Acta Materialia 289 (2025) 120860

Contents lists available at ScienceDirect

Acta Materialia

journal homepage: www.elsevier.com/locate/actamat

Full length article

# Integration of multiscale simulations and machine learning for predicting dendritic microstructures in solidification of alloys

Sepideh Kavousi, Mohsen Asle Zaeem *

*Department of Mechanical Engineering, Colorado School of Mines, 1500 Illinois Street, Golden, CO 80401, USA*

ARTICLE INFO

*Keywords:*
Solidification
Microstructures
Machine learning
Molecular dynamics
Phase-field modeling

ABSTRACT

This study presents an integration of machine learning (ML) with a multiscale computational framework to predict primary dendrite arm spacing (PDAS) during alloy solidification. Analytical models, such as Hunt (HT) and Kurz-Fisher (KF), provide the basis for developing parametric and non-parametric ML models that capture the influence of processing conditions and material properties on PDAS. The training and testing dataset is generated from high-throughput phase-field simulations across various alloy systems, incorporating material properties calculated via molecular dynamics. While non-parametric models, such as decision trees, random forests, and gradient boosting decision trees, perform well in training, they encounter overfitting challenges due to the limited size of the computational dataset. In contrast, parametric models, including linear, ridge, and lasso regression, successfully capture key PDAS features, producing predictions that align closely with experimental data. Overall, parametric ML-based models show a stronger dependence on pulling velocity, temperature gradient, and material properties compared to the HT and KF models, offering a more accurate tool for predicting PDAS and optimizing alloy solidification processes.

## 1. Introduction

Many alloy manufacturing processes involve some form of solidification or melting [1,2]. These processes includes both traditional and advanced processing methods, such as casting [3], additive manufacturing [4], welding [5], and powder metallurgy [6]. At the nanoscale, the solid-liquid (SL) interfacial properties dictate the microscale morphology, pattern selection, and growth dynamics of asymmetric solid structures known as dendrites [7–9]. The anisotropic nature of SL interfacial properties makes growth along certain crystallographic directions energetically more favorable [10,11]. Previous theoretical [12,13], experimental [14,15], and computational [9,16–18] studies have shown that the spacing between dendrites decreases with increasing solidification velocity or temperature gradient. Thus, a homogeneous initial composition in the melt transforms into a complex distribution of solute ahead of the advancing dendrites and in the interdimeric regions. This results in the growth of important microstructural characteristics affecting the mechanical properties of the product [19–22] and solutioning heat treatment times [23–25].

Early mathematical models of dendritic growth are focused on predicting the morphology and stability of a single dendritic crystal growing in undercooled melt [26–31]. Hunt (HT) [12] and Kurz-Fisher (KF) [13] models were among the first studies to consider an array of dendrites when providing theoretical predictions of their morphology, spacing, and resulting thermodynamic microstate at their interface. These models relate primary dendritic arm spacing (PDAS) to a material-related term and processing conditions, namely pulling solidification velocity (V) and temperature gradient (G) at the dendrite tip. Hunt considered only the geometry of the cell tip and obtained the following relation:

$$PDAS_{HT} = 2.83 \times (\mathrm{k_e}\Gamma\Delta\,\mathrm{T_0}D_L)^{0.25} \times V^{-0.25} \times G^{-0.5}, \quad (1)$$

where $\mathrm{k_e}$ is the equilibrium partition coefficient, $\Gamma$ is the Gibbs-Thompson coefficient, $\Delta\,\mathrm{T_0}$ is the equilibrium freezing range, and $D_L$ is the liquid diffusion coefficient. KF model considered the geometry of the cell tip and trunk and obtained:

$$PDAS_{KF} = 4.3 \times (\Gamma\Delta\,\mathrm{T_0}D_L/\mathrm{k_e})^{0.25} \times V^{-0.25} \times G^{-0.5}. \quad (2)$$

While both models exhibit similar functional forms, they differ primarily in their material-related term and the pre-factor. There are other analytical models for PDAS in the literature [23,32–36], some of these, such as the one proposed by Trivedi [32], feature a slightly different

---

* Corresponding author.
*E-mail address:* zaeem@mines.edu (M. Asle Zaeem).

https://doi.org/10.1016/j.actamat.2025.120860
Received 14 November 2024; Received in revised form 18 February 2025; Accepted 21 February 2025

pre-factor coefficient. Some other models have more complicated forms. For example, Hunt-Lu [36] developed a model in which they used the velocity-dependent distribution coefficient, liquidus slope and diffusion coefficient. This model, as represented by Eq. (3), determines the complex correlation between $PDAS$ and $V$ or $G$ based on their non-dimensional form of $(PDAS)'$, $V'$ and $G'$ given by Eq. (4).

$$(PDAS)' = 0.7798 \times 10^{-1}\, V'^{(a-0.75)} (V' - G')^{0.75} G'^{-0.6028}, \tag{3}$$

$$(PDAS)' = \frac{PDAS \times \Delta T_0}{\Gamma k},\ \mathrm{G}' = \frac{\mathrm{G}\Gamma k}{(\Delta T_0)^2},\ V' = \frac{\mathrm{V}\Gamma k}{D_L \Delta T_0},\ \text{and}$$

$$a = -1.13 - 0.1[\log G'] - 0.0076[\log G']^2 \tag{4}$$

Uncovering how pulling velocity and temperature gradient play a role in PDAS calculations is more complicated for these complex models. There are also other models which incorporate scaling factors specifically chosen to fit to the experimental results [23,37,38]. There are specific some studies which support the accuracy of these models through validation by experimental results for limited material systems. However, discrepancies arise between these model predictions and experimental results when further comparisons are conducted for various material systems [23,37,39–47].

Physics-based models, such as phase-field (PF), which have also been used to study the dynamic growth of morphology and pattern formation during solidification, predict microstructural characteristics like PDAS differently from analytical models under various processing conditions [41,48–51]. Due to the intricate interplay between factors influencing dendritic spacing obtained from both experimental and computational results, optimizing mechanical properties through microstructural control is challenging. This complexity requires adjusting one input at a time, making the process both expensive and time consuming. Machine learning (ML) have recently been used to accelerate the material design [52]. ML have been used in predicting the mechanical properties of the alloys [53] and developing new alloy systems [54]. Different studies have applied ML-based regression methods to correlate the processing conditions to the mechanical properties of the final product [19–22].

In this work, we present a comprehensive ML-integrated multiscale framework to predict PDAS across a wide range of processing conditions and alloy compositions. PF simulations of solidification for Al-Cu, Mg-Al, and Ti-Ni alloy systems, with corresponding face-centered cubic (FCC), hexagonal close-packed (HCP), and body-centered cubic (BCC) crystal structures, provide the necessary data for training an ML model. This model uncovers the effects of chemical concentration and solidification conditions on the PDAS. Key properties required for PF simulations, such as material properties ($\mathrm{k_e}$, $\Gamma$, $\Delta\,\mathrm{T}_0$, $D_L$) and other SL interfacial properties (SL interface free energy, kinetic coefficient, diffusive interface velocity) are calculated by atomistic scale molecular dynamics (MD) simulations. Subsequently, the results from high throughput PF simulations of solidification are used to train different ML methods capable of providing an analytical expression for the PDAS. Finally, we validate our proposed ML-guided PDAS models which are trained by computational data against experimental data for various alloys and processing conditions.

## 2. Methodology

The workflow depicted in Fig. 1 outlines our ML-integrated multiscale computational framework to identify the correlation between PDAS, processing conditions and material properties. The methodology comprises four key sub-sections. The proposed ML model is trained and tested only by PDAS obtained from MD-guided PF simulations of solidification. The accuracy and reliability of the proposed ML models will be tested using a separate set of experimental data. Sub-Section 2.1 summarizes the details of the PF modeling followed by parameterizing of the PF model based on material properties in Sub-Section 2.2. Details of SL interfacial property calculations with MD are covered in Sub-Section 2.3. Finally, ML regression models that we used to find the correlation between PDAS, processing conditions and material properties are presented in Sub-Section 2.4.

### 2.1. Phase-field modeling

A quantitative PF model [55] is used to simulate directional solidification. In this model, the PF order parameter $\phi$ takes the value $+1$ in the solid and $-1$ in the liquid, and it varies smoothly across the SL interface. The time evolution equation of $\phi$ is given by:

$$\tau(n)\frac{\partial\phi}{\partial t} = \left[\nabla.\left[W(n)^2\,\nabla\phi\right] + \sum_{i=1}^{2}\partial_i\left[W(n)\frac{\partial W(n)}{\partial(\partial_i\phi)}\,|\nabla\phi|^2\right]\right] + f'(\phi) - \frac{\lambda}{1-\mathrm{k_e}}g'(\phi)\left(\exp(u) - 1 + \frac{T - T_0}{\frac{\mathrm{m_L}^{\mathrm{e}}\mathrm{c}_0^{\mathrm{L}}}{\mathrm{k_e}}}\right). \tag{5}$$

$W(n)$ and $\tau(n)$ are the anisotropic interface width and PF characteristic time, respectively, which both depend on the orientation of the normal to the SL interface $n$. $\partial_i$ is the partial derivative in the x ($i = 1$) and y ($i = 2$) directions. $f(\phi)$ and $g(\phi)$ are given by:

$$f(\phi) = \phi^4/4 - \phi^2/2, \tag{6}$$

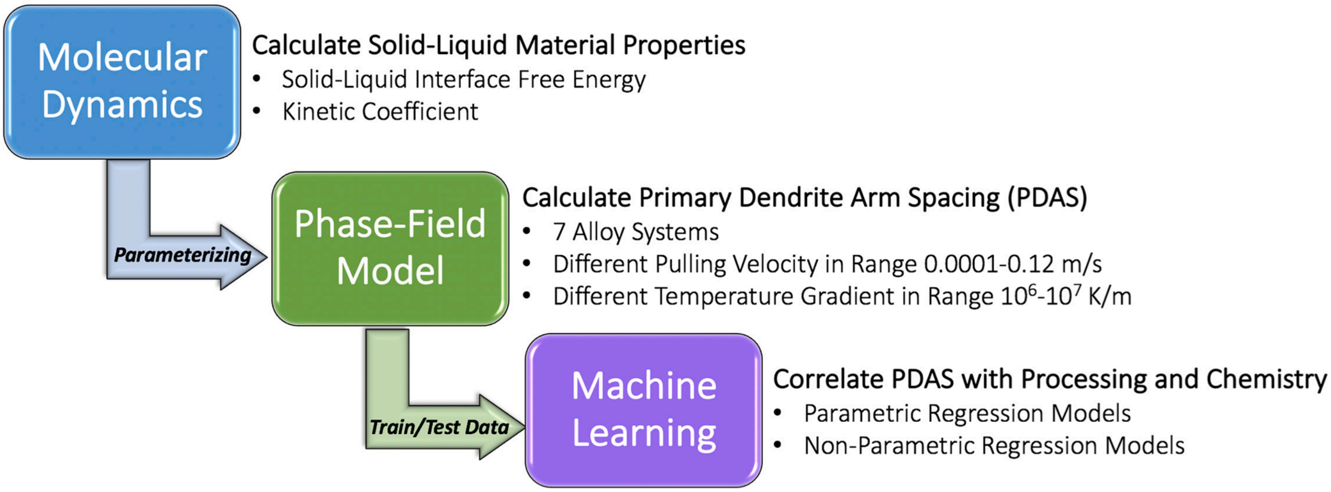


**Fig. 1.** The workflow for ML modeling and predicting PDAS based on the training and testing dataset generated by MD-integrated PF simulations of solidification.

