Article:
Han, X., Li, K., Wang, C. et al. (2025) Transient heat conduction modelling for real-time operation and control of steel-making reheating furnaces - a physics-informed knowledge distillation-assisted EngGeneNet approach. Control Engineering Practice, 164. 106472. ISSN: 0967-0661

https://doi.org/10.1016/j.conengprac.2025.106472

Transient heat conduction modelling for real-time operation and control of steel-making reheating furnaces - a physics-informed knowledge distillation-assisted EngGeneNet approach

Xiaofei Han[a], Kang Li[a,*], Chuan Wang[b], Jiayang Zhang[a], Yukun Hu[c,*]

[a]*School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, United Kingdom*
[b]*Swerim AB, Box 812, Luleå, SE-97125, Sweden*
[c]*Department of Civil, Environmental & Geomatic Engineering, University College London, London, WC1E 6BT, United Kingdom*

Abstract

Accurate and fast transient temperature distribution prediction is a long-standing technically challenging open problem in real-time operation and control of many energy-intensive industrial processes involving massive heat conduction. The powerful fitting capabilities of deep learning models to perform parallel computations make them ideal surrogate models to meet the requirements for real-time applications. However, collecting and processing a large amount of labelled data is tedious and challenging if not possible. Furthermore, most neural models are black-box models, hence suffer from a few well-known problems such as poor generalization performance and slow convergence speed. This paper proposes a physics-informed deep learning modelling framework, namely EngGeneNet to capture the salient features and functional relationships of system variables to predict the transient temperature distribution of large-scale intermediate steel products in reheating furnaces. The network can learn a mapping between the current and the future transient 2D temperature field at a given ambient temperature, equivalent to solving the partial differential equations in real-time. The heat conduction governing equations and the boundary conditions are formulated as the loss function to guide the accurate and efficient training of the proposed EngGeneNet model. Then, an 'Eng-Gene' module is embedded into the deep learning model to accelerate the training convergence and enhance generalization performance. The 'Eng-Gene' is the salient physical relationship among variables that are extracted from the first-principle knowledge of the target system. Furthermore, the knowledge distillation approach is adopted, where a computationally expensive but more accurate numerical method namely alternating direction implicit (ADI) is applied to generate sufficient training data for training the deep learning models. To improve the adaptability of the EngGeneNet model to varying product batches, transfer learning is adopted to mitigate the dataset feature space variations under different operating conditions. The proposed method has been validated on a pilot-scale walking-beam furnace with a range of steel bloom batches under different operating conditions. The results suggest that the EngGeneNet framework can effectively improve the generalization and convergence performance of the deep learning model. In terms of processing time for predicting each frame of the heat distribution map, the proposed model achieves an improvement of approximately 96% in computational efficiency with comparable accuracy of the conventional method used in real-life applications, paving the way for real-time applications in many energy-intensive engineering processes.

Keywords: EngGeneNet, Heat conduction, Data knowledge fusion modelling, Physics-informed neural network, Reheating furnaces

1. Introduction

Many energy-intensive engineering applications involve massive conduction heat transfer, from mineral processing such as iron and steel making to food, cement, glass, and plastics processing, to just name a few. Reducing energy consumption in these energy-intensive industries requires more accurate and fast models for predicting temperature profiles under varying boundary conditions in real-time operation and control. This is however a long-standing open problem as solving the Fourier's law governing the relations of flow of heat and temperature field is computationally formidable, making it deemed unsuitable for real-time operation and control, especially when addressing large-scale heat conduction problems across diverse application scenarios.

The reheating furnaces are widely used in the steel industry to heat intermediate steel products, such as blooms, billets, or slabs (known as stock), to a target temperature with desired temperature uniformity through the thickness of the material before hot-rolling (Hu et al., 2016). Accurate control of the temperature heating profile of the stock during the entire heating process is essential for achieving optimal results, and reducing product waste and energy waste. A major challenge is that the temperature of the stock is usually collected through radiometric imaging cameras, and the inside distribu-

*Corresponding authors
Email addresses: K.Li1@leeds.ac.uk (Kang Li), yukun.hu@ucl.ac.uk (Yukun Hu)

Nomenclature

Abbreviations

ADI	Alternating Direction Implicit
CFD	Computational Fluid Dynamics
CNNs	Convolution Neural Networks
CZ	Control Zone
DNNs	Deep Neural Networks
EngGeneNet	Engineering-genes Network
FCN	Fully Connected Network
FDEs	Finite Difference Equations
FDM	Finite Difference Method
FPN	Feature Pyramid Network
OC	Operating Condition
PDEs	Partial Differential Equations
PINN	Physics-Informed Neural Network
SB	Steel Bloom

Symbols

β	loss function weighting factor [-]
ϵ	Crank-Nicolson method coefficient[-]
∂D	boundary domain [-]
ρ	density (kgm^{-3})
σ	Stefan-Boltzmann constant ($Wm^{-2}K^{-4}$)
τ	Total sampling time (s)
$\widetilde{\mathcal{L}}$	loss function of the proposed EngGeneNet[-]
$\widetilde{T}$	temperature distribution predicted by the EngGeneNet ($^{\circ}C$)
$C_p(T)$	specific heat ($Jkg^{-1}K^{-1}$)
d	space interval (m)
D_I	interior computation domain [-]
e	emissivity [-]
Fo	fourier number[-]
H	height of the stock(m)
h_c	convection heat transfer coefficient ($Wm^{-2}K^{-1}$)
$k(T)$	thermal conductivity ($Wm^{-1}K^{-1}$)
L	length of the stock(m)
m	number of nodes in x dimension[-]
n	number of nodes in y dimension[-]
T	2D temperature distribution of the stock ($^{\circ}C$)
t	time step (s)
T_A	ambient temperature ($^{\circ}C$)
T_m	mean value of the temperature distribution
$w, \widetilde{W}$	input and output of the Eng-Gene module[-]

tion cannot be directly measured, leading to poor process control and considerable waste of energy and product defects. To address this challenge, significant efforts have been made to develop online thermal conduction models to predict the temperature distribution of the stock as it passes through the reheating furnace. For instance, Wang et al. (2021) developed a mathematical model using the zonal method to study heat conduction within the beam blank. Ahmed et al. (2019) constructed a steady-state model for approximating the environment of the reheating furnace to predict the slab temperature distribution. However, the computational cost of such large-scale industrial applications is high, hindering its effective integration into the real-time control framework. One of the most computationally expensive steps in estimating the temperature profile of an object in the heat conduction process is to solve the partial differential equations (PDEs) relating to the heat transfer mechanisms. Various numerical or analytical methods have been proposed to solve the governing equations for heat conduction, such as the Maxwell-Cattaneo theory (McCarthy, 1975), developed so far for solving these PDEs or their simplified variants need to calculate the transient and/or steady heat conduction in order to yield more accurate results. In addition, the Gauss-Seidel iterative solution technique (Milaszewicz, 1987), the Binder-Schmidt explicit method and the alternating direction implicit (ADI) method (Croft and Lilley, 1977) are proposed for more accurate solutions by splitting the time step. However, these methods suffer from several major limitations. Generally speaking, solving partial differential equations is inherently complex and computationally intensive, especially when addressing large-scale heat conduction problems across diverse application scenarios. This makes them impractical for real-time applications due to the high computational cost. Consequently, data-driven machine-learning approaches have emerged to simulate and model the heat conduction process and estimate the temperature distribution.

Deep learning models have been increasingly utilised in various fields, such as industrial process data analysis and scientific computing and modelling (Yin et al., 2023). For example, Fonda et al. (2019) used the UNet model to explore the impact of changes in the Rayleigh number on heat transfer in the Rayleigh–Benard convection process. Yuan et al. (2020) proposed a dynamic convolution neural network strategy that transforms 1D process data into 2D samples to extract local nonlinear spatial-temporal features in soft sensor modelling of an industrial hydrocracking process. Long et al. (2019) investigated the PDE-Net's ability to solve partial differential equations to

obtain accurate model dynamics. Then, they successfully applied the trained model to predict the system behaviours over a long period. However, most deep learning models are considered as 'Black-Box' models, and their accuracy relies heavily on the quantity and quality of the labelled data, and the interpretability issue and the generalization problem can not be addressed effectively (Liang et al., 202).

In recent years, physics-informed Neural Networks (PINNs) as DNN solver for PDEs have attracted substantial interest for their effectiveness in solving complex engineering problems without relying on large amounts of labelled data (Zhao et al., 2023). For example, Mao et al. (2020) used PINN to approximate the feasible solution to the forward problem and identify the significant parameters of the reverse problem in different dimensions. Cai et al. (2020) applied the PINNs model to solve the heat convection in thermal and velocity fields under unknown boundary conditions. Zhang et al. (2023) proposed a multi-domain PINN that consists of multiple sub-networks to address the heat conduction problem of multi-layer materials. Wang et al. (2023) utilised PINNs to address the forward and inverse problems of variable-order space-fractional PDEs in advection-diffusion. Xu et al. (2024) models the dehydration stage of ternary cathode material sintering using a two-sphere model and physics-informed neural networks, improving accuracy in predicting reaction states and addressing temperature variations through simulations and empirical validation. These studies have significantly advanced the application of PINNs in various industrial scenarios. However, the conventional PINNs employ fully connected neural networks as the backbone, which face limitations in terms of computational efficiency and scalability when dealing with large-scale, high-dimensional problems. Furthermore, another open problem for training deep learning models for large-scale industrial processes is the lack of sufficient training samples due to the formidable time and cost to conduct sufficient experiments under different operating conditions.

