Journal of Applied Artificial Intelligence
https://jaai.sabapub.com
ISSN: 2709-5908

2023 Volume 4, Issue 1: 21–44

DOI: https://doi.org/10.48185/jaai.v4i1.838

Hybrid Neural Network Models for the Optimization of Induction Hardening Processes

Salman Lari[1], Jong-Moon Kim [2], Hyock Ju Kwon[1,*]

[1]Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, N2L 3G1, Canada
[2]Innotown Planning Department, Korea Electrotechnology Research Institute, Korea

Received: 02.08.2023 • Accepted: 25.09.2023 • Published: 07.10.2023 • Final Version: 07.10.2023

Abstract: We describe a simple hybrid methodology to simulate an induction heating process that combines observational (black-box) and physics-based (white-box) methodologies. This method uses a neural network to predict the process' physical characteristics, which were previously unknown. A primary emphasis is placed on monitoring temperature variations within a subsurface layer of a bolt sample. The hybrid model incorporates an ordinary differential equation for the heating rate, leading to improved data accuracy compared to a standalone black-box model. Implementing hybrid models results in higher accuracy and less total error in final temperature prediction ($\simeq 10℃$) which outperforms residual neural network with higher total error ($\simeq 32℃$). This innovative approach not only improves predictive precision but also simplifies interpretability, ultimately serving as a pivotal instrument for the effective management and advancement of induction heating operations.

Keywords: Induction heating, Electromagnetic, Hybrid modeling, Neural networks

1. Introduction

Induction hardening stands a specialized surface treatment, involving the application of induction heating to a metal component's surface using induction, followed by rapid quenching. This sequential procedure triggers a martensitic transformation, resulting in the hardening of a thin layer (typically ranging from 0.25 to 2.5 mm) in the workpiece, commonly composed of steel or cast iron. Notably, this transformation exclusively affects the surface while preserving the overall properties of the component. The process intricately intertwines electromagnetic, thermal, mechanical, and metallurgical aspects, thereby presenting considerable challenges for accurate modeling and simulation. Addressing the complex nonlinear equations and establishing the right boundary conditions necessitates an in-depth grasp of the various involved factors. However, crucial information is often missing, adding complexity to refining and optimizing the induction hardening process.

The finite element method (FEM) stands as a predominant technique in simulating intricate physical challenges, notably in contexts like induction hardening. Commonly referred to as "White-Box modeling" [1], FEM provides an in-depth, physics-driven representation of intricate interactions within multifarious systems. It employs constitutive equations to comprehensively detail various physical domains and their material

* Corresponding Author: hjkwon@uwaterloo.ca

behaviors. The strength of FEM lies in its adeptness at illustrating processes in intricate structures and systems, thus *allowing* for a deeper understanding and meticulous optimization of each interaction.

Despite its advantages, the physical modeling approach using FEM does have certain limitations. One of the main drawbacks is the intensive computational demands it entails. Simulation times can often exceed desirable time frames, making real-time or fast-response applications challenging. Additionally, while the accuracy achieved with FEM is commendable, it may still fall short in certain scenarios, especially when dealing with highly nonlinear or complex phenomena. Another challenge is deriving inverse models for process control, which can be particularly difficult and time-consuming. Inverse modeling involves finding the inputs that lead to a desired output, and it may not always have unique or straightforward solutions. Also, making correct physical models requires a lot of knowledge about the system being modeled and the parts that make it up. This requirement can be problematic for industrial induction hardening, where information on power supply design and instrumental characteristics may be scarce or proprietary. The exact measurement of irregular and temperature-dependent magnetic hysteresis data is one of the problems that have not been answered in physical models of induction hardening. This knowledge is crucial for the electromagnetic subsystem, particularly for the generation of volumetric heat close to the sample surface, which serves as an input for the thermal component of the multiphysics model.

Physical modeling is hard, as shown by models of induction hardening using the FEM. This makes people turn to "Black-Box modeling" methods like artificial neural networks (ANNs). ANNs are adept at mapping intricate nonlinear relationships between processing parameters and subsequent material outcomes [2]. However, while ANNs offer considerable advantages, they demand vast quantities of high-quality data for effective calibration and training, and their potential for extrapolation remains curtailed. Endeavors to optimize and pinpoint parameters within pure black-box models can be intimidating, especially in the absence of mathematical guiding principles [3].

In response to the aforementioned limitations, there has been an emergence of Hybrid Models (HMs). These models combine data-based approaches with deep-rooted system understanding based on physical and conservation laws [4-7]. Such a synthesis not only bolsters the model's extrapolative prowess but also alleviates the data-intensive demands typically associated with neural network training. Intriguingly, despite their applicability in various domains, the exploration of HMs within the context of induction hardening processes remains in its infancy.

In this study, we develop and evaluate dynamic hybrid models (HMs) tailored for predicting temperature variations within a subsurface layer of a cylindrical bolt during induction hardening. The illusive process parameter is crucial to the overall heat equation and is estimated using an ANN as part of the HM architecture's many interconnected parts. This inference is then applied to a simplified version of the heat equation, which is handled as an ordinary differential equation inside the physical framework. Interestingly, HM’s parameter is temperature sensitive and shows nonlinearity with regard to several process factors, including operating power and material parameters. These determinants are extrapolated and pinpointed from training datasets. Subsequently, the physical equation is subjected to mathematical integration in order to ascertain the temporal variations in temperature inside a specific region of the object, sometimes referred to as the deep layer.

Utilizing empirical data derived from an induction heating experimental setup, we systematically trained and fine-tuned two distinct variants of HMs. A comprehensive evaluation of these models ensues, quantifying their effectiveness in explaining and predicting outcomes. Parallel to this, we contrast the prediction skills of these improved HMs with those of a model that relies only on data,

namely recurrent neural networks (RNNs). The discussion is concluded with a comparison of the benefits and drawbacks of HMs to both black-box models and purely physical models. This analysis is particularly relevant within the domain of induction heating process design, optimization, and control.

The following is how the paper is structured: Section 3 discusses the induction hardening technique as well as the mathematical model utilized to solve the concerns with electromagnetic-thermal coupling. The numerical simulations used for data harvesting, which is essential for model training, are described in more detail in Subsection 3.2. Transitioning to Section 4, a detailed exposition of both the HM and RNNs is provided. Section 5 delineates the strategies adopted for model optimization and parameter calibration. Section 6 contains the findings and a comprehensive assessment of the learnt hybrid and RNN models. In Section 7 of this work, we look at how useful the hybrid modeling approach is for models and improvements of the induction hardening process. Concluding, Section 8 encapsulates salient insights and proffers research conclusions.

2. Related Work

Historically, induction hardening processes have predominantly been scrutinized using white box modeling methodologies like mathematical models, FEM, and Multiphysics methods. A notable contribution in this domain was made by Hömberg [8], where the author formulated a mathematical representation for the induction hardening of steel. This model accommodated electromagnetic influences responsible for the heating and thermomechanical impacts leading to the hardening. A more intricate analysis was performed by Barglik et al. [9], wherein they approached induction hardening of gear wheels through nonlinear partial differential equations. This method encompassed the magnetic and temperature field distributions with an assumption that all involved material parameters are temperature dependent.

In the realm of FEM and Multiphysics, Yuan et al. [10] designed a comprehensive finite element analysis model, which encompassed various facets, from the electromagnetic induction heating to the quenching process. Their primary objective was to identify the ideal hardness distribution in the workpiece by modulating the AC current density distribution. In a similar thread, Candeo et al. [11] embarked on a coupled electromagnetothermal simulation targeting the hardening of gear contours. They employed a 3-D FEM Multiphysics technique to achieve this.

Apart from traditional approaches, black-box modeling has also carved a niche in induction hardening research. A significant work in this context was by Wang et al. [12] where they harnessed neural networks that were trained on FE simulation data. Their core objective was to finetune the distribution of eddy currents for optimal temperature control on the work strip's surface. Further, Urbanek et al. [13] ventured into approximating the convection coefficient using ANNs. This coefficient is pivotal in modeling the heat exchange dynamics between the heated sample and its ambient environment. Penha et al. [14] extended the applicability of ANN models to predict the steel's hardness, factoring in tempering time and temperature.

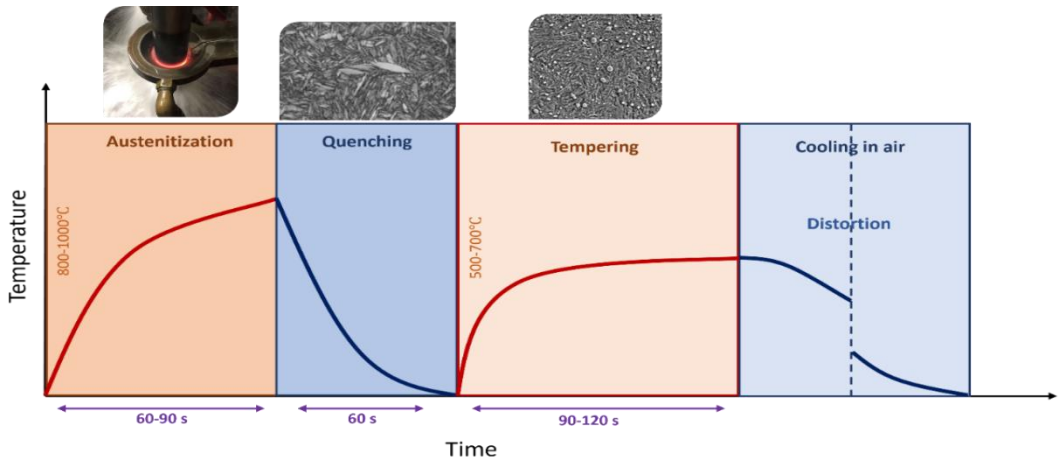
Compared to the previous works, our study claims novelty in its application of hybrid semiparametric models for induction hardening, resonating with the pioneering concept proposed by Psichogios and Ungar [3]. We believe this to be the inaugural endeavor in leveraging such models for this specific process.

