



# Prediction of the critical temperature of a superconductor by using the WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques

Paulino José García-Nieto<sup>1</sup> · Esperanza García-Gonzalo<sup>1</sup> · José Pablo Paredes-Sánchez<sup>2</sup>

Received: 31 December 2020 / Accepted: 1 July 2021 / Published online: 20 July 2021  
© The Author(s) 2021, corrected publication 2021

## Abstract

This study builds a predictive model capable of estimating the critical temperature of a superconductor from experimentally determined physico-chemical properties of the material (input variables): features extracted from the thermal conductivity, atomic radius, valence, electron affinity and atomic mass. This original model is built using a novel hybrid algorithm relied on the multivariate adaptive regression splines (MARS) technique in combination with a nature-inspired meta-heuristic optimization algorithm termed the whale optimization algorithm (WOA) that mimics the social behavior of humpback whales. Additionally, the Ridge, Lasso and Elastic-net regression models were fitted to the same experimental data for comparison purposes. The results of the current investigation indicate that the critical temperature of a superconductor can be successfully predicted using this proposed hybrid WOA/MARS-based model. Furthermore, the results obtained with the Ridge, Lasso and Elastic-net regression models are clearly worse than those obtained with the WOA/MARS-based model.

**Keywords** Critical temperature · Superconductivity · Multivariate adaptive regression splines (MARS) · Whale optimization algorithm (WOA)

## 1 Introduction

Superconducting materials (materials that conduct current with zero resistance) have significant practical applications [1–4]. Perhaps the best-known application is in the Magnetic Resonance Imaging (MRI) systems widely employed by healthcare professionals for detailed internal body imaging. Other prominent applications include the superconducting coils used to maintain high magnetic fields in the Large Hadron Collider at CERN and the extremely sensitive magnetic field measuring devices called SQUIDS (Superconducting Quantum Interference Devices). Furthermore, superconductors could revolutionize the energy

industry as frictionless (zero resistance) superconducting wires and electrical system may transport and deliver electricity with no energy loss.

A superconductor conducts current with zero resistance only at or below its *superconducting critical temperature* ( $T_c$ ) [5–9]. Moreover, the scientific model and theory that predicts  $T_c$  is an open problem, which has been baffling the scientific community since the discovery of superconductivity in 1911 by Heike Kamerlingh Onnes [1–9]. In the absence of any theory-based prediction models, we take here an entirely data-driven approach to create a statistical model that predicts  $T_c$  based on its chemical formula. Indeed, an alternative approach for the superconducting critical temperature prediction problem is the machine learning (ML) approach, which builds data-driven predictive models by exploring the relationship between material composition similarity and critical temperature. Machine learning methods need a sufficient amount of training data to be available [10–14], but the availability of an increasing number of materials databases with experimental

✉ Paulino José García-Nieto  
lato@orion.ciencias.uniovi.es

<sup>1</sup> Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo, Spain

<sup>2</sup> Department of Energy, College of Mining, Energy and Materials Engineering, University of Oviedo, 33004 Oviedo, Spain

properties allows the application of these methods for materials property prediction.

In this investigation, a new hybrid regressive model based on the multivariate adaptive regression splines (MARS) technique has been used to successfully predict the superconducting critical temperature  $T_c$  for different types of superconductors. This novel procedure, which combines the MARS approximation [15–19] with the whale optimization algorithm (WOA) [20–22], could be an attractive methodology that has not been tackled as of yet. For comparative purposes, the Ridge, Lasso, and Elastic-net regression models were also fitted to the same experimental dataset to estimate the  $T_c$  and compare the results obtained [23–29]. However, the MARS technique is a statistical learning methodology built up in accordance with the statistics and mathematical analysis which has the ability to deal with nonlinearities including interactions among variables [30, 31]. It is a nonparametric regression technique and can be seen as an extension of linear models that automatically model nonlinearities and complex interactions between variables. MARS approximation presents some benefits in comparison with the classical and metaheuristic regression techniques, including [32–35]: (1) avoiding physical models of the superconductor; (2) providing models that are more flexible than linear regression models; (3) creating models that are simple to understand and interpret; (4) allowing for the modeling of nonlinear relationships among the physico-chemical input variables of a superconductor; (5) offering a good bias-variance trade-off; and (6) providing an explicit mathematical formula of the dependent variable as a function of the independent variables through an expansion of the basis functions (hinge functions and products of two or more hinge functions). This last feature is a fundamental and noteworthy difference compared to other alternative methods, as most of them behave like a black box. Moreover, the WOA optimizer has been used to satisfactorily calculate the optimal MARS hyperparameters. In addition, previous research has indicated that MARS is a very effective tool for use in a large number of real applications, including soil erosion susceptibility prediction [36], rapid chloride permeability prediction of self-compacting concrete [37], evaluation of the earthquake induced uplift displacement of tunnels [38], estimation of hourly global solar radiation [39], atypical algal proliferation modeling in a reservoir [40], pressure drop estimation produced by different filtering media in microirrigation sand filters [41], assessing frost heave susceptibility of gravelly soils [42] and so on. However, it has never been used for evaluating superconducting critical temperature  $T_c$  from the input physico-chemical parameters in most types of superconductors.

