{\partial n}$
Magnetostatic	Magnetic potential $M$	Magnetic permeability ( $\mu$ )	$M = \bar{M}$	Magnetic flux density $q = -k\frac{\partial M}{\partial n}$

**Table 2** Governing equation deduced by considering various forms of  $k(x)$

$k(x)$	Differential equation
$k_0(a_1 + a_2z)^2$	$(a_1 + a_2z)\nabla^2\phi + 2a_2\phi_z = 0$
$k_0(a_1e^{\beta z} + a_2e^{-\beta z})^2$	$(a_1e^{\beta z} + a_2e^{-\beta z})^2\nabla^2\phi + 2\beta(a_1^2e^{2\beta z} + a_2^2e^{-2\beta z})\phi_z = 0$
$k_0(a_1\cos\beta z + a_2\sin\beta z)^2$	$(a_1\cos\beta z + a_2\sin\beta z)^2\nabla^2\phi + 2\beta(0.5(a_2^2 - a_1^2)\sin 2\beta z + a_1a_2\cos 2\beta z)\phi_z = 0$



**Fig. 1** Architecture of a fully connected feedforward back-propagation neural network

### 3 Physics-informed deep learning-based collocation method

#### 3.1 Feed forward neural network

The basic architecture of a fully connected feedforward neural network is shown in Fig. 1. It comprises multiple layers: an input layer, one or more hidden layers and an output layer. Each layer consists of one or more nodes called neurons, shown in Fig. 1 by the small colored circles. For an interconnected structure, every two neurons in neighboring layers have a connection, where the weights between neuron  $k$  in hidden layer  $l - 1$  and neuron  $j$  in hidden layer  $l$  is denoted by  $w_{jk}^l$ , see Fig. 1. No connection exists among neurons in the same layer as well as in the non-neighboring layers. Input data, defined from  $x_1$  to  $x_N$ , flow through this neural network via connections between neurons, starting from the input layer, through the hidden layers  $l - 1, l$ , to the output layer, which eventually outputs data from  $y_1$  to  $y_M$ .

The activation function is defined for an output of each neuron in order to introduce a non-linearity into the neural network and make the back-propagation possible, where gradients are supplied along with an error to update weights and biases. The activation function in layer  $l$  will be denoted by  $\sigma$  here.

There are many activation functions  $\sigma$  proposed for inference and identification with neural networks, such as

sigmoids function [41], hyperbolic tangent function (*Tanh*) [41], Rectified linear units (*Relu*), to name a few. And some recent smooth activation functions, such as Swish [42], LeCuns Tanh [41], Bipolar sigmoid [41], Mish [42], Arctan [43], listed in Appendix B Table 9 have been studied and compared in the numerical example section. All selected activation functions must be smooth enough to avoid gradient vanishing during backpropagation, since the governing equation is introduced in the loss which includes the second-order derivatives of the field variable. Afterward, the value on each neuron in the hidden layers and output layer can be yielded by adding the weighted sum of values of output values from previous layer to basis. An intermediate quantity for neuron  $j$  on hidden layer  $l$  is defined as

$$a_j^l = \sum_k w_{jk}^l y_k^{l-1} + b_j^l, \tag{4}$$

and its output is given by the activation of the above weighted input

$$y_j^l = \sigma(a_j^l) = \sigma\left(\sum_k w_{jk}^l y_k^{l-1} + b_j^l\right), \tag{5}$$

where  $y_k^{l-1}$  is the output from previous layer.

Based on the previous derivation and description, we can draw a definition which will be used in Section 3.3:

**Definition 3.1** (*Feedforward Neural Network*) A generalized neural network with activation function can be written in a tuple form  $((f_1, \sigma_1), \dots, (f_n, \sigma_n))$ ,  $f_i$  referring to an affine-line function ( $f_i = W_i x + b_i$ ) that maps  $R^{i-1} \rightarrow R^i$ . The tuple formed neural network in all defines a continuous bounded function mapping  $R^d$  to  $R^n$ :

$$FNN : \mathbb{R}^d \rightarrow \mathbb{R}^n, \text{ with } F^n(x; \theta) = \sigma_n \circ f_n \circ \dots \circ \sigma_1 \circ f_1 \tag{6}$$

where  $d$  indicates the dimension of the inputs,  $n$  the number of field variables,  $\theta = \{W; b\}$  consisting of hyperparameters such as weights and biases and  $\circ$  denotes the element-wise operator.

The universal approximation theorem [22, 23] states that this continuous bounded function  $F$  with nonlinear activation  $\sigma$  can be adopted to capture the nonlinear property of the

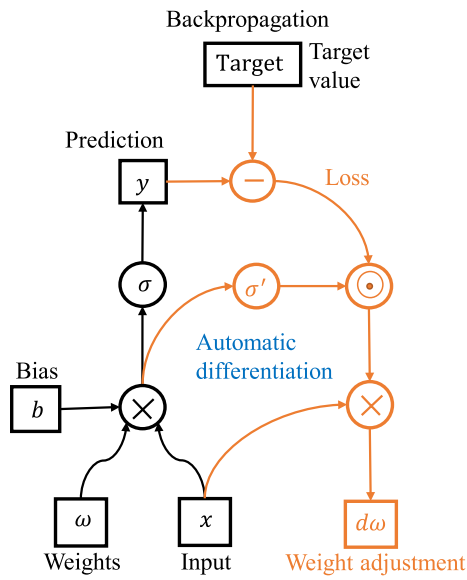


Fig. 2 The ‘compute graph’ for the feedforward neural network

system, in our case the potential problem. With this definition, we can define [44]:

**Theorem 1** *If  $\sigma^i \in C^m(\mathbb{R}^i)$  is non-constant and bounded, then  $F^n$  is uniformly  $m$ -dense in  $C^m(\mathbb{R}^n)$ .*

### 3.2 Backpropagation

Backpropagation (*backward propagation*) can be used to train multilayer feed-forward networks by calculating the gradient of a loss function and finding the minimum value of the loss function. The backward (output-to-input) flow determines how to adjust each weight as shown in Fig. 2.

Backpropagation is based on the chain rule, which is used to calculate the derivative of loss function with regard to the weight in the network. The governing equation in our problem requires the second partial derivatives of the potential function  $\phi(x)$ . To find the weights and biases, a loss function  $\text{Loss}(f, \theta)$  is defined. The backpropagation algorithm for computing the gradient of this loss function  $\text{Loss}(f, \theta)$ , the weight coefficients  $w$  and thresholds of neurons  $b$  can be written as follow:

---

**Algorithm 1** Backpropagation algorithm

---

**Input:** Create random sampling points  $x = (x_1, \dots, x_m)^T$  inside the physical domain. **Output:** The gradient of the loss function. **Set:** the corresponding activation  $a^l = x^i, i = 1, \dots, m$  for the input layer

**Feedforward:**

for  $l$  from 2 to  $L$  do compute  $z^l = w^l a^{l-1} + b^l$  and  $a^l = \sigma(z^l)$   
end for

**Output error:** Compute the output error  $\delta^L = \nabla_a \text{Loss} \odot \sigma'(z^L)$

**Backpropagate error:**

for  $l$  from  $L-1$  to 2 do compute  $\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$   
end for

**Output gradient:** The gradient of the loss function  $\frac{\partial \text{Loss}}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l$  and

$$\frac{\partial \text{Loss}}{\partial b_j^l} = \delta_j^l$$


---

### 3.3 Physics-informed deep collocation method

To train the network, we place collocation points in the physical domain and at the boundaries denoted by  $\mathbf{x}_\Omega = (x_1, \dots, x_{N_\Omega})^T$  and  $\mathbf{x}_\Gamma(x_1, \dots, x_{N_\Gamma})^T$ , respectively. Then the potential function  $\phi$  is approximated with the aforementioned deep feedforward neural network  $\phi^h(\mathbf{x};\theta)$ . Thus, a loss function related to the underlying BVP is constructed. Substituting  $\phi^h(\mathbf{x}_\Omega;\theta)$  into governing equation, we obtain

$$G(\mathbf{x}_\Omega;\theta) = k(\mathbf{x})\phi_{,ii}^h(\mathbf{x}_\Omega;\theta) + k_{,i}(\mathbf{x})\phi_{,i}^h(\mathbf{x}_\Omega;\theta), \tag{7}$$

which results in a physics-informed deep neural network  $G(\mathbf{x}_\Omega;\theta)$ . The boundary conditions illustrated in Section 2 can also be expressed by the neural network approximation  $\phi^h(\mathbf{x}_\Gamma;\theta)$  as: On  $\Gamma_D$ , we have

$$\phi^h(\mathbf{x}_{\Gamma_D};\theta) = \tilde{\phi}, \tag{8}$$

On  $\Gamma_N$ ,

$$q^h(\mathbf{x}_{\Gamma_N};\theta) = \tilde{q}. \tag{9}$$

where  $q^h(\mathbf{x}_{\Gamma_N};\theta)$  can be obtained from Eq. (2) by combing  $\phi^h(\mathbf{x}_{\Gamma_N};\theta)$ . Note the induced physics-informed neural network  $G(\mathbf{x};\theta)$ ,  $q(\mathbf{x};\theta)$  share the same parameters as  $\phi^h(\mathbf{x};\theta)$ . Considering the generated collocation points in domain and on boundaries, they can all be learned by minimizing the mean square error loss function [45]:

$$\text{Loss}(\theta) = \text{MSE} = \text{MSE}_G + \text{MSE}_{\Gamma_D} + \text{MSE}_{\Gamma_N}, \tag{10}$$

with

$$\begin{aligned} \text{MSE}_G &= \frac{1}{N_d} \sum_{i=1}^{N_d} \left\| G(\mathbf{x}_\Omega;\theta) \right\|^2 \\ &= \frac{1}{N_\Omega} \sum_{i=1}^{N_\Omega} \left\| k(\mathbf{x}_\Omega)\phi_{,ii}^h(\mathbf{x}_\Omega;\theta) + k_{,i}(\mathbf{x}_\Omega)\phi_{,i}^h(\mathbf{x}_\Omega;\theta) \right\|^2, \\ \text{MSE}_{\Gamma_D} &= \frac{1}{N_{\Gamma_D}} \sum_{i=1}^{N_{\Gamma_D}} \left\| \phi^h(\mathbf{x}_{\Gamma_D};\theta) - \tilde{\phi} \right\|^2, \\ \text{MSE}_{\Gamma_N} &= \frac{1}{N_{\Gamma_N}} \sum_{i=1}^{N_{\Gamma_N}} \left\| q(\mathbf{x}_{\Gamma_N};\theta) - \tilde{q} \right\|^2 \\ &= \frac{1}{N_{\Gamma_N}} \sum_{i=1}^{N_{\Gamma_N}} \left\| -k(\mathbf{x}_{\Gamma_N}) \frac{\partial \phi(\mathbf{x}_{\Gamma_N};\theta)}{\partial n} - \tilde{q} \right\|^2. \end{aligned} \tag{11}$$

where  $\mathbf{x}_\Omega \in R^N$ ,  $\theta \in R^K$  are the neural network parameters.  $\text{Loss}(\theta) = 0$ ,  $\phi^h(\mathbf{x};\theta)$  is then a solution to potential function. Here, the defined loss function measures how well the approximation satisfies the physical law (governing equation), boundaries conditions. Our goal is to find a

set of parameters  $\theta$  that the approximated potential  $\phi^h(\mathbf{x};\theta)$  minimizes the loss  $\text{Loss}$ . If  $\text{Loss}$  is a very small value, the approximation  $\phi^h(\mathbf{x};\theta)$  is very closely satisfying governing equations and boundary conditions, namely

$$\phi^h = \text{argmin}_{\theta \in R^K} \text{Loss}(\theta) \tag{12}$$

The solution of heat conduction problems by deep collocation method can be reduced to an optimization problem. In the deep learning Tensorflow framework, a variety of optimizers are available. One of the most widely used optimization methods is the Adam optimization algorithm, which is also adopted in the numerical study. The idea is to take a descent step at collocation point  $\mathbf{x}_i$  with Adam-based learning rate  $\eta_i$ ,

$$\theta_{i+1} = \theta_i + \eta_i \nabla_{\theta} \text{Loss}(\mathbf{x}_i;\theta_i) \tag{13}$$

and then the process in Eq. (13) is repeated until a convergence criterion is satisfied.