In this paper, a physics-informed EngGeneNet framework is proposed to address the aforementioned research challenges. Within this framework, a physics-informed 2D loss function derived from the governing equations of heat conduction is formulated to guide the training of deep learning models. This approach enables the use of deep learning models to effectively approximate the solutions to the underlying physical equations. The hard-constrained boundary conditions are integrated into the model loss function, thereby enforcing strict compliance with the physical environments. Furthermore, to improve the model training convergence speed and generalization ability, an Engineering-genes ('Eng-Gene') module is designed to embed 'a priori' physics knowledge of heat conduction into the deep learning model. The 'Eng-Gene' are fundamental functions and mathematical operators that reflect the unique non-linearity and functional operations present in the governing first principle laws and empirical equations. They may be viewed in analogy to the genes in the biological system. This approach is therefore named 'Eng-Gene' - an abbreviation of 'Engineering Genes' (Li et al., 2005). The advantage of such a physics-informed modelling method is that it can be flexible and easily applied to other large-scale engineering systems where data samples collected from the real system are limited.

To further address the lack of sufficient training data for complex industrial processes, a knowledge distillation approach is adopted to transfer the insights from a more accurate model to the proposed model. Specifically, the ADI method is employed to solve the original PDEs to obtain sufficient and accurate samples for training the deep learning models. Moreover, during the processing of intermediate steel products using a reheating furnace, the product properties can vary significantly for different production batches. To reduce training time and enhance modelling efficiency, transfer learning is employed to address product property variations under diverse operating conditions. Additionally, the integration of the 'Eng-Gene' module can further enhance the effectiveness and efficiency of transfer learning. The proposed physics-informed EngGeneNet framework is validated on a pilot-scale walking beam furnace model with a range of product batches under different operating conditions (Hu et al., 2016).

In summary, the main contributions of this paper are summarised as follows:

1) A physics-informed 2D convolution loss function is formulated by utilizing customized convolution kernels derived from governing equations with hard-constrained boundary conditions to guide the training of deep learning models for predicting the heat conduction of large-scale intermediate steel products.

2) An 'Eng-Gene' module is developed to incorporate 'a priori' physics knowledge about the temperature-dependent material properties into the deep learning model to enhance the model generalization ability and accelerate the convergence speed.

3) The knowledge distillation approach is adopted to generate sufficient training samples for the deep learning model, while transfer learning is used to address the dataset feature space variations under different operating conditions and expedite the model training process.

The remainder of this paper is organized as follows: Section 2 introduces the thermal conduction modelling of intermediate steel products in the reheating furnace. Section 3 presents the details of the physics-informed EngGeneNet framework and the transfer learning. The experiment setup and implementation details are introduced in Section 4. Section 5 presents and discusses the results. Finally, Section 6 concludes the paper.

2. Thermal modelling of intermediate steel products in reheating furnaces

As shown in Figure 1, the intermediate steel products considered in this study are primarily standard rectangular prisms with dimensions (x, y, z), and have an initial temperature distribution upon passing through the reheating furnace. For modelling simplicity and in line with practical considerations, the temperature gradient variation in the z-dimension of the stock in the

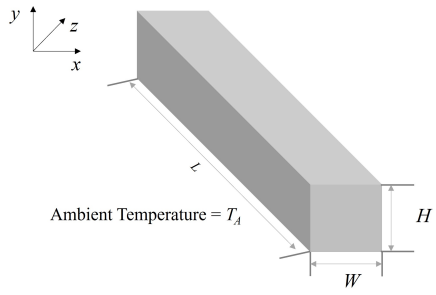

Figure 1: Schematic of the intermediate steel product

reheating furnace is typically regarded as negligible compared with the other 2D directions. Therefore, the heat conduction of steel products in the reheating furnace can be modelled as a 2D process. Heat exchange between the stock and the furnace environment occurs through the four boundaries, while thermal energy diffuses inward from the surface toward the interior of the stock.

Under the assumption of 2D transient heat conduction, the thermal behaviour of the stock is governed by Fourier's law, expressed as:.

$$\frac{\partial T}{\partial t} = \frac{k(T)}{\rho C_p(T)}(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}), \tag{1}$$

$$T = T_t(x, y), \tag{2}$$

where x is the steel product width direction, y is the steel product height direction. $k(T)$ and $C_p(T)$ are temperature-dependent material properties, representing the thermal conductivity and specific heat capacity, respectively. ρ is the material density.

To obtain accurate and stable solutions to the governing equation, appropriate boundary conditions must be specified. In the reheating furnace, the steel stock exchanges heat with the surrounding environment through both radiation and convection at its surface. Accordingly, a hard boundary heat transfer condition can be described by the following expression:

$$-k(T)\frac{\partial T}{\partial \vec{n}} = h_c(T - T_A) + \sigma e(T^4 - T_A^4), \tag{3}$$

where $k(T)$ is the thermal conductivity, h_c denotes the heat convection heat transfer coefficient, σ is the Stefan-Boltzmann constant ($5.6687 \times 10^{-8} Wm^{-2}K^{-4}$), e stands for the emissivity. $-k(T)\frac{\partial T}{\partial \vec{n}}$ is the heat flux normal to the surface. T_A is the ambient temperature.

3. Physics-informed EngGeneNet

In this study, a physics-informed EngGeneNet framework is proposed to approximate the solutions to the governing PDEs for heat conduction by formulating a 2D loss function using the finite difference method. The deep learning model is then employed as a surrogate PDEs solver to predict the transient temperature distribution of steel stocks in reheating furnaces. To implement the proposed physics-informed 2D loss, customized convolution kernels are designed to efficiently approximate spatial derivatives. Moreover, hard-constrained boundary conditions are embedded within the loss function to ensure stable model training and to account for environmental variations across different operating conditions. In addition, an 'Eng-Gene' module is integrated into the deep learning model architecture to enhance generalization capability and accelerate convergence. The overall framework is illustrated in Figure 2.

3.1. Formulation of the 2D loss function

First of all, a square mesh is constructed to discretise the cross-section of the rectangular steel products, with the objective of predicting the temperature evolution at each mesh node. Owing to the image-like spatial temperature distribution, a CNN architecture is adopted as the backbone of the surrogate model for efficient approximation of PDEs solutions. Then, the developed physics-informed loss function plays an important role in guiding the model training direction and shaping the model behaviour. Therefore, the accurate formulation of the loss function based on the underlying physical laws is essential for ensuring both stability and fidelity in model predictions.

In this study, the 2D temperature distribution $T_t(x, y)$ is divided into the interior domain D_I and boundary domain ∂D, as follows:

$$D_I = \{(x, y)|0 < x < W, 0 < y < H\}, \tag{4}$$

$$\partial D = \{(x, y)|0 \le x \le W, y = 0\, or\, H; 0 \le y \le H, x = 0\, or\, W\}, \tag{5}$$

where W and H are the width and height of the stock, respectively. The discretised computation domain with dimensions $m \times n$ for the model input $T_t(x, y)$ and the shape of model output $\widetilde{T}_{t+1}(x, y)$ is illustrated in Figure 3, where, $m = W/\Delta x$ and $n = H/\Delta y$, Δx and Δy are the space intervals in two dimensions.

To approximate the time derivative in the heat conduction governing equation, the forward difference operator is adopted, as follows:

$$\frac{\partial T}{\partial t} = \frac{T_{t+1}(x, y) - T_t(x, y)}{\Delta t} + o(\Delta t), \tag{6}$$

for interior domain nodes D_I, the second-order spatial derivative in the x-axis can be expressed as follows:

$$\frac{\partial^2 T}{\partial x^2} = \frac{T_t(x + 1, y) - 2T_t(x, y) + T_t(x - 1, y)}{(\Delta x)^2} + o(\Delta x)^2, \tag{7}$$

similarly,

$$\frac{\partial^2 T}{\partial y^2} = \frac{T_t(x, y + 1) - 2T_t(x, y) + T_t(x, y - 1)}{(\Delta y)^2} + o(\Delta y)^2, \tag{8}$$

where $o(\Delta t)$, $o(\Delta x)^2$ and $o(\Delta y)^2$ are the truncation errors, and Δt, Δx, and Δy are the step intervals for time, and x and y dimensions. By substituting the above equations into the governing Equation (1), the temperature evolution on each node within the domain can be computed. To further enhance the accuracy