$$g(\phi) = \left(1-\phi^2\right)^2. \tag{7}$$

Directional solidification is a reasonable approximation in modeling temperature field evolution in additive manufacturing [55–58]. Thus, instead of solving the heat diffusion equation, the temperature field, as shown in Eq. (5), is assumed to translate linearly along the growth direction and is described by $T(y,t) = \mathrm{T}_0 + G(y - Vt)$, where $\mathrm{T}_0$ is a reference temperature. The dimensionless chemical potential $u$ is defined as:

$$u = \ln\left(\frac{2c\mathrm{k_e}/\mathrm{c}_0^{\mathrm{L}}}{1+\mathrm{k_e}-\phi\times(1-\mathrm{k_e})}\right). \tag{8}$$

where $c$ is the alloy composition, $\mathrm{c}_0^{\mathrm{L}}$ is the equilibrium liquid composition, and $\mathrm{m_L}^{\mathrm{e}}$ is the liquidus line slope, which can be obtained from the phase diagram. The parameter $\lambda$ is a coupling constant representing the correlation between the interface width and the capillary length, discussed further in Section 2.2. To enhance computational efficiency and numerical stability, we will replace the concentration evolution equation with Eqs. (9)-(11) where $U$ is dimensionless supersaturation. The antitrapping flux $j_{at}$ is incorporated in Eq. (9) to account for the physical effects of solute-trapping in the SL interface.

$$\left(\frac{1+\mathrm{k_e}}{2}-\frac{1-\mathrm{k_e}}{2}\phi\right)\frac{\partial U}{\partial t} = \nabla\cdot\left(D_L\frac{1-\phi}{2}\nabla U + j_{at}\right) + [1+(1-\mathrm{k_e})U]\frac{1}{2}\frac{\partial\phi}{\partial t}. \tag{9}$$

$$j_{at} = \frac{1}{2\sqrt{2}}W(n)[1+(1-\mathrm{k_e})U]\frac{\partial\phi}{\partial t}\frac{\nabla\phi}{|\nabla\phi|}. \tag{10}$$

$$U = \frac{e^u-1}{1-\mathrm{k_e}}. \tag{11}$$

The PF simulations are performed for Ti-Ni, Mg-Al, and Al-Cu alloys with corresponding body-centered-cubic (BCC), hexagonal-closed-pack (HCP) and face-centered-cubic (FCC) crystal structures, respectively. We have adjusted the temperature gradient G and the puling velocity V to cover a wide range of solidification conditions. Table 1 summarizes the details of the PF simulations including the alloy composition, V, G, and the total number of simulations for each alloy system. Each simulation starts with a thin layer of solid phase at the bottom of the domain, and during the directional solidification, the solid phase grows as the temperature gradient G moves with the puling velocity V inside the liquid. Our prior knowledge [57,58] on the direct correlation between the thermal diffusion length $l_T$ (Eq.(12)) and the transient time to reach steady-state regime, leads us to choose a system size in the y-direction to be at least $20 \times l_T$. The simulation system size along the x-direction is chosen such that at least 8 dendrites grow in the steady-state regime.

$$l_T = |m_L{}^e|(1-\mathrm{k_e})c_0^L/(\mathrm{k_e}G). \tag{12}$$

The system of governing equations for the evolution of $\phi$ and $U$ are solved numerically in the two-dimensional domain using a finite-difference algorithm, and the PETSC package is used as the solver [59]. The governing equations are discretized with second order approximation. The time stepping is performed using adaptive time step based on the backward differentiation formula method, which is suitable for stiff problems due to its implicit nature [60]. Using adaptive time stepping enhances stability by adjusting the time step dynamically based on error estimates. We used generalized minimal residual algorithm [61] as the linear solver and bypassed the nonlinear solver and only used linear iterations for solving the nonlinear equations. Setting block Jacobi preconditioner, which is effective for parallel computing, transforms the original linear system into a form that is easier for iterative solvers to handle. Thus, it improves the convergence rates and reduces the number of iterations required to reach the solution. The results from PF simulations are used as training and testing datasets for the ML modeling, as discussed in Section 2.4 and presented in Fig. 2.

### 2.2. *Parameterizing the phase-field model*

The initial PF studies produced solidification structures that qualitatively resemble those observed in experiments [62,63]. Karma et al. advanced the PF models towards quantitative predictions by using thin interface asymptotic analysis [55,56]. This analysis maps the PF equations to the sharp interface equations, allowing for the parameterization of the PF model based on material properties to represent the physical system. Specifically, the thin interface analysis relates the PF parameters $\tau$, $W$, and $\lambda$ to SL material properties, namely, SL interface free energy $\gamma$, and kinetic coefficient $\mu$, as shown in Eqs. (13), and (14) [55,56]:

$$d_0(\vec{n}) = \frac{\gamma(\vec{n})\,T_m}{L\Delta T_0} = a_1\frac{W(\vec{n})}{\lambda} \tag{13}$$

$$\mu(\vec{n}) = a_1\frac{\tau_0(\vec{n})}{\lambda\,W(\vec{n})} - a_1a_2\frac{W(\vec{n})}{D_L}[1+(1-k)U_0] \tag{14}$$

$\mathrm{a}_1$ and $\mathrm{a}_2$ are equal to 0.8839 and 0.6867, respectively. $d_0$ is the capillary length, the characteristics length over which surface tension effects become significant compared to gravitational forces. In this study, we determine the PF model parameters using the material properties calculated by MD simulations. Researchers interested in PF parameterization based on material properties can refer to our previous works [64–66].

### 2.3. *Molecular dynamics simulations*

MD simulation provides a unique advantage by capturing the movements and dynamic interactions of atoms that are often challenging to observe experimentally. This capability allows for a deeper understanding of transformation mechanisms [67,68] and the calculation of material properties that are otherwise difficult, if not impossible, to obtain experimentally. In this section, we provide calculation details of the SL interfacial material properties that are essential in providing valuable insights into the fundamental mechanisms of alloy solidification.

#### 2.3.1. *Interatomic potential*

The interactions between particles in MD simulations are described

**Table 1**
Details of PF simulations, including the alloy systems, concentrations, puling velocity (V), temperature gradient (G), and the total number of PF simulations for each system.

| Alloy | Composition | V (m/s) | G (K/m) | Number of PF simulations |
|---|---|---|---|---|
| Ti-x at% Ni | $x = 3.4$ | 0.01–0.09 | $5 \times 10^6$–$10^7$ | 45 |
| | $x = 7.1$ | 0.01–0.09 | $5 \times 10^6$–$10^7$ | 42 |
| | $x = 10.4$ | 0.0001–0.02 | $5 \times 10^6$–$10^7$ | 22 |
| Mg-x at% Al | $x = 9$ | 0.005–0.12 | $1 \times 10^6$–$5 \times 10^6$ | 21 |
| Al-x at% Cu | $x = 6$ | 0.001–0.06 | $5 \times 10^6$–$10^7$ | 47 |
| | $x = 8.2$ | 0.01–0.08 | $5 \times 10^6$–$10^7$ | 43 |
| | $x = 10.6$ | 0.008–0.005 | $6 \times 10^6$–$10^7$ | 30 |

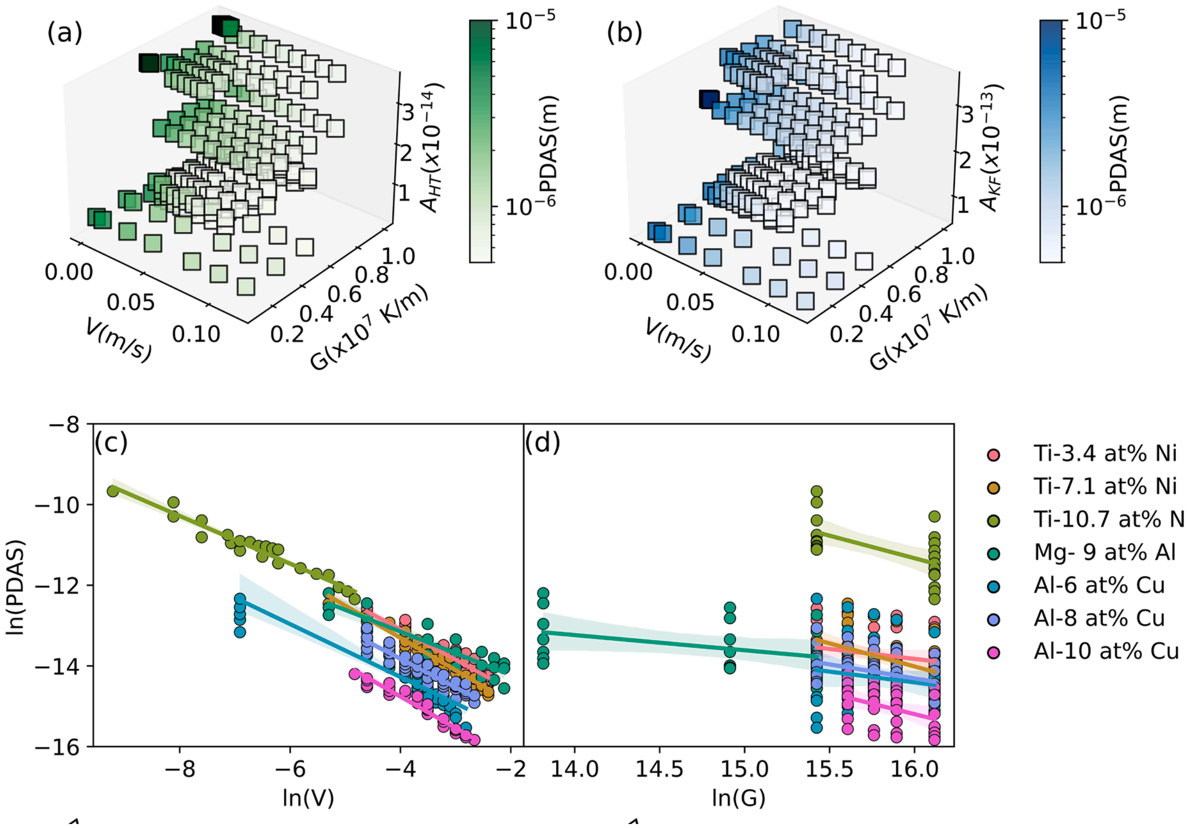


**Fig. 2.** 3D scatterplot of the PF-calculated PDAS distributed across various $V$ and $G$ for (a) HT-based and (b) KF-based models. The linear distribution of $\ln(PDAS)$ versus (a) $\ln(V)$, and (b) $\ln(G)$.

by interatomic potentials. Among the various interatomic potential models, the embedded atomic method (EAM) [69] and modified-EAM (MEAM) [70] are particularly well-suited for accurately representing the atomic interactions in metallic systems. Recent advancements in this field have focused on developing interatomic potentials that can precisely model atomic interactions in both the solid and liquid phases. These potentials are designed to capture the essential thermodynamic and physical properties and are rigorously validated against experimental data, such as phase diagram, enthalpy of fusion, density measurements. The interatomic potentials used in this study can accurately predict the high-temperature solid and liquid properties for Ti-Ni [66], Al-Cu [71] and Mg-Al [72] systems.