3. Induction Hardening

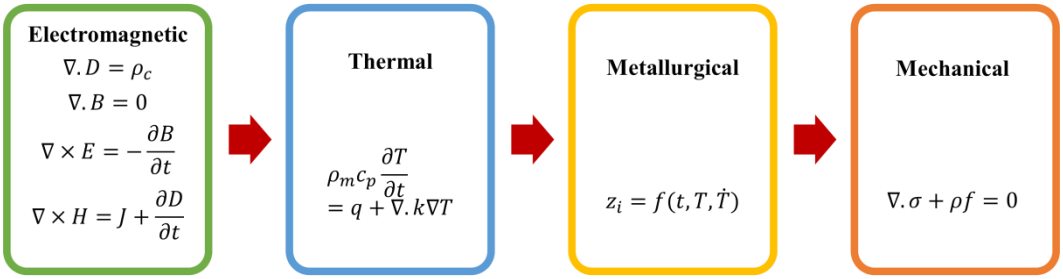
3.1. Background and Principle

Induction hardening is a widely utilized technology in the metals production industry, represented for its efficiency and resource-saving capabilities. Central to this technique is the electromagnetic induction mechanism, which generates heat in metallic objects by inducing eddy currents. The successful execution of induction hardening hinges on the precise calibration of several process parameters, including frequency, current, voltage, dwell time, and the stipulated quenching condition. The choice of these parameters is intrinsically linked to the sample's geometry and the targeted results of the heat treatment. Notably, the dwell time exhibits variability: it spans mere seconds in applications like surface hardening, extends to tens of seconds for specialized applications such as bearing journals, and might even stretch to several minutes when performing through hardening and the following tempering on steel rods [1].

Figure 1(a) illustrates the stages of heating, quenching, and tempering intrinsic to steel rod processing. When the sample is heated above the austenitization temperature, the microstructure undergoes a first-stage transformation into austenite (γ -phase). During this phase, grain expansion, redistribution, and carbon dissolution might take place. Following rapid quenching, martensite is created, which has a high strength but poor toughness. To achieve the desired toughness, a further inductive heating process is used to temper the martensite. Time and temperature during this cooling process have an impact on how precipitates develop and how the internal tensions in the martensitic structure relax, which is crucial for determining the final characteristics of the material.



(a)



(b)

Figure 1. (a) The through-hardening of steel rods as an example using the induction hardening phases of iron-based alloys; (b) the linked physical issues in the induction hardening process.

As shown in Figure 1(b), modeling the induction hardening process requires careful consideration of four interrelated physical issues: electromagnetic, thermal, metallurgical, and mechanical. For the temperature part of models, there are established methods that use the concepts of heat transfer, radiation, and convection. Plasticity models are used to assess residual stresses and distortion for the mechanical facet [15]. On the metallurgical front, challenges revolve around phase transformations, a topic often tackled within the realm of material modeling [16, 17]. On the other hand, there is no conventional method for solving the electromagnetic issue, which involves the electric circuit including the power source, inductor, and the heated component. The main difficulties result from a limited knowledge of the electrical components involved and their operating behavior. Further challenges come from understanding how the circuit, inductor, and sample interact; this interaction may be thought of as a single circuit. This intricate relationship poses substantial obstacles in synchronizing with the other three physical domains. Conventional physics-driven approaches, particularly the finite element methods, falter in precisely modeling the pivotal quantity, q, which is essential for transferring from the electromagnetic domain to the thermal sphere, leading to a deficit in both certainty and quantitative accuracy.

In this research, we propose a solution to these challenges using a hybrid modeling methodology, with an exclusive focus on the primary heating phase, deliberately omitting the quenching and tempering processes. The study delves into the intertwined issues of electromagnetic and thermal domains. Concurrently, this paper elaborates on the intricate endeavor of adeptly regulating temperature for real-time implementations. We leverage the merits of the streamlined, efficient, and reversible architecture of the HM, making it apt for these specific applications.

3.2. The Coupled Electromagnetic Thermal Problems

The interaction between electromagnetic and thermal components in a networked system may be shown using the heat equation:

$$4.\ \rho_m C_P \frac{\partial T}{\partial t} = q + \vec{\nabla}.(\kappa \vec{\nabla} T) \qquad 5.\ (1)$$

The mass density (ρm), specific heat (C_P), and thermal conductivity (κ) of the material are all relevant to linked electromagnetic and thermal issues. Connecting the two problems is the volumetric heat production source term (q), which is a nonlinear function of process factors such as temperature (T), specific heat (C_P), thermal conductivity (κ), mass density (ρ_m), inductor current (I), and generator working power level (P).

Maxwell's equations, which explain the electromagnetic phenomena, must be solved in order to calculate q. FEM models, which include the discretization of the system's governing differential equations, are often employed for this purpose. These techniques, however, often need significant processing resources, which renders them unsuitable for simulating large-scale industrial processes. Moreover, the accuracy of FEM results is contingent upon the underlying assumptions, selected boundary conditions, and incorporated constraints. Specifically for induction heating simulations, the inherent non-linearity of boundary conditions, especially those associated with heat radiation and convection, exacerbates the modeling challenge. Achieving precise representation under these conditions often becomes a meticulous endeavor. Therefore, alternative modeling techniques are required that might offer a balance between computational efficiency and accuracy. Such

methodologies, ideally, should mitigate the limitations of FEM, paving the way for more streamlined modeling and optimization in electromagnetic heating systems.

To predict the temperature development of a bolt sample heated by induction, we suggest a HM in this work. The basic idea is to combine the heat equation's physics-based modeling with a data-driven framework that calculates the process parameter q, which cannot be measured. Subsequent sections will delve into the details of this approach.

3.2 Numerical Simulation and Training Data

This section elucidates the numerical simulation procedure and the data acquisition process undertaken in our research to better understand induction heating dynamics.

Simulation Framework. EMWorks 2023 software was implemented to simulate the interplay between electromagnetic and thermal phenomena during induction heating process. Figure 2 provides a visual representation of the simulated sample's geometry subjected to induction heating, which was used to collect data on the heating process. In power control mode, the generator produces 30 kW at 15 kHz. Adjustments may be made to process factors such operating power, heating time, inductor movement, and quenching procedures using the program's user interface.

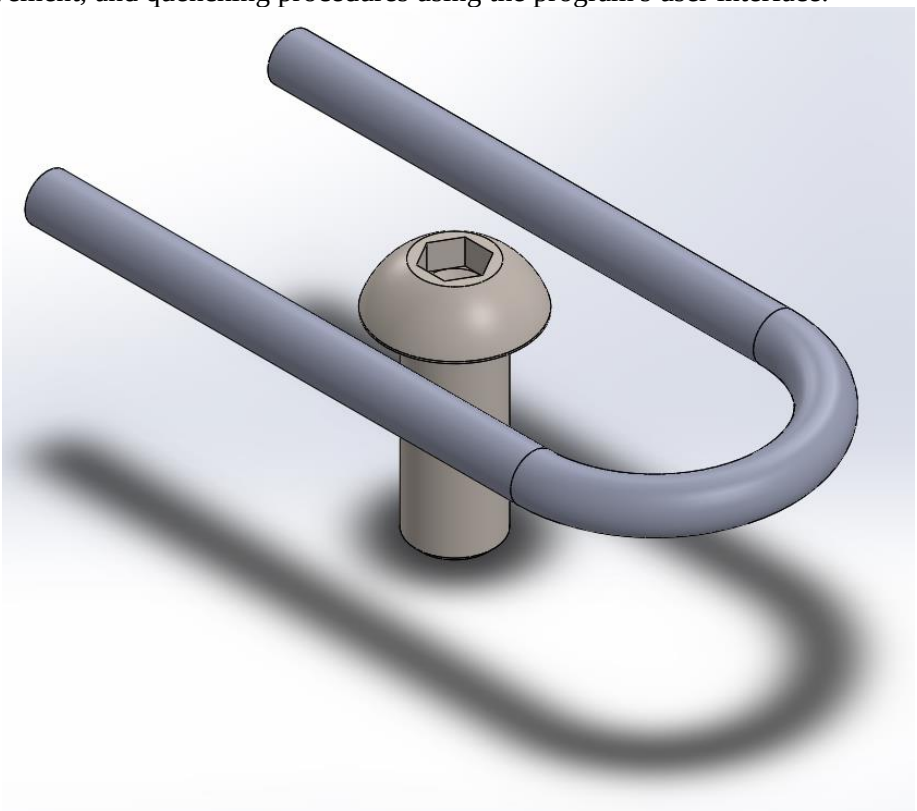


Figure 2. Example geometry for simulating induction heating.

Sample Specifications. For the simulation, we used a bolt sample with a length (L) of 50 mm and a diameter (D) of 10 mm. The sample is made of 50CrMo4 steel grade, with its chemical composition detailed in Table 1 [18]. During the simulation, the sample remained stationary, thus eliminating any dynamic interplay between it and the inductor. The heating mainly happened in a specific section of the sample right below the elliptical inductor, which had a length (l) of 100 mm and an internal

diameter (d) of 50 mm. The average temperature was taken from the sample's top surface in order to quantify temperature change.

Table 1. 50CrMo4's chemical make-up (in weight%)

C	Mn	Cr	Mo	Si	P	S	Fe
0.49	0.71	1.05	0.18	0.27	0.016	0.01	balance

Data Gathering Process. The power level values in Table 2 were used for simulations. The Curie temperature was surpassed before the temperature measurements were recorded. Additionally, we employed water quenching as a cooling method only after the sample had naturally cooled to a certain degree in ambient air. By doing so, the samples weren't subjected to intense thermal and mechanical stress.

Table 2. Simulation power levels and accompanying temperature data collecting.