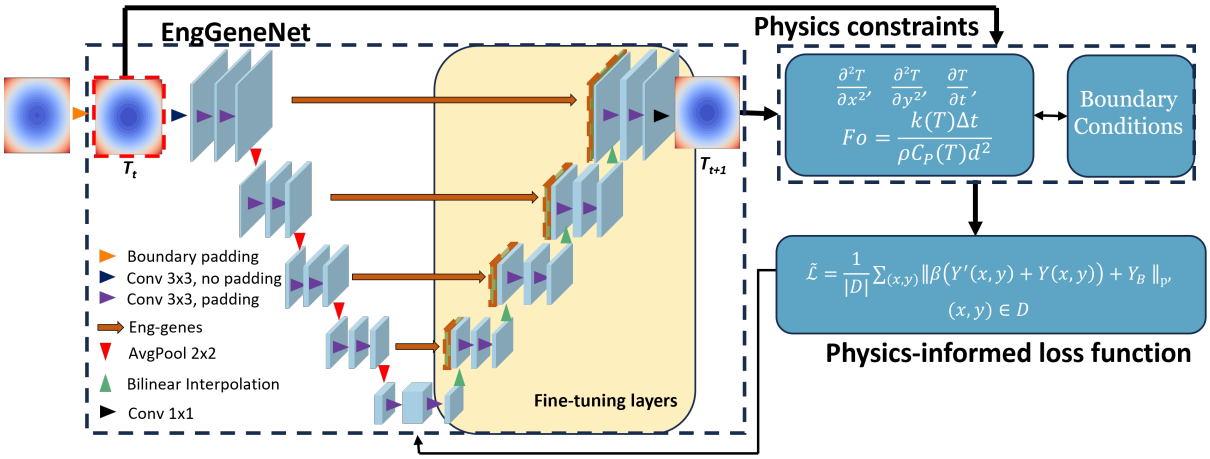

Figure 2: The framework of the proposed physics-informed EngGeneNet

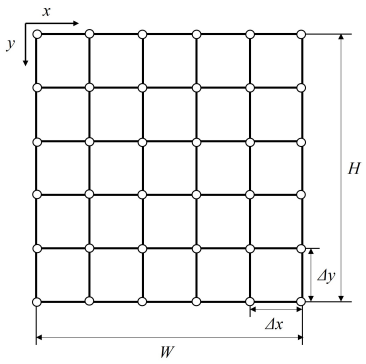

Figure 3: Square mesh covering a cross-section of the stock

and stability of the solutions and facilitate the design of the proposed physics-informed 2D loss function, the Crank-Nicolson method (Crank and Nicolson, 1947) is employed to improve the accuracy of the discretisation of the second-order spatial derivative. This method approximates the second-order spatial derivative at an intermediate time step between t and $t+1$. The corresponding formulations are described as follows:

$$\frac{\partial^2 T}{\partial x^2} = \epsilon \frac{\partial^2 T}{\partial x^2}\bigg|_{t+1} + (1-\epsilon)\frac{\partial^2 T}{\partial x^2}\bigg|_{t}, \tag{9}$$

$$\frac{\partial^2 T}{\partial y^2} = \epsilon \frac{\partial^2 T}{\partial y^2}\bigg|_{t+1} + (1-\epsilon)\frac{\partial^2 T}{\partial y^2}\bigg|_{t}, \tag{10}$$

where ϵ is a constant which determines the weight of the time-step t and $t+1$, and $0< \epsilon \leq 1$. As ϵ approaches 1, the accuracy decreases progressively for a given Δt, Δx, and Δy. When $\epsilon = 0.5$, the minimum accumulated truncation error for the systems is achieved, and the resulting finite difference formulation is known as the Crank–Nicolson method, which offers high stability. Accordingly, the second-order spatial derivative can be reformulated as follows:

$$\begin{aligned}\frac{\partial^2 T}{\partial x^2} =& \frac{1}{2}\frac{T_{t+1}(x+1,y) - 2T_{t+1}(x,y) + T_{t+1}(x-1,y)}{(\Delta x)^2} \\ &+ \frac{1}{2}\frac{T_t(x+1,y) - 2T_t(x,y) + T_t(x-1,y)}{(\Delta x)^2}\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{\partial^2 T}{\partial y^2} =& \frac{1}{2}\frac{T_{t+1}(x,y+1) - 2T_{t+1}(x,y) + T_{t+1}(x,y-1)}{(\Delta y)^2} \\ &+ \frac{1}{2}\frac{T_t(x,y+1) - 2T_t(x,y) + T_t(x,y-1)}{(\Delta y)^2}\end{aligned} \tag{12}$$

The proposed 2D physics-informed loss function, designed to guide the training of the deep learning model, is constructed based on the discrete form of the governing heat conduction equations. In this work, the spatial intervals are set to be equal $\Delta x = \Delta y = d$. Under this assumption, the discretised form of the governing equation in the time domain is expressed as follows:

$$\begin{aligned}&T_{t+1}(x+1,y) + T_{t+1}(x-1,y) + T_{t+1}(x,y+1) \\ &+ T_{t+1}(x,y-1) - 4T_{t+1}(x,y) \\ &+ T_t(x+1,y) + T_t(x-1,y) + T_t(x,y+1) \\ &+ T_t(x,y-1) - 4T_t(x,y)) \\ &= 2(T_{t+1}(x,y) - T_t(x,y))/Fo,\end{aligned} \tag{13}$$

where

$$Fo = \frac{k(T)\Delta t}{\rho C_p(T) d^2}, \tag{14}$$

By rearranging all terms to one side and grouping like terms, the equation becomes:

$$\begin{aligned}&T_{t+1}(x+1,y) + T_{t+1}(x-1,y) + T_{t+1}(x,y+1) \\ &+ T_{t+1}(x,y-1) + 4(-1 - \frac{1}{2Fo})T_{t+1}(x,y) \\ &+ T_t(x+1,y) + T_t(x-1,y) + T_t(x,y+1) \\ &+ T_t(x,y-1) + 4(-1 + \frac{1}{2Fo})T_t(x,y) = 0,\end{aligned} \tag{15}$$

which forms the foundational expression for the physics-informed 2D loss function used to supervise the training of the deep learning model.

3.2. Customized loss convolution kernels

To implement the proposed physics-informed 2D loss function within the deep learning framework, a customized convolution kernel is designed to compute the approximation error during model training. Specifically, the discrete-time governing Equation (15) is decomposed into two components: T_t, representing the temperature distribution at the current time step, and T_{t+1}, representing the temperature distribution at the subsequent time step. The latter corresponds to the model's predicted output and is denoted as $\widetilde{T}_{t+1}$. This formulation enables the training process to be guided by minimizing the residual of the left-hand side of the equation, thereby enforcing consistency with the underlying physical law. As a result, the model training can be considered as the following optimization problem:

$$\begin{aligned} &\min_{\widetilde{T}_{t+1}(x,y)} \frac{1}{|D_I|} \sum_{(x,y)} \left\| \beta \cdot [Y'(x,y) + Y(x,y)] \right\|_p \\ &\quad s.t. \\ &\quad (x,y) \in D_I, \\ &\quad -k(T)\frac{\partial T}{\partial \vec{n}} = h_c(T - T_A) + \sigma e(T^4 - T_A^4) \end{aligned} \tag{16}$$

where D_I is the computation domain, p is the order of the norm, and β is the weighting factor to regulates the relative importance assigned to the interior computational domain. $Y'(x,y)$ is defined to include all the terms needed to predict the temperature distribution at the next time step in Equation (15):

$$\begin{aligned} Y'(x,y) =& \widetilde{T}_{t+1}(x+1,y) + \widetilde{T}_{t+1}(x-1,y) \\ &+ \widetilde{T}_{t+1}(x,y+1) + \widetilde{T}_{t+1}(x,y-1) \\ &+ 4(-1 - \frac{1}{2Fo})\widetilde{T}_{t+1}(x,y), \end{aligned} \tag{17}$$

$Y(x,y)$ is defined to include all the terms from the current temperature distribution at the current time-step t in Equation (15):

$$\begin{aligned} Y(x,y) =& T_t(x+1,y) + T_t(x-1,y) \\ &+ T_t(x,y+1) + T_t(x,y-1) \\ &+ 4(-1 + \frac{1}{2Fo})T_t(x,y) \end{aligned} \tag{18}$$

Finally, the objective function in the optimization problem is reformulated as the inner domain loss function for the neural model training, as follows:

$$\widetilde{\mathcal{L}} = \frac{1}{|D_I|} \sum_{(x,y)\in D_I} \left\| \beta [Y'(x,y) + Y(x,y)] \right\|_p \tag{19}$$

where $Y'(x,y)$ and $Y(x,y)$ represent the residual components associated with the predicted and input temperature fields, respectively. Subsequently, to ensure consistency between model training and loss evaluation, two customized convolution kernels are designed to efficiently compute $Y'(x,y)$ and $Y(x,y)$ based on the discretised physics-informed formulation.

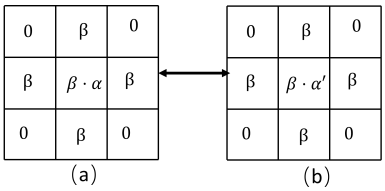

Figure 4: (a) and (b) are the customized convolution kernels for computing $Y'(x,y)$ and $Y(x,y)$ respectively

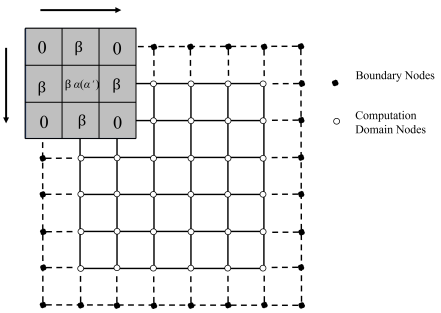

Figure 5: Customized loss function convolution process

The structure of these convolution kernels is illustrated in Figure 4, with their corresponding coefficients α' and α computed as:

$$\alpha' = 4(-1 - \frac{1}{2Fo}), \tag{20}$$

$$\alpha = 4(-1 + \frac{1}{2Fo}) \tag{21}$$

where Fo is the Fourier number as previously defined.

Figure 5 presents a schematic diagram illustrating the application of the customized convolution kernels to both the model input and output. The interior-domain loss value is subsequently computed based on the resulting convolution outputs.