### 2.3.2. Phase diagram

Performing PF simulations of solidification of alloys require precise knowledge of their freezing temperature range, equilibrium solidus and liquidus concentration, and the resulted partition coefficient. Moreover, MD simulations for determining the SL interface energy and kinetic coefficient require prior knowledge on the capability of interatomic potentials in predicting the equilibrium thermodynamic state of materials. The hybrid MD-Monte Carlo (MC) semi-grand canonical (SGC) ensemble method [73] is often used to determine the free energy and chemical potential differences between atomic species in an alloy system, which are crucial for calculating the phase diagram. We used this method to calculate the Ni-Ti phase diagram. Coexistence method is the alternative approach where different atomic species are allowed to swap across the simulation box while the whole simulation system is being equilibrated at the target thermodynamic state. Using this method, we calculated the Al-Cu and Mg-Al phase diagrams. Researchers interested in computational details of both methods can refer to our previous works [64,74].

### 2.3.3. SL interface free energy

To calculate SL interface free energy, we used the Capillary fluctuation method (CFM) [75]. In this method, a solid-liquid (SL) interface is generated within the MD simulation box using composition-temperature values obtained from the phase diagrams (Section 2.3.2), and this interface is allowed to fluctuate under thermodynamic equilibrium conditions. We analyze these fluctuations to extract the capillary wave spectrum, which provides insights into the interfacial stiffness. By fitting the spectrum to theoretical models, the solid-liquid interface energy can be determined.

For a crystal with cubic symmetry, the SL interface free energy is an anisotropic property, depending on the interface orientation, given by:

$$\gamma = \gamma_0\left[1+\delta_1\left(\sum_{i=1}^{3} n_i^4-\frac{3}{5}\right)+\delta_2\left(3\sum_{i=1}^{3} n_i^4+66n_1^2n_2^2n_3^2-17/7\right)\right] \tag{15}$$

where $\gamma_0$ is the average interfacial free energy, $\Delta_1$ and $\Delta_2$ are the anisotropy parameters, and $n_i$ are the components of the unit vector, $\widehat{n}$, normal to the interface plane [76]. We perform MD simulations for different SL interface orientations and fit the MD-calculated stiffness to the equation derived from Eq. (15) to determine the average SL interface energy and corresponding anisotropy parameters. Researchers interested in computational details of the CFM can refer to our previous works on Al-Cu [77] and Ti-Ni [66,78].

### 2.3.4. Kinetic coefficient

The solidification kinetic coefficient characterizes the rate at which a liquid transforms into a solid under different thermodynamic conditions [79]. To determine this coefficient, a MD simulation of the alloy undergoing solidification is performed. The solidification driving force is considered as any condition that results in deviation of the alloy from the thermodynamic equilibrium condition predicted by the phase diagram. The solidification driving force for solidification of pure metals is undercooling below the melting point. By analyzing the relationship between the SL interface velocity and undercooling, the solidification kinetic coefficient can be extracted. There are well-developed MD methods for calculating the kinetic coefficient for pure metals [65,79]. For alloys, either undercooling below the liquidous temperature or decreasing the alloy composition below the liquidous composition will initiate solidification. However, due to the lack of an established method for calculating kinetic coefficient for alloys and complexity of MD post-processing steps, the original PF models considered this property to be zero [55] . However, recent studies of directional solidification at a relatively higher velocity range emphasized the importance of kinetic coefficient on microstructural evolution during the solidification [64, 80]. Kavousi et al. developed an algorithm based on hybrid MD and MC simulations to investigate the nonequilibrium interface kinetics during the solidification to calculate the kinetic coefficient; computational details of this algorithm are provided in [74]. For a crystal with cubic symmetry, the kinetic coefficient $\mu$ is an anisotropic property, depending on the interface orientation, given by:

$$\mu = \mu_0\left(1 + 3\varepsilon_1 - 4\varepsilon_1\sum_{i=1}^{3} n_i^4\right) \tag{16}$$

where $\mu_0$ is the average kinetic coefficient, and $\varepsilon_1$ is the anisotropy parameter [80]. We will perform MD simulations for different SL interface orientations and fit the corresponding kinetic coefficient to the equation derived from Eq. (16) to determine the average SL interface energy and corresponding anisotropy parameters.

### *2.4. Machine learning regression modeling*

We use two different approaches, known as parametric and nonparametric regression models, to perform data analytics. The parametric regression models follow a predetermined functional form. Either linear [81] or non-linear [82,83] functions can be used to describe the relationship between the response (dependent) and explanatory (independent) variables. Linear regression, polynomial regression, and non-linear regression models are some examples of this category. On the contrary, the predictors in the nonparametric models do not take predetermined forms. These models are more flexible and may better capture the unexpected or unusual features of the data [84]. However, they need a larger sample size than the regression parametric models because the data must both determine the model structure and estimate its parameters.

#### *2.4.1. Parametric regression models*

Parametric linear regression models construct the best-fit regression line through data points such that the overall distance between the data points and the line itself (also referred as error or residual) is minimized. Although there are various types of regression models, in this study we focus on Linear (LR), Lasso Regression (LaR), and Ridge Regression (RR) models. The LR model simply fits the correlation between the features $x_i$ ($i = 1..m$) and target variable ($y$) into a line [81]:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_mx_m + b \tag{17}$$

The parameters $a_i$ ($i = 1..m$) and $b$ are selected to minimize the loss function $Loss_{LR}$, also known as residual sum of squares (RSS). $Y$ is the observed values, $\widehat{Y}$ is the model predictions. The loss function is defined by:

$$Loss_{LR} = \sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \tag{18}$$

$n$ is the sample size. LR model does not have any limitations on the magnitudes of $a_i$ and $b$ coefficients. Thus, large coefficients can lead to a model that overfits the training dataset. To overcome this shortcoming, we will investigate the regularized LaR and RR models, which introduce penalties on the magnitude of regression coefficients.

RR and LaR models enhance the accuracy and promote making informed decisions in data The L1 regularization employed in LaR adds a penalty of absolute values of the coefficients to the loss function $Loss_{LaR}$ [85,86].

$$Loss_{LaR} = \sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 + \omega\left(\sum_{i=1}^{m}|a_i| + b\right) \tag{19}$$

The L2 regularization employed in RR, adds a penalty of squared of magnitude of coefficients to the loss function [87,88]

$$Loss_{RR} = \sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 + \omega\left(\sum_{i=1}^{m} a_i^2 + b^2\right) \tag{20}$$

A larger value for $\omega$ adds a more aggressive penalization to the LaR and RR models and setting $\omega$ value of zero produces results similar to the LR model. The optimal value for parameter $\alpha$ is obtained through hyperparameter tuning based of minimizing the mean squared error (MSE). We evaluate the accuracy of each method using two metrics: R$^2$ value and MSE, given by [81]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \tag{21}$$

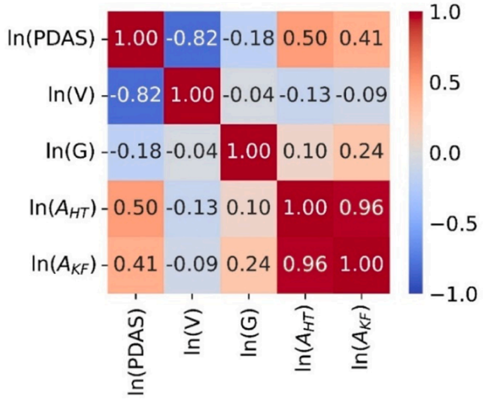
$$R^2 = 1 - \frac{RSS}{TSS} \tag{22}$$

$$TSS = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 \tag{23}$$

**Table 2**

The MD-calculated material properties, including equilibrium partition coefficient $k_e$, equilibrium freezing range $\Delta T_0$ (K), Gibbs–Thomson coefficient $\Gamma(\times 10^{-7}\mathrm{Km})$, Liquid diffusion coefficient $D_L$ $(\times 10^{-9}\ \mathrm{m}^2/\mathrm{s})$, capillary length $d_0$(nm), solid-liquid interfacial free energy $\gamma_0(\mathrm{J}/\mathrm{m}^2)$ and its corresponding anisotropy $\delta_1$ and $\delta_2$.

| Alloy | $k_e$ | $\Delta T_0$ | $\Gamma$ | $D_L$ | $d_0$ | $\gamma_0$ | $\delta_1$ | $\delta_2$ |
|---|---|---|---|---|---|---|---|---|
| Ti-3.4 at% Ni | 0.28 | 63.1 | 2.91 | 3.92 | 4.61 | 0.185 | 0.007 | −0.003 |
| Ti-7.1 at% Ni | 0.31 | 126 | 2.75 | 3.12 | 2.18 | 0.175 | 0.01 | −0.002 |
| Ti-10.4 at% Ni | 0.36 | 177 | 2.69 | 2.05 | 1.52 | 0.171 | 0.025 | −0.002 |
| Mg-9 at% Al | 0.315 | 36.4 | 2.0 | 2.56 | 5.48 | 0.153 | 0.02 | 0.001 |
| Al-6 at% Cu | 0.19 | 31.9 | 0.95 | 4.00 | 2.98 | 0.096 | 0.014 | −0.004 |
| Al-8.2 at% Cu | 0.236 | 41.1 | 0.91 | 4.90 | 2.22 | 0.092 | 0.008 | −0.004 |
| Al-10 at% Cu | 0.246 | 47.9 | 0.89 | 4.88 | 1.86 | 0.090 | 0.01 | −0.008 |



| | ln(PDAS) | ln(V) | ln(G) | ln($A_{HT}$) | ln($A_{KF}$) |
|---|---|---|---|---|---|
| ln(PDAS) | 1.00 | -0.82 | -0.18 | 0.50 | 0.41 |
| ln(V) | -0.82 | 1.00 | -0.04 | -0.13 | -0.09 |
| ln(G) | -0.18 | -0.04 | 1.00 | 0.10 | 0.24 |
| ln($A_{HT}$) | 0.50 | -0.13 | 0.10 | 1.00 | 0.96 |
| ln($A_{KF}$) | 0.41 | -0.09 | 0.24 | 0.96 | 1.00 |

**Fig. 3.** Heatmap of the Pearson correlations between PF-calculated PDAS, the processing conditions and material properties. Off-diagonal non-zero elements imply statistical dependency in the data. Colors indicate the strength of the correlation, where dark red and dark blue indicate perfect correlation and anti-correlation, respectively.