This paper is structured as follows: Sect. 2 contains the experimental arrangement, all the variables included in this research and MARS, Ridge, Lasso, and Elastic-net methodologies; Sect. 3 presents the findings acquired with this novel technique by collating the MARS results with the observed values as well as the significance ranking of the input variables, and Sect. 4 concludes this study by providing an inventory of principal results of the research.

## 2 Materials and methods

### 2.1 Dataset

The SuperCon database [43] is currently the biggest and most comprehensive database of superconductors in the world. It is free and open to the public, and it has been used in almost all ML studies of superconductors [44–46]. The SuperCon dataset was pre-processed for further research by Hamidieh [7], and this database is deposited in the University of California Irvine data repository [47]. As a result of the pre-treatment, materials that had some missing features were removed. Also, preliminary processing included the formation of new features based on existing ones. Atomic mass, density, first ionization energy, atomic radius, density, electron affinity, fusion heat, thermal conductivity, and valence were taken as the initial 8 features (see Table 1). That is, the chemical formula of the material was considered and based on statistical parameters of each features: mean, weighted mean, geometric mean, weighted geometric mean, entropy, entropy weighted, range, weighted range, standard deviation, and weighted standard deviation were calculated (see Table 2). This gives us  $8 \times 10 = 80$  features. One additional feature, a numeric variable counting the number of elements in the superconductor, is also extracted. We end up with 81 features. Thus, we have data with 83 columns: 1 column corresponding to the name of the material (identification), 81 columns corresponding to the features extracted, and 1 column of the observed critical temperature ( $T_c$ ) values. The dataset contains information for 21,263 superconductors so that we have 21,262 rows of data. All 82 attributes for each material are numeric. The 81 features extracted are used as independent predictors (input variables) of the critical temperature ( $T_c$ ), which is the dependent variable of the model. This approach to the formation of features is quite general and suitable for the study of superconducting materials due to the general uncertainty of the dependence of the critical temperature.

**Table 1** The physico-chemical properties of an element that are employed for building its features in order to forecast  $T_c$

Variable	Units	Description
Atomic Mass	Atomic mass units (AMU)	Total proton and neutron rest masses
First Ionization Energy	Kilo-Joules per mole (kJ/mol)	Energy required to remove a valence electron
Atomic Radius	Picometer (pm)	Calculated atomic radius
Density	Kilograms per meters cubed (kg/m <sup>3</sup> )	Density at standard temperature and pressure
Electron Affinity	Kilo-Joules per mole (kJ/mol)	Energy required to add an electron to a neutral atom
Fusion Heat	Kilo-Joules per mole (kJ/mol)	Energy to change from solid to liquid without temperature change
Thermal Conductivity	Watts per meter-Kelvin (W/(m K))	Thermal conductivity coefficient $\kappa$
Valence	No units	Typical number of chemical bonds formed by the element

### 2.2 Multivariate adaptive regression splines (MARS) approach

In statistical machine learning, multivariate adaptive regression splines (MARS) is a regression method first conceived by Friedman in 1991 which is appropriate for problems containing a large number of input variables [15–19]. The technique uses a nonparametric approach that can be understood as a prolongation of linear models which allows for considering *interactions* among input variables and nonlinearities.

The MARS technique constructs models according to the following expansion [15–19]:

$$\hat{f}(x) = \sum_{i=0}^M c_i B_i(x) \tag{1}$$

Therefore, this technique approximates the dependent output variable  $y$  by means of an averaged addition of  $B_i(x)$  so that the coefficients  $c_i$  are constant.  $B_i(x)$  can be [15–19]:

- constant and equal to 1. This term is called intercept and corresponds to the term  $c_0$ ;
- a hinge or hockey stick function: this function is  $\max(0, \text{constant} - x)$  or  $\max(0, x - \text{constant})$ . The constant value is termed knot. The MARS technique chooses variables and knot values for these according to the procedure indicated later;
- the multiplication of hinge functions: in this case, these functions model nonlinear relationships between variables.

For instance, Fig. 1 shows a couple of splines for  $q = 1$  at the node  $t = 3.5$ .

Two steps provide the base of the MARS method. First, it constructs a very complex model in the forward phase and then it simplifies it in the backward stage [19, 30, 34, 48]:

- Forward stage: MARS starts with the intercept term, which is calculated by averaging the values of the dependent variable. Next, it adds linear combinations of pairs of hinge functions with the aim of minimizing the least-square error. These new hinge functions depend on a knot and a variable. Thus, to add new terms MARS has to try all the different combinations of variables and knots with the previous terms, called *parent terms*. Then, the coefficients  $c_i$  are determined using linear regression. Finally, it adds terms until a certain threshold for the residual error or a maximum number of terms is reached.
- Backward stage: the previous stage usually constructs an overfitted model. In order to construct a better model with greater generalization skill, this new stage simplifies the model by removing terms, using the generalized cross-validation (*GCV*) criterion described below by first removing the terms that add more *GCV* to the model.