3.3. Establishment of boundary conditions

Accurate boundary condition constraints are essential for ensuring physically consistent solutions to the governing equa-

1	1	1	1	1	1	1
1	0	0	0	0	0	1
1	0	0	0	0	0	1
1	0	0	0	0	0	1
1	0	0	0	0	0	1
1	0	0	0	0	0	1
1	1	1	1	1	1	1

Figure 6: Mask of boundary constraints

tions in heat conduction problems. In this study, a hard-constrained boundary condition is incorporated into the model loss to explicitly enforce both convective and radiative boundary conditions. This design enhances the robustness and stability of the neural network, particularly under diverse thermal environments. Specifically, by applying the inward finite difference approximation to the boundary condition and rearranging the terms into a residual form, the boundary condition loss Y_B at each boundary node is defined as follows:

$$Y_B = k(T)\frac{T_B - T_P}{d} + h_c(T_B - T_A) + \sigma e(T_B^4 - T_A^4) \tag{22}$$

where T_B denotes the boundary node temperature, and T_P is the adjacent interior node. To account for the different spatial orientations of the boundary, the loss terms are formulated separately for each edge of the computational domain:

$$\begin{cases} Y_B(0,y) = k(T)\dfrac{T(1,y) - T(0,y)}{d} + h_c(T(0,y) - T_A) \\ \qquad + \sigma e(T(0,y)^4 - T_A^4), 0 \le y \le n \\ Y_B(m,y) = k(T)\dfrac{T(m,y) - T(m-1,y)}{d} + h_c(T(m,y) - T_A) \\ \qquad + \sigma e(T(m,y)^4 - T_A^4), 0 \le y \le n \\ Y_B(x,0) = k(T)\dfrac{T(x,1) - T(x,0)}{d} + h_c(T(x,0) - T_A) \\ \qquad + \sigma e(T(x,0)^4 - T_A^4), 0 \le x \le m \\ Y_B(x,n) = k(T)\dfrac{T(x,n) - T(x,n-1)}{d} + h_c(T(x,n) - T_A) \\ \qquad + \sigma e(T(x,n)^4 - T_A^4), 0 \le x \le m \end{cases} \tag{23}$$

To efficiently integrate this hard-constrained boundary loss into the CNN training framework, a binary computation mask (illustrated in Figure 6) is employed to localize the loss computation strictly to the boundary region ∂D. This mask-based approach enables tensor-wise parallelization and is fully compatible with modern deep learning models.

The resulting boundary loss Y_B is then incorporated into the physics-informed 2D loss function to formulate the final loss function for model training, as follows:

$$\widetilde{\mathcal{L}} = \frac{1}{|D|}\sum_{(x,y)\in D} \left\| \beta[Y'(x,y) + Y(x,y)] + Y_B \right\|_p \tag{24}$$

where β is the weighting factor to regulates the relative importance assigned to the interior computational domain and the boundary conditions. The integration of Y_B from the hard-constrained boundary conditions not only enforces physical consistency at the domain boundaries but also ensures that the learning process remains stable and grounded in domain-specific laws. This hybrid formulation, combining data-driven learning with physically meaningful constraints, significantly enhances the model robustness and generalization under various thermal boundary conditions.

3.4. Development of EngGene module

In this work, the CNN architecture is adopted as the model backbone to approximate the proposed 2D loss function, which can significantly reduce the computation time leveraging the parallel computing capabilities of convolution operations. This approach has significant potential for the real-time operation and control of the reheating furnace. The UNet is a U-shaped CNN architecture that includes both contracting and expansion paths (Ronneberger et al., 2015). The contraction path captures global contextual information, while the expansion path provides high-resolution spatial details. These feature maps generated by the two paths are then concatenated together through a skip connection. This unique structure has achieved significant success in various image-related tasks. In contrast to the conventional applications of UNet in image segmentation, this study reformulates the task as an image-to-image regression problem, wherein the objective is to predict continuous temperature fields.

In addition, key material properties such as specific heat capacity $C_p(T)$ and thermal conductivity $k(T)$ exhibit strong nonlinearity and temperature dependence during the heat conduction process. In this study, these dependencies are modelled using polynomial functions of temperature, expressed as:

$$\begin{cases} C_p(T_m) = a_1 + a_2 T_m + a_3 T_m^2 + a_4 T_m^3 + a_5 T_m^4 + a_6 T_m^5, \\ k(T_m) = b_1 + b_2 T_m \end{cases} \tag{25}$$

where T_m is the mean value on the whole computation domain at the current time step. The polynomial coefficients a_i and b_i determine the heat transfer properties of different materials and are provided in Table 1 (Hu et al., 2016).

Furthermore, the Fourier number (Fo) is a dimensionless parameter that characterizes the ratio of the heat diffusion rate to the time-scale of a transient thermal process. It plays a critical role in characterizing the transient temperature evolution of materials and is directly influenced by $C_p(T)$ and $k(T)$. As shown in Equation (13), the governing equation depends on $T(x,y)/Fo$, indicating the presence of a salient fundamental physical relationship within the physics model.

When solving PDEs analytically, explicitly incorporating such temperature-dependent material properties can substantially increase computational complexity, especially for large-scale problems. Conversely, relying solely on data-driven models such as CNNs to learn these complex physical relationships may result in poor generalization and unreliable performance.

To address this limitation, 'a priori' engineering knowledge is embedded into the neural network training process through the design of specialized convolution layers that encode such explicit physical relationships. This approach, referred to as the 'Eng-Gene' method, extracts fundamental functional components that capture nonlinearities inherent in the governing physical laws of engineering systems. These components act as thermal 'genetic' information, enriching the learning process.

In this work, the corresponding 'Eng-Gene' module associated with the Fourier number can be defined as follows:

$$\widetilde{W} = \frac{w}{Fo} = \frac{w\rho C_p(T)d^2}{k(T)\Delta t}, \tag{26}$$

where w is the model input data, while $\widetilde{W}$ is the output of the designed 'Eng-Gene' module, $C_p(T)$ and $k(T)$ are temperature-

Table 1: Polynomial coefficients a_i for $C_p(T)$ and b_i for $k(T)$

	$a_1(b_1)$	$a_2(b_2)$	a_3	a_4	a_5	a_6
$C_p(T)$	0.377E+03	0.125E+00	-0.197E-04	-0.253E-06	0.398E-09	-0.180E-12
$k(T)$	0.385E+03	-0.481E-01	-	-	-	-

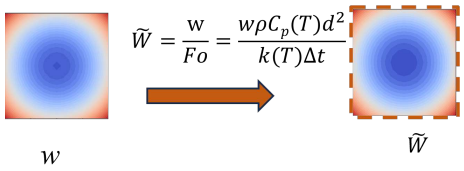

Figure 7: 'Eng-Gene' module

dependent properties. $\widetilde{W}$ can be regarded as an intermediate variable of the neural model, as illustrated in Figure 7. It helps to extract the underlying nonlinear relationship between the input and output, thereby enabling the model to more accurately and quickly minimize the physics-informed loss function.

The overall framework of the proposed physics-informed EngGeneNet is illustrated in Figure 2, where the input is a 2D temperature distribution T_t at time-step t, and the output is the predicted transient temperature distribution $\widetilde{T}_{t+1}$ at time-step $t+1$.The ambient conditions are implemented around the computational domain through a designed padding function. Consequently, the first convolution layer is left unpadded to preserve the original data shape. As the input and the model output are required to calculate the physics-informed loss function during the model training. To facilitate this computation, a skip connection is introduced between the two ends of the neural model. Considering that the convolution layers at different levels in the UNet architecture are responsible for extracting features of varying complexities, the 'Eng-Gene' module is embedded into each level of the convolution layers in UNet architecture to enhance the model prediction performance and the generalization ability, as shown on the orange path in Figure 2. Notably, due to the plug-and-play feature, the 'Eng-Gene' module can not only be easily integrated into the UNet framework but can also be flexibly integrated with other neural network architectures.

3.5. Transfer learning

Most deep learning techniques face challenges when addressing the issue of feature space deviation among different datasets. A commonly adopted solution to mitigate this issue is transfer learning (TL). In this study, the variability in operating conditions is also considered. Specifically, differences in size, density, and other material properties across production batches can lead to shifts in the feature space of the temperature distribution profiles. To reduce model training time while maintaining comparable prediction accuracy, the transfer learning approach is employed.

In the implementation, the material property parameters within both the physics-informed loss function and the 'Eng-Gene' module are updated to adapt to the varying operating conditions. Subsequently, only the parameters in the expansion path of the UNet architecture (highlighted in yellow in Figure 2) are fine-tuned using the pre-trained baseline model, while the parameters in the contraction path are kept frozen. This design leverages the role of the contraction path in capturing global contextual features, while the expansion path is responsible for reconstructing higher-resolution spatial details. Then this fine-tuning process is performed using a relatively small subset of data samples and training epochs, making the adaptation process efficient. In addition, the integration of the 'Eng-Gene' module significantly enhances the effectiveness of the transfer learning strategy by embedding the physical variation into the model.

4. Experiment setup and implementation

4.1. Furnace operating conditions and product batches

The furnace investigated in this study is a large-scale real walking-beam bloom reheating furnace with dimensions of 36 m × 10 m (length × width) and a height ranging from 4.0 m to 4.7 m, as illustrated in Figure 8 (Hu et al., 2016). The furnace model is divided into six control zones (CZs), where CZ1 and CZ3 are the master thermal control zones, while CZ2 and CZ4 are subordinate zones. The thermal inputs to the subordinate zones are proportionally regulated based on the thermal inputs of the corresponding master control zones. During operation, blooms enter the furnace from one end, progress through the heating zone, and are discharged from the opposite end.

In the experimental evaluation, two distinct batches of mild steel blooms, designated SB1 and SB2, were tested, with their material properties summarized in Table 2. Furthermore, to assess the model prediction performance under varying physical conditions, a range of operating conditions (OCs) was introduced, as detailed in Table 3. Specifically, OC1 serves as the baseline operating condition and is used to configure the hyperparameters of the baseline model. OC2 is designed to evaluate the transfer learning performance when adapting to different product batches. OC3 is employed to assess model robustness under varying data sampling intervals. OC4 is introduced to further validate the robustness and adaptability of the proposed EngGeneNet under diverse environmental conditions. These varying operating conditions and product types provide a comprehensive framework for evaluating the proposed EngGeneNet across several key performance metrics, including generalization capability, transfer learning effectiveness, and training efficiency.

