

Numerical Solution and Uncertainty Quantification of Bioheat Transfer Equation Using Neural Network Approach

Ante Lojić Kapetanović*, Anna Šušnjara†, Dragan Poljak‡

*Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
University of Split
Split, Croatia*

*alojic00@fesb.hr, †asusnja@fesb.hr, ‡dpoljak@fesb.hr

Abstract— The paper deals with a novel approach to carry out uncertainty quantification in modeling of bioheat transfer equation using neural networks and deep learning. The output uncertainty is achieved via Monte Carlo (MC) dropout procedure using Bayesian inference, while the input uncertainty propagation is achieved using MC simulation of the ensemble of physics-informed neural networks (PINNs). The proposed approach uses a neural network with integrated physical knowledge without need of any prior assumptions and mesh generation, respectively. Deterministic modeling is related to both analytical and numerical (finite element method) solution of bioheat transfer equation. The computational example considered in this work pertains to the solution of rather simple one-dimensional problem of bioheat transfer, as a kind of opener to the subject.

Index Terms—bioelectromagnetism, bioheat transfer equation, numerical modeling, physics-informed deep learning, uncertainty quantification

I. INTRODUCTION

The Pennes' bioheat transfer equation (PBHTE) is a standard model to tackle the problem of the unknown temperature distribution in the human body tissue [1]. The thermal input parameters of the PBHTE are not unique, i.e. they differ from one individual to another and, consequently, could be considered as stochastic. Therefore, throughout this study, input parameters of the PBHTE are modeled as random variables of a predetermined statistical distribution thus resulting in related output uncertainty. An example of the real-world application of the bioheat modeling is hyperthermia therapy. Hyperthermia therapy (also known as thermal therapy or thermotherapy in literature) is a malignant disease treatment where patient's body is locally exposed to the appropriate amount of heat generated by an external source [2] and is usually an adjunct to chemotherapy and/or radiotherapy. Research has shown that such treatment can kill cancerous cells by damaging proteins and other indispensable structures inside cells [3], keeping the healthy surrounding tissue intact [4]. However, one of the challenges is to expose only local ill area while the rest of the body should not be affected in order to avoid further complications. To ensure a secure application of hyperthermia therapy, bioheat transfer has to be modeled as accurate as possible. Here, the PBHTE is employed even though it has

drawbacks such as relatively crude approximation of heat transfer due to its assumption of simple geometry of heated tissue and compliance to variability of flowrates in thermally significant vessels [5]. Nevertheless, by employing the PBHTE with the knowledge of the output uncertainty, the decision making process in hyperthermia therapy will be more secure and precise.

This study presents a novel uncertainty quantification approach using neural network modeling and deep learning. The last decade has been marked by the rapid development of numerous branches in deep learning - object recognition in computer vision [6], natural language processing and speech recognition, bioinformatics [7] and reinforcement learning [8]. Unlike mentioned regimes where the state of the art performance is achieved using deep neural networks and large sets of data, big data acquisition for scientific modeling of physical dynamic processes, such as modeling bioheat transfer in the human body tissue, is computationally expensive. However, there is an ongoing research activity aiming to overcome such problem proposing a novel approach named scientific machine learning [9]. The ultimate idea is to integrate knowledge about prior physical domain into a learning algorithm. The information content of the data an algorithm sees is enriched and may quickly converge to the right solution even with a few training examples. The study is further expanded with an uncertainty quantification based on approximate inference using Monte Carlo (MC) dropout, firstly introduced in [10] with the fundamentals carved in research by MacKay, Neal, Barber and Bishop [11]–[14]. Additionally, the input uncertainty is propagated through ensemble of neural networks each performing an MC dropout using MC simulation.

The paper is organized as follows: Section II outlines the formulation of the PBHTE; in Section III, the classical approach of solving the PBHTE is described; Section IV defines stochastic modeling; Section V gives an overview of the neural network based numerical model and stochastic approach; Section VI is dedicated to the computational results and finally conclusion and discussion together with possible further work are presented in Section VII.