Total sum of squares (TSS) measures the total variance in the response Y by comparing it to the average value $\overline{Y}$. TSS determines the variability in the response, and a lower MSE and a higher $R^2$ results in a better model predictive accuracy and a better regression model.

### 2.4.2. Nonparametric regression models

Nonparametric regression models are powerful techniques for modeling complex relationships in data without assuming a specific parametric form. Nonparametric regression analysis relaxes the assumption of linearity, substituting the much weaker assumption of a smooth population regression function [82]. They are flexible and can capture complex, non-linear patterns in data, making them highly effective in many ML tasks. This drastically increases the computational time, but meanwhile, makes the model predictions more accurate. The fact that nonparametric regression models do not follow a specific function form to correlate features to the dependent variables, also makes them more sensitive to the outliers than the parametric regression models [82,83]. In this study we will use decision tree (DT) [89], random forest (RF) [90]and gradient boosting decision tree (GBDT) [91] as the three nonparametric models. We focus on these models based on an initial evaluation of different models' performance on our dataset. The initial evaluation included other models, such as K-Nearest Neighbor and AdaBoost regression models, which did not pass the initial evaluation based on MSE and $R^2$. It should be noted that the initial evaluation step did not include either the hyperparameter tuning nor k-fold evaluation of the models. The refinement of models in this step was just made based on the default parameter settings and only one set of train-test dataset assignments.

DT regression analyzes the features of an object and trains a tree-structured model to make predictions, yielding meaningful continuous output [92,93]. This method is particularly effective for capturing non-linear patterns in data. The tree is built by a sequence of step-by-step questions, with each answer directing the model along a particular branch. The tree grows during the learning process depending on the complexity of the fed-in data [84]. This model operates based on "if-then" rules, ultimately resulting in a specific prediction [92,93]. However, individual DTs can be prone to overfitting [94]. To address this, random forest regression builds an ensemble of multiple DTs, each trained on a random subset of the data and features and averaging their predictions to improve accuracy and reduce overfitting [95]. GBDT is another ensemble method, which builds sequential trees with each new tree correcting the errors of the previous ones by focusing on the residuals. This results in a more accurate model by iteratively refining predictions. We utilize the scikit-learn library [96] for ML tasks, and employ Seaborn [97] and Matplotlib [98] for data visualization.

## 3. Results and discussions

Table 2 provides comprehensive summary of the material properties derived from MD simulations, which are essential to accurately model the solidification process and quantitatively predict dendritic growth characteristic. The dataset includes the following parameters: $k_e$, $\Delta\ T_0$, $\Gamma$, $D_l$, $d_0$, $\gamma_0$, $\Delta_1$, and $\Delta_2$.

Widely adopted approaches in the literature [39,40,45,50,51] employ analytical models to elucidate the relationship between the PDAS and the processing-material properties. In this study, we use models in the following power-law form:

$$PDAS = \mathrm{C} \times (A)^{\gamma} \times V^{-\alpha} \times G^{-\beta} \tag{24}$$

Here, C, $\gamma$, $\alpha$, and $\beta$ are the unknown parameters to be determined through data analysis. The dimensionless parameter A encapsulates material-specific properties and exhibit distinct definitions for the HT ($A_{HT}$) and KF ($A_{KF}$) models. Thus, the proposed parametric models in this study will be classified as HT-based and KF-based models.

$$A_{HT} = \mathrm{k_e}\Gamma\Delta\ \mathrm{T}_0 D_L \tag{25}$$

$$A_{KF} = \Gamma\Delta\ \mathrm{T}_0 D_L / \mathrm{k_e} \tag{26}$$

Fig. 2 presents a 3D scatterplot of the PF-calculated PDAS distributed across various velocities and temperature gradients for the investigated material systems. Fig. 2(a) and (b) illustrate the dependence of PDAS on $V$ and $G$ for $A_{HT}$ and $A_{KF}$, respectively, with the color bar indicating the arm spacing values. These results present a direct relationship between PDAS and the parameter $A$, and an inverse relationship with $V$ and $G$. To linearize the data and simplify the fitting process, a logarithmic transformation is applied to both sides of the Eq. (24), thereby converting the power-law model into:

$$\ln(PDAS) = \ln(\mathrm{C}) + \gamma \times \ln(A) - \alpha \times \ln(V) - \beta \times \ln(G) \tag{27}$$

Fig. 2(c) and (d) reveal an almost linear correlation between $\ln(PDAS)$ with both $\ln(V)$ and $\ln(G)$ across seven investigated alloy systems, validating the effectiveness of the power-law model in describing the interaction between material properties, processing parameters and PDAS. Therefore, Eqs. (25)-(27) will be used to determine the unknown parameters for both HT-based and KF-based ML models. Additionally, Fig. 2(c) and (d) provide insights into the impact of processing parameters on PDAS. The deviation from linear behavior is greater for $\ln(PDAS) - \ln(G)$ than $\ln(PDAS) - \ln(V)$, suggesting that $V$ have a more pronounced effect on PDAS compared to $G$, and $\alpha$ is expected to be a larger value than $\beta$. This is also supported by the heatmap represented by Fig. 3. This figure is a graphical representation of the Pearson correlation coefficient, a statistical measure quantifying the strength and direction of linear relationships between $\ln(PDAS)$ and $\ln(V)$, $\ln(G)$, $\ln(A_{HT})$, and $\ln(A_{KF})$. Heatmaps use color gradients to represent the variations in the data intensity, making patterns and trends easy to spot. A positive Pearson coefficient in Fig. 3 relating $\ln(PDAS)$ to either $\ln(A_{HT})$ or $\ln(A_{KF})$ represent a positive linear impact of material properties on arm spacing. On the other hand, there is a negative linear correlation between $\ln(PDAS)$ and processing conditions, namely $\ln(V)$ and lnG. Despite their advantages on facilitating trend comparisons within a dataset, heatmaps are limited to two-dimentional representations, making it challenging to effectively analyze multi-dimentional relationships. Comparing heatmap predictions with those from parametric ML models (Section 3.1) will help understanding their shortcommings.

### 3.1. Parametric ML models

In this section, we use parametric ML models based on the analytical HT and KF models to determine the correlation between PDAS, pro-

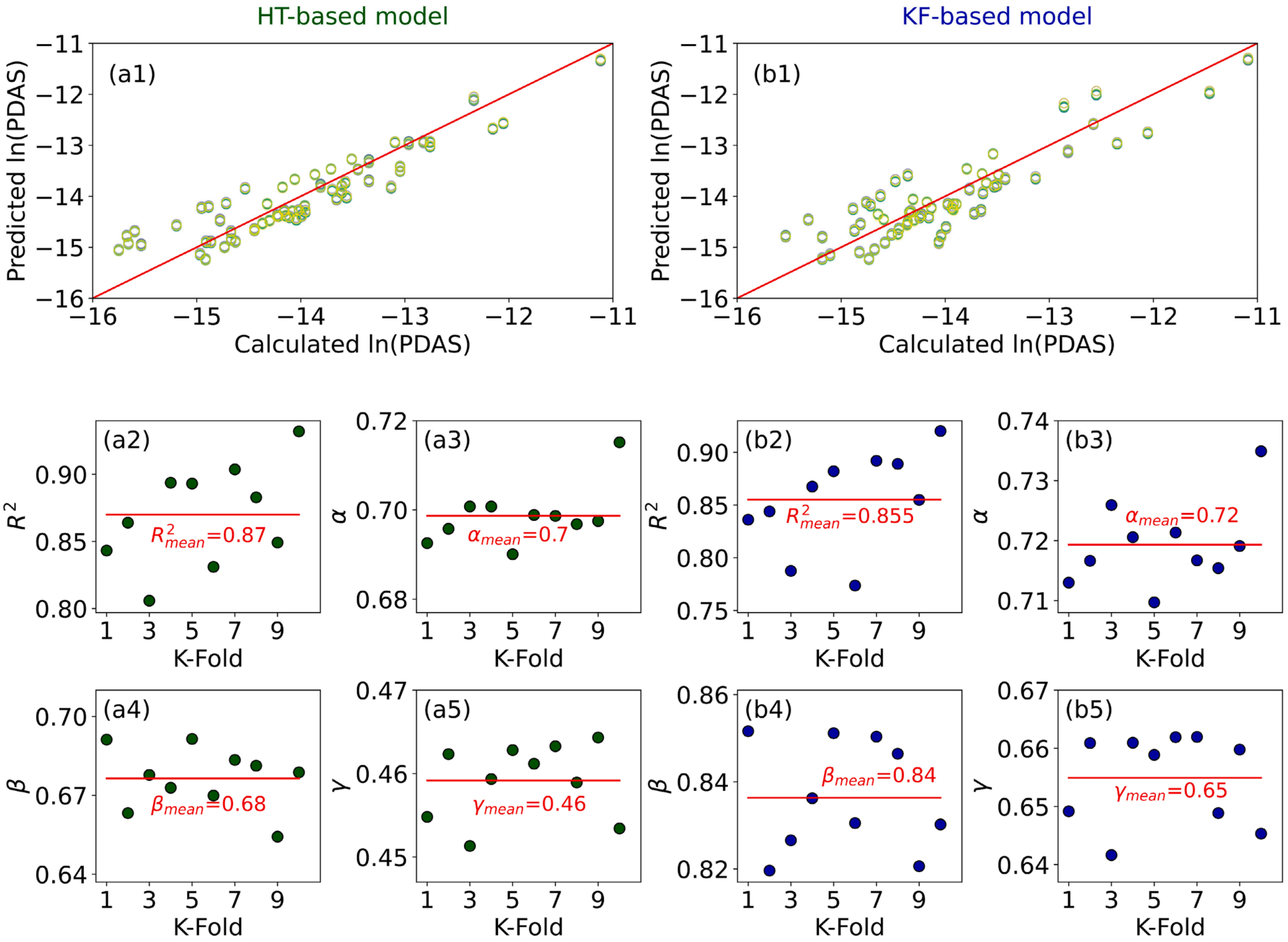


**Fig. 4.** Comparison of the variations of: (a1) and (b1) the ML predicted versus calculated PDAS, (a2) and (b2) $R^2$, (a3) and (b3) $\alpha$, (a4) and (b4) $\beta$, and (a5) and (b5) $\gamma$ for (a) HT-based and (b) KF based models across each fold in cross validation. The average values of $R^2$, $\alpha$, $\beta$, and $\gamma$ are represented on the corresponding figures.

cessing conditions and material properties. This approach allows us to quantify how variations in $V$, $G$, and $A_{HT}$ or $A_{KF}$ affect the PDAS. To ensure robustness and generalizability of the model, a 10-fold cross-validation process was employed during training. This technique involved dividing the data into 10 subsets, iteratively training on 9 subsets and validating on the remaining 1. Each iteration results in a different set of model parameters (C, $\gamma$, $\alpha$, and $\beta$), and $R^2$ value, capturing variations in the ML-predicted values for both the HT-based and KF-based models. This process helps to minimize overfitting and provides a more reliable estimate of the model's predictive accuracy on unseen data [99,100]. Fig. 4 highlights these slight variations among the k-fold models, reflecting the differences in the correlation between ln(PDAS) and the input parameters for each model. These variations are also reflected in $R^2$ values and the predicted model parameters given by Eqs. (24)-(27), and the proposed process helps to minimize overfitting and improve the model's predictive accuracy on unseen data.