### 3.4 Convergence of deep collocation method for non-homogeneous PDEs

With the universal approximation theorem of neural networks, a feedforward neural network is used to approximate the potential function as  $\phi^h(\mathbf{x};\theta)$ . The approximation power of neural networks for a quasilinear parabolic PDEs is shown by Sirignano et al. [29]. For non-homogeneous elliptic PDEs, the convergence study can be boiled down to:

$$\exists \phi^h \in F^n, \text{ s.t. as } n \rightarrow \infty, \text{ Loss}(\theta) \rightarrow 0, \phi^h \rightarrow \phi. \tag{14}$$

The non-homogeneous PDEs has a unique solution, s.t.  $\phi \in C^2(\Omega)$  with its derivatives uniformly bounded. Also, the conductivity function  $k(\mathbf{x})$  is assumed to be  $C^{1,1}$  ( $C^1$  with Lipschitz continuous derivative).

**Theorem 2** Assume that  $\Omega$  is compact with measures  $\ell_1, \ell_2$ , and  $\ell_3$  whose supports are constrained in  $\Omega, \Gamma_D$ , and  $\Gamma_N$ . Furthermore, the governing Eq. (1) subject to 2 has a unique classical solution and material function  $k(\mathbf{x})$  being  $C^{1,1}$  ( $C^1$  with Lipschitz continuous derivative). Then,  $\forall \epsilon > 0$ ,  $\exists K > 0$ , which may dependent on  $\sup_{\Omega} \|\phi_{ii}\|$  and  $\sup_{\Omega} \|\phi_{,i}\|$ , s.t.  $\exists \phi^h \in F^n$ , satisfies  $\text{Loss}(\theta) \leq K\epsilon$

**Proof** For governing Eq. (1) subject to 2, according to Theorem 1,  $\forall \epsilon > 0, \exists \phi^h \in F^n$ , s.t.

$$\sup_{\mathbf{x} \in \Omega} \left\| \phi_{,i}(\mathbf{x}_\Omega) - \phi_{,i}^h(\mathbf{x}_\Omega) \right\|^2 + \sup_{\mathbf{x} \in \Omega} \left\| \phi_{,ii}(\mathbf{x}_\Omega) - \phi_{,ii}^h(\mathbf{x}_\Omega) \right\|^2 < \epsilon \tag{15}$$

Recalling that the loss is constructed by Eq. (10), for  $\text{MSE}_G$  and applying triangle inequality, we obtain:

$$\|G(\mathbf{x}_\Omega; \boldsymbol{\theta})\|^2 \leq \|k(\mathbf{x}_\Omega)\phi_{,ii}^h(\mathbf{x}_\Omega; \boldsymbol{\theta})\|^2 + \|k_{,i}(\mathbf{x}_\Omega)\phi_{,i}^h(\mathbf{x}_\Omega; \boldsymbol{\theta})\|^2 \tag{16}$$

Let us consider the  $C^{1,1}$  conductivity function  $k(\mathbf{x})$ ,  $\exists M_1 > 0, M_2 > 0, \exists x \in \Omega, \|k(\mathbf{x})\| \leq M_1, \|k_{,i}(\mathbf{x})\| \leq M_2$ . From Eq. (15), we can then obtain:

$$\int_\Omega k_i^2(\mathbf{x}_\Omega) (\phi_i^h - \phi_{,i})^2 d\ell_1 \leq M_2^2 \varepsilon^2 \ell_1(\Omega)$$

$$\int_\Omega k^2(\mathbf{x}_\Omega) (\phi_{,ii}^h - \phi_{,ii})^2 d\ell_1 \leq M_1^2 \varepsilon^2 \ell_1(\Omega) \tag{17}$$

On boundaries  $\Gamma_D$  and  $\Gamma_N$ , we can obtain:

$$\int_{\Gamma_D} (\phi^h(\mathbf{x}_{\Gamma_D}; \boldsymbol{\theta}) - \phi(\mathbf{x}_{\Gamma_D}; \boldsymbol{\theta}))^2 d\ell_2 \leq \varepsilon^2 \ell_2(\Gamma_D)$$

$$\int_{\Gamma_N} k^2(\mathbf{x}_{\Gamma_N}) (\phi_{,n}^h(\mathbf{x}_{\Gamma_N}; \boldsymbol{\theta}) - \phi_{,n}(\mathbf{x}_{\Gamma_N}; \boldsymbol{\theta}))^2 d\ell_3 \leq M_1^2 \varepsilon^2 \ell_3(\Gamma_N) \tag{18}$$

Therefore, using Eqs. 17 and 18, as  $n \rightarrow \infty$ , we obtain

$$\begin{aligned} \text{Loss}(\boldsymbol{\theta}) &= \frac{1}{N_\Omega} \sum_{i=1}^{N_\Omega} \|k(\mathbf{x}_\Omega)\phi_{,ii}^h(\mathbf{x}_\Omega; \boldsymbol{\theta}) + k_{,i}(\mathbf{x}_\Omega)\phi_{,i}^h(\mathbf{x}_\Omega; \boldsymbol{\theta})\|^2 \\ &+ \frac{1}{N_{\Gamma_D}} \sum_{i=1}^{N_{\Gamma_D}} \|\phi^h(\mathbf{x}_{\Gamma_D}; \boldsymbol{\theta}) - \bar{\phi}\|^2 \\ &+ \frac{1}{N_{\Gamma_N}} \sum_{i=1}^{N_{\Gamma_N}} \left\| -k(\mathbf{x}_{\Gamma_N}) \frac{\partial \phi(\mathbf{x}_{\Gamma_N}; \boldsymbol{\theta})}{\partial n} - \bar{q} \right\|^2 \\ &\leq \frac{1}{N_\Omega} \sum_{i=1}^{N_\Omega} \|k(\mathbf{x}_\Omega)\phi_{,ii}^h(\mathbf{x}_\Omega; \boldsymbol{\theta})\|^2 \\ &+ \frac{1}{N_\Omega} \sum_{i=1}^{N_\Omega} \|k_{,i}(\mathbf{x}_\Omega)\phi_{,i}^h(\mathbf{x}_\Omega; \boldsymbol{\theta})\|^2 \\ &+ \frac{1}{N_{\Gamma_D}} \sum_{i=1}^{N_{\Gamma_D}} \|\phi^h(\mathbf{x}_{\Gamma_D}; \boldsymbol{\theta}) - \bar{\phi}\|^2 \\ &+ \frac{1}{N_{\Gamma_N}} \sum_{i=1}^{N_{\Gamma_N}} \left\| -k(\mathbf{x}_{\Gamma_N}) \frac{\partial \phi(\mathbf{x}_{\Gamma_N}; \boldsymbol{\theta})}{\partial n} - \bar{q} \right\|^2 \\ &\leq (M_2^2 + M_1^2)\varepsilon^2 \ell_1(\Omega) + \varepsilon^2 \ell_2(\Gamma_D) + M_1^2 \varepsilon^2 \ell_3(\Gamma_N) = K\varepsilon \end{aligned} \tag{19}$$

□

With Theorem 2 and the condition that  $\Omega$  is a bounded open subset of  $\mathbb{R}^n, \forall n \in \mathbb{N}_+, \phi^h \in F^n \in L^2(\Omega)$ , it can be concluded from Sirignano et al. [29] that:

**Theorem 3**  $\forall p < 2, \phi^h \in F^n$  converges to  $\phi$  strongly in  $L^p(\Omega)$  as  $n \rightarrow \infty$  with  $\phi$  being the unique solution to the potential problems.

In summary, for feedforward neural networks  $F^n \in L^p$  space ( $p < 2$ ), the approximated solution  $\phi^h \in F^n$  will converge to the solution to the non-homogeneous PDE.

### 3.5 Collocation points generation

Model training is an important process in machine learning and the quality of training datasets determines the reliability of the machine learning model to a large extent. The deep collocation method (DCM) utilizes physics-informed neural networks for solving PDEs with randomly generated training points in the physical domain. To test the influence of training points on the stability and accuracy, different sampling methods are compared. The Halton and Hammersley sequences generate random points by a constructing the radical inverse [46]. They are both low discrepancy sequences. The method of Korobov Lattice creates samples from Korobov lattice point sets [47]. Sobol Sequence is a quasi-random low-discrepancy sequence to generate sampling points [48]. Latin hypercube sampling (LHS) is a statistical method, where a near-random sample of parameter values is generated from a multidimensional distribution [49]. Monte Carlo methods can create points by repeated random sampling [50]. The distribution plots of different sampling points inside a cube is listed in Appendix B Table 10 (Fig. 3).

### 3.6 Material transfer learning

To improve the generality and robustness of the DCM, transfer learning is exploited, which makes use of the information from an already trained model yielding to training with less data and a reduced training time. The basic idea can be found in Fig. 4. For different material variations in nonhomogeneous media, the ‘knowledge’ of one material model can be exploited as the pretrained model resulting in a two-stage paradigm. The material transfer learning model is divided

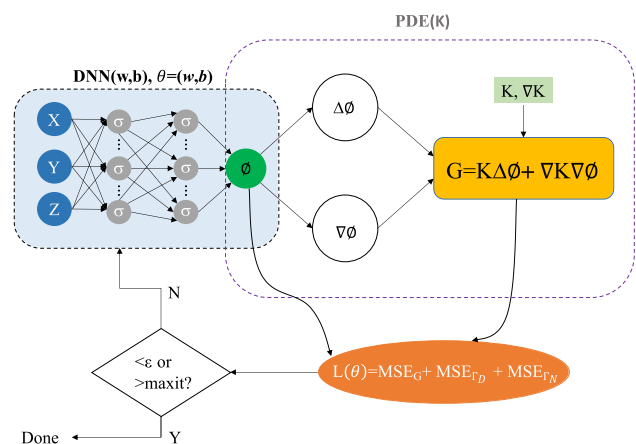
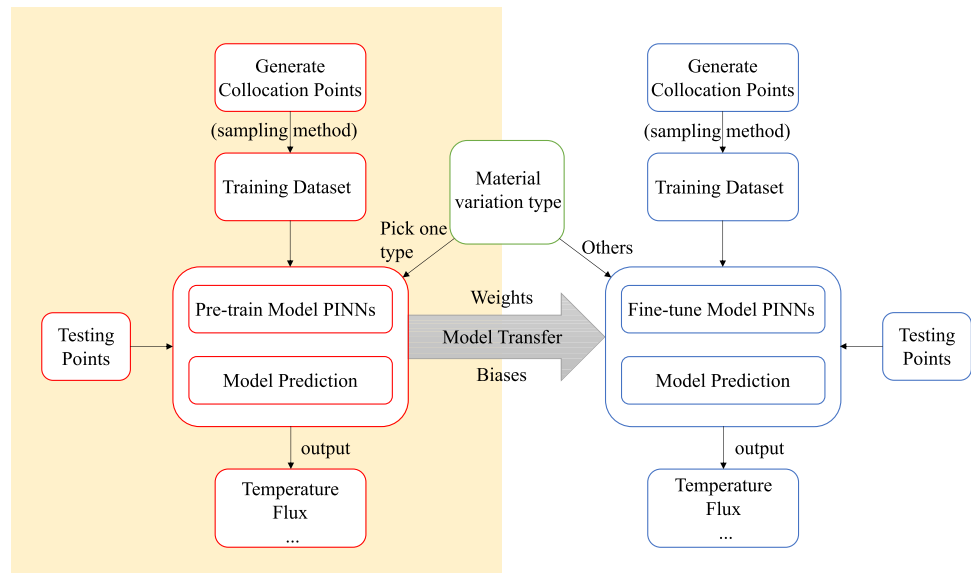


Fig. 3 Schematic diagram of physics-informed neural networks

Fig. 4 Transfer learning



into two parts, i.e. pretraining, where the network is trained on a large dataset and longer iterations for one material variation type. The remaining part is the fine-tuning, where the pretrained model is trained on other material variations with few data and number of epochs. Consequently, the weights and biases and network configurations from a trained model are passed to other relevant models.