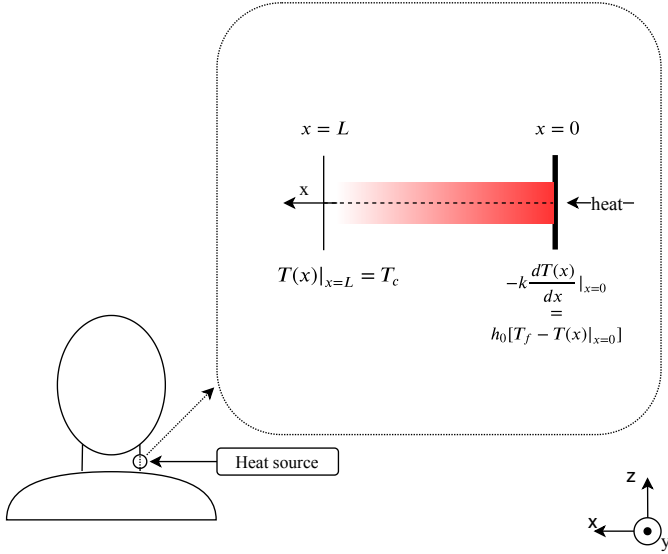


Fig. 1. The geometry of the problem - heat generation onto the localized body area.

II. FORMULATION

The PBHTE, governing the temperature distribution in the human body tissue, is given with the following expression:

$$\nabla(k\nabla T) + \omega_b \cdot T_a - \omega_b \cdot T + Q_m + Q_{ext} = c_v \rho \frac{\partial T}{\partial t} \quad (1)$$

where ρ , c , k are respectively the density, the specific heat, and the thermal conductivity of the tissue; ρ_b and c_b denote density and specific blood heat; ω_b is the blood perfusion; T_a is the arterial temperature, and T stands for the tissue temperature. Q_m is the metabolic heat and Q_{ext} is the heat source due to spatial heating [15].

For the sake of simplicity, instead of the general form PBHTE given in (1), the steady state will be considered. Therefore, the time-invariant PBHTE can be written as:

$$\nabla(k\nabla T) + \omega_b \cdot T_a - \omega_b \cdot T + Q_m + Q_{ext} = 0 \quad (2)$$

Finally, as an opener to the subject, a rather simple one-dimensional case is analyzed. Hence, a solution domain of interest Ω is defined as the one-dimensional space $x \in [0, L]$, $x \subseteq \Omega$, $\Omega \subset \mathbb{R}^1$, and L is the tissue depth, here considered to be 3 cm. The geometry of the problem is depicted in Fig. 1. Note that the heat is generated perpendicular to the skin and there is no influence of the external heat source at tissue depth L . The corresponding one-dimensional equivalent of (2), with no external heat generated onto the skin, is given by:

$$k \frac{d^2 T(x)}{dx^2} + \omega_b \cdot T_a - \omega_b \cdot T(x) + Q_m = 0 \quad (3)$$

with Robin and Dirichlet boundary conditions:

$$\begin{cases} -k \frac{dT_0(x)}{dx} = h_0[T_f - T_0(x)], & \text{for } x = 0 \\ T_L(x) = T_c, & \text{for } x = L \end{cases} \quad (4)$$

where h_0 , T_c and T_f are the convection coefficient, the body core temperature and the atmospheric temperature, respectively.

The input thermal parameters k , ω_b , T_a , Q_m , h_0 and T_f are treated as uniformly distributed random variables with 20% variation coefficient [16]. The nominal values of all input parameters are given in Table I. Within the framework

TABLE I
THE EXPECTED VALUES FOR THE THERMAL INPUT PARAMETERS

Parameter	Expected value
ρ, ρ_b	1000 kg/m ³
c, c_b	4200 J/kg°C
T_a, T_c	37°C
T_f	25°C
k	0.5 W/m°C
ω_b	2100 kg/sm ³
Q_m	33800 W/m ³
h_0	10 W/m ² °C
h_f	100 W/m ² °C

of the deterministic modeling the one dimensional problem formulated by (3) and (4) can be solved via analytical and numerical procedure, respectively.

III. ANALYTICAL AND NUMERICAL SOLUTION

The accuracy of the deterministic model and the novel neural network optimization method is verified by the corresponding analytical solution. For the sake of completeness, the analytical solution is given in Appendix A, while the mathematical details could be found elsewhere, e.g. in [15]. Furthermore, the numerical solution of (3) via Finite Element Method (FEM) is outlined in Appendix B. The full mathematical description is available in a variety of references, e.g. in [17].