Fig. 5 illustrates the comparison between the PF-calculated PDAS versus the predictions of original analytical models, and the parametric and nonparametric ML models. As shown in this figure, the original analytical models, given by Eqs. (1) and (2), cannot accurately predict the PF simulation results. The $R^2$ for HT and KF analytical models, as given in Table 3, are 0.549 and −0.68, respectively. The poor performance of the analytical model, especially the KF model, highlights the need for obtaining a new model for predicting PDAS. We began investigating the parametric models focusing on HT-based and KF-based LR models. The performance of the models is evaluated based on $R^2$ and MSE values. For the HT-based LR model, with the model parameters given in Table 4, with an $R^2$ of 0.89 and MSE of 0.137 for the training data, and an $R^2$ of 0.87 and MSE of 0.141 for the testing data, demonstrating strong predictive accuracy. The KF-based LR model is represented by model parameters given in Table 4, achieving $R^2$ of 0.879 and MSE of 0.151 for the training data, and $R^2$ of 0.866 and MSE of 0.156 for the testing data. Overall, the results indicate that the HT-based model is more reliable than the KF-based model. Additionally, the comparison of LR model parameters with the original analytical models reveals that the PF simulation results exhibit a stronger linear correlation between PDAS and key factors such as pulling velocity, temperature gradient, and the material parameter A. A similar statistical analysis is performed for both LaR and RR models. Tuning the hyperparameter $\omega$ is crucial in finding the optimal balance between model complexity and prediction accuracy, preventing overfitting in both LaR and RR models. One common method for tuning $\omega$ is through grid search [101,102], where a predefined range of $\omega$ values is systematically explored to identify the one that minimizes the error. We coupled hyperparameter tuning with k-fold cross-validation to ensure robustness and mitigate the risk of overfitting. The best grid search outcome for HT-based and KF-based RR models are 1.86 and 1.83, respectively. Similarly, the optimum values for LaR models are 0.01 for both the HT-based and KF-based models. We used these optimized $\omega$ values were used for further analysis to enhance model performance.

Table 4 provides a summary of the comparison between parameters in the HT-based and KF-based linear regression models, as described by Eqs. (24)-(26), along with their corresponding performance metrics. The standard deviation for the model parameters is calculated based on their

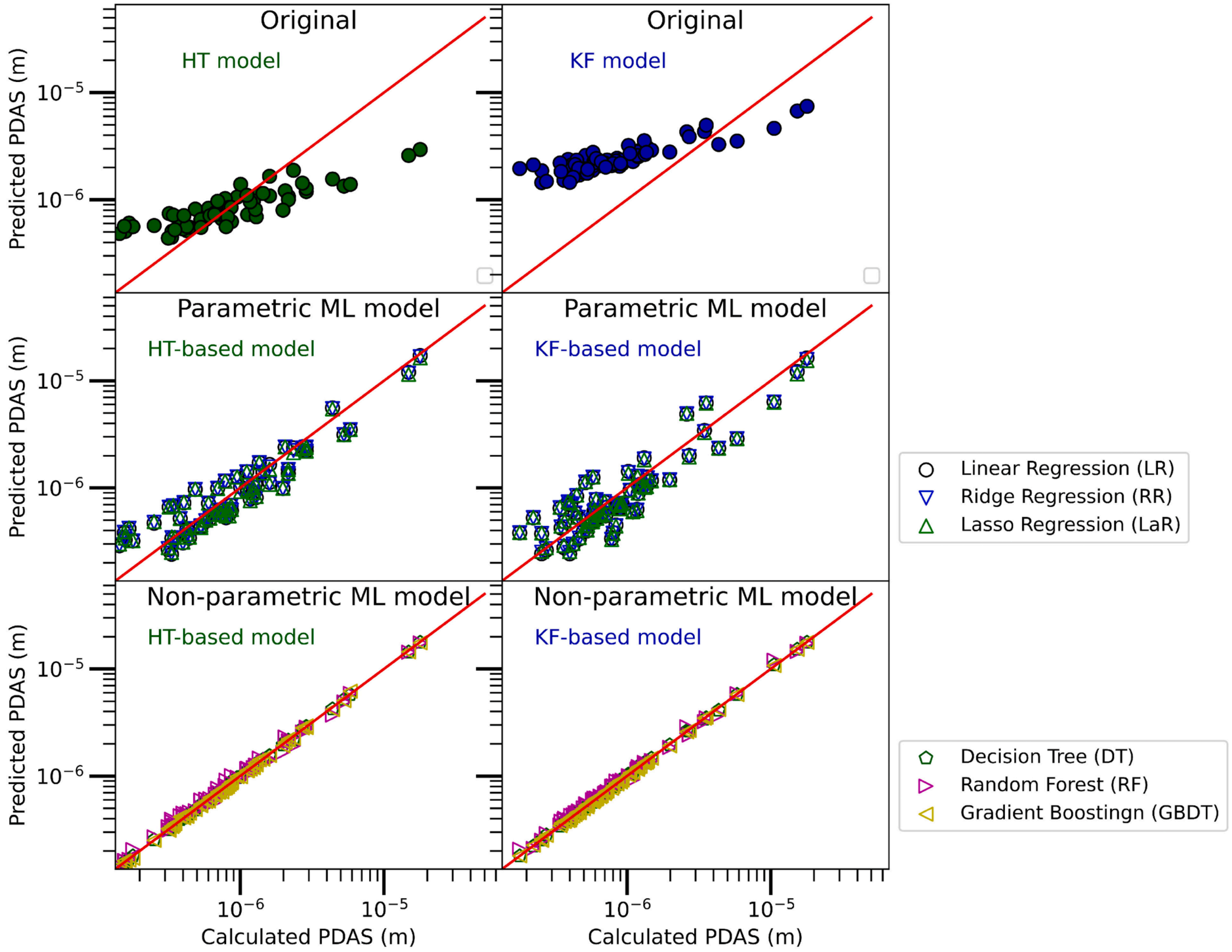


**Fig. 5.** The comparison of PF-calculated PDAS with the predictions based on original analytical, parametric (LR, LaR, and RR), and nonparametric (DT, RF, and XGDT) models for HT-based and KF-based ML models.

**Table 3**
The performance metrics ($R^2$ and MSE) for HT-based and KF-based parametric (LR, RR, and LaR) and nonparametric (DT, RR, and GBDT) models.

| Model | Detail | HT-based | | KF-based | |
|---|---|---|---|---|---|
| | | $R^2$ | MSE | $R^2$ | MSE |
| Analytical | Original model | 0.549 | 0.452 | −0.68 | 1.514 |
| Parametric ML modeles | Linear Regression (LR) | 0.87 | 0.141 | 0.855 | 0.156 |
| | Ridge Regression (RR) | 0.87 | 0.141 | 0.855 | 0.156 |
| | Lasso Regression (LaR) | 0.869 | 0.142 | 0.854 | 0.157 |
| Nonparametric ML modeles | Decision tree (DT) | 0.958 | 0.044 | 0.96 | 0.041 |
| | Random forest (RF) | 0.967 | 0.038 | 0.967 | 0.035 |
| | Gradient boosting (GBDT) | 0.991 | 0.011 | 0.985 | 0.016 |

variation across the 10-fold cross validation. In general, the results reveal the superiority of HT-based models, which exhibit slightly higher $R^2$ and lower MSE compared to the KF-based models. The parameter α, which governs the influence of velocity on PDAS, is slightly higher for KF-based models (∼0.72) compared to the HT-based models (∼0.7). Both values exceed the original analytical model's value of 0.25, indicating a greater sensitivity to pulling velocity in the PF simulations. Similarly, parameter β, which measures the effect of temperature gradient on PDAS, is set to 0.5 in the analytical models but is estimated at ∼0.65 for HT-based models and ∼0.79 for KF-based models, suggesting a stronger impact of temperature gradient in PF simulations, particularly in the KF-based models. A comparable trend is observed for parameter γ, which governs the influence of parameter A. Furthermore, for both HT-based and KT-based models, the additional regularizations applied to the RR and LaR models did not make a noticeable effect on the model parameters α, β, and γ.

Compared to the values represented by heatmap (Fig. 3), the HT-based and KF-based models followed the expected trend, showing a positive correlation with the material properties and negative correlation with the processing conditions. While the parameters α and γ for both of the parametric models are somewhat comparable to those in the heatmap, the heatmap predicted a much weaker dependence of ln (PDAS) on ln(G). In contrast, the ML-based models indicated a significantly stronger dependence, with β values of 0.67 and 0.85 for HT-based and KF-based LR models, respectively. Furthermore, the prefactor C calculated in the ML models are substantially larger than the original analytical models (Eqs. 1 and 2). To further investigate the differences, we conducted ML modeling with various constraints on the parameter C. The results, summarized in Tables S1 and S2 in the Supplementary Material, indicate that changes in β are more pronounced compared to α and γ. This suggests that the factors influencing the effect of material properties and velocity on PDAS are independent of the prefactor C,

**Table 4**
Comparison of regression model parameters (C, α, β, and γ) for HT-based and KF-based LR, RR, and LaR models.

| | Detail | C | α | β | γ |
|---|---|---|---|---|---|
| HT-based | Linear Regression (LR) | 8159± 1656 | 0.7 ± 0.006 | 0.67±0.01 | 0.46±0.004 |
| | Ridge Regression (RR) | 4408±862 | 0.7 ± 0.006 | 0.65±0.01 | 0.46±0.004 |
| | Lasso Regression (LaR) | 2319±474 | 0.69±0.006 | 0.62±0.01 | 0.45±0.004 |
| KF-based | Linear Regression (LR) | $(9.20\pm2.6)\times10^6$ | 0.72±0.007 | 0.84±0.012 | 0.66±0.007 |
| | Ridge Regression (RR) | $(2.91\pm0.8)\times10^6$ | 0.72±0.006 | 0.79±0.011 | 0.64±0.007 |
| | Lasso Regression (LaR) | $(1.38\pm0.4)\times10^6$ | 0.71±0.006 | 0.77±0.012 | 0.63±0.007 |

whereas the impact of temprature gradient is not. Expanding the range for parameter C improved the model's performance by increasing $R^2$ and reducing MSE, reaching optimal values at approximately $8\times10^3$ and $9\times10^6$ for the HT-based and KF-based models, respectively. This is in great agreement with the ML-estimated optimum value of parameter C (represented in Table 4).