There are still some unresolved limitations in the literature. Most importantly, physics-informed deep learning algorithms lack a more systematic procedure to prevent overfitting and finding global minima.

## 4 Sensitivity analysis

Algorithm-specific parameters, such as the neural architecture configurations, parameters related to optimizers and number of collocation points significantly influence the model's accuracy. To quantify their influence on the accuracy, a global sensitivity analysis (GSA) is performed. Classical GSA including regression methods, screening approaches, such as Morris method [51], the variance-based measures, such as Sobol's method [52], and the Fourier amplitude sensitivity test (FAST) [53], or the extended FAST (EFAST) [54].

Variance-based methods are usually more computationally expensive than the derivative-based methods as well as the regression methods. If the model or the parameters in analysis is large, the use of variance-based method can be costly. The method of Morris is generally robust to correctly screen the most and least sensitive parameters for a highly parameterized model with 300 times fewer model evaluations than the Sobol' method [55]. Therefore, the

computational cost of a sensitivity analysis can potentially be reduced by first performing parameter screening using the Morris method to identify non-influential parameters, reducing the dimension of the parameter space to be studied in further analysis, then filter them again, but with the eFAST method. In this way, we can quantify the effects of inputs more accurately with a relatively small amount of time.

### 4.1 Method of Morris

The method of Morris [56] is a screening technique used to rank the importance of parameters by averaging coarse difference relations termed elementary effects. Given a model with  $n$  parameters,  $\mathbf{X} = X_1, X_2, \dots, X_n$  denoting a vector of parameter values, we can specify an objective function  $y(x) = f(X_1, X_2, \dots, X_n)$ , change the variables  $X_i$  by specific ranges and then calculate the distribution of elementary effects (EE) of each input factor with respect to the model outputs, i.e.

$$EE_i = \frac{f(x_1, \dots, x_i + \Delta_i, \dots, x_n) - f(x)}{\Delta_i} \quad (20)$$

where  $f(x)$  represents the prior point in the trajectory. Using the single trajectory shown in Eq. (20), the elementary effects of each parameter can be calculated with  $p + 1$  model evaluations. After sampling the trajectories, the resulting sets of elementary effects are then averaged to obtain the total-order sensitivity of the  $i$ -th parameter  $\mu_i^*$ :

$$\mu_i^* = \frac{1}{n} \sum_{j=1}^n |EE_i^j| \quad (21)$$



Similarly, the variance of the set of EEs can be calculated as

$$\sigma_i^2 = \frac{1}{n-1} \sum_{j=1}^n (EE_i^j - \mu_i)^2 \tag{22}$$

The mean value  $\mu^*$  quantifies the individual effect of the parameters on an output while the variance  $\sigma^2$  indicates the influence of parameter interactions. We rank the parameters according to  $\sqrt{\sigma^2 + \mu^{*2}}$ .

### 4.2 eFAST method

The eFAST method [54] is based on Fourier transformations. The spectrum is obtained by each parameter and the output variance of model results due to interactions. Employing a suitable search function, the model  $y(x) = f(X_1, X_2, \dots, X_n)$  can be transformed by the Fourier transform into  $y = f(s)$

$$y = f(s) = \sum_{j=-\infty}^{+\infty} (A_j \cos(js) + B_j \sin(js)), \tag{23}$$

with

$$A_j = \frac{\pi}{2} \int_{\frac{\pi}{2}}^{-\frac{\pi}{2}} f(s) \cos(js) ds, \tag{24}$$

$$B_j = \frac{\pi}{2} \int_{\frac{\pi}{2}}^{-\frac{\pi}{2}} f(s) \sin(js) ds. \tag{25}$$

The spectral curve of the Fourier progression is defined as  $\Lambda_j = A_j^2 + B_j^2$ . The variance of the model results due to the uncertainty in the parameter  $X_i$  is given by

$$D_i = \sum_{p \in Z_0} \Lambda_p \omega_i, \tag{26}$$

with the parametric frequency  $\omega_1$ , the spectrum of the Fourier transform  $\Lambda$ , and the non-zero integers  $Z_0$ . The total variance can be obtained by cumulatively summing the spectra at all frequencies

$$D = 2 \sum_{j=1}^{\infty} \Lambda_j. \tag{27}$$

The fraction of the total output variance caused by each parameter apart from interactions with other parameters is measured by the first-order index

$$S_i^{FAST} = \frac{D_i}{D}. \tag{28}$$

To find the total sensitivity of  $X_i$ , the frequency of  $X_i$  is set to  $\omega_i$ , while a different frequency  $\omega'$  is set for all other parameters. By calculating the frequency  $\omega_i$  and its higher harmonics  $p\omega_i$  spectra, the output variance  $D_{-i}$  due to the influence of all parameters except  $X_i$  and their interrelationships can be obtained. Thus, the total-order sensitivity indices can be obtained:

$$S_{T_i} = \frac{D - D_{-i}}{D}. \tag{29}$$

## 5 Numerical examples

In this section, several cases are considered testing the accuracy and efficiency of our DCM including the influence of suitable NN configurations, sampling methods and optimizers taking advantage of GSA. Also, different material variations using material transfer learning are studied.

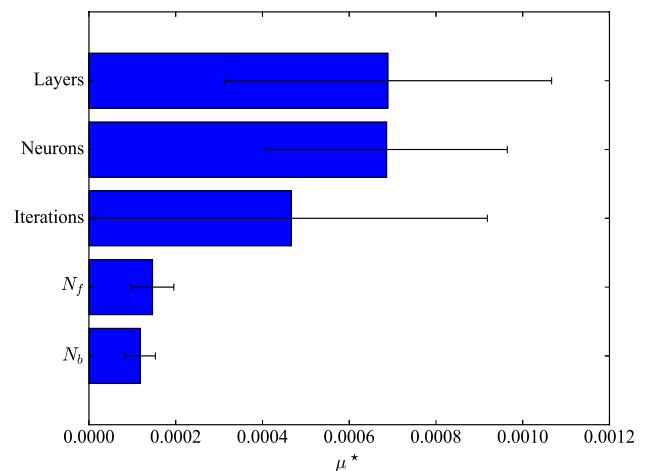


Fig. 5 Horizontal barplot of the mean absolute elementary effects  $\mu^*$

Table 3 Hyper-parameters settings in training

Model	Hyper-parameters	Values
Adam optimizer	Learning rate	0.001
	Maximum number of iterations to perform	50,000
L-BFGS-B optimizer	Maximum number of function evaluations	50,000
	Maximum number of variable metric corrections	50
	Maximum number of line search steps (per iteration)	50

The accuracy is measured in the relative error between the predicted solution and the analytical solution:

$$e = \frac{\|E_{pred} - E_a\|}{\|E_a\|} \tag{30}$$

where  $E_a$  is the analytical solution and  $E_{pred}$  is the predicted solution while  $\|\cdot\|$  refers to the  $L_2$ -norm. All simulations are done on a 64-bit macOS Catalina computer with Intel(R) Core(TM) i7-8850H CPU, 32GB memory. The parametric settings for training are summarised in Table 3.

### 5.1 Case 1: Sensitivity analysis

First, we perform a SA to determine the key parameters of the deep collocation method.

#### 5.1.1 Parameters screening with Morris method

The sensitivity indices computed by the Morris screening method with 30 trajectories and 4 grid levels are listed in

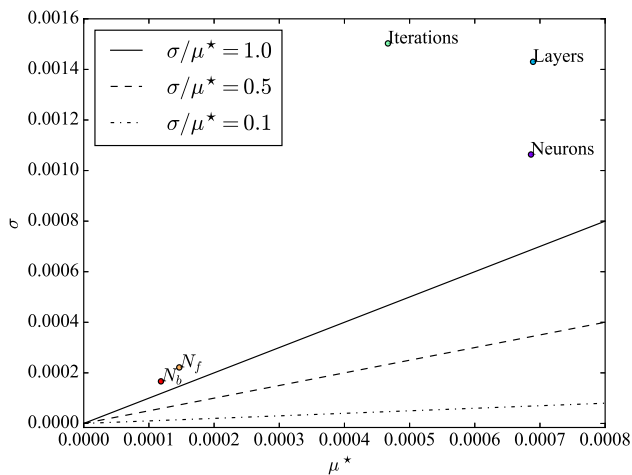
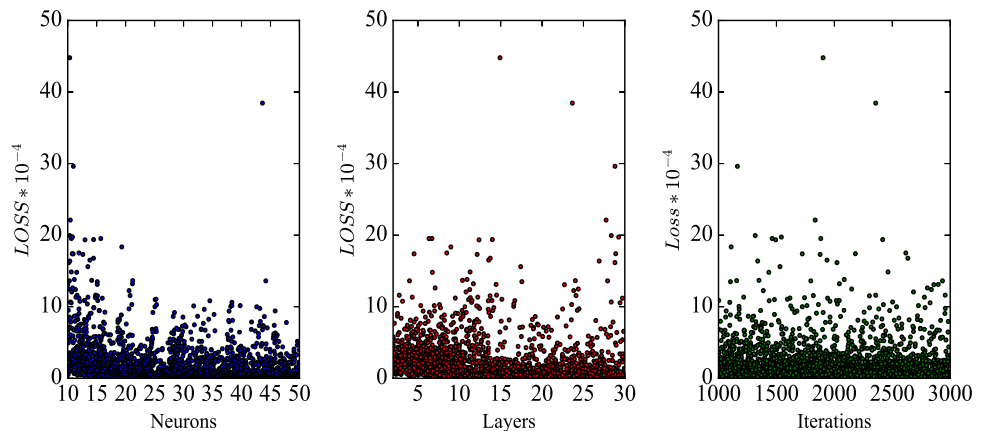