It is well-known fact that FEM, as very powerful numerical tool, requires a generation of mesh (which appears to be pretty demanding for highly irregular and non-homogeneous computational domains), a fine grained solution domain approximation where linear combination of polynomials is used to express a local solution over a finite element whose accuracy appreciably depends on the resolution of the mesh for a particular problem of interest. The temperature distribution obtained both analytically and via 20 finite elements using linear interpolation over an element is shown in Fig. 2. Here, the nominal values of input parameters from Table I are used.

IV. STOCHASTIC MODELING VIA MONTE CARLO METHOD

The PBHTE output uncertainty is quantified by implementing the MC sampling method, widely shown to be robust and accurate approach in uncertainty quantification. Generally, the MC sampling method is purely statistical method that relies on repeated random sampling of input parameters, based on some predefined probability distribution. For each sample, the input parameters are fixed and the problem becomes deterministic. The ensemble of solutions is obtained by iteratively applying

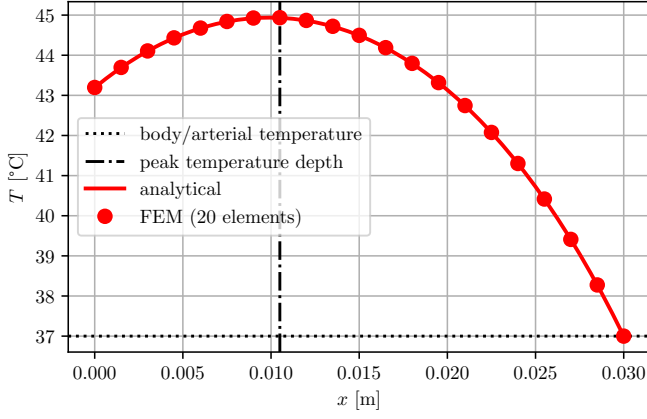


Fig. 2. FEM solution using linear test and base functions.

a numerical method to each sample of randomly generated input parameters. From here, the statistics such as mean and variance can be extracted. The biggest disadvantage of the MC sampling method lies in its iterative nature where, in order to acquire high accuracy on the output, a large number of samples is required [18].

In this paper, 100 realizations with 100 samples each has been used within the framework of the MC method and for each method the results have been obtained using analytical solution. The total uncertainty quantification of the ensemble of results is depicted in Fig. 3. It can be observed that the results obtained via MC sampling method are normally distributed, even though, in this case, samples have been obtained from uniformly distributed input thermal parameters. This phenomenon, shown in Fig. 4, is well-known in statistical analysis and stems from the *central limit theorem*.

V. NUMERICAL SOLUTION AND STOCHASTIC MODELING USING PHYSICS-INFORMED NEURAL NETWORKS

The most of engineering and physics dynamical real-world problems are described through non-linear partial differential equations (PDEs). Even though neural network based numerical solvers for ordinary differential equations (ODEs), a system of ODEs and PDEs were developed in late 1990s by Lagaris et al. [19], assuming that neural networks in fact are universal function approximators [20], only recently physics-informed neural networks (PINNs), introduced by Raissi et al. in [21], have reached full capacity by exploiting advances of the automatic differentiation.

In this study, DeepXDE Python library [22] is used because of its high-level abstraction syntax, which closely resembles the mathematical formulation without losing the usability and customizability. The first step is to represent the solution $T(x)$ with a neural network $\hat{T}(x)$ where x is the spatial coordinate of the PBHTE and the input data for the PINN. This PINN is a fully connected multi-layer perceptron (MLP), which applies linear and nonlinear transformations to the inputs recursively. It is consisted of a single unit input layer, three hidden layers with fifty units each, and a single unit output layer. Each unit