### 3.2. Nonparametric models

DT is one of the non-parametric regression models that we investigated in this study. Unlike parametric models, DT makes fewer assumptions about the underlying data distribution, making it flexible. DT model is sensitive to hyperparameters such as maximum tree depth and the minimum samples required to split a node. To optimize the DT's performance and prevent overfitting or underfitting, hyperparameter tuning is essential. We have paired grid search with the k-fold cross-validation to ensure that the hyperparameters generalize well to unseen data while controlling for variance and bias. The optimal hyperparameters were identified with a maximum depth of 13 and a minimum sample split of 2. Fig. 5 compares the PDAS determined by PF simulations with the predictions of the DT regression models. Similar to the parametric models, we employed 10-fold cross-validation for training the model and presented the model's performance with both $R^2$ and MSE values. In general, the nonparametric DT model for both HT-based and KF-based models presents a very good performance for both training and testing datasets. However, obtaining $R^2=1$ and MSE=0 for the training raises the question of model overfitting. Hyperparameter tuning of RR includes determining maximum depth, number of trees, and bootstrapping status. We perform hyperparameter tuning for GBDT model based on learning rate, number of trees, and maximum depth. $R^2$ and MSE presented in Table 5 show that, similar to the DT model, RR and GBDT for both HT-based and KF-based models show a very high performance.

One of the primary challenges in using DT, RF, or GBDT algorithms is determining when to stop the tree-growing process. Most of the tree building methods produce very complex models which may memorize the training data instead of learning and forming patterns, which often result in overfitted models [103,104]. The overfitting is especially more pronounced with small datasets [103,105], as overfitting becomes more severe and negatively impacts the performance on unseen data. This is a key barrier against the adoption of these models in practical applications [103,104]. In this study, the limited dataset steaming from the computational costs of running MD and PF simulations made the nonparametric models more prone to overfitting. Although we observed that non-parametric models such as DT, RF, and GBDT outperformed parametric models like linear regression in terms of $R^2$ values, we remain cautious about overfitting. To address this concern, we further evaluate the model's transferability by testing its performance on unseen experimental data, ensuring that the models are not only effective for the training and testing datasets but also generalizable to real-world applications.

**Table 5**
The performance metrics ($R^2$ and MSE) for testing the HT-based and KF-based parametric (LR, RR, LaR) and nonparametric (DT, RR, GBDT) models with unseen experimental data.

| Model | Detail | HT-based | | KF-based | |
|---|---|---|---|---|---|
| | | $R^2$ | MSE | $R^2$ | MSE |
| Analytical | Original model | 0.686 | 0.431 | 0.381 | 0.851 |
| Parametric ML models | Linear regression | 0.89 | 0.151 | 0.882 | 0.163 |
| | Ridge Regression | 0.924 | 0.104 | 0.835 | 0.227 |
| | Lasso Regression | 0.924 | 0.104 | 0.835 | 0.227 |
| Non-Parametric ML models | Decision tree | 0.296 | 0.969 | −0.273 | 1.751 |
| | Random forest | 0.282 | 0.988 | −0.491 | 2.051 |
| | Gradient boosting | 0.737 | 0.361 | −0.566 | 2.154 |

In order to validate and assess the transferability of our ML models, we tested them on unseen experimental data [106–108], using models previously trained with PF-calculated data. Fig. 6 compares the experimentally measured PDAS with ML-predicted values based on the original analytical, parametric (LR, RR, LaR) and nonparametric (DT, RR, GBDT) models. The results show that the analytical HT and KF models perform poorly, with $R^2$ values of 0.686 and 0.381, respectively (as listed in Table 5). This confirms the previously recognized inefficiency of the analytical models and highlights the need for development of robust ML-based predictive models. However, contrary to the strong performance shown in Fig. 5 for non-parametric models, the DT, RF, and GBDT models exhibit poor performance when applied to experimental data, especially for the KF-based model. This suggests that high $R^2$ obtained earlier for these models were likely due to overfitting. In contrast, the parametric models, LR, RR, and LaR, demonstrate much stronger transferability, achieving $R^2$ values of 0.89, 0.924, and 0.924 for HT-based, and 0.882, 0.835, and 0.835 for KF-based models, respectively. These results highlight the parametric models' ability to generalize and reliably predict PDAS based on processing parameters and material properties, reinforcing their utility as predictive tools in practical applications.

In summary, our proposed ML-based models uncovering the interplay between the processing parameters, material properties, and PDAS are given as Eqs. (28) and (29), which are chosen from the LR modeling due to its simplicity and reasonable accuracy. Compared to original HT and KF models, these models predict a stronger sensitivity to the processing conditions ($V$ and $G$) and material-dependent parameter $A$. Several studies in literature have also discussed that the role of temperature gradient [39,40,45,50] and velocity [37,39,40,43,45–51] on the geometry of dendrite tip and arm spacing more pronounced than the HT and KF models.

$$PDAS_{\mathrm{ML-HT}} = (8159\ ) \times (A_{HT})^{0.459} \times V^{-0.699} \times\ G^{-0.676} \tag{28}$$

$$PDAS_{\mathrm{ML-KF}} = \left(9.20 \times 10^6\right) \times (A_{KF})^{0.655} \times V^{-0.719} \times\ G^{-0.836} \tag{29}$$

## 4. Conclusion

This paper presented a comprehensive approach for predicting primary dendrite arm spacing (PDAS) in solidification processes by combining ML models with MD-informed PF simulations, using established HT and KF analytical frameworks. The training and testing dataset generated by PF modeling of directional solidification covered

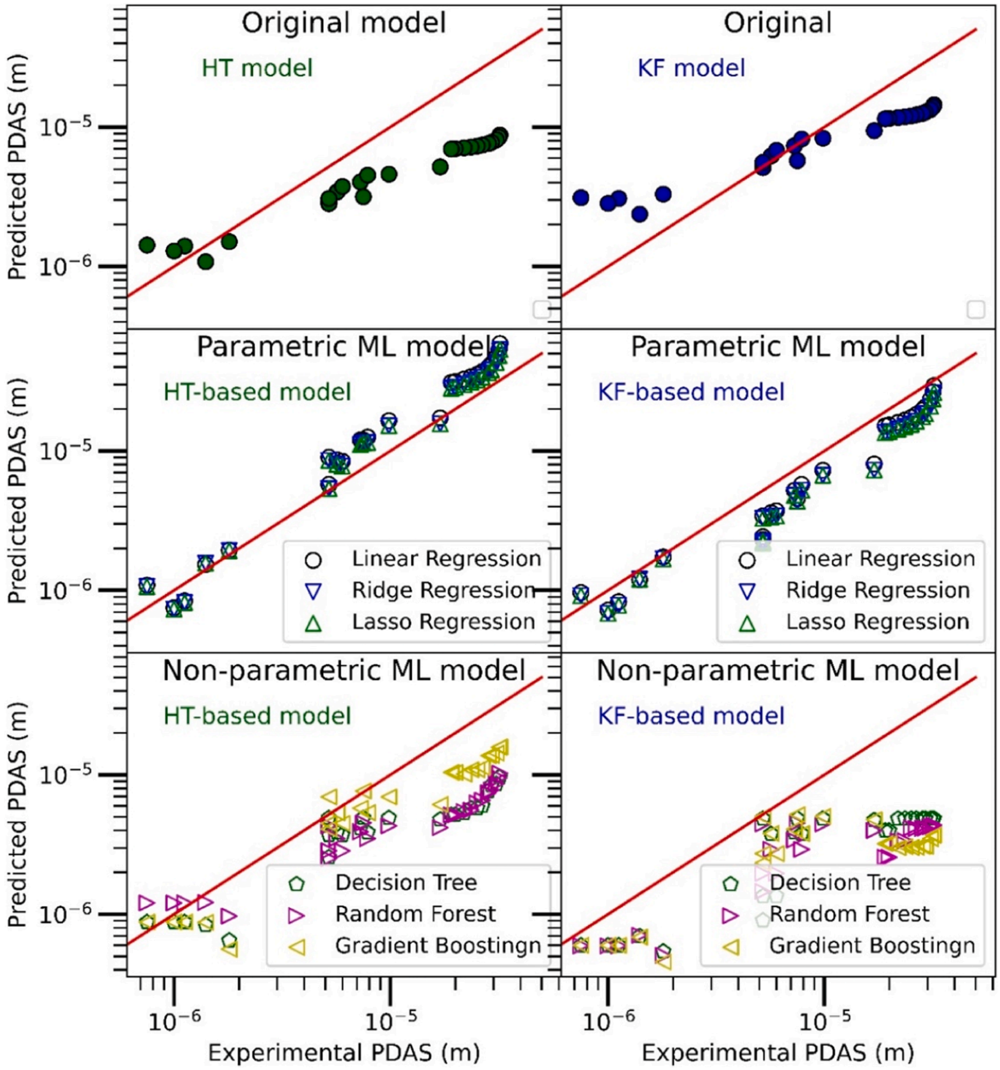


**Fig. 6.** The regression analysis on the unseen testing dataset between experimental data [106–108] and predictions of original analytical, parametric (LR, LaR, and RR), and nonparametric (DT, RF, and XGDT) HT-based and KF-based models.

seven different compositions of Mg-Al, Al-Cu, and Ti-Ni material systems for a wide range of processing conditions. This study focused on understanding the relationship between PDAS, processing conditions, and material properties through parametric and non-parametric ML models. Specifically, parametric models like linear, ridge, and lasso regression, are non-parametric models such as decision trees, random forests, and gradient boosting were examined. The model's accuracy and transferability were tested on unseen experimental data for some other alloys.

Parametric models demonstrated greater reliability and generalizability, showing consistent performance on both PF simulation data and experimental data, particularly when tested on unseen experimental results. In contrast, non-parametric models, while accurate on training data, suffered from overfitting and performed poorly on unseen experimental data. The results indicated that HT-based parametric models, with higher $R^2$ values and lower mean squared error (MSE), offer slightly better predictive performance than KF-based models. In general, analytical HT and KF models represent an underestimation of the impact of processing parameters. Pulling velocity and temperature gradient effects are almost similar in ML-guided analytical models, but in the original analytical models a much higher sensitivity of PDAS to temperature gradient than pulling velocity was considered. The ML-derived PDAS relationships also revealed a stronger sensitivity to pulling velocity, temperature gradient, and material-dependent parameters compared to the original analytical models.

The findings of this work highlight the value of ML, especially parametric regression models, for optimizing alloy manufacturing processes and controlling the microstructures.

## CRediT authorship contribution statement

**Sepideh Kavousi:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohsen Asle Zaeem:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

## Declaration of competing interests

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.