Fig. 6  $\sigma$  versus  $\mu^*$  for parameter screening with Morris method

Fig. 7 Scatter plot of loss value against parameter values with 3000 runs



Figs. 5 and 6, showing the effect of the numbers of neurons, layers, iterations and collocation points on the loss values. Figure 5 depicts the horizontal barplot of the GSA measure  $\mu^*$ . The highest  $\mu^*$  value is found for the numbers of layers and neurons. The numbers of collocation points barely have an effect on the loss value. According to a classification scheme proposed by Garcia Sanchez et al. [57], the ratio  $\sigma/\mu^*$  allows the characterisation of the model parameters in terms of (non-)linearity ( $\sigma/\mu^* < 0.1$ ), (non-)monotony ( $0.1 < \sigma/\mu^* < 0.5$ ) or possible parameter interactions ( $1 < \sigma/\mu^*$ ), see also Fig. 6. For our test models, all parameters are in the range  $\sigma/\mu^* > 1$  suggesting that most parameters exhibit either non-linear behaviour, interaction effects with each other or both. The plot of the mean value and standard deviation ( $\sigma, \mu^*$ ) in Fig. 6 reveals that the most influential parameter with largest  $\sqrt{\sigma^2 + \mu^{*2}}$  is the numbers of layers. The number of neurons and iterations is less important. The collocation points inside the physical domain and on the surface do not have a significant impact neither. Thus, while tuning the parameters of the model,

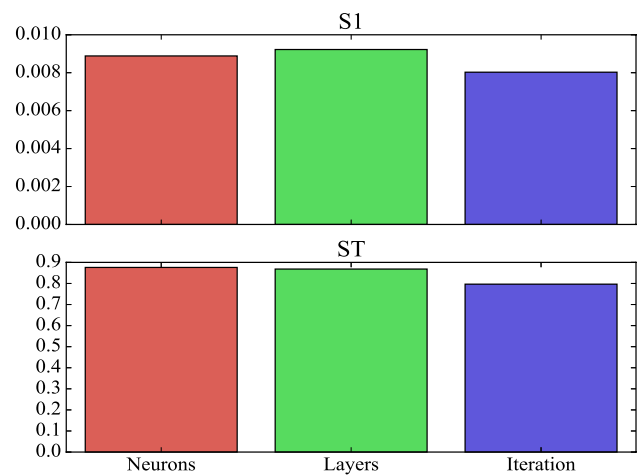


Fig. 8 Results for first-order and total indices with eFAST method

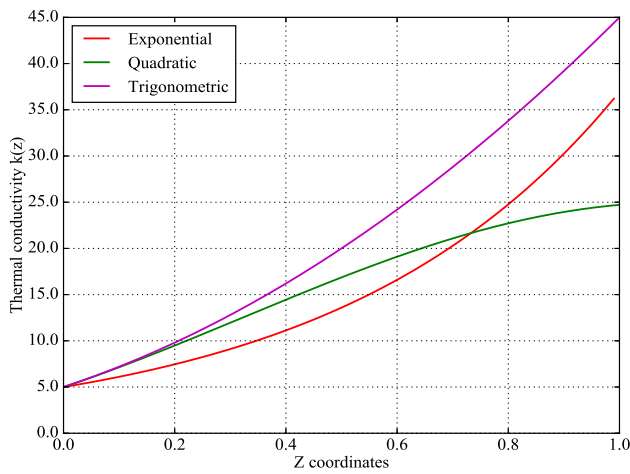


Fig. 9 Thermal conductivity variation along the z direction

more attention should be paid on the numbers of layers, neurons and iterations.

5.1.2 Variance-based sensitivity indices

We now take advantage of the variance-based eFAST method to compute sensitivity indices. The independent first-order sensitivity indices  $S_i$  and dependent total order sensitivity indices  $S_{T_i}$  can be found in Fig. 8. Due to the high computational costs, with 3000 simulations run and 1000 generated samples, no analyses concerning the variation of in  $S_i$  and  $S_{T_i}$  with different sample sizes were performed.

The associated scatter plots are shown in Fig. 7. The more randomly the loss values are distributed, the less sensitive the parameters is. According to Fig. 7, the number of layers is the most influential parameter, followed by the number of neurons and number of iterations.

The first-order sensitivity index  $S_i$  represents the parameter importance. The number of layers affects the model most, followed by the numbers of neurons and the least influential parameter is the number of iterations, which agrees well with the results of Morris Method. However, the first-order indices are all beyond 0.01, which manifest that those algorithm-specific parameters individually do not have too much influence on the loss value of the model. The total effects index  $S_{T_i}$  greater than 0.8 can be regarded very important parameters. Again, the number of layers and neurons is greater than 0.8. For the number of iterations, it is between 0.5 to 0.8. However, there is a big difference between the value of total and first-order sensitivity indices, which quantifies the effects of the parameter’s interactions. It can be concluded that the output variance can be attributed to their interactions with other parameters rather than their nonlinear effects and all interactions between these three parameters are noteworthy.

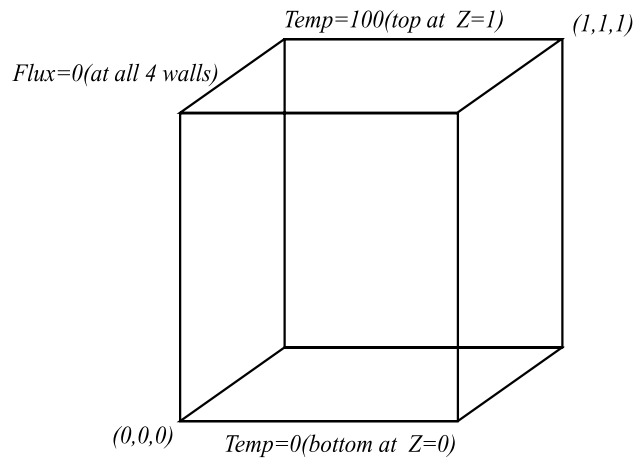


Fig. 10 The boundary conditons of the unit cube

Table 4 Analytical solutions for various forms of thermal conductivity  $k(x)$

$k(x)$	Analytical solution for potential function
$5(1 + 2z)^2$	$\phi = \frac{300z}{1+2z}$
$5e^{2z}$	$\phi = 100 \frac{1-e^{-2z}}{1-e^{-2L}}$
$5(\cos z + 2\sin z)^2$	$\phi = 100 \frac{\cot(L)+2*\sin z}{(\cos z+2\sin z)}$

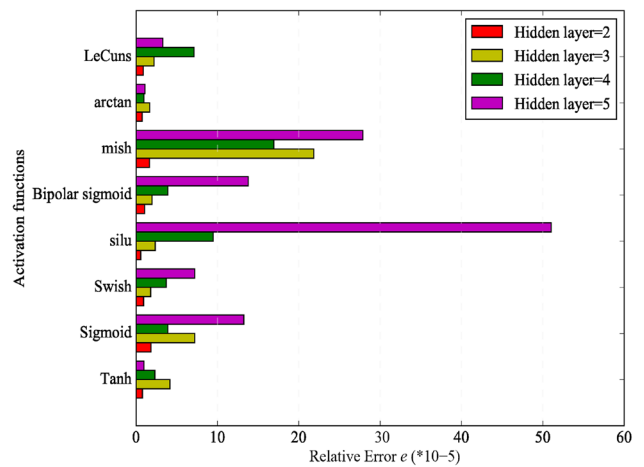


Fig. 11 Comparison of results predicted by DCM with different activation functions

5.2 Case 2. Cube with material gradation along the z-axis

Let us consider a unit cube ( $L = 1$ ) with prescribed constant temperature on two sides. The top surface of the cube at  $z = 1$  is maintained at a temperature of  $T = 100$  while the bottom

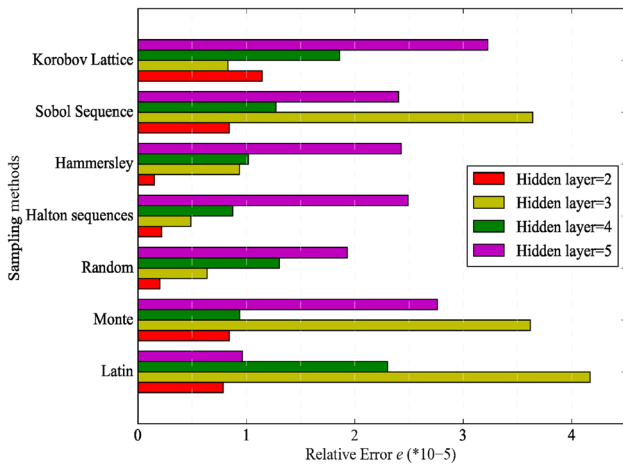


Fig. 12 Comparison of results predicted by DCM with different sampling methods

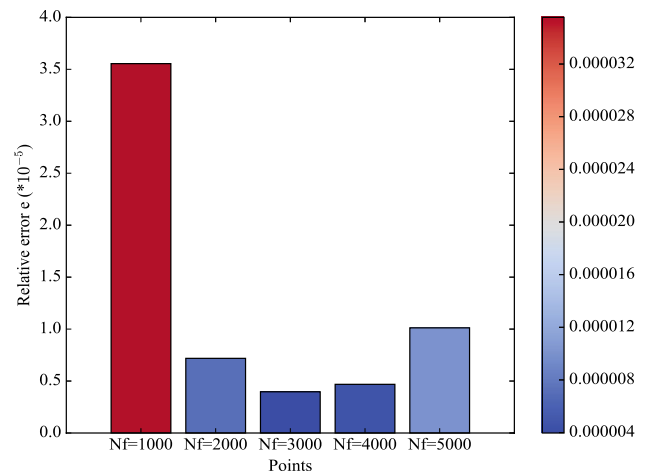


Fig. 14 Comparison of predicted results for different collocation points in the cube

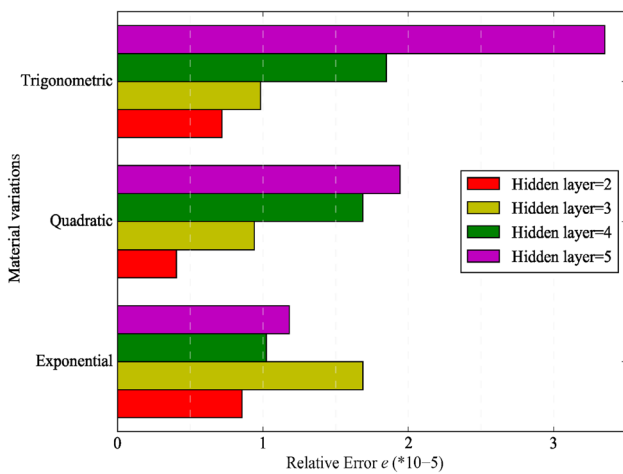


Fig. 13 Comparison of predicted results for different material variations

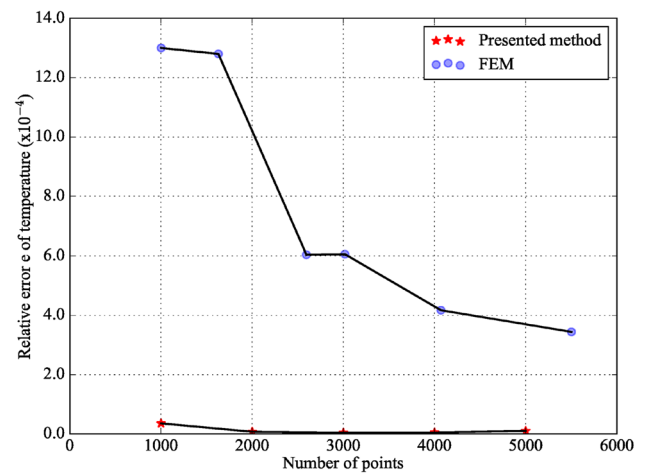


Fig. 15 Comparison of predicted results with FEM versus numbers of points

temperature at  $z = 0$  is zero. The remaining four faces are insulated (zero normal flux). Three different classes of variations shown in Table 4 are considered [58]. The profiles of the thermal conductivity  $k(z)$  of the three material variation cases are illustrated in Fig. 9, and the boundary conditions of the unit cube can be found in Fig. 10. For each nonhomogeneous thermal conductivity, the analytical solution is summarized in Table 4.