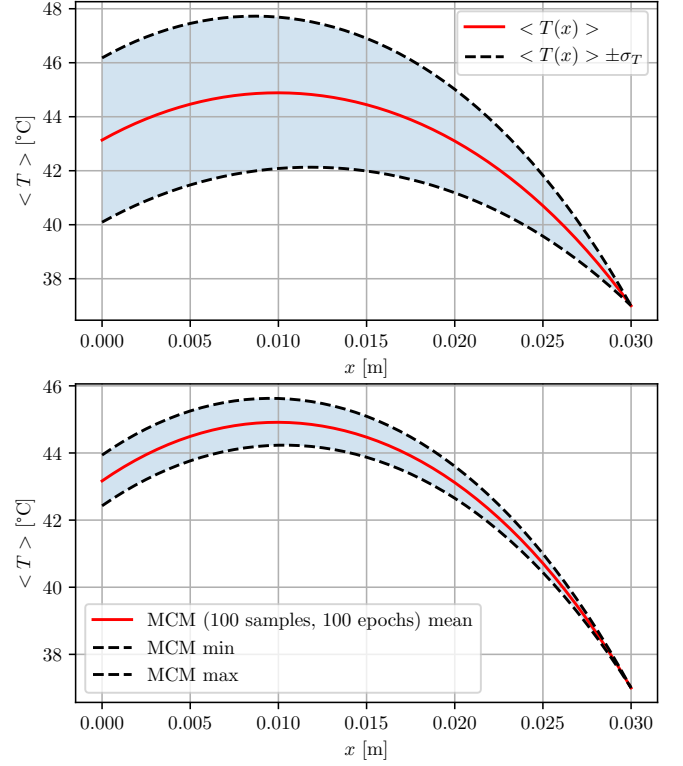


Fig. 3. The upper plot depicts single sample expectation inside 95% confidence interval and lower plot depicts Monte Carlo simulation with 100 iterations.

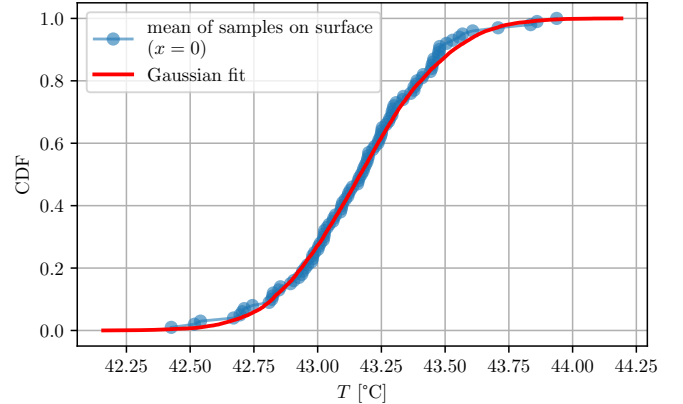


Fig. 4. The visual proof of central limit theorem: expected values of each sample for the $x = 0$ are shown as cumulative distributed function and are compared with Gaussian distribution.

activates its input using hyperbolic tangent activation function. Regularization is applied through the embedded underlying physical knowledge and by adding dropout after each layer. The latter regularization procedure enables quantification of the output uncertainty, which is discussed subsequently in the section. The full neural network $\hat{T}(x; \theta)$ architecture, where θ is the vector of neural network parameters, along with the restriction of the neural network to satisfy the physics that