4.2. Knowledge distillation approach to generate training data

A conventional method for capturing the temperature profile of steel blooms is the use of radiometric imaging cam-

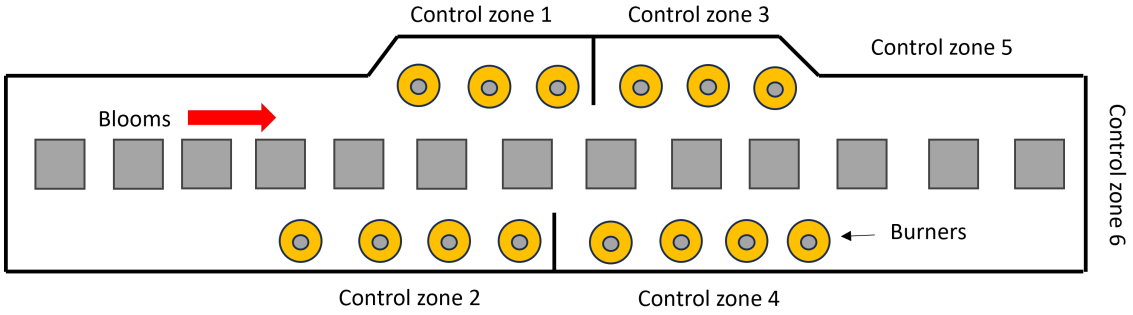

Figure 8: Schematic of the reheating furnace

Table 2: Material properties in the experiment

	Properties	Value	Unit
SB1	Mass	5353	kg
	Thermal conductivity($k(T)$)	Eqn. (25)	$Wm^{-1}K^{-1}$
	Specific heat capacity($C_p(T)$)	Eqn. (25)	$Jkg^{-1}K^{-1}$
	Convection heat transfer coefficient(h_c)	25	$Wm^{-2}K^{-1}$
	Stephan-Boltzmann constant(σ)	5.6687×10^{-8}	$Wm^{-2}K^{-4}$
	Emissivity	0.8	-
	Width	0.3	m
	Height	0.3	m
	Length	7.625	m
	d	0.01875	m
SB2	Mass	8827	kg
	Thermal conductivity($k(T)$)	Eqn. (25)	$Wm^{-1}K^{-1}$
	Specific heat capacity($C_p(T)$)	Eqn. (25)	$Jkg^{-1}K^{-1}$
	Convection heat transfer coefficient(h_c)	25	$Wm^{-2}K^{-1}$
	Stephan-Boltzmann constant(σ)	5.6687×10^{-8}	$Wm^{-2}K^{-4}$
	Emissivity	0.8	-
	Width	0.4	m
	Height	0.4	m
	Length	7.028	m
	d	0.025	m

Table 3: Operating conditions in experiments

	Blooms	Initial Temp ($^\circ C$)	Vertical Temp ($^\circ C$)	Horizontal Temp ($^\circ C$)	Δt (s)
OC1	SB1	20	1250	1250	1
OC2	SB2	20	1200	1200	2
OC3	SB1	35	1250	1250	2
OC4	SB2	20	**1200**	**1000**	1

eras. However, this technique only provides surface temperature measurements and fails to capture the internal thermal distribution throughout the entire stock. Moreover, acquiring sufficient experimental data for training pure data-driven deep learning models in complex industrial settings is often prohibitively expensive and time-consuming. To overcome these limitations, numerical methods such as the ADI method have been widely employed to generate operation data. This approach was also adopted in previous work by Hu et al. (2016).

The ADI method decomposes the energy balance equation at each time step into two sequential sets of finite difference equations (FDEs), solving them alternately along different spatial directions. In the first half-step, the terms involving the y-direction are treated as unknowns. Applying the energy balance principle, the governing equation for this half-time step is formulated as:

$$\begin{aligned}&(T_{t+1}(x,y-1)-T_{t+1}(x,y))+(T_{t+1}(x,y+1)-T_{t+1}(x,y))\\&+(T_t(x-1,y)-T_t(x,y))+(T_t(x+1,y)-T_t(x,y))\\&=\left(\frac{\rho C_p(T)d^2}{k(T)}\right)\left(\frac{T_{t+1}(x,y)-T_t(x,y)}{\Delta t/2}\right)\end{aligned} \tag{27}$$

In the second half-step, the roles of spatial directions are reversed, and the x-direction terms are treated as unknowns. The

Table 4: Performance comparison of various deep learning backbones with and without 'Eng-Gene' in the physics-informed framework

Models	Training dataset			Test dataset		
	MAE (Reduction)	RMSE (Reduction)	MAE-MaxE (Reduction)	MAE (Reduction)	RMSE (Reduction)	MAE-MaxE (Reduction)
PI-VGG	1.257	1.534	4.200	1.541	1.821	4.621
PI-EngGeneVGG	0.595 (↓52%)	0.727 (↓52%)	2.405 (↓42%)	0.598 (↓61%)	0.730 (↓59%)	2.416 (↓51%)
PI-FPN	0.477	0.530	1.193	0.699	0.753	1.437
PI-EngGeneFPN	0.169 (↓64%)	0.215 (↓59%)	0.715 (↓40%)	0.171 (↓75%)	0.217 (↓71%)	0.720 (↓49%)
PI-UNet	0.276	0.298	0.557	0.515	0.538	0.822
PI-EngGeneUNet	**0.026 (↓90%)**	**0.044 (↓85%)**	**0.305 (↓45%)**	**0.029 (↓94%)**	**0.048 (↓91%)**	**0.313 (↓61%)**

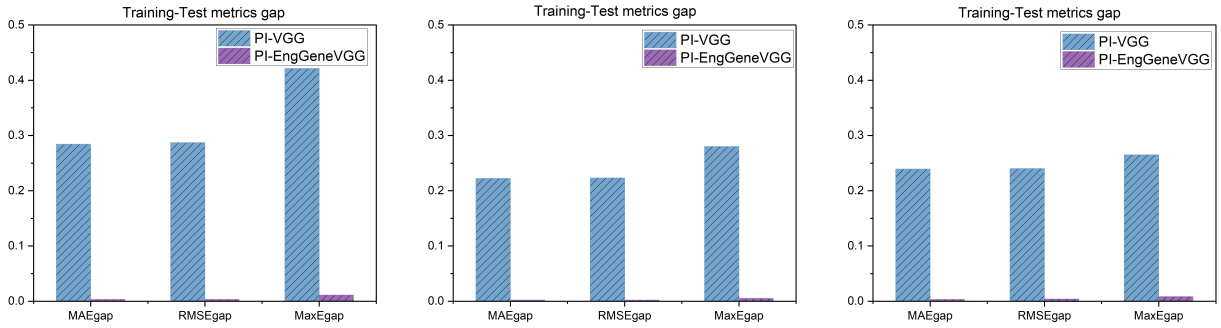

Figure 9: Metrics gap between Training&Test datasets of various deep learning models with and without the 'Eng-Gene' module

corresponding equation is given by:

$$\begin{aligned}&(T_t(x,y-1)-T_t(x,y))+(T_t(x,y+1)-T_t(x,y))+\\&(T_{t+1}(x-1,y)-T_{t+1}(x,y))+(T_{t+1}(x+1,y)-T_{t+1}(x,y))\\&=\left(\frac{\rho C_p(T)d^2}{k(T)}\right)\left(\frac{T_{t+1}(x,y)-T_t(x,y)}{\Delta t/2}\right)\end{aligned} \tag{28}$$

The boundary conditions are consistent with those defined in Equation (23). The final temperature distribution is obtained by alternately solving these finite difference equations using the Gaussian elimination method at each time step.

4.3. Implementation

In the knowledge distillation experiments, a total of 12,000 two-dimensional temperature distribution samples under operating condition OC1 were generated using the ADI method to develop the baseline model. Additionally, 6,000 samples under OC2 and OC4, and 5,000 samples under OC3 were generated to fine-tune the physics-informed EngGeneNet and to evaluate model performance under varying operating conditions. The full dataset was divided into training, validation, and testing sets with a ratio of 8:1:1. In the model configuration, the padding values in the convolution layers are set to 'zeros'. The Adam optimiser is used to train the model. The base model was trained in 50 epochs. However, during the transfer learning, only 30 epochs of training were conducted on a relatively small training dataset, achieving the same accuracy as the base model. All experiments were conducted on a workstation equipped with 8 Intel Core processors, 16GB of RAM, and a GTX2080 GPU with 8GB of graphics memory.

To quantitatively assess the performance of the proposed physics-informed EngGeneNet for predicting the temperature profile of blooms in the reheating furnace, the following evaluation metrics are used: the mean absolute error (MAE), the root of the mean square error for all nodes (RMSE), and the mean absolute error of maximum absolute errors for the dataset (MAE-MaxE).