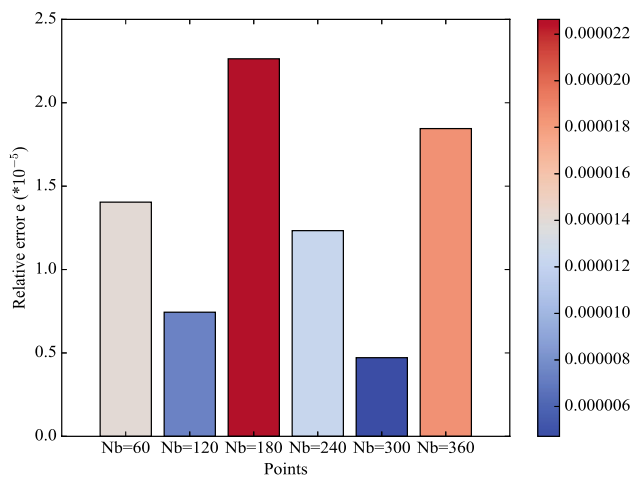
### 5.2.1 Deep collocation method configurations

First, different NN configurations are investigated. Figure 11 shows the relative error for various activation functions and layers. The *arctan* function yields the most stable and accurate results. Both *arctan* and *Tanh* function outperform the

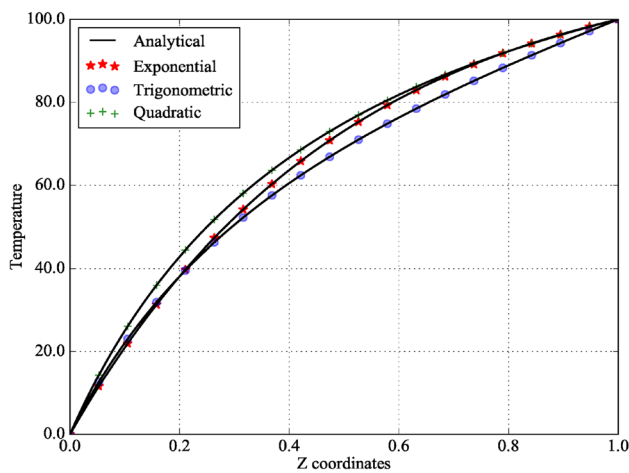
other activation functions. Figure 12 depicts the influence of different sampling methods on the relative error. Random sampling method obtained most stable and accurate potentials with increasing layers. *Korobov*, *Hammersley*, *LatinHypercube* sampling methods also provide reasonable results.

Next, we focus on various material variations, see Fig. 13. All material variations can be predicted accurately, but the most accurate results are obtained for the exponential conductivity. The results from Figs. 11, 12 and 13 suggest that 2 hidden layers are a good choice for the underlying problem.

We study now different numbers of collocation points (inside the cube and on its surface). The relative error in the temperature is depicted in Figs. 14 and 16. We also compared our results to results from FEM in Fig. 15. The temperature profiles along the  $z$ -axis for three material

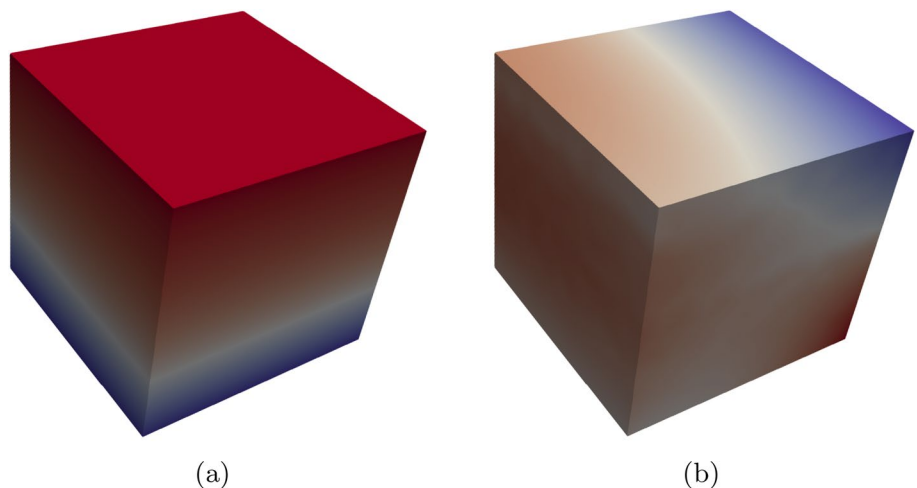


**Fig. 16** Comparison of predicted results for different collocation points on each boundary



**Fig. 17** Temperature profile in the Z direction for different material variations

**Fig. 18 a** Predicted temperature and **b** Predicted flux for exponential material variation for the functionally graded unit cubic



variations are plotted with the corresponding analytical solutions in Fig. 17.

The predicted temperature and flux distributions for three material variations inside the cube are shown in Figs. 18, 19 and 20. The heat distribution varies with graded variation in the  $z$  coordinates which is consistent with the material property of the FGMs.

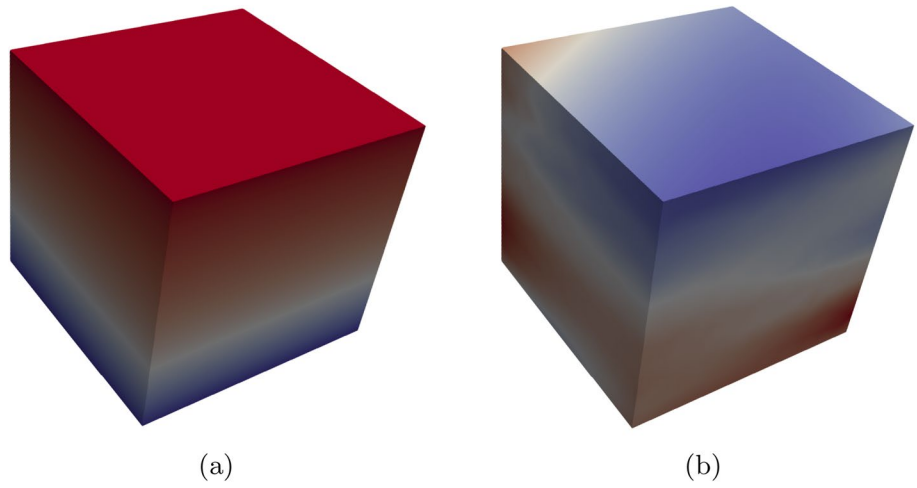
Let us now test the influence of the optimizer on the results. First-order methods minimize the function using its gradient, while second-order methods minimize the loss function using the second-order derivatives (Hessian information). In this application, a combination of these two optimizers is employed. The used first-order method is the Adam algorithm while L-BFGS is the tested second-order method. The convergence history for different optimizers is illustrated in Fig. 21. Although the first-order optimizer can be faster, they require more iterations. The L-BFGS optimizer needs less iterations, but there is the risk in being trapped in local minima. Using the combined optimizers, the loss reaches a significant smaller value with acceptable number of iterations and simultaneously ensures the solution being close to the global minima. The results for different number of layers are illustrated in Fig. 22.

### 5.2.2 Material transfer learning

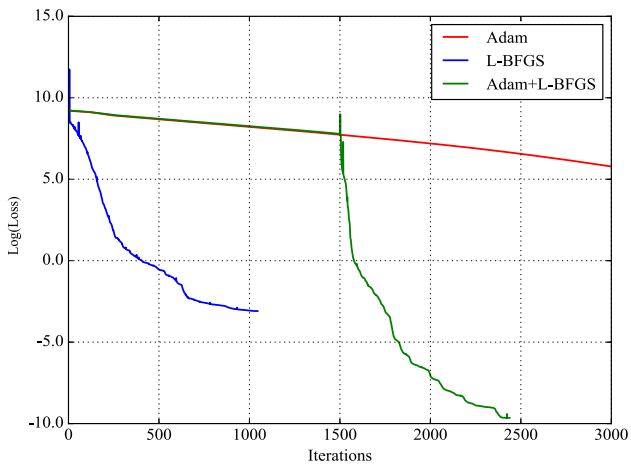
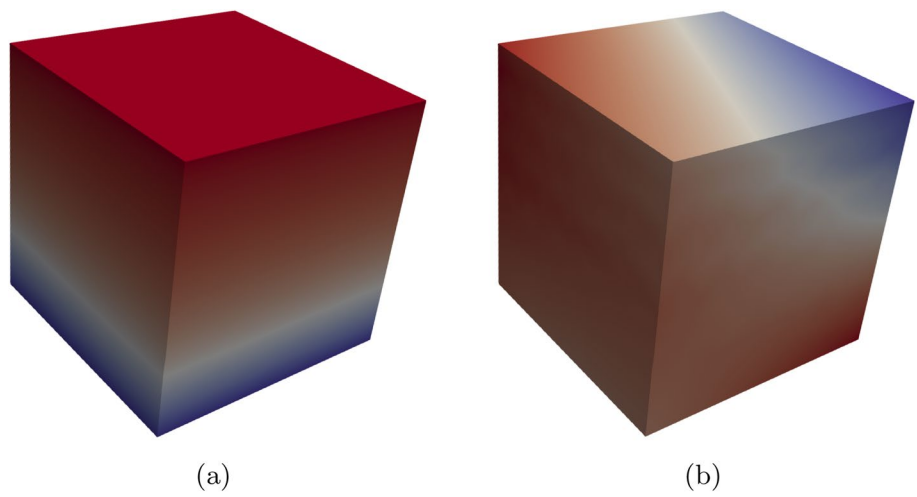
The loss vs number of iterations is shown in Fig. 23. After funetuning, the loss decreases to a smaller value in less iterations for all three material variations. The numerical results are summarized in Table 5 demonstrating that the computational effort can be drastically reduced with transfer learning.

Figure 24 shows the loss vs iteration using transfer learning for different material parameters while Tables 6 and 7 list the accuracy and CPU time with and without transfer learning.