Power (kW)	**0.9**	**1.5**	**2.1**	**3**	**4.5**	**6**	**7.5**	**9**	**10.5**	**12**	**13.5**	**15**

4. Neural Network and Hybrid Model Structures

For trained models to be accurate and of high quality, feature selection is essential. Several factors play a crucial role in the induction heating process, such as the material properties of the sample, the geometric dimensions of both the inductor and the sample, the power level at which the operation is carried out, the current flowing through the inductor, and the frequency and voltage applied. However, we restricted our attention to power level and temperature for the input layer of characteristics. Given their ease of use, low cost of training, and more generality, we think models trained with these properties are more useful. Our numerical model also allows the operator to enter merely the power level into the control system. So, models that are learned with controlled factors are more useful and can be used in reverse for optimization or smart induction heating [19,20]. Furthermore, because exact material data are sometimes unavailable, it is desirable to train a model without explicitly depending on them.

4.1 Recurrent Neural Network

Mathematical structures known as neural networks may learn and build connections between input and output variables. By using the available data, neural networks have the capability to discover the mapping function that exists between these variables. This allows them to understand complex nonlinear processes, since they possess intrinsic nonlinear characteristics. For many modeling tasks involving sequential data like time series prediction, RNNs are particularly valuable due to their capacity to capture dynamics through internal memory cells. Figure 3 provides an illustration of an RNN cell and its evolution over time. Similar to feedforward networks, RNNs also have neurons link in a sequence to form a directed graph, but they also include an internal loop. A recurrent neuron gets the output from the preceding time step, $h(t-1)$, at each time step t. So, each recurrent neuron has two separate sets of weights: one for the input $X(t)$ and one for the output $h(t-1)$. While acting on unrolled neurons over time, RNN training uses a technique similar to conventional backpropagation. The approach includes forwarding the unrolled network and backpropagating the cost function gradients over time to change network parameters.

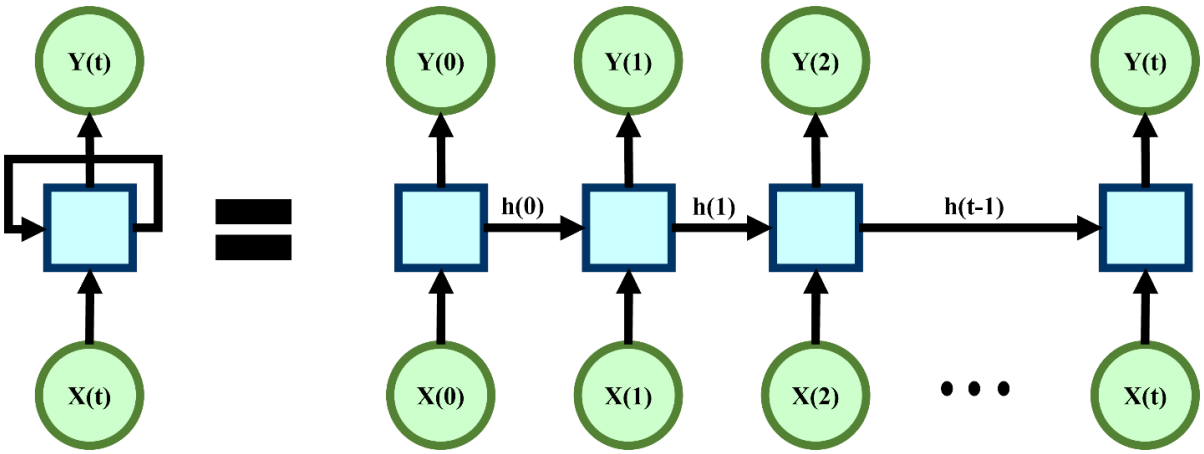


Figure 3. RNN cell and its temporal unrolling.

Depending on how complicated their memory cell is, RNNs display a range of structural variations. In Figure 3, h stands for a single RNN cell's output, which is influenced by both the input received and the state of the cell at the previous time step. Basic RNNs produce the same memory cell as its state, as seen by $h(t) = Y(t)$. Lipton et al. [21] present a thorough and up-to-date overview on RNNs, including all of the many types of RNNs that have been developed recently, including LSTM and GRU. The LSTM and GRU architectures have been specifically developed to effectively handle long sequences, whereby the impact of prior memory plays a crucial role in making accurate predictions. These structures have found use in a variety of fields, including stock market forecasting and natural language processing.

But in this study, we concentrate on modeling a process where the system's initial state is fixed, which is best characterized as an initial value problem. So, it would not make sense to use an RNN model learned on long sequences to extend a starting state that is only known at the start of the dynamics. As a result, we have decided to use deep basic RNNs, which involve stacking many layers of basic RNN cells. This approach enables us to effectively capture the temperature changes within a particular region of a cylindrical sample undergoing induction heating.

4.2 Hybrid Model Structure

Hybrid modeling involves combining the existing physical understanding of a process with data-driven structures [3]. The physical information given in Equation (1) completely controls the temperature changes within a bolt sample during the induction heating method described in Section 3. However, Equation (1)'s heat source term q represents a process variable that cannot be directly identified by study.

Solving the heat equation requires knowledge of q, either through numerical or analytical means. One approach is to employ FEM simulations to solve the electromagnetic problem. Yet, confusing input process control parameters, elusive boundary conditions, and unknown material qualities might cause discrepancies between simulation results and actual observations. To address these challenges, we propose adopting a hybrid modeling approach that melds experimental data with fundamental first principles knowledge. The black-box component of the hybrid model learns the function q, thus complementing the existing knowledge. While it is possible to incorporate all available physical knowledge to model the system's behavior, it is common practice to progressively modulate the extent of knowledge integration. To facilitate the simplification of Equation (1), we neglect the

spatial derivative component associated with conduction, resulting in a modified form of Equation (1).

$$5.\ \frac{dT}{dt} = \alpha(t) \qquad 6.\ (2)$$

The equation α = q / ($\rho m\ C_P$) represents a relationship between variables. The Equation (2) represents a physical model that encompasses the interconnections among process variables. This might provide difficulties or need a substantial amount of data in order to accurately encapsulate these relationships when relying only on data-driven models. The parameter α in the model is gleaned from simulation data. The streamlined model does not explicitly depend on specific material qualities, which is important to note. This is helpful when it is challenging to get reliable material data of the highest quality. Through the appropriate use of the parameter α, the trained hybrid model successfully accommodates dependence on material attributes from experimental examples.

The simplified lumped model of a bolt sample layer's temperature variations in Equation (2) implies black body radiation's cooling effect on dynamics. It is customary to take into account the effect of black body radiation together with the boundary conditions while analyzing Eq. (1). Nevertheless, inside our physical model, we include this component directly into the reduced heat equation (Equation 2) by using a lumped mass method. This inclusion shows that partial process information improves model performance while maintaining the same training data. The addition of this additional term complies with the Stefan-Boltzmann equation and becomes significant in slow operations, as those carried out at low power levels. If the heating rate stays low, the sample will have plenty of time to cool by radiation under these conditions. The incorporation of this modicum of knowledge diminishes the necessity for conducting or amassing numerous simulations at low power levels for model training. The benefits and drawbacks of physics-based and data-driven models are skillfully balanced by hybrid models. The more advanced physical knowledge is what:

$$7.\ \frac{dT}{dt} = \alpha(t) - B((T + 273.15)^4 - (T_{amb} + 273.15)^4) \qquad 8.\ (3)$$

In this context, T_{amb} denotes the prevailing ambient temperature, while B represents the coefficient that characterizes the rate of heat dissipation from the surface of the bolt. For our specific sample shape and numerical simulation, the value of B is about 63 × 10-14 °C/(s K4). This estimate is based on a heuristic evaluation of heat loss. Theoretically, B relies on temperature and process variables, therefore optimization like α is required. However, for the sake of simplicity in this study, B is approximated as a constant. The precise value of B is calculated as follows:

$$9.\ B = \frac{4\epsilon\sigma_B}{D\rho_m C_P} \qquad 10.\ (4)$$

The symbol ϵ represents the emissivity of the dark body, while B is the Stefan-Boltzmann constant. D refers to the width of the sample, ρ_m represents the mass density, and C_P denotes the specific heat. Equation (4) has a temperature-dependent B component that might alter when the material is heated. Additionally, differences in emissivity with temperature are possible. Therefore, if realistic results are required, optimization should be carried out taking into account the temperature dependence of both ϵ and B.

To derive the temperature profile, it is necessary to determine the unknown parameter α, representing the temperature rate. The value α of is calculated by multiplying the network output by a scaling factor αm, which is part of the HM framework. Figure 4 shows a picture of how the series HM was put to use. A feedforward neural network acts as an estimator for the HM, as shown in

Figure 5. Using the previous time step's process parameters, the network consults the physical model to produce a prediction about the system's condition in the next time step.

In the remainder of the study, the hybrid model using Equation (2) will be designated as HM_1, while the hybrid model using Equation (3) will be termed as HM_2.

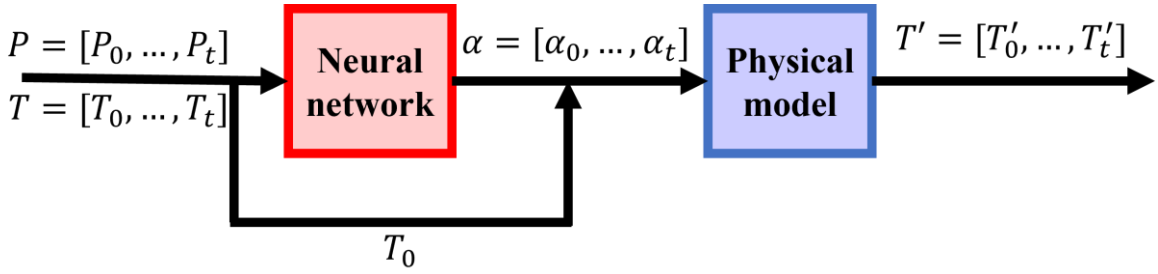


Figure 4. Structure for a serial hybrid model. The physical model's input process parameter, α, is estimated by the neural network.