Generalized cross-validation (*GCV*) is the goodness-of-fit index utilized to assess the suitability of the terms of the model in order to prune them from the model. *GCV* not only takes into account the residual error but also how complex the model is. High values of *GCV* mean high residual error and complexity. The formula of this index is [15–19, 30, 34, 48]:

$$GCV(M) = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_M(\mathbf{x}_i))^2}{(1 - C(M)/n)^2} \tag{2}$$

where the parameter  $C(M)$  increases with the number of terms in the regression function and thus, the value of the *GCV* index rises. It is given by [15–19]:

$$C(M) = (M + 1) + dM \tag{3}$$

where  $d$  is a coefficient that determines the importance of this parameter and  $M$  is the number of terms in Eq. (1).

**Table 2** Description of the procedure for features extraction from material’s chemical formula. (The last column serves as an example: features relied on thermal conductivities for  $\text{Re}_7\text{Zr}_1$  are derived and reported to two decimal places; Rhenium and Zirconium’s thermal conductivity coefficients are  $t_1 = 48$  and  $t_2 = 23 \text{ W/(m K)}$ , respectively. Here:  $p_1 = \frac{2}{3}$ ;  $p_2 = \frac{1}{3}$ ;  $w_1 = \frac{48}{71}$ ;  $w_2 = \frac{23}{71}$ ;  $A = \frac{p_1 w_1}{p_1 w_2 + p_2 w_2} \approx 0.867$ ;  $B = \frac{p_2 w_2}{p_1 w_2 + p_2 w_2} \approx 0.193$ )

Feature and description	Formula	Sample value ( $\text{Re}_7\text{Zr}_1$ )
Mean	$\mu = (t_1 + t_2)/2$	35.5
Weighted mean	$v = (p_1 t_1) + (p_2 t_2)$	39.67
Geometric mean	$= \sqrt{t_1 t_2}$	33.23
Weighted geometric mean	$= (t_1)^{p_1} (t_2)^{p_2}$	37.56
Entropy	$= -w_1 \ln(w_1) - w_2 \ln(w_2)$	0.63
Weighted entropy	$= -A \ln(A) - B \ln(B)$	0.44
Range	$= t_1 - t_2 (t_1 > t_2)$	25
Weighted range	$= p_1 t_1 - p_2 t_2$	24.33
Standard deviation	$= \left[ (1/2) \left( (t_1 - \mu)^2 + (t_2 - \mu)^2 \right) \right]^{\frac{1}{2}}$	12.5
Weighted standard deviation	$= \left[ p_1 (t_1 - v)^2 + p_2 (t_2 - v)^2 \right]^{\frac{1}{2}}$	11.79

The relative importance of the independent variables that appear in the regression function (as only some of these variables remain in the final function) can be assessed using different criteria [15–19, 30, 34, 48]: (a) the *GCV* attached to a variable can be one of these criteria, and it is measured taking into account how much this index increases if the variable is erased from the final function; (b) the same criterion can be applied using the *RSS* index; (c) another criterion is the number of subsets (*Nsubset*) of which the variable is a part. If it is part of more terms, its importance is greater.

### 2.3 Whale optimization algorithm (WOA)

The whale optimization algorithm (WOA) is a new technique for solving optimization problems that was first proposed by Mirjalili and Lewis in order to optimize numerical problems [20]. The algorithm simulates the highly intelligent hunting behavior of humpback whales. This foraging behavior is called the bubble-net feeding method and is only observed in humpback whales, which create bubbles to encircle their prey while hunting. The whales dive approximately 12 m deep and then create the bubble spiral around their prey and then swim upward the surface following the bubbles. The mathematical model for spiral bubble-net feeding behavior is given as follows [20–22]:

- Encircling prey

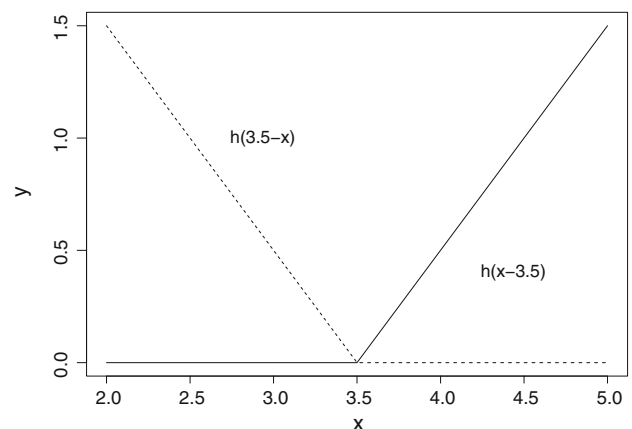
Humpback whales can recognize the location of prey and encircle them. Since the position of the optimum design in the search space is not known a priori, the WOA algorithm assumes that the current best candidate solution

is the target prey or is close to the optimum. After the best search agent is defined, the other search agents will hence try to update their positions toward the best search agent. This behavior is represented by the following equations:

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \tag{4}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D}$$

where  $t$  indicates the current iteration,  $\vec{A}$  and  $\vec{C}$  are coefficient vectors,  $\vec{X}_p$  is the position vector of the prey, and  $\vec{X}$  indicates the position vector of a whale. The vectors  $\vec{A}$  and  $\vec{C}$  are calculated as follows:



**Fig. 1** An example of linear basis functions (hinge function)

$$\begin{aligned} \vec{A} &= 2\vec{a} \cdot \vec{r}_1 - \vec{a} \\ \vec{C} &= 2\vec{r}_2 \end{aligned} \tag{5}$$

where components of  $\vec{a}$  are linearly decreased from 2 to 0 over the course of iterations and  $\vec{r}_1, \vec{r}_2$  are random vectors in  $[0,1]$ .