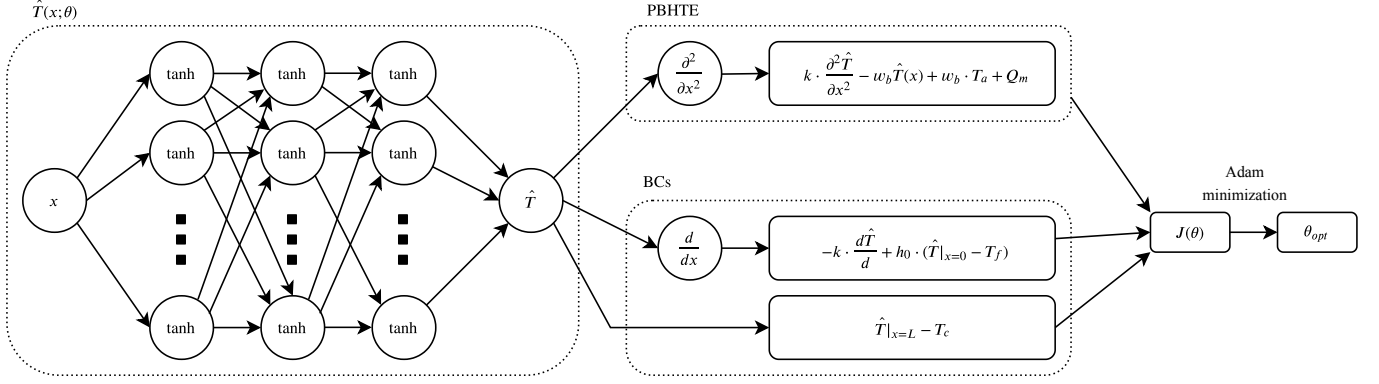


Fig. 5. PINN configuration for solving the PBHTE with mixed boundary conditions, where the multi-layer perceptron is function $T(x)$ approximator, after which PDE and boundary conditions are defined and finally loss function $J(\theta)$ is constructed.

comes from underlying dynamics of the PBHTE and known boundary conditions is shown in Fig. 5.

PINNs, at their core, are mesh-free PDE solvers and are learned by a gradient-based optimizer, where the end goal is the minimization of the loss function. The loss function can be written as the weighted summation of the l^2 norm of residual points for the solution domain τ_d and for the boundary τ_b :

$$J(\theta; \tau) = w_d \cdot \frac{1}{\tau_d} \sum_{x \in \tau_d} \left\| k \cdot \frac{\partial^2 \hat{T}}{\partial x^2} - w_b \cdot \hat{T} + w_b \cdot T_a + Q_m \right\|_2^2 + w_b \cdot \frac{1}{\tau_b} \sum_{x \in \tau_b} \left(\left\| -k \cdot \frac{d\hat{T}}{dx} + h_0 \cdot (\hat{T}|_{x=0} - T_f) \right\|_2^2 + \left\| \hat{T}|_{x=L} - T_c \right\|_2^2 \right) \quad (5)$$

where w_d and w_b are weights initialized using Xavier initialization scheme [23]. τ_d represents the set of randomly scattered data points in the solution domain Ω , captured as the training data. Having constructed the loss function, training is performed using Adam, the first-order stochastic gradient-based optimization algorithm, introduced in [24].

Since the problem of solving the PBHTE is transformed to the optimization problem, solutions will slightly differ for each run. Absolute precision is not possible due to the non-convex nature of the loss function, but high precision is possible using fine tuned hyperparameters such as the learning rate, the number of residual points, the number of layers and units per layer.

Uncertainty quantification of the output is achieved using dropout method which "drops" units from the network randomly with some probability p . Dropout procedure is performed for the regularization purposes and is carried out during the training period. By using dropout during test time, the stochastic prediction on the output is achieved, where the mean of the output is directly estimated by the MC method:

$$\mathbb{E}(T|x; \theta) \approx \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \hat{T}(x) \quad (6)$$

where N_{test} is the total number of test samples.

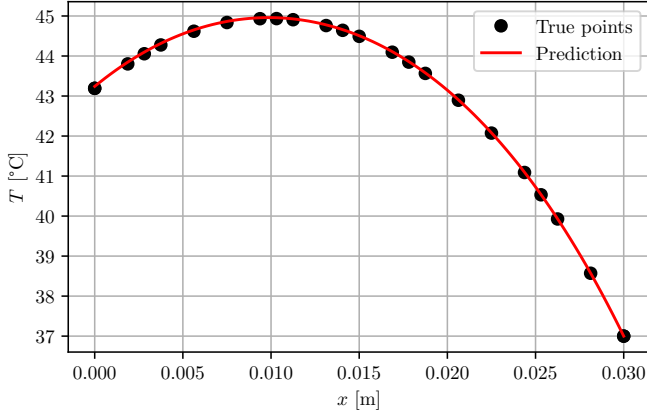
VI. RESULTS AND DISCUSSION

The number of training points is chosen in order to match the number of finite elements for FEM, Section III of the paper. The neural network was trained using 20 training points in the solution domain Ω with integrated underlying physical knowledge for 10,000 epochs. The obtained results are shown in Fig. 6. The regression at the defined interval across 100 test points is shown in Fig. 6a, while Fig. 6b shows the expectation of the neural network output located within the 95% confidence interval (CI). The regression line is in excellent agreement with true values obtained analytically. The learning curves, depicted in Fig. 7, show that a small number of training points is sufficient to achieve low values of the loss function on test set.