5. Results and discussion

5.1. Ablation studies

To validate the effectiveness and adaptability of the proposed physics-informed framework, a series of ablation experiments involving different deep learning backbones were conducted, including VGG (Simonyan and Zisserman, 2014), Feature Pyramid Network (FPN) (Lin et al., 2017), and UNet. These architectures were selected to cover a broad spectrum of convolutional network designs, from the sequential convolutional layers in VGG, to the multi-scale feature fusion strategy in FPN, and the encoder-decoder structure in UNet. In each case, the model performance was evaluated both with and without the integration of the proposed 'Eng-Gene' module. Due to its modular and plug-and-play design, the 'Eng-Gene' module can be seamlessly incorporated into these networks without altering the overall architecture. The module preserves the spatial receptive field while introducing 'a priori' engineering

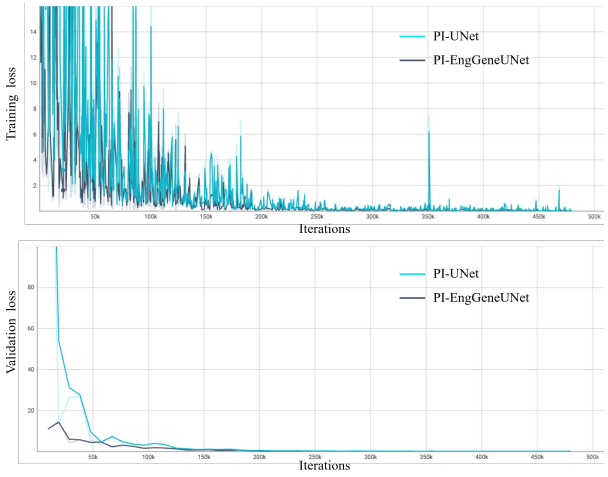

Figure 10: Training and validation loss evolution of PI-EngGeneUNet and PI-UNet

physics in the heat conduction problem. Specifically, in the VGG model, the 'Eng-Gene' module was introduced into the standard convolution layers in selected blocks. In the FPN architecture, the 'Eng-Gene' module was inserted at multiple levels to enhance the multi-scale representation learning. Furthermore, the loss function design remains unchanged across all backbone variants, as the physics-informed loss is formulated independently of the specific network architecture.

The results for both the training and test datasets, as presented in Table 4, demonstrate the superior prediction performance of the proposed physics-informed EngGeneNet model. First of all, integrating the 'Eng-Gene' module into the baseline models (VGG, FPN, and UNet) consistently leads to substantial performance improvements compared to their original backbone without the module. Furthermore, among the three deep learning models, EngGeneUNet exhibits the highest accuracy on both datasets. achieving at least 90% MAE reduction, 85% RMSE reduction, and 45% MAE-MaxE reduction on both the training and test datasets compared to the physics-informed UNet without the 'Eng-Gene' module. This confirms the significance of the 'Eng-Gene' module in improving the prediction performance of deep learning models. In addition, it is noteworthy that the values of MAE and RMSE on test datasets for EngGeneUNet are relatively low, both below 0.1. This shows that the proposed EngGeneUNet can achieve equivalent accuracy as the conventional ADI method.

To further evaluate the generalization ability of the proposed physics-informed EngGeneNet, Figure 9 presents a comparison of the performance gaps between training and test datasets in terms of three standard evaluation metrics: MAE, RMSE, and MAE-MaxE. These gaps are calculated as the absolute differences between the training and test errors, defined as follows:

$$\mathrm{MAE}_{\mathrm{gap}} = |\mathrm{MAE}_{\mathrm{test}} - \mathrm{MAE}_{\mathrm{train}}| \tag{29}$$

$$\mathrm{RMSE}_{\mathrm{gap}} = |\mathrm{RMSE}_{\mathrm{test}} - \mathrm{RMSE}_{\mathrm{train}}| \tag{30}$$

$$\mathrm{MaxE}_{\mathrm{gap}} = |\mathrm{MAE\text{-}MaxE}_{\mathrm{test}} - \mathrm{MAE\text{-}MaxE}_{\mathrm{train}}| \tag{31}$$

Notably, all the physics-informed EngGeneNet models with the 'Eng-Gene' module combined with the deep learning models exhibit far better generalization performance on all three performance metrics than the deep learning models without the 'Eng-Gene' module. This indicates that the proposed EngGeneNet effectively enhances the generalization ability on unfamiliar data. Furthermore, Figure 10 illustrates the evolution of the loss function for both the physics-informed EngGeneUNet and physics-informed UNet across training and validation datasets under the same number of training epochs. Notably, the initial training and validation losses of EngGeneUNet are consistently lower than the model with the UNet backbone. The PI-EngGeneUNet model training also exhibits a significantly faster convergence speed. These analyses confirm that the 'Eng-Gene' module plays a key role in significantly enhancing both the generalization ability and the convergence speed of deep learning models.

Table 5: Comparison results of the integration of the 'Eng-Gene' module or an extra convolution layer

Module	Metrics	Training	Test	epochs
Extra conv	MAE	0.099	0.106	50
	RMSE	0.139	0.148	
	MAE-MaxE	0.670	0.690	
Eng-Gene	MAE	0.026	**0.029(↓72%)**	50
	RMSE	0.044	**0.048(↓67%)**	
	MAE-MaxE	0.305	**0.313(↓54%)**	
	MAE	0.090	0.093	**23**
	RMSE	0.105	0.109	
	MAE-MaxE	0.311	0.331	

In addition, a comparison experiment was conducted to further evaluate the effectiveness of the proposed EngGeneUNet. In this experiment, the 'Eng-Gene' module in the proposed physics-informed EngGeneUNet was replaced with an extra 1x1 convolution layer so that both models have the same complexity. The two models were then trained for the same epochs. Table 5 shows the performance of the two models for all performance metrics. The results on both the training and test datasets show that the performance of the UNet model with an 'Eng-Gene' module is much better than the model with an extra convolution layer when both models were training with the same epochs, achieving 72% MAE reduction, 67% RMSE reduction and 54% MAE-MaxE reduction on the test dataset. Furthermore, the results show that the EngGeneUNet requires only 23 epochs to achieve a MAE comparable to that of the UNet model with an extra convolution layer trained for 50 epochs. This again confirms that the 'Eng-Gene' module can accelerate the training speed with significantly improved model performance.

To further validate the independent contribution of the 'Eng-Gene' module, an ablation study was conducted to investigate the role of different physical variables incorporated within the module. In this experiment, the variables d^2, ρ, $k(T)$, and

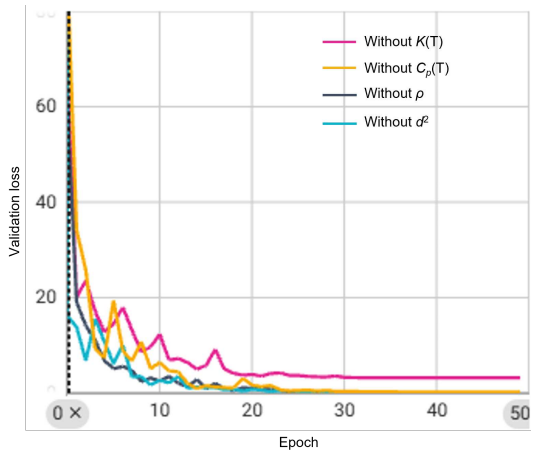

Figure 11: Validation loss evolution for 'Eng-Gene' module ablation study

Table 6: Ablation study results on different variables in 'Eng-Gene' module

Removed variables	Metrics	Training dataset	Test dataset
d^2	MAE	0.037	0.040
	RMSE	0.051	0.054
	MAE-MaxE	0.231	0.243
ρ	MAE	0.047	0.050
	RMSE	0.066	0.070
	MAE-MaxE	0.301	0.319
$C_p(T)$	MAE	0.065	0.068
	RMSE	0.095	0.097
	MAE-MaxE	0.472	0.496
$k(T)$	MAE	0.354	**0.555**
	RMSE	0.377	**0.577**
	MAE-MaxE	0.676	**0.891**

$C_p(T)$ were individually removed from the model to assess their respective importance and contribution to the overall performance. The performance degradation after the removal of each variable demonstrates its unique role in capturing the underlying physical dynamics. Figure 11 illustrates the evolution of the validation loss after excluding each of the aforementioned variables. The results show that the thermal conductivity $k(T)$ and the specific heat capacity $C_p(T)$ have a notable impact on both the convergence process and the final predictive accuracy. In particular, the removal of thermal conductivity $k(T)$ significantly degrades the convergence speed and overall model performance. This finding confirms the critical role of thermal conductivity in modelling heat conduction problems, particularly due to the strong nonlinearity of the term $1/k(T)$ within the proposed 'Eng-Gene' module. The final results on both training and test datasets are summarized in Table 6, which further confirms this conclusion. Furthermore, the increased performance gap between the training and test datasets when thermal conductivity $k(T)$ is removed indicates its substantial contribution to the generalization ability of the model.

5.2. Sensitivity analysis of model parameters

To evaluate the sensitivity and effectiveness of the proposed model, a sensitivity analysis of model parameters was conducted to investigate the impact of key hyperparameters, including the initial learning rate, the decay rate in the Adam optimizer, and the weighting factor applied in the loss function. Among these, the initial learning rate directly influences the convergence speed and stability of training. Figure 12(a) illustrates the results of the experiments conducted with initial learning rates ranging from 10^{-1} to 10^{-3}, covering a range of commonly used values (Kingma and Ba, 2014; Ronneberger et al., 2015). Larger learning rates accelerated convergence but often caused instability or divergence. Among the tested values, 10^{-2} and 10^{-3} demonstrated a comparable trade-off between convergence speed and stability, yielding similar and satisfactory performance.

The effect of the decay factor in the Adam optimizer, which gradually decreases the learning rate throughout the training process, was further investigated. This decay mechanism helps to stabilize updates in later stages of training, thereby improving generalization. As illustrated in Figure 12(b), moderate decay rates led to smoother convergence and better generalization performance. In contrast, a very small decay factor resulted in slow convergence. These results suggest that both the initial learning rate and the decay factor significantly influence the convergence behaviour of the model. However, when chosen within a reasonable range, their impact on the final performance is comparable.