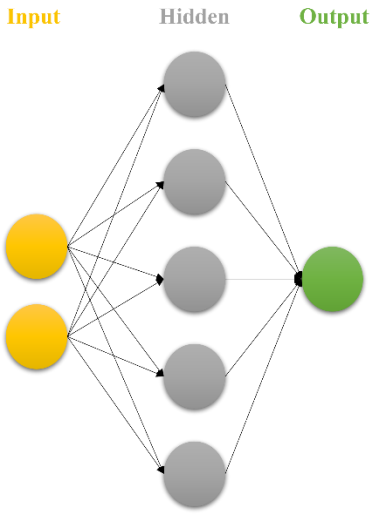


Figure 5. A feed-forward neural network with one output neuron, a hidden layer of size 5, and an input layer of size 2.

5. Model Training

Based on the descriptions provided in Section 4 we constructed three distinct models, RNN, HM_1, and HM_2. Relevant hyperparameters were carefully chosen, including the number of layers, the number of neurons per layer, and the activation functions of each neuron. Subsequently, these models underwent comprehensive training utilizing the furnished dataset. The ensuing content delineates the intricate particulars of this training process. Ultimately, the adeptness of these trained models was utilized for the prediction of temperature variations within the rod.

Finding a mapping function that connects input variables, also known as features, to the desired output variables is the goal of training and improving a model. The sample's temperature is the important output variable in the context of induction heating. A supervised learning problem is one in which the goal variable—in this example, the sample temperature—is already known as a result of simulation observations.

5.1 Data Preparation

Input data is often scaled to fit within a certain range in neural network applications. This is because normalizing the training data often results in better performance from artificial neural networks (ANNs). In our case, linear scaling is employed for data normalization:

$$X' = X/X_{max} \quad (5)$$

Here, X' represents the normalized value of variable X, achieved by dividing X by its maximum value, $Xmax$. For the sake of standardization, we fixed the maximum values to $Pmax$ = 30 kW for power and $Tmax$ = 1200 °C for temperature.

All models (RNN, HM_1, HM_2)'s inputs and outputs are scaled using the method described in Equation (5). The data is arranged into batches for the RNN model using the following format: [batch-size, seq-length, input-size]. In this configuration, batch size is the first dimension, time sequence length is the second, and output size or feature size is the third. The original time series is advanced one-time step backward to produce the output vector, which represents temperature.

The RNN model uses temperature and power level records to anticipate the temperature at time t. As seen in Figure 3, this data is supplied into the RNN cell. The benefit of batch data preparation is that the algorithm automatically collects each batch's gradients while the RNN's parameters are being optimized. The sequence length, referred to as seq-length in our methodology, is defined as a duration equivalent to two-time steps. Following that, the batch size is decided based on the total amount of data that is then available.

The data from Table 2 is divided into training and test sets for the simulation. Table 3 displays the creation of the two divisions. Partition 1's training set contains only high-power simulations, whereas partition 2's training set contains two low-power simulations. The overfitting issue is addressed by generating a new validation set from the training data. During the tuning process, this validation set is used to keep an eye on the mean squared error (MSE).

Table 3. Considered simulation data partitions from Table 2 for model training and testing.

Partition number	Training data	Test data
1	15, 9, 6 (kW)	13.5, 12, 10.5, 7.5, 4.5 (kW)
2	15, 9, 6, 3, 0.9 (kW)	13.5, 12, 10.5, 7.5, 4.5, 2.1, 1.5 (kW)

Since each simulation takes a different amount of time, the data is split into training and confirmation sets for each simulation on its own. The training set to validation set ratio was decided upon as 70% to 30%. For instance, data from the 15-kW simulation is randomized, with 70% earmarked for training and 30% for validation. This approach ensures accurate representation of data percentages for each experiment (at various power levels) within both the training and validation sets.

5.2 Hyperparameter Identification

Researchers had to consider a wide variety of factors in order to discover the optimal values for the RNN model's hyperparameters. Some of these factors were the number of layers in the RNN, the number of neurons included inside each RNN cell, and the activation process of neurons. This search was conducted within predefined bounds, spanning from 1 to 10 layers and 2 to 50 neurons.

Additionally, the exploration encompassed two activation functions: hyperbolic tangent (tanh) and rectified linear unit (ReLU). Our goal was to create a sufficiently large search area that would make it easier to identify the RNN network with the most promising performance.

Upon completing the search, the optimal RNN model was identified, characterized by 7 fundamental cell layers, and each cell was equipped with 20 neurons. In the realm of activation functions, ReLU exhibited superior performance in comparison to the tanh activation function within the RNN model. Achieving the lowest MSE, which is determined by dividing the total number of data points by the sum of squared errors (SSE), was the key criteria for choosing the best model.

The HM utilizes a NN that consists of a single hidden layer, which is distinguished by linear activation functions for both input and output. Both the tanh and ReLU activation functions were studied throughout the evaluation process to determine whether they were acceptable for usage in the hidden layer. The input and exit layers were made so that they could handle data with two dimensions and one dimension, respectively. Through systematic experimentation, the dimensions of the hidden layer spanned from 1 to 10. Intriguingly, the observation surfaced that a hidden layer dimension of 10, coupled with a tanh activation function, delivered commendable performance. It is worth noting that introducing deeper network architectures into the HMs did not yield significant performance enhancements.

5.3 RNN Training

The RNN optimization procedure seeks to reduce the total squared errors in estimating the sample's temperature over all training instances. This is pursued through the utilization of a predefined error function:

$$SSE = \sum_i^N (T_i' - T_{i,exp}')^2 \qquad (6)$$

Here, $T_{i,exp}'$ represents the target output, while T_i' corresponds to the output obtained from the network. During the training process, the back-propagation algorithm takes center stage, effectively deciphering the optimal weights and biases for the neurons within the network. This task is achieved by computing the error signal, as delineated in Equation (6), through a meticulous comparison between the network's output and the actual targets of the process. Following the network's exposure to the entire spectrum of training samples, the accrued error traverses in a retrogressive fashion, thereby facilitating the update of the network's parameters.

In order to mitigate the risk of overfitting during the training process, a vigilant approach was adopted. After every 20 iterations, the model's prediction error was painstakingly measured using the validation data. If the validation error showed signs of a rising trend, an immediate halt was imposed on the training process. To counter the potential entrapment in local minima, a strategic maneuver was embraced: the training process underwent ten reinitializations, with network parameters being reset during each iteration. The restart method used in this study was enhanced by randomly reshuffling both the training and validation sets during each restart iteration. This introduced a degree of variation into the training process.

Following the execution of ten optimization cycles, a precisely selected trained model was identified, which had the lowest MSE. Note that the MSE was computed as the SSE divided by the number of data points. The next phase entails using the selected model to make predictions and test them, especially using simulation data that has not been seen before.

5.4 HM NN Training

HM NN optimization is performed in a manner similar to that of RNN training as discussed in Section 5.3, with a few key differences. When using the HM, the temperature output vector is created without changing the initial time series. Also, figuring out how to optimize the network settings for the mixed structure (shown in Figure 4) is hard because the neural network output α' does not show up directly in Equation (6). This is due to the unavailability of direct measurements for the network output.

However, by leveraging the information from physical knowledge, it is still possible to generate an error signal for the neural network. This methodology draws upon sensitivity equations [3]. The objective of training the hybrid models is to minimize the MSE as defined in Equation (6) across all training examples. Even though α' is fixed at each sample moment, the SSE gradient may still be assessed in relation to this internal parameter.

15. $error\ signal = \frac{\partial SSE}{\partial \alpha_i'} = 2g_i(T_i' - T_{i,exp}')$ 16. (7)

The term g_i, which stands for the slope of the HM output, shows how the output is related to the internal constant α_i'.

17. $g_i = \frac{\partial T_i'}{\partial \alpha_i'}$ 18. (8)

The error signal described by Equation (7) is necessary for the HM neural network in order to modify internal settings and lessen hybrid network error. Using the selected data normalization mapping process, the scaled gradients g_i are linked to the original gradients $G_i = \frac{\partial T_i}{\partial \alpha_i}$. So, Equation (2)'s sensitivity equation may be expressed as follows:

19. $\frac{dG}{dT} = 1$ 20. (9)

In a similar vein, for Equation (3), the corresponding expression becomes:

21. $\frac{dG}{dT} = 1 - 4BG(T + 273.15)^3$ 22. (10)

By using the Equations (9) and (10) in conjunction with the initial condition $G(t = t_0) = 0$, together with their corresponding Equations (2) and (3), we can deduce the fundamental components necessary for calculating the SSE and its gradient with respect to the network output. For every sample data, the error signals produced from Equation (7) must be acquired and added, allowing backpropagation algorithms to update the weights.

Using a series of 10-time steps (Figure 4), the HM trains the network. One benefit of this method is that it can be used to train a model on many time series with different starting points, as shown in Figure 4 by the different T_0 and P_0 numbers. Following the neural network's determination of the feature vector's (P, T) corresponding vector, the physical model passes along the input vector's (T_0) starting state by way of a set of simplified heat equations.

In order to mitigate the issue of overfitting during the training process, the model's prediction error is assessed on the validation data at regular intervals of 20 iterations. When the validation error begins to rise, the training is stopped. To reduce the chance of being stuck in local minima, the training procedure is repeated ten times with the network settings reset each time. Also, each time the process starts over, both the training set and the proof set are given to the network by chance.

Upon doing 10 optimization cycles, a precisely selected trained model is identified based on its lowest MSE. This selected model is subsequently enlisted for prediction and testing on previously unexplored experimental data. This time, the MATLAB Adam optimizer's learning rate is set at 0.01.