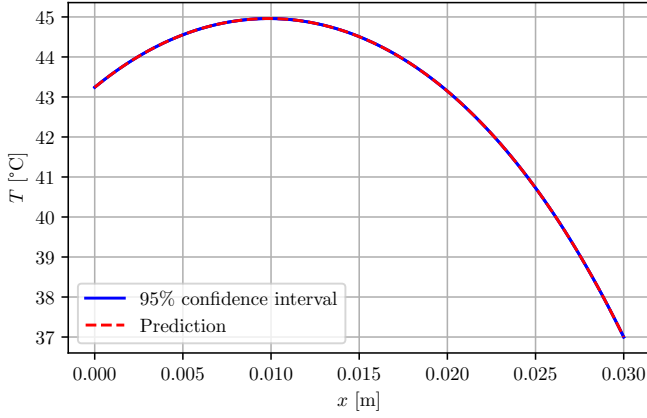
MC dropout, when applied during the testing time provides uncertainty quantification of the output, is not satisfactory method for uncertainty analysis when the input thermal parameters of the PBHTE are treated as random variables, as stated in Section II. In order to tackle this problem, the neural network ensemble is trained using an MC method where input parameters are considered uniformly distributed with respect to the nominal mean value and 20% variation coefficient, and are sampled 5 times with 5 different expectation values in each sampling. Overall, 25 neural networks were trained on the sampled data and the final output is shown in Fig. 8.

VII. CONCLUSION

This paper deals with a novel approach to uncertainty quantification in modeling of PBHTE using neural networks and deep learning. The work aims to demonstrate the ability of the neural network as the universal function approximator to successfully converge to the correct solution on the given solution domain, with extremely low error measured on a test set. The R2 score is 99.9863% measured on 100 test points. Uncertainty of the output is achieved using MC dropout procedure. This method uses Bayesian inference where the probability distribution of the output is based on the prior



(a)



(b)

Fig. 6. The neural network output within 95% CI after 10,000 training iterations. The hyperparameter tuning is performed manually, where the best results are obtained using the learning rate of 0.001 and l_2 relative error as loss metrics.

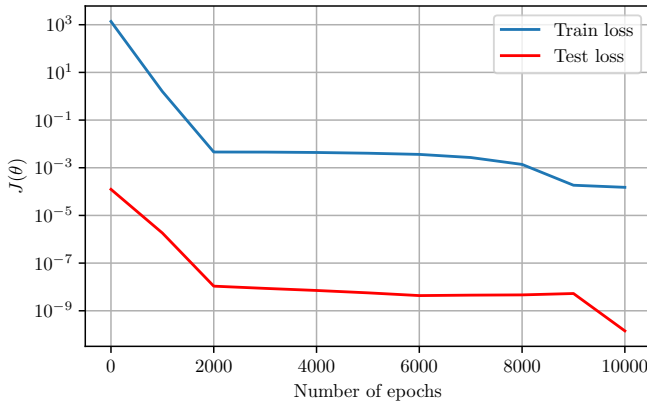


Fig. 7. Learning curves show the values of the loss function during training and testing with respect to number of training epochs.

distribution of hyperparameters and the likelihood of the output. The input uncertainty propagation is achieved using the ensemble of PINNs trained on uniformly distributed input

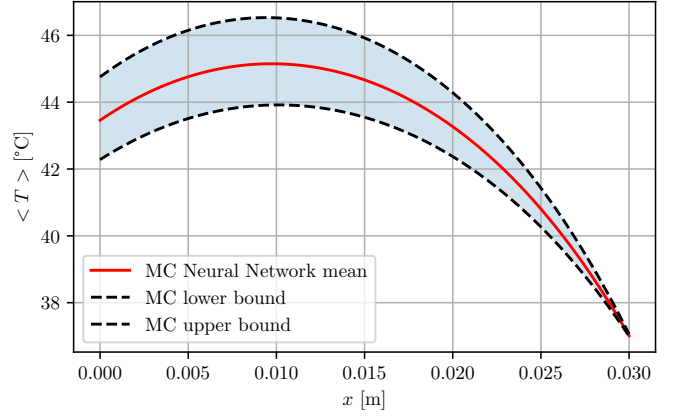


Fig. 8. Uncertainty propagation via MC method. Resulting graph shows expected mean value using red regression line inside the confidence interval constructed as area between the outputs with the lowest and highest value.

parameters using MC method.