## Acknowledgments

This study was supported by the National Science Foundation, NSF-CMMI 2031800. Authors are grateful for the supercomputing time allocation provided by the NSF's ACCESS (Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support), Award No TG-DMR140008.

## References

[1] A. Faghri, Y. Zhang, Melting and Solidification, Fundamentals of Multiphase Heat Transfer and Flow, Springer International Publishing, Cham, 2020, pp. 257–321.

[2] M.U. Kumar, S. Hrushikesh, Solidification of metals and alloys, in: A. Zakaria, A. Nada (Eds.), Casting Processes and Modelling of Metallic Materials, IntechOpen, Rijeka, 2020. Ch. 2.

[3] G. Azizi, B.G. Thomas, M. Asle Zaeem, Review of peritectic solidification mechanisms and effects in steel casting, Metallurg. Mater. Trans. B 51 (5) (2020) 1875–1903.

[4] L.E. Murr, A metallographic review of 3D printing/additive manufacturing of metal and alloy products and components, Metallogr. Microstruct. Anal. 7 (2) (2018) 103–132.

[5] M.M. Ahmed, M.M.E.-S. Seleman, D. Fydrych, Ç. Gürel, Review on friction stir welding of dissimilar magnesium and aluminum alloys: scientometric analysis and strategies for achieving high-quality joints, J. Magn. Alloys (2023).

[6] M. Abedi, S. Sovizi, A. Azarniya, D. Giuntini, M.E. Seraji, H.R.M. Hosseini, C. Amutha, S. Ramakrishna, A. Mukasyan, An analytical review on Spark Plasma Sintering of metals and alloys: from processing window, phase transformation, and property perspective, Critic. Rev. Solid State Mater. Sci. 48 (2) (2023) 169–214.

[7] P.C. Grignon, Memoires de physique sur l'art de fabriquer le fer: d'en fondre & forger des canons d'artillerie, sur l'histoire naturelle et sur divers sujets particuliers de physique et d'économie, Delalain (1775).

[8] W. Kurz, D.J. Fisher, R. Trivedi, Progress in modelling solidification microstructures in metals and alloys: dendrites and cells from 1700 to 2000, Int. Mater. Rev. 64 (6) (2019) 311–354.

[9] S. Mao, Y. Cao, W. Chen, D. Sun, An anisotropic lattice Boltzmann - phase field model for dendrite growth and movement in rapid solidification of binary alloys, npj Comput. Mater. 10 (1) (2024) 63.

[10] E. Brener, H. Müller-Krumbhaar, D. Temkin, Structure formation and the morphology diagram of possible structures in two-dimensional diffusional growth, Phys. Rev. E 54 (3) (1996) 2714.

[11] E. Brener, H. Müller-Krumbhaar, D. Temkin, Kinetic phase diagram and scaling relations for stationary diffusional growth, Europhys. Lett. 17 (6) (1992) 535.

[12] J. Hunt, Solidification and Casting of metals: Proceedings of an International Conference On Solidification, Sheffield Metallurgical and Engineering Association U of S, Editor. Metals Society, 1979.

[13] W. Kurz, D.J. Fisher, Dendrite growth at the limit of stability: tip radius and spacing, Acta Metallurg. 29 (1) (1981) 11–20.

[14] D. Tourret, J. Klemm-Toole, A.E. Castellanos, B. Rodgers, G. Becker, A. Saville, B. Ellyson, C. Johnson, B. Milligan, J. Copley, R. Ochoa, A. Polonsky, K. Pusch, M. Haines, K. Fezzaa, T. Sun, K. Clarke, S. Babu, T. Pollock, A. Karma, A. Clarke, Morphological stability of solid-liquid interfaces under additive manufacturing conditions, Acta Mater. 250 (2023) 118858.

[15] J. Lamb, R. Ochoa, A. Eres-Castellanos, J. Klemm-Toole, M.P. Echlin, T. Sun, K. Fezzaa, A. Clarke, T.M. Pollock, Quantification of melt pool dynamics and microstructure during simulated additive manufacturing, Scr. Mater. 245 (2024) 116036.

[16] P.K. Galenko, Convection effect on solidification microstructure: dendritic arm spacing and microsegregation, Eur. Phys. J. Spec. Top. 232 (8) (2023) 1261–1271.

[17] P.K. Galenko, D.V. Alexandrov, From atomistic interfaces to dendritic patterns, Philos. Trans. R. Soc. A 376 (2113) (2018) 20170210.

[18] D. Tourret, Y. Song, A.J. Clarke, A. Karma, Grain growth competition during thin-sample directional solidification of dendritic microstructures: a phase-field study, Acta Mater. 122 (2017) 220–235.

[19] S. Kwak, J. Kim, H. Ding, X. Xu, R. Chen, J. Guo, H. Fu, Using multiple regression analysis to predict directionally solidified TiAl mechanical property, J. Mater. Sci. Technol. 104 (2022) 285–291.

[20] S. Dong, R. Chen, J. Guo, H. Ding, Y. Su, H. Fu, Effect of power on microstructure and mechanical properties of Ti44Al6Nb1. 0Cr2. 0V0. 15Y0. 1B alloy prepared by cold crucible directional solidification, Mater. Des. 67 (2015) 390–397.

[21] R. Chen, S. Dong, J. Guo, H. Ding, Y. Su, H. Fu, Microstructure evolution and mechanical properties of directionally-solidified TiAlNb alloy in different temperature gradients, J. Alloys. Compd. 648 (2015) 667–675.

[22] Y. Wang, J.M. Lamim Ribeiro, P. Tiwary, Machine learning approaches for analyzing and enhancing molecular dynamics simulations, Curr. Opin. Struct. Biol. 61 (2020) 139–145.

[23] J. Strickland, B. Nenchev, H. Dong, On directional dendritic growth and primary spacing—a review, Crystals 10 (7) (2020) 627.

[24] D. Welton, N. D'Souza, J. Kelleher, S. Gardner, Z. Dong, G. West, H. Dong, Discontinuous precipitation in Ni-base superalloys during solution heat treatment, Metallurg. Mater. Trans. A 46 (2015) 4298–4315.

[25] H.T. Pang, N. D'souza, H. Dong, H.J. Stone, C.M. Rae, Detailed analysis of the solution heat treatment of a third-generation single-crystal nickel-based superalloy CMSX-10K®, Metallurg. Mater. Trans. A 47 (2016) 889–906.

[26] J.S. Langer, H. Müller-Krumbhaar, Theory of dendritic growth—I. Elements of a stability analysis, Acta Metallurg. 26 (11) (1978) 1681–1687.

[27] R. Trivedi, W.A. Tiller, Interface morphology during crystallization—I. Single filament, unconstrained growth from a pure melt, ACTA METALLURG. 26 (5) (1978) 671–678.

[28] G.R. Kotler, L.A. Tarshis, An extension to the analysis of dendritic growth in pure systems, J. Cryst. Growth 5 (2) (1969) 90–98.

[29] R. Trivedi, Growth of dendritic needles from a supercooled melt, Acta Metallurg. 18 (3) (1970) 287–296.

[30] G. Horvay, J.W. Cahn, Dendritic and spheroidal growth, Acta Metallurg. 9 (7) (1961) 695–705.

[31] J.S. Langer, H. Müller-Krumbhaar, Theory of dendritic growth—II. Instabilities in the limit of vanishing surface tension, Acta Metallurg. 26 (11) (1978) 1689–1695.

[32] R. Trivedi, Interdendritic spacing: part II. A comparison of theory and experiment, Metallurg. Mater. Trans. A 15 (1984) 977–982.

[33] J. Kirkaldy, D. Venugopalan, Pattern selection relations for deep-rooted binary alloy cells, Scr. Metallurg. 23 (9) (1989) 1603–1608.

[34] B. Spencer, H. Huppert, On the solidification of dendritic arrays: selection of the tip characteristics of slender needle crystals by array interactions, Acta Mater. 46 (8) (1998) 2645–2662.

[35] J. Hunt, Pattern formation in solidification, Sci. Technol. Adv. Mater. 2 (1) (2001) 147–155.

[36] S.-Z. Lu, J. Hunt, P. Gilgien, W. Kurz, Cellular and dendritic growth in rapidly solidified Al Fe and Al Cu alloys, Acta Metallurg. Mater. 42 (5) (1994) 1653–1660.

[37] D. Bouchard, J.S. Kirkaldy, Prediction of dendrite arm spacings in unsteady-and steady-state heat flow of unidirectionally solidified binary alloys, Metallurg. Mater. Trans. B 28 (4) (1997) 651–663.

[38] D. Bouchard, J.S. Kirkaldy, Scaling of intragranuiar dendritic microstructure in ingot solidification, Metallurg. Mater. Trans. B 27 (1) (1996) 101–113.

[39] D.G. McCartney, J.D. Hunt, Measurements of cell and primary dendrite arm spacings in directionally solidified aluminium alloys, Acta Metallurg. 29 (11) (1981) 1851–1863.

[40] G.A. Santos, P.R. Goulart, A.A. Couto, A. Garcia, Primary dendrite ARM spacing effects upon mechanical properties of an AL–3Wt%CU–1Wt%LI alloy, in: A. Öchsner, H. Altenbach (Eds.), Properties and Characterization of Modern Materials, Springer Singapore, Singapore, 2017, pp. 215–229.

[41] N. Raghavan, S. Simunovic, R. Dehoff, A. Plotkowski, J. Turner, M. Kirka, S. Babu, Localized melt-scan strategy for site specific control of grain size and primary dendrite arm spacing in electron beam additive manufacturing, Acta Mater. 140 (2017) 375–387.

[42] L. Li, R.A. Overfelt, Influence of directional solidification variables on the cellular and primary dendrite arm spacings of PWA1484, J. Mater. Sci. 37 (16) (2002) 3521–3532.

[43] K.P. Young, D.H. Kerkwood, The dendrite arm spacings of aluminum-copper alloys solidified under steady-state conditions, Metallurg. Trans. A 6 (1) (1975) 197–205.

[44] M. Imagumbai, Behaviors of manganese-sulfide in aluminum-killed steel solidified uni-directionally in steady State&mdash;dendrite structure and inclusions, ISIJ International 34(11) (1994) 896–905. https://doi.org/10.2355/isijinternational.34.896.

[45] K.S. Cruz, E.S. Meza, F.A.P. Fernandes, J.M.V. Quaresma, L.C. Casteletti, A. Garcia, Dendritic arm spacing affecting mechanical properties and wear behavior of Al-Sn and Al-Si alloys directionally solidified under unsteady-State conditions, Metallurg. Mater. Trans. A 41 (4) (2010) 972–984.

[46] H. Weidong, G. Xingguo, Z. Yaohe, Primary spacing selection of constrained dendritic growth, J. Cryst. Growth 134 (1) (1993) 105–115.