- Exploitation phase: bubble-net attack method

The bubble-net strategy is a hybrid technique that combines two approaches that can be mathematically modeled as follows [20–22]:

1. Shrinking encircling mechanism: This behavior is achieved by decreasing the value of  $\vec{a}$ . Note that the fluctuation range of  $\vec{A}$  is also decreased by  $\vec{a}$ . In other words,  $\vec{A}$  is a random value in the interval  $[-a, a]$  where  $a$  is decreased from 2 to 0 over the course of iterations. Setting random values for  $\vec{A}$  in  $[-1, 1]$ , the new position of a search agent can be defined anywhere in between the original position of the agent and the position of the current best agent.
2. Spiral updating position: This approach first calculates the distance between the whale located at  $(\vec{X}, \vec{Y})$  and prey located at  $(\vec{X}^*, \vec{Y}^*)$ . A spiral equation is then created between the position of whale and prey to mimic the helix-shaped movement of humpback whales as follows:

$$\vec{X}(t + 1) = \vec{D}' e^{bt} \cos(2\pi t) + \vec{X}^* \tag{6}$$

where  $\vec{D}' = |\vec{X}^*(t) - \vec{X}(t)|$  is the distance between the  $i$ -th whale and the prey (best solution obtained so far),  $b$  is a constant for defining the shape of the logarithmic spiral, and  $t$  is a random number in  $[-1, 1]$ . Note that humpback whales swim around the prey within an increasingly shrinking spiral-shaped path. In order to model this simultaneous behavior, we assume that there is a probability of 50% to choose between either the shrinking encircling mechanism or the spiral model to update the position of the whales during optimization. The mathematical model is as follows [20–22]:

$$\vec{X}(t + 1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}' e^{bt} \cos(2\pi t) + \vec{X}^* & \text{if } p \geq 0.5 \end{cases} \tag{7}$$

where  $p$  is a random number in  $[0, 1]$ . In addition to the bubble-net method, the humpback whales search for prey randomly. The mathematical model of the search is as follows:

- Exploration phase: search for prey

The same approach based on the variation of the  $\vec{A}$  vector can be utilized to search for prey (exploration). In fact, humpback whales search randomly according to their relative position to each other. Therefore, we use  $\vec{A}$  with the random values greater than 1 or less than  $-1$  to force the search agent to move far away from a reference whale. In contrast to the exploitation phase, the position of a search agent in the exploration phase is updated according to a randomly chosen search agent instead of the best search agent. This mechanism and  $|\vec{A}| > 1$  emphasize exploration and allow the WOA algorithm to perform a global search. The mathematical model is as follows [20–22]:

$$\begin{aligned} \vec{D} &= |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \\ \vec{X}(t + 1) &= \vec{X}_{rand} - \vec{A} \cdot \vec{D} \end{aligned} \tag{8}$$

where  $\vec{X}_{rand}$  is a random position vector (a random whale).

The WOA algorithm starts with a set of random solutions. At each iteration, search agents update their positions with respect to either a randomly chosen search agent or the best solution obtained so far. The  $a$  parameter is decreased from 2 to 0 in order to provide exploration and exploitation, respectively. A random search agent is chosen when  $|\vec{A}| > 1$ , while the best solution is selected when  $|\vec{A}| < 1$  for updating the position of the search agents. Finally, the WOA algorithm is concluded upon the satisfaction of a termination criterion.

### 2.4 Ridge regression (RR)

Typically, we consider a sample consisting of  $n$  cases (or number of observations), that is, we have a set of training data  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ , each of which consists of  $p$  covariates (number of variables) and a single outcome. Let  $y_i$  be the outcome and  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$  be the covariate vector for the  $i$ th case. The most popular estimation method is known as the *least squares* fitting procedure, in which the coefficients  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  have been selected to minimize the *residual sum of squares* (RSS) [23–25]:

$$RSS = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \tag{9}$$

*Ridge regression* is very similar to least squares, with the exception that their coefficients are estimated by minimizing a slightly different quantity. Specifically, the ridge regression coefficient estimates  $\hat{\beta}^{RR}$  are the values that minimize [18, 23–25]:



$$L^{RR}(\beta) = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2 \tag{10}$$