The main advantage of the approach proposed in this paper over the classic numerical methods is the construction of a neural network with integrated physical knowledge without any prior assumptions and without the need of mesh generation, respectively. Moreover, instead of transforming the PBHTE to an algebraic system, PINNs engrave both the equation and boundary conditions into the loss function, thus transforming the problem of modeling the heat transfer in biological tissues to the optimization problem. The method does not depend neither on the complexity of the computational domain geometry nor on the dimensionality of the problem. The future work will directly address the problems of solving multidimensional problems of complex geometries in the field of (bio)electromagnetism.

ACKNOWLEDGMENT

This research has been funded by DATACROSS project of The Centre of Research Excellence for Data Science and Advanced Cooperative Systems (CRE ACROSS-DataScience).

APPENDIX A

ANALYTICAL SOLUTION OF THE PBHTE

Analytical solution of the BPHTe, defined by (3), with mixed boundary conditions, Robin's for $x = 0$ and Dirichlet's for $x = L$, is given in [15]. This solution represents the temperature distribution for the steady state of the human body tissue and it is written in the following equation:

$$T_0(x) = T_a + \frac{Q_m}{\omega_b \rho_b c_b} + \frac{\left(T_c - T_a - \frac{Q_m}{\omega_b \rho_b c_b}\right) \left[\sqrt{A} \operatorname{Ach}(\sqrt{A}x) + \frac{h_0}{k} \operatorname{sh}(\sqrt{A}x)\right]}{\sqrt{A} \operatorname{Ach}(\sqrt{A}L) + \frac{h_0}{k} \operatorname{sh}(\sqrt{A}L)} + \frac{\frac{h_0}{k} (T_f - T_a - \frac{Q_m}{\omega_b \rho_b c_b}) \cdot \operatorname{sh}[\sqrt{A}(L-x)]}{\sqrt{A} \operatorname{Ach}(\sqrt{A}L) + \frac{h_0}{k} \operatorname{sh}(\sqrt{A}L)} \quad (\text{A.1})$$

with

$$A = \frac{\omega_b \rho_b C_b}{k} \quad (\text{A.2})$$

APPENDIX B

SOLUTION OF THE PBHTE VIA FINITE ELEMENT METHOD

According to the standard FEM algorithm, featuring weighted residual approach, the PBHTE (3) is multiplied by the set of weight functions w_j and the following expression is obtained

$$\int_0^L \left[k \cdot \frac{\partial^2 T}{\partial x^2} - \omega_b \cdot T + \omega_b \cdot T_a + Q_m \right] w_j(x) dx = 0 \quad (\text{B.1})$$

where, w_j are weight functions for $j = 0, 1, 2, \dots$

Temperature values over the solution domain can then be approximated as a polynomial expansion over a set of basis functions:

$$\hat{T}(x) = \sum_i \hat{T}_i \cdot N_i(x) \quad (\text{B.2})$$

N_i are basis functions for $i = 0, 1, 2, \dots$

Inserting (B.2) into weighted residual integral (B.1), utilizing the weak formulation of the problem and applying the FEM discretization with Galerkin-Bubnov procedure, the following expression is obtained:

$$\begin{aligned} \sum_i^M \hat{T}_i \left\{ k \cdot N_j(x) \cdot \frac{\partial N_i(x)}{\partial x} \Big|_{\text{boundary}} \right. \\ \left. - \int_{x_i}^{x_i+\Delta x} k \cdot \frac{\partial N_i(x)}{\partial x} \cdot \frac{\partial N_j(x)}{\partial x} dx \right. \\ \left. - \int_{x_i}^{x_i+\Delta x} \omega_b \cdot N_i(x) \cdot N_j(x) dx \right. \\ \left. + \int_{x_i}^{x_i+\Delta x} (\omega_b \cdot T_a + Q_m) N_j(x) dx \right\} = 0 \end{aligned} \quad (\text{B.3})$$

for $i = 1, 2, \dots, M$, and $j = 1, 2, \dots, M$ where M stands for total number of elements. N_i denotes the linear shape functions, while x_i and $x_i + \Delta x$ are local points of i -th element and $\Delta x = \frac{L}{M}$ is the length of an 1-D element.