6. Results

Figure 6 displays the MSE patterns observed throughout the optimization process for both the HM and RNN models. The patterns are shown separately for the training and validation datasets. The HMs need a reduced number of training iterations compared to the RNN model because to their lower MSE. This divergence draws attention to one benefit of HMs over wholly data-driven methods like RNN. Because they include certain aspects of the system's behavior as well as missing information, HMs have lower MSE than data-driven models. The MSE of HM_1 and HM_2 are almost identical, with HM_2 showing quicker convergence as a result of the inclusion of additional physical information. However, for this improvement to occur, the physical model must precisely depict the process. We also tested several additional physical models, which had bad results but are not included here. Importantly, even with twice as much training time, the RNN model never reaches the HMs' low error levels. Moreover, in Figure 6, it's noticeable that the validation error plateaus after around 500 iterations, while the training MSE continues to decrease. The optimization stops as soon as the validation error begins to increase in order to prevent overfitting on training data. This guarantees the trained model's proficiency with new data. Notably, the MSE difference in this example between the hybrid and RNN final models is nearly ten times bigger.

Following the completion of the training trials with a successful SSE minimization, the beginning temperature of each simulation was then transmitted across time with the help of the control process parameter vector, more especially the power level $P(t)$. Figure 7 shows a comparison of the RNN and HM_1 models that were trained using tests with 15 kW, 9 kW, and 6 kW of power. It is interesting to note that, even for the trials it was trained on, the trained RNN fails to faithfully represent temperature dynamics. But HM_1 does well not just on the dataset it was trained on, but also on data it has never seen before in the form of test data. But the performance of HM_1 gets worse when the power is lower ($P \leq 4.5$ kW). This could be because the physical model is simple or because the power levels do not change much during training. It's notable that the results of HM_1 are less satisfactory for $P = 7.5$ kW. Also, the RNN shows that it can create dynamics for the high-power scenario with $P = 15$ kW. This is because, given the increased heating rate and higher precision in propagating the beginning temperature of a rapid process, the RNN network is better able to capture this dynamic. The data was first obtained with a sample frequency of 100 Hz, with a corresponding time interval of 0.01 seconds. In order to effectively compare HMs and RNNs, the original dataset was modified by down sampling it to a frequency of 10 Hz, with a time interval of 0.1 seconds (dt).

Figure 8 shows the HM and RNN's prediction performance on the down sampled simulations. The bulk of simulations' dynamics are provided by the RNN, however it has exceedingly sluggish processes ($P \leq 4.5$ kW). Surprisingly, the performance of the RNN is on par with that of HMs in the very challenging simulation scenario, using a power capacity of 15 kW. It is worth noting that this evaluation was conducted using the same simulation data that was used to train the models. However, when applied to unfamiliar data, such as in the simulations where $P = 12$ kW and $P = 10.5$ kW, the performance of the RNN is inferior to that of HMs. Performance- and fit-quality-wise, HM_2 outperforms HM_1 for low-power simulations, notably at $P = 4.5$ kW, on both training and unobserved data. HM_2 accounts for the radiation term, which cools at high temperatures, as mentioned in Section 4.2. Black-body radiation's influence on temperature changes becomes negligible under conditions

typified by high power levels, brief process times, and large heating rates, yielding equivalent findings for both HMs. Figure 9 shows the models' root mean squared errors (RMSE = √MSE) on both the training and test sets of data.

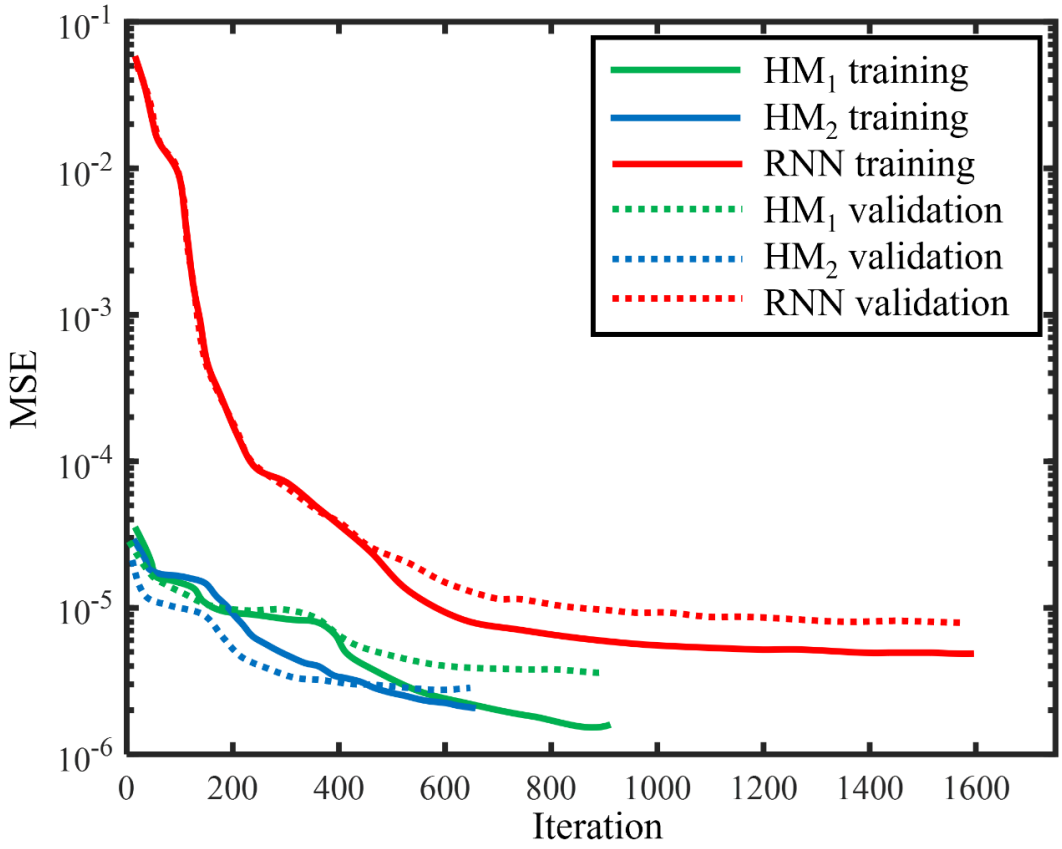


Figure 6. The MSE on the hybrid and RNN models' training and validation data. The scale on the y axis is logarithmic.

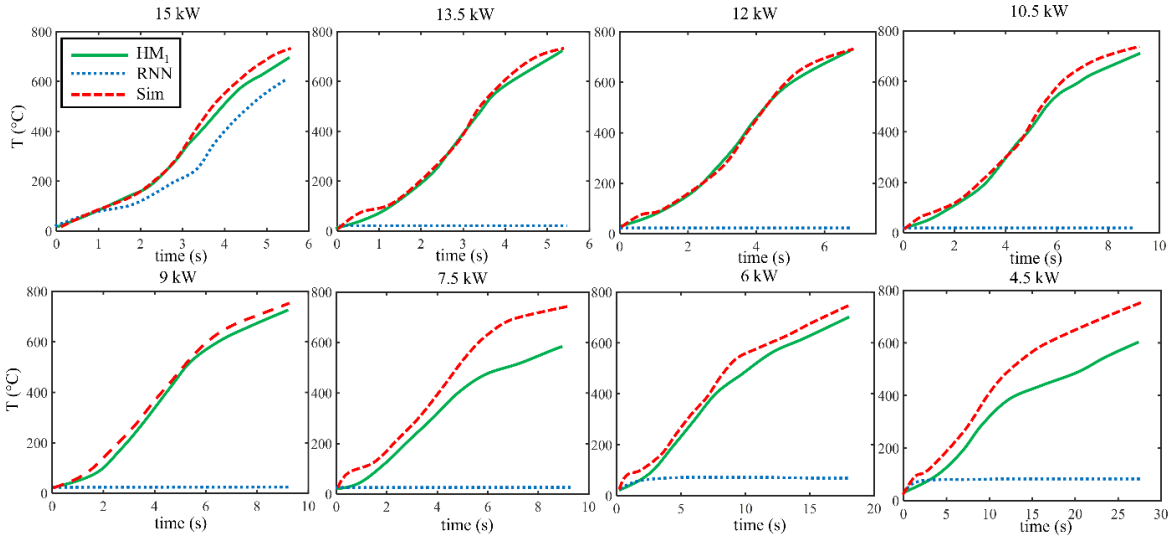


Figure 7. HM1 and RNN estimations' dynamic profiles are compared to simulations that use training data with a time step of 0.01 s. 15-, 9-, and 6-kW power level trials make up the training data set.

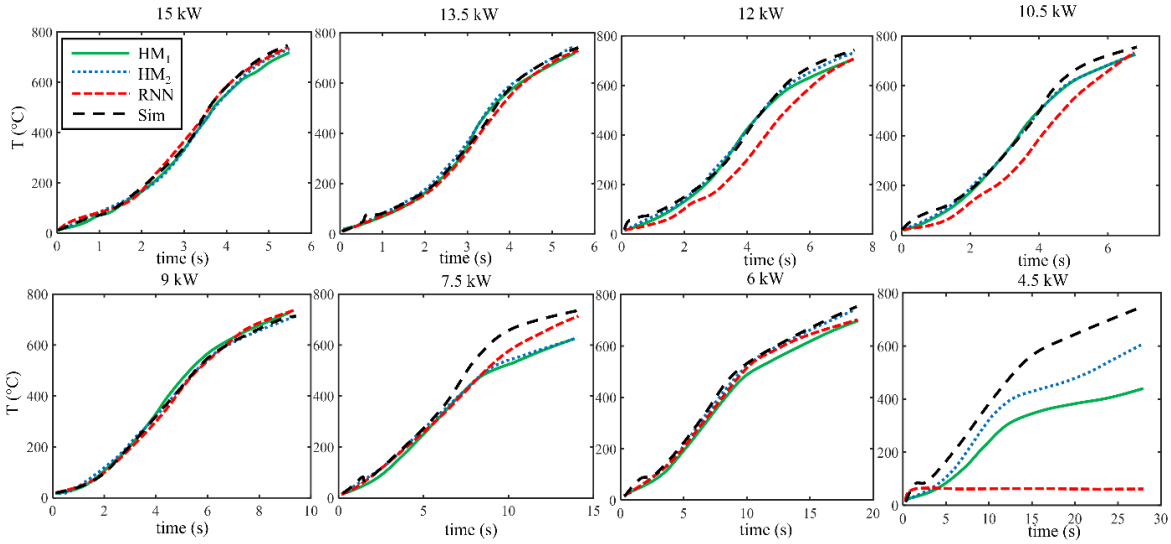


Figure 8. HM1 and RNN estimates' dynamic profiles are compared to simulations that use training data with a time step of 0.1 s. The training data set includes power levels of 15, 9, and 6 kW.