where  $\lambda \geq 0$  is the *regularization parameter* or complexity parameter to be determined separately (tuning parameter), that controls the amount of *shrinkage*: the larger the value of  $\lambda$ , the greater the amount of shrinkage. Indeed, Eq. (10) trades off two different criteria. As with least squares, Ridge regression seeks coefficient estimates that fit the data well, by making the RSS small. However, the second term,  $\lambda \sum_{j=1}^p \beta_j^2$ , called a *shrinkage penalty*, is small when  $\beta_1, \dots, \beta_p$  are close to zero, and so it has the effect of *shrinking* the estimates of  $\beta_j$  toward zero. The tuning parameter  $\lambda$  serves to control the relative impact of these two terms on the regression coefficient estimates. When  $\lambda = 0$ , the penalty term has no effect, and Ridge regression will produce the least squares estimates (as  $\lambda \rightarrow 0$ ,  $\hat{\beta}^{RR} \rightarrow \hat{\beta}^{RRS}$ ). However, as  $\lambda \rightarrow \infty$ , the impact of the shrinkage penalty grows, and the Ridge regression coefficient estimates will approach zero (as  $\lambda \rightarrow \infty$ ,  $\hat{\beta}^{RR} \rightarrow \mathbf{0}$ ). Unlike least squares, which generates only one set of coefficient estimates, ridge regression will produce a different set of coefficient estimates,  $\hat{\beta}_\lambda^{RR}$ , for each value of  $\lambda$ . Since selecting a good value for  $\lambda$  is critical, cross-validation has been used.

The advantage of Ridge regressions over least squares is rooted in the *bias-variance trade-off*. As  $\lambda$  increases, the flexibility of the ridge regression fit decreases, leading to decreased variance but increased bias. At the least squares coefficient estimates, which correspond to ridge regression with  $\lambda = 0$ , the variance is high, but there is no bias. But as  $\lambda$  increases, the shrinkage of the ridge coefficient estimates leads to a substantial reduction in the variance of the predictions, at the expense of a slight increase in bias. Ridge regression improves prediction error by shrinking large regression coefficients in order to reduce *overfitting*, but it does not perform covariate selection and therefore does not help to make the model more interpretable.

### 2.5 Least absolute shrinkage and selection operator (Lasso) regression (LR)

Ridge regression does have one obvious disadvantage: it will include all  $p$  predictors in the final model. The penalty  $\lambda \sum_{j=1}^p \beta_j^2$  in Eq. (10) will shrink all of the coefficients toward zero, but it will not set any of them exactly to zero (unless  $\lambda \rightarrow \infty$ ). This may not be a problem for prediction

accuracy, but it can create a challenge in model interpretation in situations in which the number of  $p$  variables is quite large.

The Lasso regression is a relatively recent alternative to Ridge regression that helps to overcome this disadvantage. The Lasso coefficients,  $\hat{\beta}_\lambda^{Lasso}$ , minimize the quantity [18, 25–28]:

$$L^{LR}(\beta) = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j| \tag{11}$$

Comparing Eqs. (11) to (10) demonstrates that the Lasso and Ridge regressions have similar formulations. The only difference is that the  $\beta_j^2$  term in the Ridge regression penalty in Eq. (10) has been replaced by  $|\beta_j|$  in the Lasso penalty in Eq. (11). In statistical terms, the Lasso uses an  $L_1$  penalty instead of an  $L_2$  penalty. The  $L_p$  norm of a coefficient vector  $\beta$  is given by  $\|\beta\|_p = (\sum_{i=1}^n |\beta_i|^p)^{1/p}$ .

As with Ridge regression, the Lasso shrinks the coefficient estimates toward zero. However, in the case of the Lasso, the  $L_1$  penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero when the tuning parameter  $\lambda$  is sufficiently large. Hence, then it performs *variable selection*. As a result, the models generated are generally much easier to interpret than those produced by Ridge regression. It can be said to yield *sparse models*, that is, models that involve only a subset of the variables. As in Ridge regression, selecting a good value of  $\lambda$  for the Lasso is critical. As a result, cross-validation has been employed.

### 2.6 Elastic-net regression (ENR)

Elastic-net regression (ENR) first emerged in response to critiques of the Lasso regression model, whose variable selection can be too dependent on data and thus unstable. The solution was to combine the penalties of Ridge and Lasso regressions to get the best of both worlds. Therefore, ENR is a convex combination of Ridge and Lasso regressions. Indeed, it aims at minimizing the following loss function [18, 23–29]:

$$L^{ENR}(\beta) = \frac{1}{2n} \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \left( \frac{1-\alpha}{2} \sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j| \right) \tag{12}$$

where  $\alpha$  is the mixing parameter between Ridge ( $\alpha = 0$ ) and Lasso ( $\alpha = 1$ ). Now, there are two parameters to tune:

$\lambda$  and  $\alpha$ . In short, the ENR is a regularized regression method that linearly combines both penalties i.e.  $L_1$  and  $L_2$  of the Lasso and Ridge regression methods, and it proves particularly useful when there are multiple correlated features. The essential difference between Lasso and Elastic-net regressions lies in the fact that the Lasso model is likely to pick only one of these features at random while elastic-net model is likely to pick both at once.