From (B.3), global matrix system is assembled, which can be, for convenience, written, as follows

$$\mathbf{LHS} \times \mathbf{T} = \mathbf{RHS} \quad (\text{B.4})$$

where **LHS** is

$$\begin{aligned} \mathbf{LHS}_{ji}^k = \int_{x_i}^{x_i+\Delta x} k \cdot \frac{\partial N_i(x)}{\partial x} \cdot \frac{\partial N_j(x)}{\partial x} dx \\ - \int_{x_i}^{x_i+\Delta x} \omega_b \cdot N_i(x) \cdot N_j(x) dx \end{aligned} \quad (\text{B.5})$$

and **RHS** is given by

$$\mathbf{RHS}_j^k = - \int_{x_i}^{x_i+\Delta x} (\omega_b \cdot T_a + Q_m) \cdot N_j(x) dx \quad (\text{B.6})$$

The vector of values for temperature distribution is obtained by solving the matrix system (B.4) after incorporating the prescribed boundary conditions.

REFERENCES

- [1] H.-W. Huang and T.-L. Horng, "Chapter 1 - bioheat transfer and thermal heating for tumor treatment," in *Heat Transfer and Fluid Flow in Biological Processes*, S. M. Becker and A. V. Kuznetsov, Eds. Boston: Academic Press, 2015, pp. 1 – 42.
- [2] (2011, August) Hyperthermia in cancer treatmentnational cancer institute. [Online]. Available: <https://www.cancer.gov/about-cancer/treatment/types/surgery/hyperthermia-fact-sheet>
- [3] B. Hildebrandt, P. Wust, O. Ahlers, A. Dieing, G. Sreenivasa, T. Kerner, R. Felix, and H. Riess, "The cellular and molecular basis of hyperthermia," *Critical Reviews in Oncology/Hematology*, vol. 43, no. 1, pp. 33 – 56, 2002.
- [4] J. van der Zee, "Heating the patient: a promising approach?" *Annals of Oncology*, vol. 13, no. 8, pp. 1173 – 1184, 2002.
- [5] K. R. Foster, M. C. Ziskin, and Q. Balzano, "Thermal Response of Human Skin to Microwave Energy: A Critical Review," *Health Physics*, vol. 111, no. 6, pp. 528–541, 2016.
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems* 25, 2012.
- [7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, 5 2015.
- [8] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, 2016.
- [9] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, and A. Ramadhan, "Universal differential equations for scientific machine learning," 2020.
- [10] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," 2015.
- [11] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.
- [12] R. M. Neal, "Bayesian learning for neural networks," Ph.D. dissertation, University of Toronto, 1995.
- [13] D. J. MacKay, "Choice of basis for laplace approximation," vol. 33, 1 1998.
- [14] D. Barber and C. Bishop, "Ensemble learning in bayesian neural networks," pp. 215–237, January 1998.
- [15] Z.-S. Deng and J. Liu, "Analytical Study on Bioheat Transfer Problems with Spatial or Transient Heating on Skin Surface or Inside Biological Bodies," *Journal of Biomechanical Engineering*, vol. 124, no. 6, pp. 638–649, 2002.
- [16] A. Susnjara, D. Poljak, F. Rezo, and J. Matkovic, "Stochastic sensitivity analysis of bioheat transfer equation," 5 2019.
- [17] D. Poljak, "Chapter 7 - on the use of boundary integral methods in bio-electromagnetics," in *Numerical Methods and Advanced Simulation in Biomechanics and Biological Processes*, M. Cerrolaza, S. J. Shefelbine, and D. Garzón-Alvarado, Eds., 2018, pp. 119 – 143.
- [18] D. Xiu, "Fast numerical methods for stochastic computations: A review," *Communications in Computational Physics*, vol. 5, pp. 242–272, 03 2009.
- [19] I. E. Lagaris, A. Likas, and D. I. Fotiadis, "Artificial neural networks for solving ordinary and partial differential equations," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, 1998.
- [20] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, 1989.
- [21] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, 11 2018.
- [22] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis, "DeepXDE: A deep learning library for solving differential equations," *arXiv preprint arXiv:1907.04502*, 2019.
- [23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 9, 2010.
- [24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.