Furthermore, the weighting factor β in the loss function regulates the relative importance assigned to the interior computational domain and the boundary conditions. Proper tuning of this parameter is essential to ensure balanced learning across the entire domain. In this work, an initial weighting factor $\beta = 0.25$ is introduced to balance the contributions from the interior and boundary domain. This value is chosen to normalize the coefficient associated with the convolution term $\alpha = 4(-1 + 1/(2Fo))$ and $\alpha' = 4(-1 - 1/(2Fo))$ in the customized kernels, ensuring comparable magnitudes between the inner nodes and the boundary nodes. To further investigate the sensitivity of the results to this parameter, we evaluate the model performance under varying β values of 0.1 and 1.0. A detailed comparison of the validation loss evolution is shown in Figure 12(c). The results indicate that the model achieves relatively stable and accurate performance when β is set to either 0.25 or 0.1. These findings confirm the critical role of properly weighting the boundary conditions in accurately solving heat conduction problems.

Several comparison experiments with different batch sizes (the number of training samples simultaneously fed to the model in each iteration), activation functions, padding modes, and loss functions were conducted to assess the impacts of various components in the deep learning models. Table 7 summarizes the experimental results, and it is noteworthy that significant performance degradation occurs when the padding mode

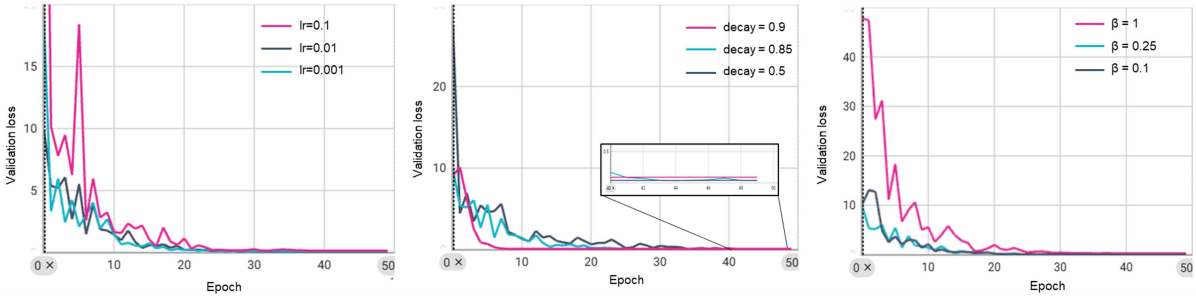

Figure 12: Training parameters analysis.(a) Initial learning rate; (b) Decay factor; (c) Loss weighting factor β

Table 7: Comparison results on test dataset for different components in the framework

Activation function	Padding mode	Loss function	Batch size	MAE	RMSE	MAE-MaxE
GELU	Replicate	L1 loss	1	0.278	0.297	0.582
GELU	Zeros	L2 loss	1	0.089	0.107	0.346
GELU	Zeros	L1 loss	1	**0.029**	**0.048**	**0.313**
RELU	Zeros	L1 loss	1	0.105	0.129	0.442
GELU	Zeros	L1 loss	15	0.048	0.070	0.372

is changed to 'Replicate', which may cause disruptions to the initially stable boundary conditions by the 'Replicate' padding mode. In addition, the activation function also has a significant impact on the prediction performance. On the other hand, the choice of different loss functions, for example, the L2 loss function has a mild impact on the accuracy of the training set. For different model training batch sizes, the results suggest that a smaller batch size leads to better performance on the test dataset. In this PDEs-solving task, a smaller batch size increases the number of iterations for approximating the physics-informed loss function, which may enhance the model training stability.

5.3. Validation under various physical constraints

A common step in machine learning is the model retraining when the feature space of the dataset varies significantly. Transfer learning has been an efficient strategy to fine-tune the pre-trained model for new applications with significantly changed feature space and limited new training data. To verify the effectiveness of the 'Eng-Gene' module in re-training speed while improving the model generalization performance in the transfer learning, this section presents the experimental results for transfer learning of physics-informed EngGeneUNet under different operating conditions, where the basic models were first trained under OC1, then re-trained under OC2, OC3, and OC4. Table 8 compares the performance of the re-trained deep learning models under different operating conditions using physics-informed EngGeneUNet and UNet, respectively. Firstly, the performance of the fine-tuned models for the varying operating conditions is comparable to the performance of the pre-trained models. Furthermore, it is noteworthy that the PI-EngGeneUNet performs better than PI-UNet under the new operating conditions, achieving more than 90% MAE and RMSE reductions and 60% or above MAE-MaxE reduction on test datasets. This finding confirms the ability of the 'Eng-Gene' module to improve the generalization ability of the deep learning models.

Figure 13 illustrates a few examples of the performance of the proposed EngGeneNet in different operating conditions. The first column shows the baseline temperature profile yielded by the conventional numerical ADI approach. The second column presents the temperature profile prediction by the proposed physics-informed EngGeneUNet. The third column shows the MAE error map between these two results. It is evident that the predicted 2D temperature profile predicted by the proposed physics-informed EngGeneUNet model is sufficiently close to the results produced by the conventional ADI approach. However, due to the formidable computational complexity of ADI method for large-scale and high-dimensional problems, the proposed EngGeneUNet model could achieve higher computational efficiency once trained, which is more suitable for real-time applications.

Table 9 summarizes the model training performance and training time with 30 training epochs and different model training batch sizes under operating condition OC2. The results show that increasing the batch size in the transfer learning does not significantly compromise the model performance for both the training and test datasets. Therefore, the training time in the transfer learning has been significantly reduced by over 75% when the batch size is increased from 1 to 15.

5.4. Transient temperature evolution performance

Evaluating the continuous prediction performance of the proposed physics-informed EngGeneNet is essential, as it poses greater challenges due to the accumulation of prediction errors over time. Therefore, experiments were conducted to compare the mean bulk temperature and temperature distribution uni-

Table 8: Comparison results under different operating conditions in transfer learning

base models	Metrics	OC2		OC3		OC4	
		Training set	Test set	Training set	Test set	Training set	Test set
PI-UNet	MAE	0.319	1.216	0.312	0.727	0.447	0.643
	RMSE	0.349	1.246	0.334	0.748	0.483	0.680
	MAE-MaxE	0.778	1.724	0.610	1.010	1.002	1.237
PI-EngGeneUNet	MAE	0.026	**0.028(↓97%)**	0.035	**0.038(↓94%)**	0.046	**0.047(↓92%)**
	RMSE	0.005	**0.012(↓99%)**	0.027	**0.060(↓91%)**	0.059	**0.062(↓90%)**
	MAE-MaxE	0.221	**0.234(↓86%)**	0.327	**0.366(↓64%)**	0.268	**0.270(↓78%)**

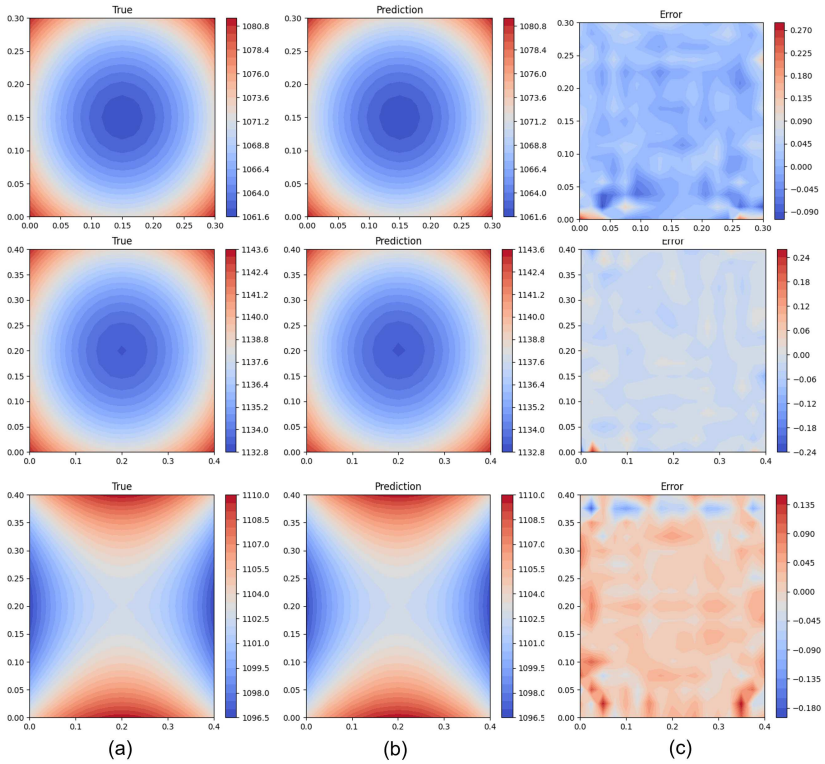

Figure 13: Several prediction results on test datasets. (a) ADI solutions; (b) PI-EngGeneNet solutions; (c) Error map

formity obtained using the conventional ADI method and the proposed physics-informed EngGeneUNet under varying operating conditions. Here, the temperature uniformity is defined as the standard error of the corresponding temperature distribution. Figure 14 (a) and (b) illustrate the evolutions of the mean temperature profile and the temperature uniformity for SB1 in a constant ambient environment of 1200°C. Notably, There is a comparable consistency in the two profiles between the two methods. Figure 14 (c) and (d) shows the mean temperature and temperature uniformity evolution for SB2 under varying physical conditions, where temperature changed from $1000°C$ to $1200°C$ at 8000s. It is noteworthy that the predicted temperature profiles by both the ADI and the EngGeneNet models exhibit consistently similar temperature profile prediction performance under both conditions. Although the temperature uniformity exhibits a slight difference of about $1°C$ in this operating condition, which is within an acceptable error range in this long-period evolution. These findings demonstrate the robustness of the proposed physics-informed EngGeneNet in predicting temperature profiles across varying operating conditions, as well as the effectiveness of model fine-tuning through transfer learning.