### 2.7 Approach accuracy

Eighty of the above-mentioned input variables from Sect. 2.1 have been employed in this study to build this novel WOA/MARS-based method. As is well known, the superconducting critical temperature  $T_c$  is the dependent variable to be predicted. In order to predict  $T_c$  from eighty variables with sufficient confidence, it is essential to select the best model fitted to the observed dataset. Although there are several possible statistics that can be used to ascertain the goodness-of-fit, the rule employed in this study was the coefficient of determination  $R^2$  [48–50], as it is a statistic employed in the scope of a statistical model whose principal objective is to predict upcoming results or to check an assumption. Next, the observed values are referred to as  $t_i$  and the values predicted by the model  $y_i$ , making it possible to define the following sums of squares given by [48–50]:

- $SS_{tot} = \sum_{i=1}^n (t_i - \bar{t})^2$ : is the overall sum of squares, proportional to the sample variance.
- $SS_{reg} = \sum_{i=1}^n (y_i - \bar{t})^2$ : is the regression sum of squares, also termed the explained sum of squares.
- $SS_{err} = \sum_{i=1}^n (t_i - y_i)^2$ : is the residual sum of squares.

where  $\bar{t}$  is the mean of the  $n$  observed data:

$$\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i \tag{13}$$

Based on the former sums, the coefficient of determination is specified by the following equation [48–50]:

$$R^2 \equiv 1 - \frac{SS_{err}}{SS_{tot}} \tag{14}$$

Further criteria considered in this study were the root-mean-square error (RMSE) and mean absolute error (MAE) [48–51]. The RMSE is a statistic that is frequently used to evaluate the predictive capability of a mathematical model. Indeed, the root-mean-square error (RMSE) [48–51] is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (t_i - y_i)^2}{n}} \tag{15}$$

If the root-mean-square error (RMSE) is zero, there is no difference between the predicted and the observed data. The MAE, on the other hand, measures the average magnitude of the errors in a set of forecasts without considering their direction. MAE is the average over the verification sample of the absolute values of the differences between a forecast and the corresponding observation. Its mathematical expression is given by [48–51]:

$$MAE = \frac{\sum_{i=1}^n |t_i - y_i|}{n} \tag{16}$$

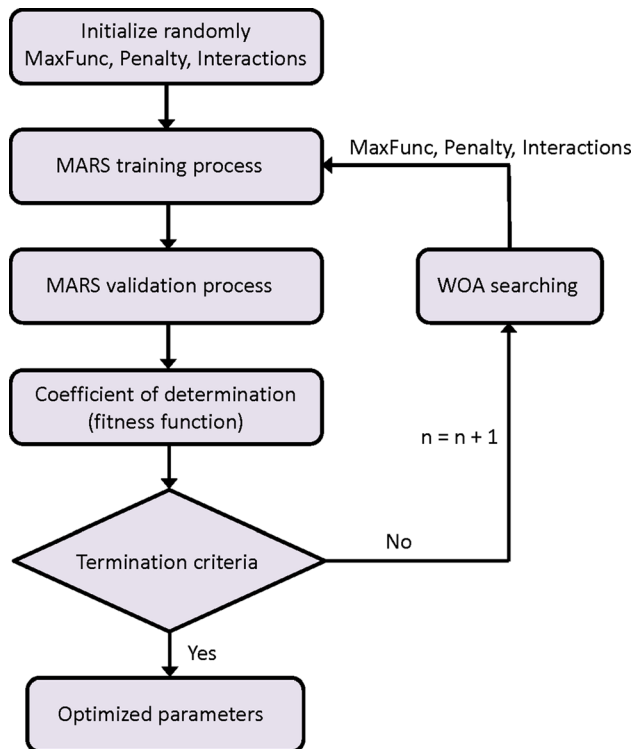
Moreover, the MARS methodology relies heavily on the three hyperparameters [15–19]:

- Maximum number of basis functions (Maxfuncs): maximum number of model terms before pruning, i.e., the maximum number of terms created by the forward pass.
- Penalty parameter ( $d$ ): the generalized cross-validation (GCV) penalty per knot. A value of 0 penalizes only terms, not knots. The value  $-1$  means no penalty.
- Interactions: maximum degree of interaction between variables.

It is important to consider that the MARS technique relies largely on the determination of all three of the aforementioned optimal hyperparameters. Some of the methods often used to determine suitable hyperparameters are [15–19, 30, 34, 48, 52]: grid search, random search, Nelder-Mead search, artificial bee colony, genetic algorithms, pattern search, etc. In this study, the numerical optimizer denominated whale optimization algorithm (WOA) [20–22] has been employed to determine these parameters based on its ability to solve nonlinear optimization problems.

Hence, a novel hybrid WOA/MARS-based method has been applied to predict the superconducting critical temperature  $T_c$  (output variable) from eighty variables (input variables) by studying their influence in order to optimize the calculation through the analysis of the coefficient of determination  $R^2$  with success. Figure 2 shows the flowchart of this new hybrid WOA/MARS-based model developed in this study.

Cross-validation was the standard technique used to find the real coefficient of determination ( $R^2$ ) [48–50]. Indeed, in order to guarantee the predictive ability of the WOA/MARS-based model, an exhaustive tenfold cross-validation algorithm was used [53], which involved splitting the sample into 10 parts and using nine of them for training and the remaining one for testing. This process was performed 10 times using each of the parties of the 10 divisions for testing and calculating the average error. Therefore, all the possible variability within the WOA/MARS-based model



**Fig. 2** Flowchart of the process of parameter optimization with WOA of the MARS model

**Table 3** Search space for each of the MARS parameters in the WOA tuning process

MARS hyperparameters	Lower limit	Upper limit
Maximum number of basis functions (MaxFuncs)	3	100
Interactions	1	4
Penalty parameter ( $d$ )	-1	4

parameters has been evaluated in order to determine the optimum point, by having first searched for the parameters, which minimize the average error.