[47] M.L.N.M. Melo, E.M.S. Rizzo, R.G. Santos, Predicting dendrite arm spacing and their effect on microporosity formation in directionally solidified Al-Cu alloy, J. Mater. Sci. 40 (7) (2005) 1599–1609.

[48] D. Ma, Modeling of primary spacing selection in dendrite arrays during directional solidification, Metallurg. Mater. Trans. B 33 (2) (2002) 223–233.

[49] W. Wang, P.D. Lee, M. McLean, A model of solidification microstructures in nickel-based superalloys: predicting primary dendrite spacing selection, Acta Mater. 51 (10) (2003) 2971–2987.

[50] C. Yang, Q. Xu, B. Liu, Primary dendrite spacing selection during directional solidification of multicomponent nickel-based superalloy: multiphase-field study, J. Mater. Sci. 53 (13) (2018) 9755–9770.

[51] S. Ghosh, L. Ma, N. Ofori-Opoku, J.E. Guyer, On the primary spacing and microsegregation of cellular dendrites in laser deposited Ni–Nb alloys, Model. Simul. Mat. Sci. Eng. 25 (6) (2017) 065002.

[52] Y. Liu, C. Niu, Z. Wang, Y. Gan, Y. Zhu, S. Sun, T. Shen, Machine learning in materials genome initiative: a review, J. Mater. Sci. Technol. 57 (2020) 113–122.

[53] J.F. Durodola, Machine learning for design, phase transformation and mechanical properties of alloys, Prog. Mater. Sci. 123 (2022) 100797.

[54] M. Hu, Q. Tan, R. Knibbe, M. Xu, B. Jiang, S. Wang, X. Li, M.-X. Zhang, Recent applications of machine learning in alloy design: a review, Mater. Sci. Eng. R 155 (2023) 100746.

[55] B. Echebarria, R. Folch, A. Karma, M. Plapp, Quantitative phase-field model of alloy solidification, Phys. Rev. E 70 (6) (2004) 061604.

[56] A. Karma, W.-J. Rappel, Quantitative phase-field modeling of dendritic growth in two and three dimensions, Phys. Rev. E 57 (4) (1998) 4323.

[57] N. Provatas, J.A. Dantzig, et al., Dendritic growth, in: K.H.J. Buschow, R.W. Cahn, M.C. Flemings, B. Ilschner, E.J. Kramer, S. Mahajan, et al. (Eds.), Encyclopedia of Materials: Science and Technology, Elsevier, Oxford, 2001, pp. 2052–2058.

[58] S. Kavousi, M. Asle Zaeem, Quantitative phase-field modeling of solute trapping in rapid solidification, Acta Mater. 205 (2021) 116562.

[59] S. Balay, S. Abhyankar, M. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. Gropp, D. Kaushik, Petsc Users Manual Revision 3.8, Argonne National Lab.(ANL), Argonne, IL (United States), 2017.

[60] C.F. Curtiss, J.O. Hirschfelder, Integration of stiff equations, Proc. Natl. Acad. Sci. U.S.A. 38 (3) (1952) 235.

[61] Y. Saad, M.H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput. 7 (3) (1986) 856–869.

[62] R. Kobayashi, A numerical approach to three-dimensional dendritic solidification, Exp. Math. 3 (1) (1994) 59–81.

[63] R. Kobayashi, Modeling and numerical simulations of dendritic crystal growth, Physica D 63 (3) (1993) 410–423.

[64] S. Kavousi, B.R. Novak, D. Moldovan, M. Asle Zaeem, Quantitative prediction of rapid solidification by integrated atomistic and phase-field modeling, Acta Mater. 211 (2021) 116885.

[65] S. Kavousi, B.R. Novak, M. Asle Zaeem, D. Moldovan, Combined molecular dynamics and phase field simulation investigations of crystal-melt interfacial properties and dendritic solidification of highly undercooled titanium, Comput. Mater. Sci. 163 (2019) 218–229.

[66] S. Kavousi, B.R. Novak, M.I. Baskes, M. Asle Zaeem, D. Moldovan, Modified embedded-atom method potential for high-temperature crystal-melt properties of Ti–Ni alloys and its application to phase field simulation of solidification, Model. Simul. Mat. Sci. Eng. 28 (1) (2020) 015006.

[67] N. Zhang, M. Asle Zaeem, Competing mechanisms between dislocation and phase transformation in plastic deformation of single crystalline yttria-stabilized tetragonal zirconia nanopillars, Acta Mater. 120 (2016) 337–347.

[68] A. Mahata, M. Asle Zaeem, M.I. Baskes, Understanding homogeneous nucleation in solidification of aluminum by molecular dynamics simulations, modelling and simulation, Mater. Sci. Eng. 26 (2) (2018) 025007.

[69] M.S. Daw, M.I. Baskes, Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals, Phys. Rev. B 29 (12) (1984) 6443.

[70] M.I. Baskes, Modified embedded-atom potentials for cubic materials and impurities, Phys. Rev. B 46 (5) (1992) 2727.

[71] A. Mahata, T. Mukhopadhyay, M. Asle Zaeem, Modified embedded-atom method interatomic potentials for Al-Cu, Al-Fe and Al-Ni binary alloys: from room temperature to melting point, Comput. Mater. Sci. 201 (2022) 110902.

[72] A. Mahata, M. Asle Zaeem, Insights on Solidification of Mg and Mg–Al Alloys by Large Scale Atomistic Simulations, Springer International Publishing, Cham, 2020, pp. 51–53.

[73] J. Rahman, Atomistic Simulations for Computing Solid Liquid Interface Properties of the Al-Mg System, 2009.

[74] S. Kavousi, B.R. Novak, J. Hoyt, D. Moldovan, Interface kinetics of rapid solidification of binary alloys by atomistic simulations: application to Ti-Ni alloys, Comput. Mater. Sci. 184 (2020) 109854.

[75] J. Hoyt, M. Asta, A. Karma, Method for computing the anisotropy of the solid-liquid interfacial free energy, Phys. Rev. Lett. 86 (24) (2001) 5530.

[76] E. Asadi, M. Asle Zaeem, S. Nouranian, M.I. Baskes, Quantitative modeling of the equilibration of two-phase solid-liquid Fe by atomistic simulations on diffusive time scales, Phys. Rev. B 91 (2) (2015) 024105.

[77] G. Azizi, S. Kavousi, M. Asle Zaeem, Interactive effects of interfacial energy anisotropy and solute transport on solidification patterns of Al-Cu alloys, Acta Mater. 231 (2022) 117859.

[78] S. Kavousi, A. Gates, L. Jin, M. Asle Zaeem, A temperature-dependent atomistic-informed phase-field model to study dendritic growth, J. Cryst. Growth 579 (2022) 126461.

[79] J. Monk, Y. Yang, M. Mendelev, M. Asta, J. Hoyt, D. Sun, Determination of the crystal-melt interface kinetic coefficient from molecular dynamics simulations, Model. Simul. Mat. Sci. Eng. 18 (1) (2009) 015004.

[80] T. Pinomaa, N. Provatas, Quantitative phase field modeling of solute trapping and continuous growth kinetics in quasi-rapid solidification, Acta Mater. 168 (2019) 167–177.

[81] G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor, Linear Regression, An introduction to Statistical Learning: With applications in Python, Springer, 2023, pp. 69–134.

[82] W. Härdle, Applied Nonparametric Regression, Cambridge university press, 1990.

[83] K. Takezawa, Introduction to Nonparametric Regression, John Wiley & Sons, 2005.

[84] E. Alpaydin, Introduction to Machine Learning, MIT press, 2020.

[85] R. Muthukrishnan, R. Rohini, LASSO: a feature selection technique in predictive modeling for machine learning, in: 2016 IEEE International Conference on Advances in Computer Applications (ICACA), 2016, pp. 18–20.

[86] V. Roth, The generalized LASSO, IEEE Trans. Neural Netw. 15 (1) (2004) 16–28.

[87] W.N. van Wieringen, Lecture notes on ridge regression, arXiv preprint arXiv: 1509.09169 (2015). DOI:10.48550/arXiv.1509.09169.

[88] P. Bühlmann, B. Yu, Boosting with the L 2 loss: regression and classification, J. Am. Stat. Assoc. 98 (462) (2003) 324–339.

[89] S. Suthaharan, Decision Tree Learning, Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, Springer US, Boston, MA, 2016, pp. 237–269.

[90] S. Suthaharan, Random Forest Learning, Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, Springer US, Boston, MA, 2016, pp. 273–288.

[91] J. Elith, J.R. Leathwick, T. Hastie, A working guide to boosted regression trees, J. Anim. Ecol. 77 (4) (2008) 802–813.

[92] O.Z. Maimon, L. Rokach, Data mining with decision trees: theory and applications, World Sci. (2014).

[93] S.B. Kotsiantis, Decision trees: a recent overview, Artif. Intell. Rev. 39 (2013) 261–283.

[94] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J.C.W. Chan, P. van den Bossche, J. Van Mierlo, N. Omar, Random forest regression for online capacity estimation of lithium-ion batteries, Appl. Energy 232 (2018) 197–210.

[95] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[96] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[97] M. Waskom, seaborn: statistical data visualization, J. Open Source Softw. 6 (2021) 3021.

[98] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (3) (2007) 90–95.

[99] A.Y. Ng, Preventing" overfitting" of Cross-Validation Data, ICML, Citeseer, 1997, pp. 245–253.

[100] B. Ghojogh, M. Crowley, arXiv preprint, 2019 arXiv:1905.12787.

[101] D. Degras, T. Chapalain, B. Thirion, On the stability of reduced-rank ridge regression, in: European Signal Processing Conference, 2024.

[102] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B 58 (1) (1996) 267–288.

[103] M. Wu, M. Hughes, S. Parbhoo, M. Zazzi, V. Roth, F. Doshi-Velez, Beyond sparsity: tree regularization of deep models for interpretability, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2018.

[104] R.G. Leiva, A.F. Anta, V. Mancuso, P. Casari, A novel hyperparameter-free approach to decision tree construction that avoids overfitting by design, IEEE Access 7 (2019) 99978–99987.

[105] H. Seto, A. Oyama, S. Kitora, H. Toki, R. Yamamoto, J.i. Kotoku, A. Haga, M. Shinzawa, M. Yamakawa, S. Fukui, T. Moriyama, Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data, Sci. Rep. 12 (1) (2022) 15889.

[106] Y.-J. Liang, A. Li, X. Cheng, X.-T. Pang, H.-M. Wang, Prediction of primary dendritic arm spacing during laser rapid directional solidification of single-crystal nickel-base superalloys, J. Alloys Compd. 688 (2016) 133–142.

[107] X. Xu, X. Lin, M. Yang, J. Chen, W. Huang, Microstructure evolution in laser solid forming of Ti–50 wt% Ni alloy, J. Alloys Compd. 480 (2) (2009) 782–787.

[108] X. Wang, K. Chou, Effects of thermal cycles on the microstructure evolution of Inconel 718 during selective laser melting process, Addit. Manuf. 18 (2017) 1–14.