5.5. Model Computational efficiency

Figure 15 compares the computational time for predicting per frame of 2D temperature distribution by both the pre-trained physics-informed EngGeneUNet model and the conventional ADI method. It is shown that the EngGeneUNet model can achieve 96% time reduction than the conventional ADI approach in predicting bloom temperature profile, confirming that the EngGeneNet models offer great potential for real-time operation and control.

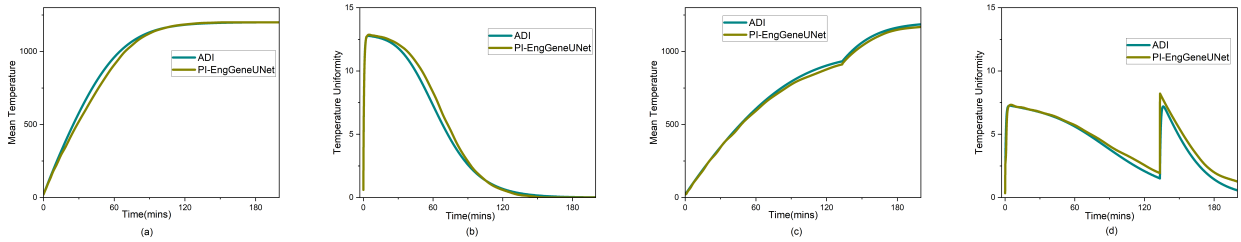

Figure 14: Predicted stock temperature history during the trial period

Table 9: The influence of batch size for transfer learning under OC2 (30 epochs)

Batch size	Metrics	Training dataset	Test dataset	Training time (saving)
1	MAE	0.038	0.040	115.62 mins
	RMSE	0.005	**0.011**	
	MAE-MaxE	0.208	**0.222**	
10	MAE	0.026	0.029	30.52 mins
	RMSE	0.005	0.012	
	MAE-MaxE	0.220	0.232	
15	MAE	0.026	**0.028**	27.94 mins (75%)
	RMSE	0.005	0.012	
	MAE-MaxE	0.221	0.234	

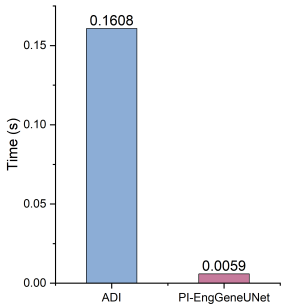

Figure 15: Computation time per frame of temperature distribution prediction

6. Conclusion

This paper has developed a physics-informed EngGeneNet for transient temperature profile prediction of large-scale intermediate steel stocks in reheating furnaces, highlighting the significance of incorporating the fundamental heat conduction physics knowledge into the backbone deep learning CNN models and the capability of the 'Eng-Gene' module in capturing salient features and interactions among the temperature-dependent material properties and variables. Formulating the 2D heat conduction modelling as an image-to-image task, the UNet is employed as the backbone of the EngGeneNet, and the physics-informed loss function was constructed by discretising the two-dimensional transient heat conduction governing equations. To improve the interpretability and generalization performance while accelerating the training of deep learning models, an 'Eng-Gene' module is extracted from first principle equations and integrated into the deep learning model. To address the lack of training samples and the material property variations under different operating conditions, both knowledge distillation and transfer learning are employed. The experimental results show that the constructed physics-informed loss function can effectively guide the training of deep learning models, achieving comparable prediction performance with the numerically expensive ADI approach. The proposed 'Eng-Gene' module can improve the prediction performance, guarantee the generalization performance, while accelerating the training and retraining of deep learning models. Physics-informed EngGeneNet is a universal function approximator used to solve the numerical solutions of the PDEs, while ensuring adherence to the underlying first-principles governing laws. Therefore, its application is not limited to the transient heat conduction problem discussed in this paper but can also deal with other physics-related problems, such as compressible and incompressible flow modelling problems and computational modelling of the forward and reverse problems. This study has further demonstrated that 'Eng-Gene' is an effective 'Grey-Box' modelling approach that can be extended to modelling physics problems to accelerate the training of deep learning models and to enhance their generalization performance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Yukun Hu wishes to acknowledge the Transforming Foundation Industries Network+ (EPSRC grant EP/V026402/1) for partially funding this work. Xiaofei Han wishes to thank Chinese Scholarship Council for supporting his research. The

EngGeneNet presented in the paper is a continual development of the work funded by EPSRC grant GR/S85191/01.

References

Ahmed, Z., Lecompte, S., de Raad, T., De Paepe, M., 2019. Steady state model of a reheating furnace for determining slab boundary conditions. Energy Procedia 158, 5844–5849. doi:https://doi.org/10.1016/j.egypro.2019.01.542.

Cai, S., Wang, Z., Chryssostomidis, C., Karniadakis, G.E., 2020. Heat transfer prediction with unknown thermal boundary conditions using physics-informed neural networks, in: Fluids engineering division summer meeting, American Society of Mechanical Engineers. p. V003T05A054.

Crank, J., Nicolson, P., 1947. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type, in: Mathematical proceedings of the Cambridge philosophical society, Cambridge University Press. pp. 50–67.

Croft, D.R., Lilley, D.G., 1977. Heat transfer calculations using finite difference equations / David R. Croft and David G. Lilley. Applied Science Publishers, London.

Fonda, E., Pandey, A., Schumacher, J., Sreenivasan, K.R., 2019. Deep learning in turbulent convection networks. Proceedings of the National Academy of Sciences 116, 8667–8672.

Hu, Y., Tan, C., Broughton, J., Roach, P.A., 2016. Development of a first-principles hybrid model for large-scale reheating furnaces. Applied Energy 173, 555–566. doi:https://doi.org/10.1016/j.apenergy.2016.04.011.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Li, K., Peng, J., Irwin, G., Piroddi, L., Spinelli, W., 2005. Estimation of nox emissions in thermal power plants using eng-genes neural networks. IFAC Proceedings Volumes 38, 115–120. doi:https://doi.org/10.3182/20050703-6-CZ-1902.01748.

Liang, Y., Li, S., Yan, C., Li, M., Jiang, C., 202). Explaining the black-box model: A survey of local interpretation methods for deep neural networks. Neurocomputing 419, 168–182. doi:https://doi.org/10.1016/j.neucom.2020.08.011.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125.

Long, Z., Lu, Y., Dong, B., 2019. Pde-net 2.0: Learning pdes from data with a numeric-symbolic hybrid deep network. Journal of Computational Physics 399, 108925. doi:https://doi.org/10.1016/j.jcp.2019.108925.

Mao, Z., Jagtap, A.D., Karniadakis, G.E., 2020. Physics-informed neural networks for high-speed flows. Computer Methods in Applied Mechanics and Engineering 360, 112789. doi:https://doi.org/10.1016/j.cma.2019.112789.

McCarthy, M.F., 1975. Singular surfaces and waves, in: Continuum Mechanics of Single-Substance Bodies. Elsevier, pp. 449–521.

Milaszewicz, J.P., 1987. Improving jacobi and gauss-seidel iterations. Linear Algebra and Its Applications 93, 161–170.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer. pp. 234–241.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Wang, J., Li, G., Wei, L., Yi, Z., Wang, X., 2021. Investigation on thermal characteristics of walking reheating furnace for beam blank. International Communications in Heat and Mass Transfer 128, 105646. doi:https://doi.org/10.1016/j.icheatmasstransfer.2021.105646.

Wang, S., Zhang, H., Jiang, X., 2023. Physics-informed neural network algorithm for solving forward and inverse problems of variable-order space-fractional advection–diffusion equations. Neurocomputing 535, 64–82. doi:https://doi.org/10.1016/j.neucom.2023.03.032.

Xu, L., Yang, C., Xu, X., Chen, N., 2024. Physics-informed data-driven model of dehydration reaction stage in the sintering process of ternary cathode materials. Control Engineering Practice 147, 105929. doi:https://doi.org/10.1016/j.conengprac.2024.105929.

Yin, X., Sun, K., Li, S., Wang, X., Dong, Y., Cui, L., 2023. Enhancing deep learning for the comprehensive forecast model in flue gas desulfurization systems. Control Engineering Practice 138, 105587. doi:https://doi.org/10.1016/j.conengprac.2023.105587.

Yuan, X., Qi, S., Wang, Y., Xia, H., 2020. A dynamic cnn for nonlinear dynamic feature learning in soft sensor modeling of industrial process data. Control Engineering Practice 104, 104614. doi:https://doi.org/10.1016/j.conengprac.2020.104614.

Zhang, B., Wang, F., Qiu, L., 2023. Multi-domain physics-informed neural networks for solving transient heat conduction problems in multilayer materials. Journal of Applied Physics 133, 245103. doi:10.1063/5.0153705.

Zhao, X., Gong, Z., Zhang, Y., Yao, W., Chen, X., 2023. Physics-informed convolutional neural networks for temperature field prediction of heat source layout without labeled data. Engineering Applications of Artificial Intelligence 117, 105516. doi:https://doi.org/10.1016/j.engappai.2022.105516.