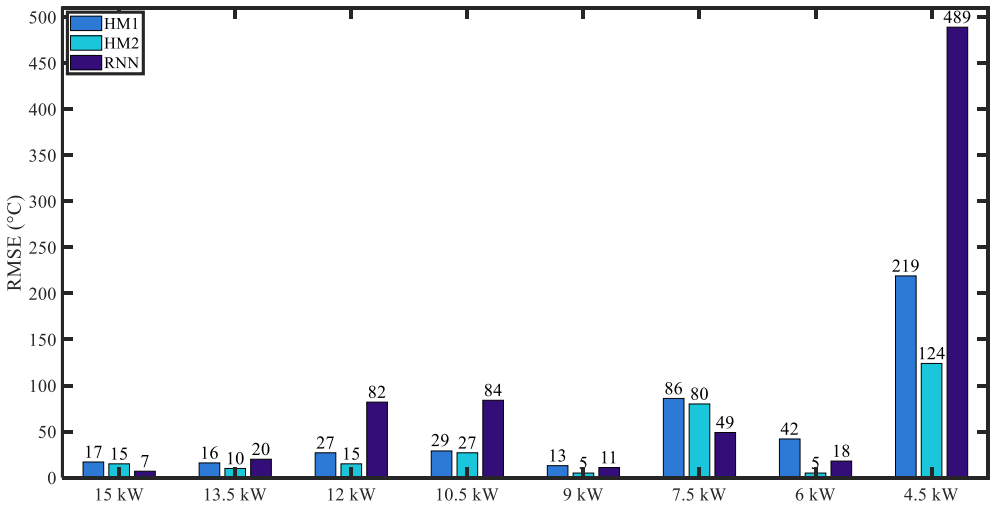


Figure 9. Comparing the RMSE of the HMs with RNN estimations across many simulations. Simulations with 15-, 9-, and 6-kW power outputs and a time step of dt = 0.1 s are included in the training data set.

In order to provide a wider range of values for the process parameter P, we expanded the original training dataset by including two more simulation trials conducted at lower power levels (P = 3.9 kW and P = 0.9 kW). This supplementation was performed alongside the existing data points for P = 15 kW, P = 6 kW, and P = 3 kW. Figure 10 presents the results of identification and testing on unseen data. Figure 11 clearly illustrates that the RNN beats the HMs when it comes to producing dynamics for low-power testing ($P \leq 3$ kW). HM_1 and HM_2 are both good predictors at moderate and high-power levels, whereas HM_2 performs better at low power levels. Figure 11's RMSE analysis shows that HM2 has superior extrapolation abilities on unseen data.

One notable benefit of HM structures in ANNs is their inherent interpretability. This characteristic makes them very helpful for the objectives of comprehending and optimizing processes. The heating rate is indicated by the internal parameter α in the HMs discussed in this study, and it has a physical significance. Figure 12 displays the fluctuation of a neural network's output parameter α for six different power levels. The development α of illustrates that several operating regimes exist at various power levels throughout the heating process. For high, midrange, and low power trials, several α profiles are seen. When the power is high, as P = 15 kW, both HMs estimate comparable values of α, but as the power falls, their courses diverge. This shows that the heat source follows distinct dynamics at lower operating powers. It is significant that the value of α always stays positive, supporting its physical interpretation as the source component in the heat equation for the process under consideration.

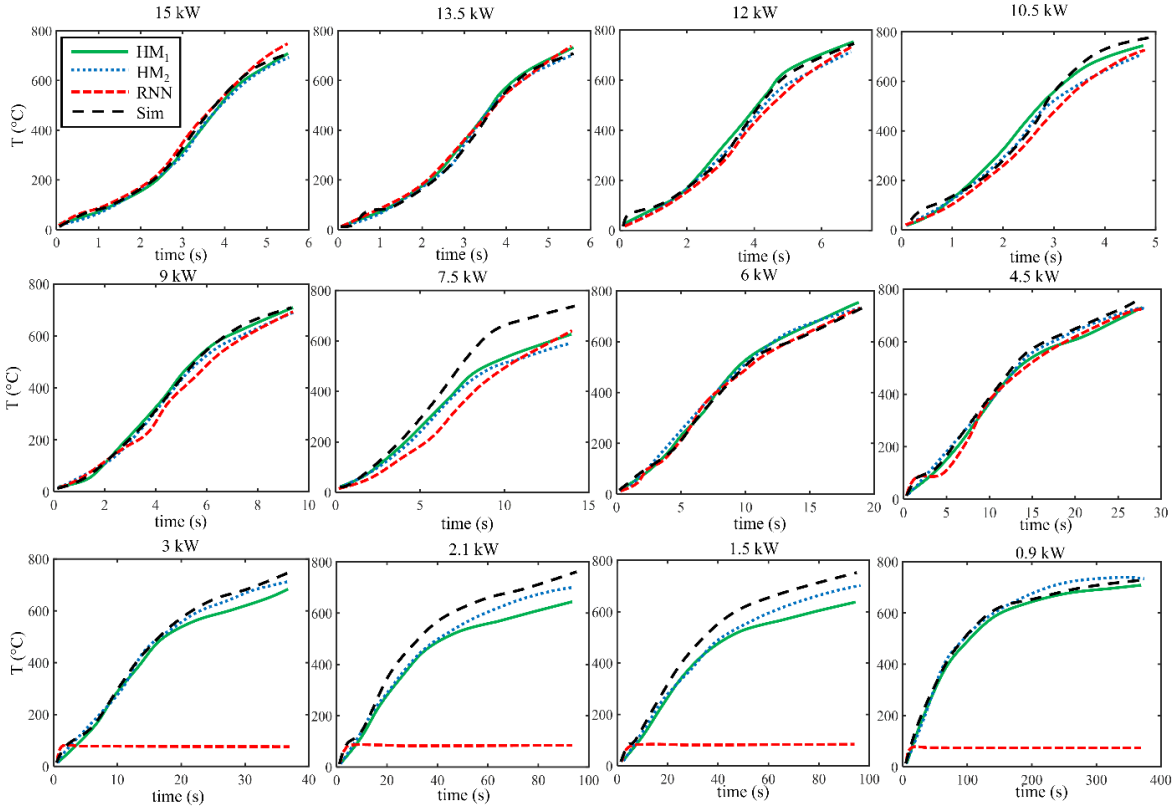


Figure 10. HM1 and RNN estimations' dynamic profiles were compared to simulation results using training data with a time step of dt = 0.1 s. Simulations with power levels of 15, 9, 6, 3, and 0.9 kW make up the training data set.

An important aspect of stand-alone neural networks is their tendency to achieve higher accuracy as training datasets grow larger. Because they only use data to infer nonlinear correlations between process variables, these models are expected to perform better with large datasets. To evaluate their identification capabilities, we conducted training using the entirety of the available simulation data. In Figure 13, we present the RMSE of the HMs in comparison to the RNN model. Overall, it is evident that HM_2, with less total error, consistently characterizes the data more correctly than HM_1 and RNN. Although RNN accuracy is increasing, it is still unable to compete with HMs, especially HM_2. Furthermore, it is evident that HM_2, which integrates the impact of black-body radiation, performs better than HM_1 in the low power domain. For high powers levels, both HMs exhibit similar performance, displaying only marginal differences in RMSE.

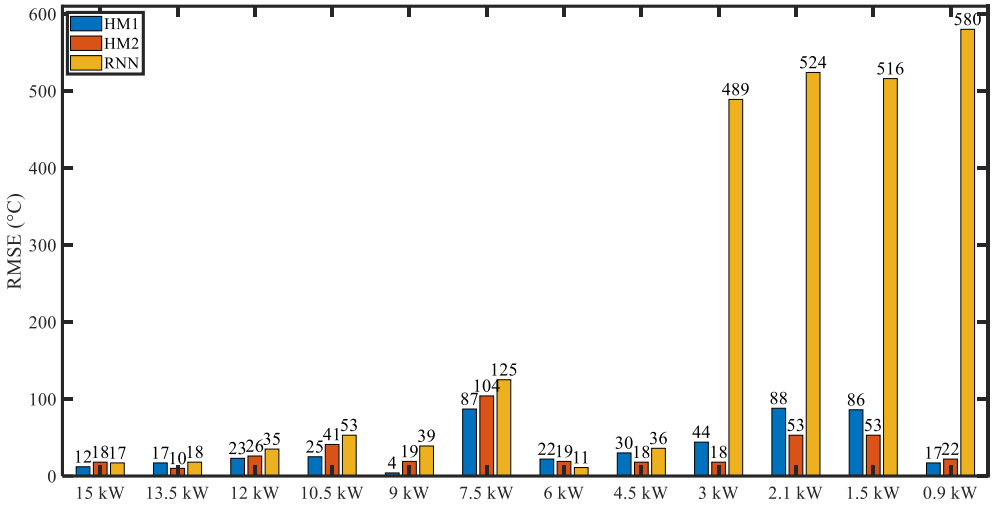


Figure 11. Comparison of the RMSE of the RNN-based estimations of the HMs for several simulations. Simulations with power levels of 15, 9, 6, 3, and 0.9 kW make up the training data set. The training set's sampling period is dt = 0.1 s.