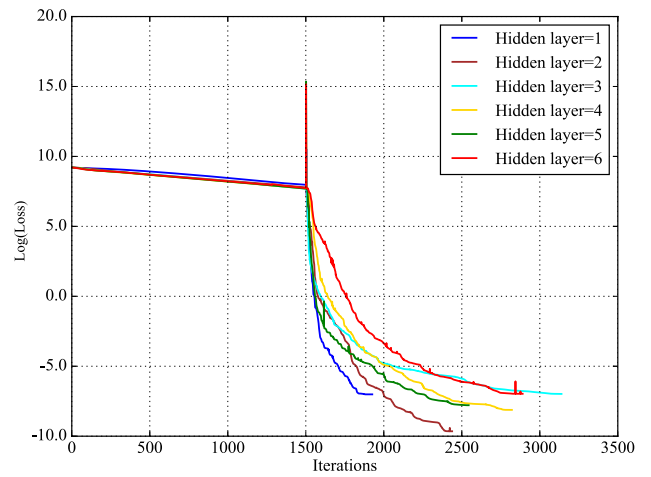
**Fig. 19** **a** Predicted temperature and **b** Predicted flux for trigonometric material variation for the functionally graded unit cubic



**Fig. 20** **a** Predicted temperature and **b** Predicted flux for quadratic material variation for the functionally graded unit cubic



**Fig. 21** Comparison of different optimizers for deep collocation method



**Fig. 22** Convergence graph for DCM with increasing hidden layers

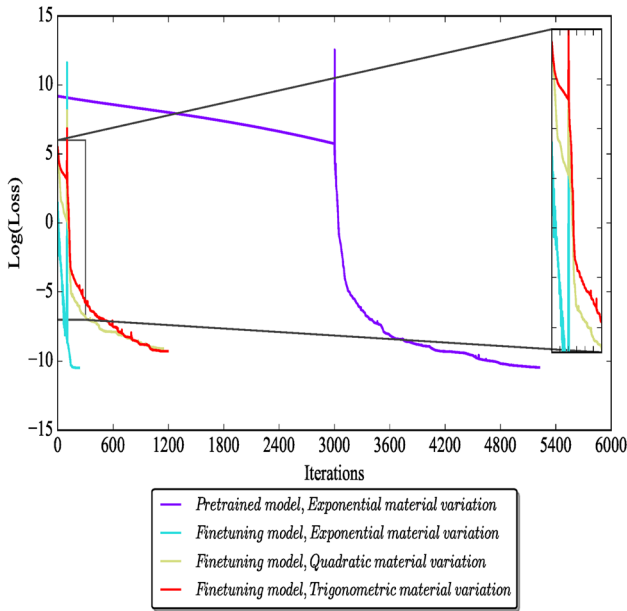
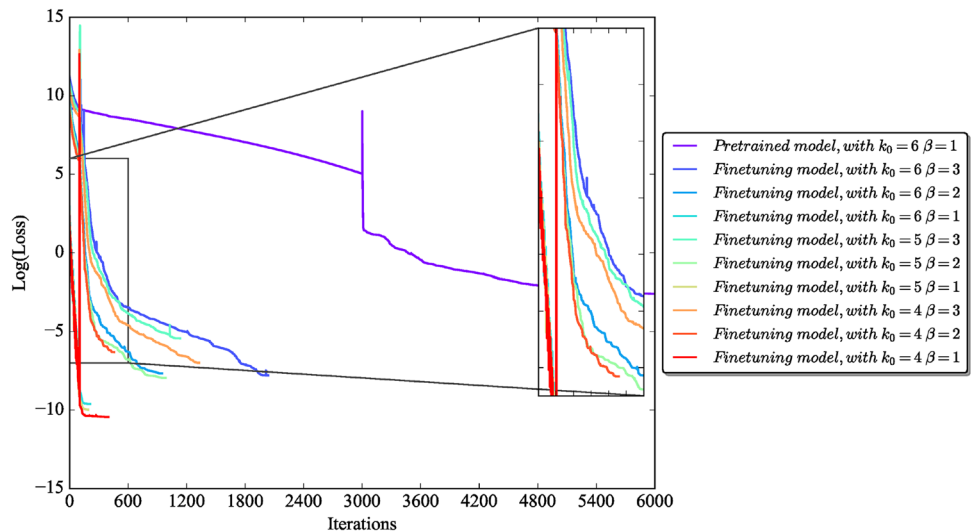


Fig. 23 Loss vs iteration using transfer learning with different material variations

Table 5 Relative error and training time for material variation with transfer learning

Results	Material variation			
	Exponential		Quadratic	Trigono- metric
	Without TL	With TL	With TL	With TL
Relative error	4.2846e-06	3.9015e-06	3.7033e-06	3.6562e-06
Training time	45.5s	9.1s	22.4s	18.3s

Fig. 24 Loss vs iteration using transfer learning with different material parameters



### 5.3 Case 3: Cube with a 3D material gradation

Now, we consider a cube with the following three-dimensional thermal conductivity variation:

$$k(x, y, z) = (5 + 0.2x + 0.4y + 0.6z + 0.1xy + 0.2yz + 0.3zx + 0.7xyz)^2 \tag{31}$$

The iso-surfaces of the 3D variation of the thermal conductivity is illustrated in Fig. 25. The analytical solution for this variation is

$$\phi(x, y, z) = \frac{xyz}{(5 + 0.2x + 0.4y + 0.6z + 0.1xy + 0.2yz + 0.3zx + 0.7xyz)} \tag{32}$$

The boundary conditions at the six faces of the cube are listed in Table 8.

The predicted temperature and flux distributions are shown in Fig. 26. The predicted relative error of the temperature across the cube is 5.215360e-03, see also Fig. 27.

### 5.4 Case 4: Irregular-shaped annular sector

Next, we present results for an irregular-shaped annular sector as depicted in Fig. 29. The inner radius is 0.3, the outer radius is 0.5, the top surface is at Z = 0.1 and the thermal conductivity for the geometry varies exponentially according to

$$k(z) = 5e^{(3z)} \tag{33}$$

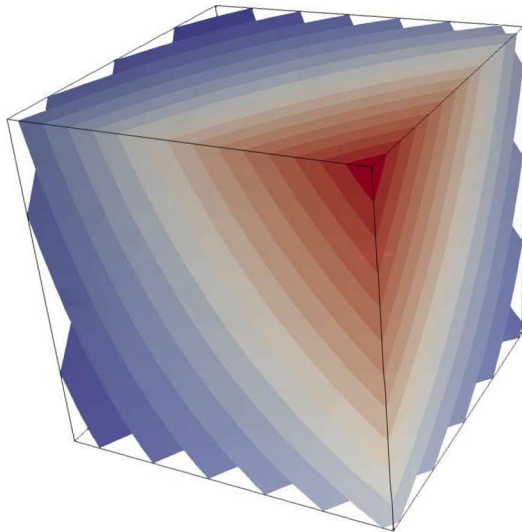
The variation of the thermal conductivity  $k(z)$  is illustrated in Fig. 28. The temperature is specified along the inner radius as  $T_{inner} = 0$ , and outer radius as  $T_{outer} = 100$ ; all

**Table 6** Relative error of temperature with varying material parameters

$k_0$	$\beta$					
	3		2		1	
	Without TL	With TL	Without TL	With TL	Without TL	With TL
6	1.9416e-05	8.6204e-06	1.7244e-05	8.2445e-06	3.8324e-06	3.1974e-06
5	1.8445e-05	1.9075e-05	6.8346e-06	9.9521e-06	1.9358e-06	2.7892e-06
4	1.4026e-05	8.0790e-06	6.8956e-06	1.6358e-05	4.8229e-06	2.3579e-06

**Table 7** Computation time with varying material parameters (s or sec)

$k_0$	$\beta$					
	3		2		1	
	Without TL	With TL	Without TL	With TL	Without TL	With TL
6	6.9715e+01	1.8325e+01	4.6163e+01	1.0890e+01	4.7989e+01	5.4962e+00
5	5.6954e+01	1.2305e+01	4.0428e+01	1.0409e+01	4.3479e+01	4.9966e+00
4	6.4699e+01	1.3876e+01	5.3908e+01	6.9735e+00	3.8583e+01	6.1923e+00

**Fig. 25** Representation of iso-surfaces for the three-dimensional variation of thermal conductivity  $k(x,y,z)$ **Table 8** The boundary conditions of cube with a 3D material gradation

Boundary condition	
Dirichlet	Neumann
$\phi(0, y, z) = 0$	$q(1, y, z) = -0.2zy(25 + 2y + 3z + zy)$
$\phi(x, 0, z) = 0$	$q(x, 1, z) = -0.1xz(50 + 2x + 6z + 3xz)$
$\phi(x, y, 0) = 0$	$q(x, y, 1) = -0.1xy(50 + 2x + 4y + xy)$

other surfaces are insulated. The boundary conditions of the geometry are shown in Fig. 29.

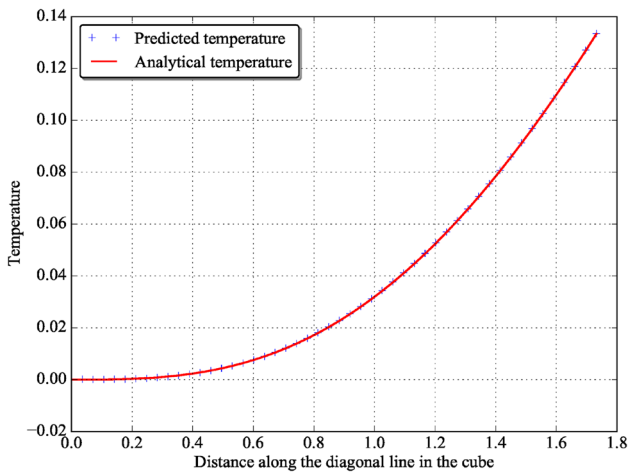
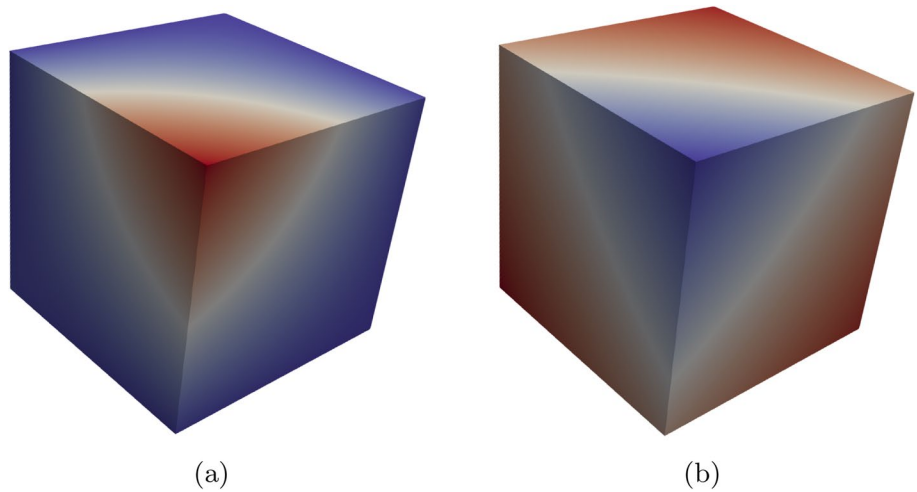
The results of the predicted temperature can be found in Figs. 30 and 31 and is compared to a FEM solution using the commercial software package ABAQUS, as no analytical solution is available for this problem. The temperature along the radial direction at the edge is plotted and compared with results obtained by ABAQUS in Fig. 32.