The implementation of the new hybrid WOA/MARS-based model has been performed using a multivariate adaptive regression splines (MARS) method, based on information obtained from the Earth library [54] together with the WOA technique with the *MetaheuristicOpt* package [20, 52] from the R Project. Additionally, the Ridge, Lasso, and Elastic-net regression models were implemented by using the *glmnet* package [55].

The bounds (initial ranges) of the space of solutions used in the WOA technique are shown in Table 3. A population of 40 whales has been used in the WOA

**Table 4** Optimal hyperparameters of the best fitted MARS model found with the WOA technique in this investigation for the training set

Hyperparameters	Optimal values
MaxFuncs	56
Interactions	2
Penalty ( $d$ )	1

optimization. The stopping criteria were the number of iterations along with at least 5 iterations with the same results. A total of fifty iterations were performed.

To optimize the MARS parameters, the WOA module is used as it searches for the best Maxfuncs, Interactions, and Penalty parameters by comparing the cross-validation error in every iteration. The search space is organized into three dimensions, one for each parameter. The main fitness factor or objective function is the coefficient of determination  $R^2$ .

### 3 Analysis of results and discussion

All of the eighty independent input variables (eighty physico-chemical variables) are indicated above in Tables 1 and 2. The total number of samples used in the present study was 21,263, which is to say that it has built and treated data from 21,263 experimental samplings. This entire dataset was split into two approximate halves and one was used as a training set while the other was used as the testing set. As the training set still contained a very large number of samples, 1000 samples were randomly extracted and the hyperparameter tuning was performed using tenfold cross-validation. Once the optimal parameters were determined, a model was constructed with the whole training dataset, which served as model validation using the testing dataset.

Based on this methodology, Table 4 identifies the optimal parameters of the best fitted MARS-relied approach that were encountered using the WOA optimizer.

Table 5 shows a list of 32 main basis functions for fitted WOA/MARS-based model and their coefficients, respectively. Note that  $h(x) = x$  if  $x > 0$  and  $h(x) = 0$  if  $x \leq 0$ . Therefore, the MARS model can be seen as an extension of linear models that automatically model nonlinearities and interactions as a weighted sum of the basis functions called *hinge functions* [15–19].

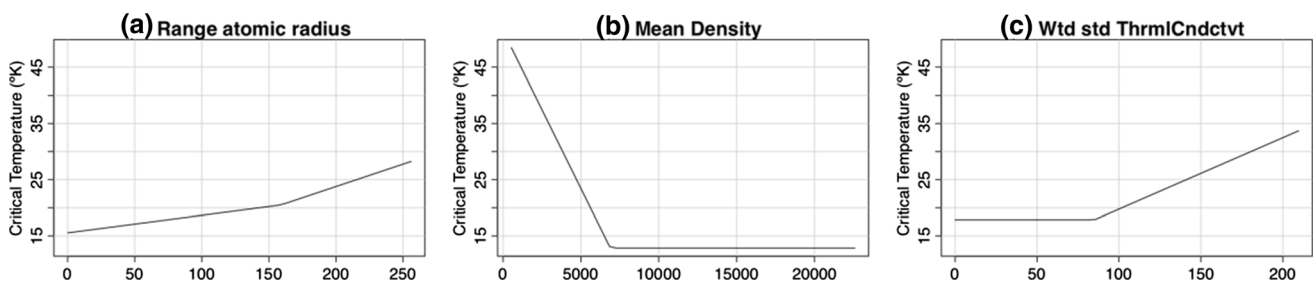
A pictorial graph of the first-order and second-order terms that create the MARS-based approach for the superconducting critical temperature  $T_c$  is shown in Figs. 3 and 4, respectively.

Based on the resulting calculations, the WOA/MARS-based technique allowed for the construction of a model