7. Discussion and Applicability of the Presented Approach

The recommended strategy's particular benefits are examined in greater depth in this section, which also acknowledges its inherent disadvantages. We also consider potential extensions to the system to get over any current drawbacks.

7.1 Material Parameter Identification

It is crucial to accurately determine a material's thermophysical characteristics. This can often involve expensive experimental methods or advanced instrumentation. The approach presented in this study offers an alternative, providing a practical and cost-efficient means of estimating specific heat properties of materials. Using a reformulated equation, this method integrates material properties into the aggregated heat equation (Equation 3).

The methodology used in this research includes the ability to calculate the specific heat of different materials. The strategy is briefly explained in the section that follows. To incorporate material properties explicitly into the combined heat equation (Equation 3), it is reformulated as:

$$\frac{dT}{dt} = \frac{\alpha}{\rho_m C_p} - \frac{B}{\rho_m C_p}\left((T + 273.15)^4 - (T_{amb} + 273.15)^4\right) \quad (11)$$

Here, the specific heat (C_p) and mass density (ρ_m) are clearly provided. When thermal conductivity is minimal or the temperature distribution is steady after brief temperature changes, Equation (11) may be employed. When the sample's volume is tiny (or its dimensions are much smaller than the inductor), a slight temperature difference exists between the sample's outside and inside.

By following the steps outlined in the parts before this one, you can train the HM to get the best fit for the temperature and figure out the model parameters $\alpha(T,P)$, $B(T)$, and $C_p(T)$. Although the material's mass density (ρ_m) is known, it may also be guessed using the same method. The sensitivity equations, which create an error to update parameters in the neural network, are formulated as:

$$\frac{dG}{dt} = \left[\frac{dG_1}{dt}, \frac{dG_2}{dt}, \frac{dG_3}{dt}\right] \quad (12)$$

Here, $G = [G1, G2, G3] = [\partial T/\partial \alpha, \partial T/\partial Cp, \partial T/\partial B]$, and can be readily obtained from Equation (11). It is important to note that two separate neural networks need to be employed in parallel to estimate the model parameters. One network is used to estimate $\alpha(T, P)$, while the other network estimates $C_p(T)$ and $B(T)$, as these two parameters are solely temperature-dependent.

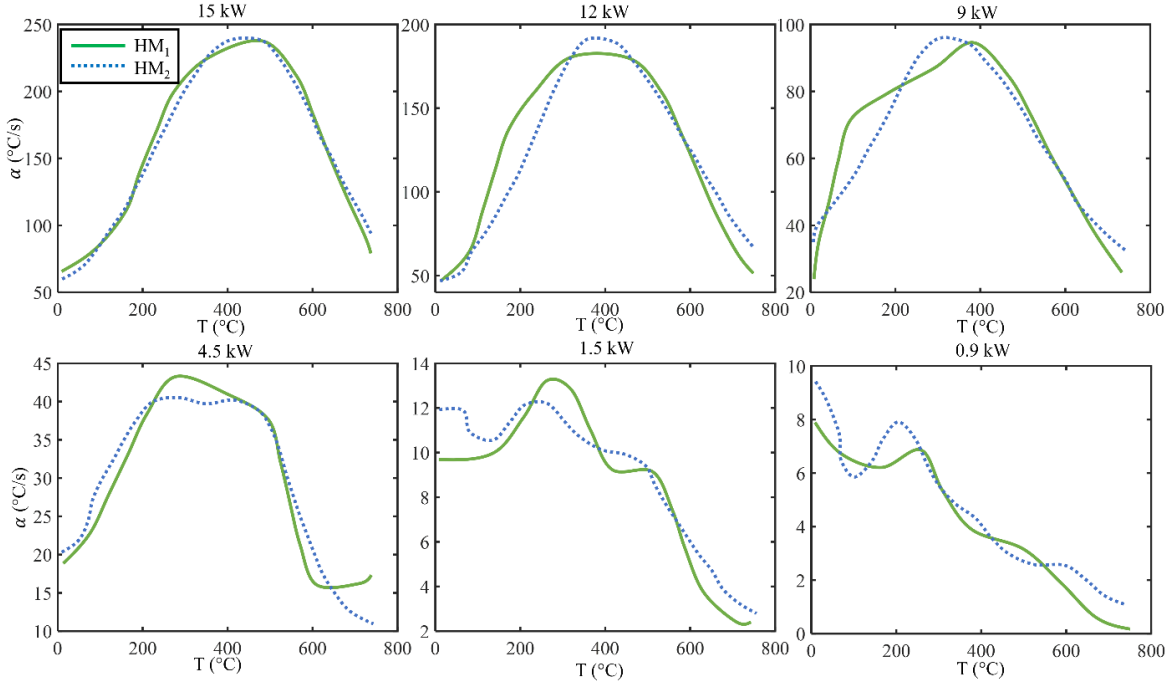


Figure 12. The neural network's output α for six power levels. Simulations with power levels of 15, 9, 6, 3 and 0.9 kW and sampling rates of 1 Hz (dt = 0.1 s) make up the training data set.

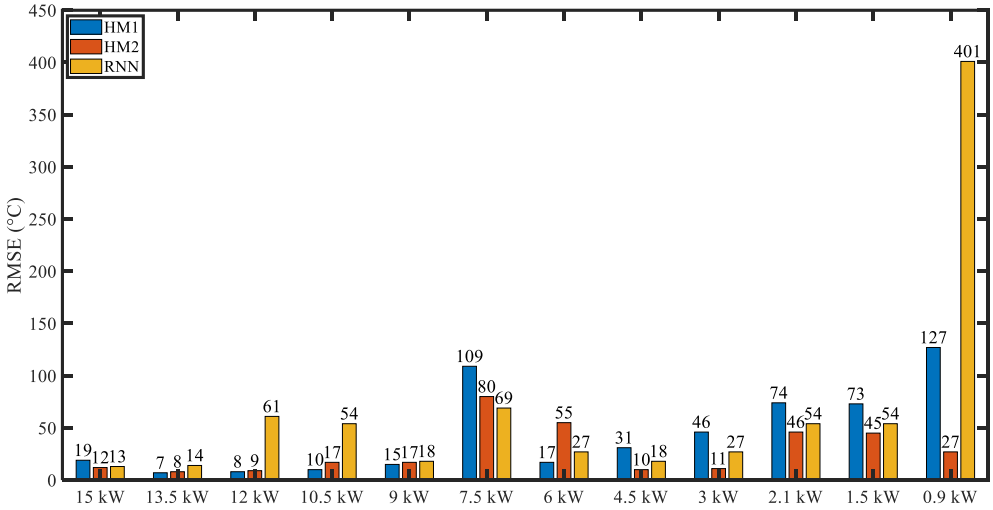


Figure 13. The RMSE of the HM_1, HM_2, and RNN models for identification. All of the simulations in the training data set have sampling times of dt = 0.1 s.

7.2 Model Prediction Error and Contingency Plans

As shown in Figure 10, the trained HM model performs well on both visible and hidden training and test data, although it still makes mistakes in some test trials. For instance, a maximum inaccuracy of 300°C is noted in the HM's predictions in the simulation when the power output is set at 7.5 kW.

Also, for the power levels of 2.1 kW and 1.5 kW, the maximum deviations are around 100°C. These deviations and potential strategies to mitigate them can be explained as follows:

- The increased error observed in low-power simulations can be partially attributed to the use of a smaller amount of training data in this power regime, as indicated in Table 3. While using HM_2 may help minimize error, more training data would improve the model's performance even more.
- Due to the use of simplified physical information, the model is not perfect. This is aided by the physical component's lack of a conduction term. Enhancing model performance and extrapolation quality can be achieved by implementing a multi-sensor HM. Alternatively, integrating a surrogate model based on real experimental data as additional physical knowledge is another viable option.
- There is room for improvement in the approach used to creating a "training set" and a "test set" from simulation data. The current approach strictly divides available simulation data into two parts: training and test sets, as shown in Table 3. However, finding the best-performing model using this approach requires exploring various combinations of power experiments in the training data. This process demands multiple partitioning and model training iterations, which can be computationally intensive. An alternative strategy worth exploring is to partition the data in a way that includes training examples from each simulation. When there are few simulations available for model calibration, this strategy is advantageous.

This effort was not intended to provide a perfect, completely trained HM for immediate use in industrial applications. The focus was on presenting and demonstrating the approach. Notwithstanding the constraints imposed by the available data, the approach was effectively shown, hence establishing the proof of concept. These difficulties must be addressed, and backup plans must be created in order to use a trained HM in actual industrial settings.

7.3 Application Ranges of the HM in Comparison to FEM

FEM plays a crucial role in comprehending the complex interactions and interdependencies of physical phenomena within intricate systems, particularly in the realm of processes like induction hardening. These models offer a qualitative understanding that can be elevated to quantitative insights with a comprehensive grasp of the multi-physical system. However, the accuracy of FE results hinges significantly on the quality of the input data and the expertise applied in constructing and interpreting the model. For instance, finding the boundary conditions of an induction heating issue may be difficult due to the intricacy and size of the geometry as well as the operating environment. This can provide a number of difficulties.

Because of how they are made, FE models can also be used to study things like temperature differences and local phase changes across the whole sample space. When computational speed is not a crucial consideration, these resources prove to be valuable in comprehending the concepts of induction hardening and process design. However, for applications that demand fast computation times, typically within minutes or even sub-seconds, FE models become too expensive and impractical for direct implementation.

The HM given demonstrates significant improvements in calculation speed, rendering it well-suited for two primary applications: a) facilitating prompt decision-making in offline process planning, and b) integrating into the system unit to enhance process control. The model is only able to provide data

for one or a small number of spots in the sample, despite being able to provide quantitative findings. The precise utilization of this information is crucial. The current study focuses on the through hardening of steel rods, which is a reference process. This process involves lower applied frequencies compared to surface hardening, resulting in a slower heating process and more uniform temperature distributions across the cross-section. This method, which requires just a single surface or, as here, subsurface assessment point, is well suited to the current iteration of the model.