## 6 Conclusion

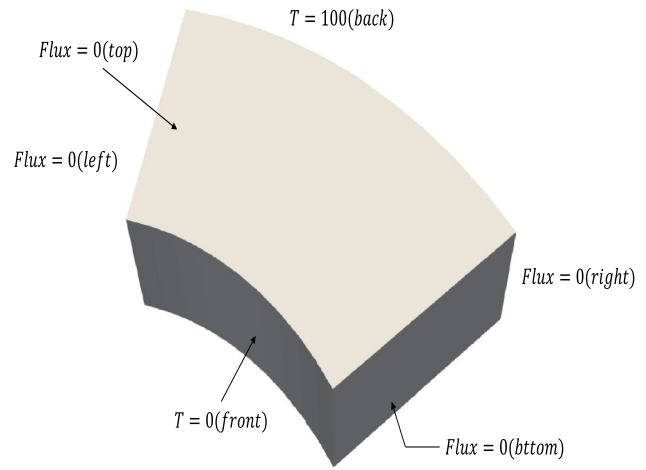
We presented a transfer learning-based deep collocation method (DCM) for solving the problems of potential in non-homogeneous media. It avoids classical discretization methods, such as FEM, and treats the problem as minimization problem, minimizing a loss function which is related to the underlying governing equation. Thanks to the nonlinear activation function, the approach enables us to discover complex nonlinear pattern. The DCM requires sampling inside the physical domain. Therefore, we obtained a suitable sampling method for selected problems. To find the most favorable configuration of the neural network for specific problems, we carried out a sensitivity analysis quantifying the influence of algorithm-specific parameters on specific outputs such as the relative error in the L2 norm. For different material variation forms and material parameters, a material transfer learning is embedded into the framework to enhance the robustness and generality of this deep collocation method. To demonstrate the performance of the proposed DCM, various benchmark problems including the heat transfer and a representative potential problem are studied.



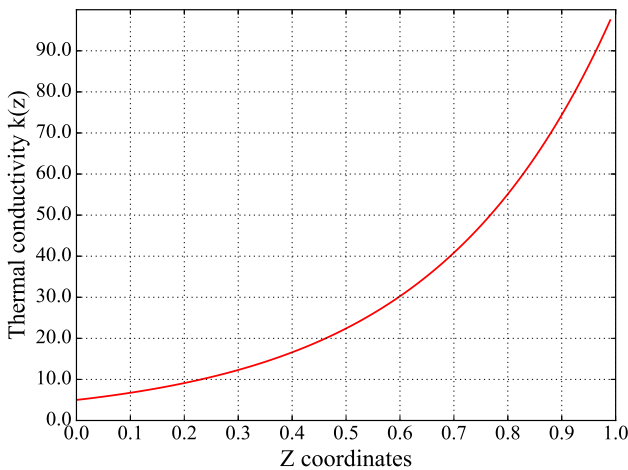
**Fig. 26** Distributions of *a* temperature; and *(b)* flux for the cube with 3D material variation



**Fig. 27** Temperature profile in diagonal line for the cube with a 3D material gradation



**Fig. 29** Annular sector subjected to thermal boundary conditions

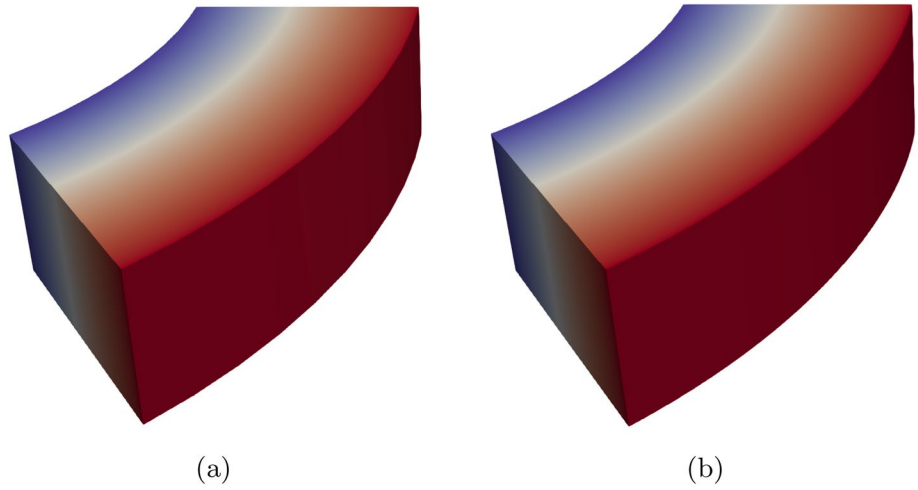


**Fig. 28** Profile of thermal conductivity in *z* direction. The exponential variation of the conductivity is  $k(z) = 5e^{(3z)}$

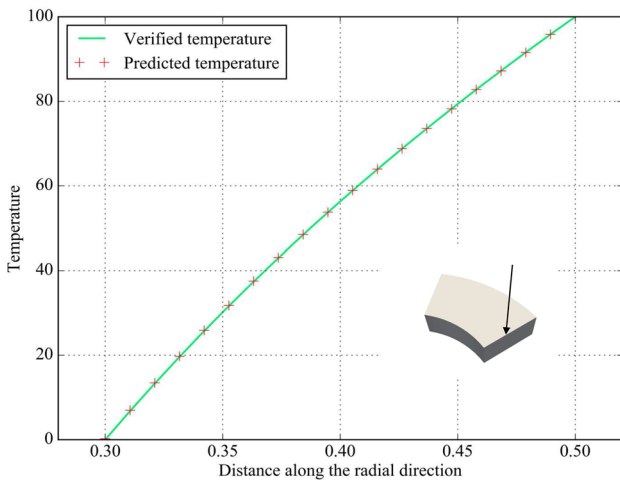
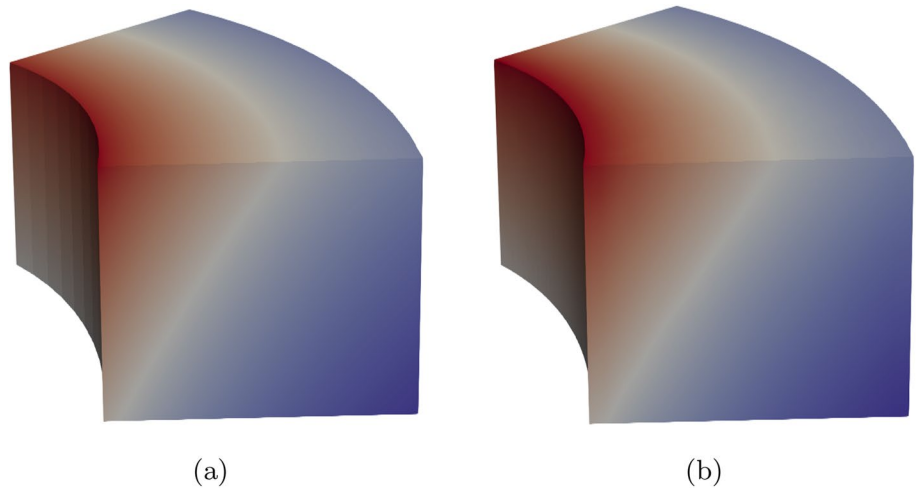
### Appendix A: Data flow for this study

The data flow inside all three modules is depicted in Fig. 33. The first module is based on the DCM. The second module includes a parametric study on the influence of algorithm-specific parameters including numbers of collocation points and parameters for the deep learning configurations on predictive accuracy, which in turn provides guidance for further applications of the DCM. For the parametric analysis, a two-step Morris screening method and EFAST method are adapted providing qualitative and quantitative measures of importance and interaction of DCM-specific parameters. To facilitate and improve the generality and robustness of the presented model, data are finally imported into a material transfer learning model to transfer and expand the learned knowledge between different material variations.

**Fig. 30** Temperature distribution for irregular-shaped FGMs obtained by **a** deep collocation method; and **b** ABAQUS



**Fig. 31** Flux distribution for irregular-shaped FGMs obtained by **a** deep collocation method; and **b** ABAQUS



**Fig. 32** Temperature comparison along the right top edge (indicated by the arrow)

### Appendix B: Activation function and sampling method for comparison

The following table shows a list of classical activation functions and its graphs that studied in this application, which will help to choose a suitable activation for physics-informed neural networks (Table 9).

Various sampling methods are used to generate sequence of points within a cube. The purpose of the sampling method is to generate training datasets for the DCM and improve the training of the network. Proper sampling will help in case the neural network is only trained on fixed points and prevent a biased trained model, which may have a better prediction on random new data (Table 10).

Table 9 Activation function

Table B1: Activation function

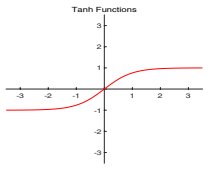
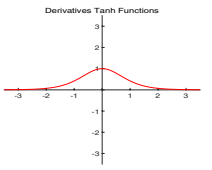
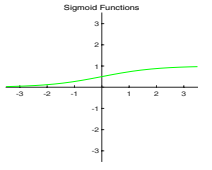
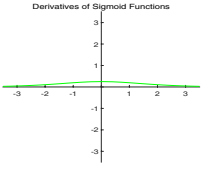
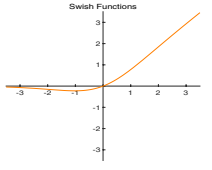
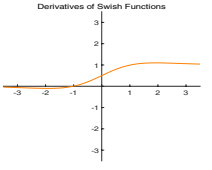
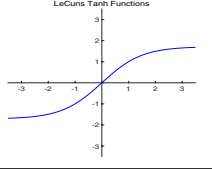
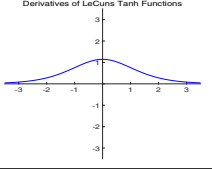
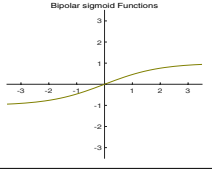
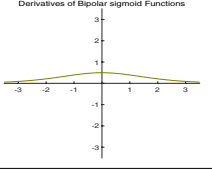
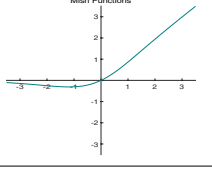
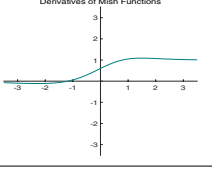
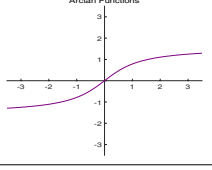
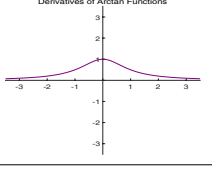
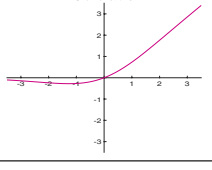

Activation function	Explicit function form	Function figure	Derivatives of function figure
Tanh	$f(x) = \frac{e^{2x}-1}{e^{2x}+1}$		
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$		
Swish	$f(x) = \frac{x}{1+e^{-\beta x}}$		
LeCuns Tanh	$f(x) = 1.7159 \times \tanh(\frac{2}{3}x)$		
Bipolar sigmoid	$f(x) = \frac{e^x-1}{e^x+1}$		
Mish	$f(x) = x \times \tanh(\ln(1 + e^x))$		
Arctan	$f(x) = \tan^{-1}(x)$		
Silu	$f(x) = x \times \text{sigmoid}(x)$		

Table 10 Sampling method

Table B2: Sampling method

Sampling method	points figure	Sampling method	points figure
Latin hypercube		Monte Carlo	
Random		Halton Sequences	
Hammersley Sequence		Korobov Lattice	
Sobol Sequence			