Understanding the temperature development at one location adequately defines the temperature distribution throughout the whole cross-section for stable processes such as via hardening. The acquisition of temperature data at a single spot is of significant importance in the regulation of an inductive heating process with the aim of attaining predetermined goals. The ultimate desired qualities, such hardness, are significantly influenced by temperature profiles. By modifying the power generator's input values and the necessary holding time, the proposed technique enables the generation of a desired temperature profile. This makes it possible to test different power and holding time combinations to reduce power usage while preserving the correct temperature profile and ultimate hardness.

The steady heating procedure and the deliberate keeping at a certain temperature make the modeling idea for induction tempering well suited. Additionally, the cooling phase of the heat treatment process, as seen in Figure 1(a), may be included into the HM, enabling a comprehensive description of the thermal aspects of induction hardening.

When applying the HM to surface hardening, which involves highly non-uniform temperature distributions, two approaches can be considered:

- A FE-based substitute model can be used as a physical model in the HM. This has two main benefits. By including a precise physical model, it first improves the HM's extrapolation skills by allowing it to give temperature distribution, metallurgical characteristics, and solutions for all associated multi-physics issues. The black-box portion of the HM makes predictions about the FEM proxy model's unknown parameters using modeling data. Material information, border circumstances, B-H features, and other factors are examples of these criteria. Second, the FE-based substitution model used in the taught HM has a fast execution time, making it a viable option for real-time decision making and monitoring, managing, and optimizing operations. Although this approach involves the incorporation of complex physical knowledge, it also presents a cost-benefit ratio. Given the financial implications, the use of an FE based surrogate model is seen more advantageous in terms of augmenting the overall functionalities and boosting the generalizability of the existing HM.
- The presented HM has the potential to be extended beyond single-point temperature modeling by integrating a heat conduction term, allowing for the inclusion of many points throughout the cross-section. This may be accomplished by using a larger set of samples during the training phase (described in Section 7.4) that include temperature data at varied radii.

7.4 Model Generalization

In this research, we developed a novel method to accurately replicate the temperature of a specific location within a specimen. While understanding the temperature profile of a single point remains important in various situations with well-defined objectives (as discussed in Section 7.3), the idea of comprehending the broader temperature distribution is increasingly appealing and pertinent. The

fundamental approach we employed has the potential to be extended theoretically to calculate temperature variations across the entire specimen, including gradients.

However, achieving this goal requires a more extensive data collection effort that involves multiple points distributed throughout the specimen. To illustrate, it is recommended to consider how the heat equation is applied when examining a cylindrical sample with axial symmetry:

$$\rho_m C_p \frac{dT}{dt} = \frac{\kappa}{r}\frac{\partial T}{\partial r} + \kappa \frac{\partial^2 T}{\partial r^2} + q(r, P, T) \quad (13)$$

This extended data collection process will provide a more comprehensive and insightful understanding of temperature distribution and variation within the specimen.

Assuming a uniform thermal conductivity, the process involves training an HM by leveraging the physical equation defined in Equation (13). Within this model lie several unknown parameters, which encompass the heat source term q, as well as the first and second spatial derivatives of temperature. Neural networks are employed to determine these unknowns, using pertinent features such as location r, material properties, current temperature, and operating power.

The creation and organization of the sequence of black-box components within this process require experimentation and optimization, which present initial challenges. One effective approach for predicting the second derivative function is to incorporate the first derivative of temperature from one black box as a feature in another black box. The use of radial activation functions by these black-box neurons is well-suited for approximating continuous functions. However, obtaining a substantial volume of high-quality data, along with careful experimental design, is essential to accurately estimate spatial derivative functions.

Notably, it is worth highlighting that the geometric dimensions can be adjusted and employed as additional features to train the hybrid model. Such a model can be scaled and applied to a variety of dimensions, accommodating different geometric shapes with ease.

8. Conclusion

To show how heat induction works in a piece of bolt steel, a mixed modeling method is used. This approach makes use of a feed-forward neural network and an ODE to represent a condensed physical equation. The physical model's unknown process parameter, α, which serves as a gauge for the heating rate, is determined using the neural network. This hybrid modeling strategy's efficacy is contrasted with that of making predictions just utilizing neural network models like RNN. The research reveals that the accuracy of the RNN model falls short in describing highly intricate dynamic data. Moreover, even when the size of the training dataset is increased, the RNN model cannot outperform the hybrid models, particularly HM_2. In contrast, the hybrid model is capable of interpreting data regardless of the sampling rate. This observation also suggests that hybrid models are less susceptible to being excessively influenced by noise or overfitting compared to pure neural network models. The study demonstrates that training the hybrid model using a small number of simulations already yields satisfactory predictive capability for the model.

The HM's predicting abilities are improved by better understanding the underlying physics. We showed that the model can extrapolate and simulate low-power simulations more effectively than HM_1, which just employs the heating rate α in the physical equation, by including a radiation component into the reduced heat equation. By augmenting or enhancing the physical insights, the neural network may select a function from a reduced parameter space that can estimate the process variable α with greater precision. By enhancing stability and durability, this phenomenon distinguishes the HM's training phase from RNN training. For instance, the optimization process is

constantly steady even with a higher learning rate. Overall, this results in more straightforward HMs that are nevertheless capable of estimating crucial process parameters. In contrast to neural networks that operate independently, neural networks integrated into a sub model exhibit a more constrained search space, resulting in expedited identification of hyperparameters for the sub model.

In this research, a single process parameter, indicated as α, was predicted using a neural network. Several parameters, including the coefficient B in the HM_2 radiation term as a function of temperature and power, may be predicted theoretically. This sort of hybrid model is essential for taking both the heating and quenching processes into account. We are thinking about incorporating these studies into our next research projects.

Acknowledgment

This research was supported by the Korea-Canada Artificial Intelligence Joint Research Center at the Korea Electrotechnology Research Institute (Operation Project: No. 22A03007), which is provided by Busan City, Korea

References

[1]Rudnev, Valery, and George E. Totten. "ASM Handbook Volume 4C: Induction Heating and Heat Treatment." ASM International, Ohio (2014).

[2]Shanmuganathan, Subana. Artificial neural network modelling: An introduction. Springer International Publishing, 2016.

[3]Psichogios, Dimitris C., and Lyle H. Ungar. "A hybrid neural network-first principles approach to process modeling." AIChE Journal 38.10 (1992): 1499-1511.

[4]Kumar Akkisetty, Pavan, et al. "Population balance model-based hybrid neural network for a pharmaceutical milling process." Journal of Pharmaceutical Innovation 5 (2010): 161-168.

[5]Mao, Zhi-zhong, Yu-qing Chang, and Lu-ping Zhao. "Soft-sensor for copper extraction process in cobalt hydrometallurgy based on adaptive hybrid model." Chemical Engineering Research and Design 89.6 (2011): 722-728.

[6]Luo, Na, et al. "Development of a hybrid model for industrial ethylene oxide reactor." Industrial & engineering chemistry research 51.19 (2012): 6926-6932.

[7]Wu, Zhe, David Rincon, and Panagiotis D. Christofides. "Process structure-based recurrent neural network modeling for model predictive control of nonlinear processes." Journal of Process Control 89 (2020): 74-84.

[8]Hömberg, Dietmar. "A mathematical model for induction hardening including mechanical effects." Nonlinear Analysis: Real World Applications 5.1 (2004): 55-90.

[9]Barglik, Jerzy, et al. "3D modeling of induction hardening of gear wheels." Journal of Computational and Applied Mathematics 270 (2014): 231-240.

[10]Yuan, J., et al. "FEM modeling of induction hardening processes in steel." Journal of materials engineering and performance 12 (2003): 589-596.

[11]Candeo, Alessandro, et al. "Multiphysics modeling of induction hardening of ring gears for the aerospace industry." IEEE Transactions on Magnetics 47.5 (2011): 918-921.

[12]Wang, Youhua, et al. "A neural network combined with a three-dimensional finite element method applied to optimize eddy current and temperature distributions of traveling wave induction heating system." Journal of Applied Physics 109.7 (2011).

[13]Urbanek, Piotr, Jacek Kucharski, and A. Frączyk. "Neural network modelling of chosen nonlinearities in induction heated industrial plants." Automatyka/Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie 15 (2011): 625-635.

[14]Penha, Renata, Lauralice Canale, and Antonio Canale. "Modeling of hardness of low alloy steels by means of neural networks." Journal of ASTM International 8.8 (2011): 1-10.

[15]Leitner, Silvia, et al. "Model-based residual stress design in multiphase seamless steel tubes." Materials 13.2 (2020): 439.

[16]Brunbauer, Silvia, et al. "Residual stress and microstructure evolution in steel tubes for different cooling conditions–Simulation and verification." Materials Science and Engineering: A 747 (2019): 73-79.

[17]Schemmel, Manuel, et al. "Size effects in residual stress formation during quenching of cylinders made of hot-work tool steel." Advances in Materials Science and Engineering 2015 (2015).

[18]Vieweg, Annika, et al. "Phase evolution and carbon redistribution during continuous tempering of martensite studied with high resolution techniques." Materials & Design 136 (2017): 214-222.

[19]Asadzadeh, Mohammad Zhian, et al. "Inverse model for the control of induction heat treatments." Materials 12.17 (2019): 2826.

[20]Asadzadeh, Mohammad Zhian, et al. "Hybrid modeling of induction hardening processes." Applications in Engineering Science 5 (2021): 100030.

[21]Lipton, Z. C., Berkowitz, J., Elkan, C., 2015. A Critical Review of Recurrent Neural Net- works for Sequence Learning. arXiv:1506.00019v4.