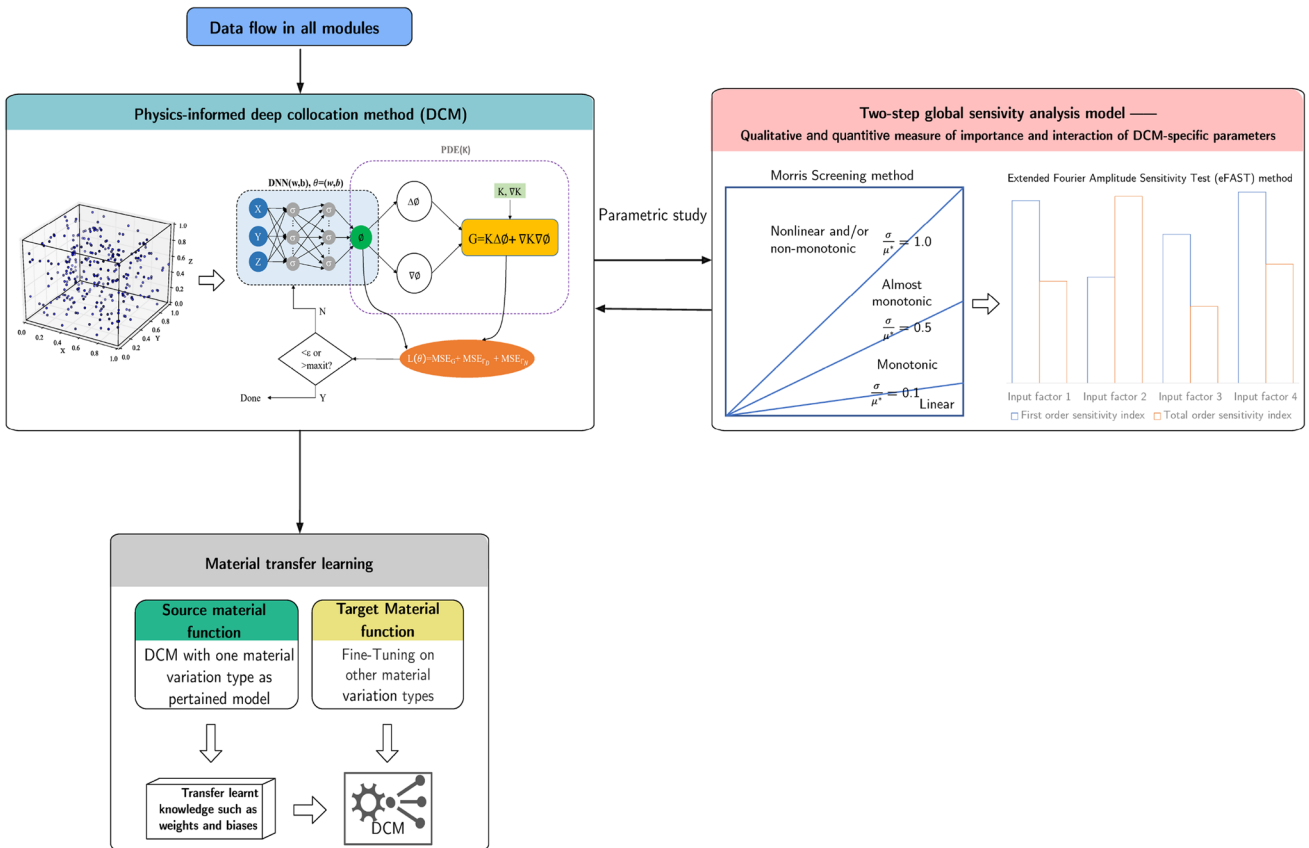


Fig. 33 Data flow in physics-informed deep collocation method and sensitivity analysis

**Acknowledgements** The authors extend their appreciation to the Distinguished Scientist Fellowship Program (DSFP) at King Saud University for funding this work.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
- Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
- Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. In: *Advances in neural information processing systems*, pp 153–160
- Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge
- Patterson J, Gibson A (2017) *Deep learning: a practitioner's approach*. O'Reilly Media, Inc.
- Yang L, MacEachren A, Mitra P, Onorati T (2018) Visually-enabled active deep learning for (geo) text and image classification: a review. *ISPRS Int J Geo-Inf* 7(2):65
- Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172(5):1122–1131
- Ouyang W, Wang X, Zeng X, Qiu S, Luo P, Tian Y, Li H, Yang S, Wang Z, Loy C-C et al (2015) Deepid-net: Deformable deep convolutional neural networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2403–2412
- Zhao Z-Q, Zheng P, Shoutao X, Wu X (2019) Object detection with deep learning. A review. *IEEE Trans Neural Netw Learn Syst*
- Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C, Casper J, Catanzaro B, Cheng Q, Chen G et al (2016) Deep speech 2: end-to-end speech recognition in English and mandarin. In: *International conference on machine learning*, pp 173–182
- Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K (2019) Speech recognition using deep neural networks: a systematic review. *IEEE Access*
- Yue T, Wang H (2018) Deep learning for genomics: a concise overview. [arXiv:1802.00810](https://arxiv.org/abs/1802.00810)
- Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow P-M, Zietz M, Hoffman MM et al (2018) Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 15(141):20170387
- Heaton JB, Polson NG, Witte JH (2017) Deep learning for finance: deep portfolios. *Appl Stochastic Models Bus Ind* 33(1):3–12
- Fischer T, Krauss C (2018) Deep learning with long short-term memory networks for financial market predictions. *Eur J Oper Res* 270(2):654–669
- Gyrya V, Shashkov MJ, Skurikhin AN, Tokareva S Machine learning approaches for the solution of the Riemann problem in fluid dynamics: a case study
- McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5(4):115–133
- Dias FM, Antunes A, Mota AM (2004) Artificial neural networks: a review of commercial hardware. *Eng Appl Artif Intell* 17(8):945–952
- Lagaris IE, Likas A, Fotiadis DI (1998) Artificial neural networks for solving ordinary and partial differential equations. *IEEE Trans Neural Netw* 9(5):987–1000
- Lagaris IE, Likas AC, Papageorgiou DG (2000) Neural-network methods for boundary value problems with irregular boundaries. *IEEE Trans Neural Netw* 11(5):1041–1049
- Kevin SM, James RM (2009) Artificial neural network method for solution of boundary value problems with exact satisfaction of arbitrary boundary conditions. *IEEE Trans Neural Netw* 20(8):1221–1233
- Funahashi K-I (1989) On the approximate realization of continuous mappings by neural networks. *Neural Netw* 2(3):183–192
- Hornik K, Stinchcombe M, White H (1989) Multilayer feed-forward networks are universal approximators. *Neural Netw* 2(5):359–366
- Mhaskar HN, Poggio T (2016) Deep vs. shallow networks: an approximation theory perspective. *Anal Appl* 14(06):829–848
- Weinan E, Han J, Jentzen A (2017) Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun Math Stat* 5(4):349–380
- Han J, Jentzen A, Weinan E (2018) Solving high-dimensional partial differential equations using deep learning. *Proc Natl Acad Sci* 115(34):8505–8510
- Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686–707
- Beck C, Weinan E, Jentzen A (2019) Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *J Nonlinear Sci*
- Sirignano J, Spiliopoulos K (2018) Dgm: a deep learning algorithm for solving partial differential equations. *J Comput Phys* 375:1339–1364
- George EK, Ioannis GK, Lu L, Paris P, Sifan W, Liu Y (2021) Physics-informed machine learning. *Nat Rev Phys* 3(6):422–440
- Anitescu C, Atroshchenko E, Alajlan N, Rabczuk T (2019) Artificial neural network methods for the solution of second order boundary value problems. *Comput Mater Continua* 59(1):345–359
- Guo H, Zhuang X, Rabczuk T (2019) A deep collocation method for the bending analysis of Kirchhoff plate. *Comput Mater Continua* 59(2):433–456
- Guo H, Zhuang X, Chen P, Alajlan N, Rabczuk T (2022) Stochastic deep collocation method based on neural architecture search and transfer learning for heterogeneous porous media. *Eng Comput* 1–26
- Samaniego E, Anitescu C, Goswami S, Nguyen-Thanh VM, Guo H, Hamdia K, Zhuang X, Rabczuk T (2020) An energy approach to the solution of partial differential equations in computational mechanics via machine learning: concepts, implementation and applications. *Comput Methods Appl Mech Eng* 362:112790
- Nguyen-Thanh VM, Zhuang X, Rabczuk T (2020) A deep energy method for finite deformation hyperelasticity. *Eur J Mech A Solids* 80:103874

36. Goswami S, Anitescu C, Chakraborty S, Rabczuk T (2020) Transfer learning enhanced physics informed neural network for phase-field modeling of fracture. *Theor Appl Fract Mech* 106:102447
37. Zhuang X, Guo H, Alajlan N, Zhu H, Rabczuk T (2021) Deep autoencoder based energy method for the bending, vibration, and buckling analysis of Kirchhoff plates with transfer learning. *Eur J Mech A Solids* 87:104225
38. Wenzhen Q, Chen W, Zhuojia F (2015) Solutions of 2d and 3d non-homogeneous potential problems by using a boundary element-collocation method. *Eng Anal Bound Elem* 60:2–9
39. Alves CJS, Chen CS (2005) A new method of fundamental solutions applied to nonhomogeneous elliptic problems. *Adv Comput Math* 23(1–2):125–142
40. Paris F, Canas J (1997) *Boundary element method: fundamentals and applications*, vol 1. Oxford University Press, Oxford
41. Dhingra A Activation functions in neural networks
42. Misra D (2019) Mish: A self regularized non-monotonic neural activation function. [arXiv:1908.08681](https://arxiv.org/abs/1908.08681)
43. Zhang H, Weng T-W, Chen P-Y, Hsieh C-J, Daniel L (2018) Efficient neural network robustness certification with general activation functions. In: *Advances in neural information processing systems*, pp 4939–4948
44. Hornik K (1991) Approximation capabilities of multilayer feed-forward networks. *Neural Netw* 4(2):251–257
45. Raissi M, Perdikaris P, Karniadakis GE (2017) Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. [arXiv:1711.10561](https://arxiv.org/abs/1711.10561)
46. Rafajłowicz E, Schwabe R (2006) Halton and Hammersley sequences in multivariate nonparametric regression. *Stat Prob Lett* 76(8):803–812
47. Wang X, Sloan IH, Dick J (2004) On Korobov lattice rules in weighted spaces. *SIAM J Numer Anal* 42(4):1760–1779
48. Dick J, Pillichshammer F, Waterhouse BJ (2007) The construction of good extensible Korobov rules. *Computing* 79(1):79–91
49. Shields MD, Zhang J (2016) The generalization of Latin hypercube sampling. *Reliab Eng Syst Saf* 148:96–108
50. Shapiro A (2003) Monte Carlo sampling methods. *Handb Oper Res Manag Sci* 10:353–425
51. Iooss B, Lemaître P (2015) A review on global sensitivity analysis methods. In: *Uncertainty management in simulation-optimization of complex systems*. Springer, pp 101–122
52. Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comput Simul* 55(1–3):271–280
53. Cukier RI, Fortuin CM, Shuler Kurt E, Petschek AG, Schai-bly JH (1973) Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory. *J Chem Phys* 59(8):3873–3878
54. Saltelli A, Tarantola S, Chan KP-S (1999) A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41(1):39–56
55. Herman JD, Kollat JB, Reed PM, Wagener T (2013) Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. *Hydrol Earth Syst Sci Discuss* 10(4)
56. Morris MD (1991) Factorial sampling plans for preliminary computational experiments. *Technometrics* 33(2):161–174
57. Garcia Sanchez D, Lacarrière B, Musy M, Bourges B (2014) Combining first-and second-order elementary effects methods. Application of sensitivity analysis in building energy simulations. *Energy Build* 68:741–750
58. Sutradhar A, Paulino GH (2004) A simple boundary element method for problems of potential in non-homogeneous media. *Int J Numer Methods Eng* 60(13):2203–2230

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.