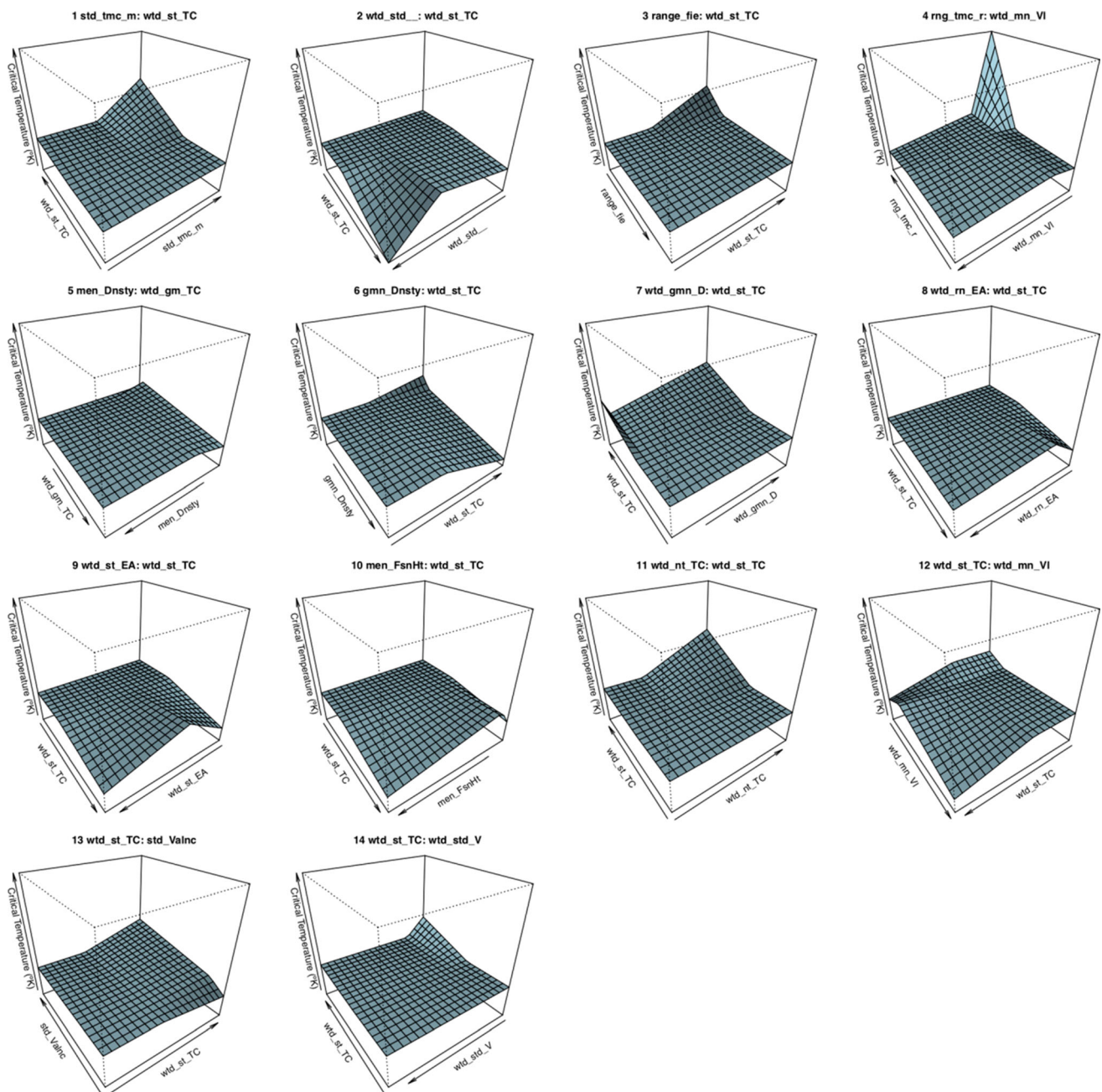


**Table 5** List of basis functions of the best fitted WOA/MARS-based model for the superconducting critical temperature ( $T_c$ ) and their coefficients  $c_i$

$B_i$	Definition	$c_i$
$B_1$	1	8.2954365
$B_2$	$h(159\text{-range\_atomic\_radius})$	-0.0310443
$B_3$	$h(\text{range\_atomic\_radius}-159)$	0.1093384
$B_4$	$h(6889.5\text{-mean\_Density})$	0.0056201
$B_5$	$h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	0.3020147
$B_6$	$h(54.1925\text{-std\_atomic\_mass}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	0.0014886
$B_7$	$h(\text{std\_atomic\_mass}-54.1925) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	0.0263762
$B_8$	$h(64.2578\text{-wtd\_std\_atomic\_mass}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0080287
$B_9$	$h(\text{wtd\_std\_atomic\_mass}-64.2578) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0379316
$B_{10}$	$h(310.6\text{-range\_fie}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	0.0027402
$B_{11}$	$h(\text{range\_fie}-310.6) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0001969
$B_{12}$	$h(\text{range\_atomic\_radius}-159) \times h(\text{wtd\_mean\_Valence}-2.26857)$	-0.0748170
$B_{13}$	$h(\text{range\_atomic\_radius}-159) \times h(2.26857\text{-wtd\_mean\_Valence})$	3.1894155
$B_{14}$	$h(6889.5\text{-mean\_Density}) \times h(\text{wtd\_gmean\_ThermalConductivity}-6.09074)$	-0.0000326
$B_{15}$	$h(6889.5\text{-mean\_Density}) \times h(6.09074\text{-td\_gmean\_ThermalConductivity})$	-0.0009077
$B_{16}$	$h(2006.63\text{-gmean\_Density}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	0.0002240
$B_{17}$	$h(\text{gmean\_Density}-2006.63) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0000403
$B_{18}$	$h(79.0562\text{-wtd\_gmean\_Density}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	0.0078285
$B_{19}$	$h(\text{wtd\_gmean\_Density}-79.0562) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	0.0000419
$B_{20}$	$h(46.9714\text{-wtd\_range\_ElectronAffinity}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0138414
$B_{21}$	$h(\text{wtd\_range\_ElectronAffinity}-46.9714) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0016589
$B_{22}$	$h(60.1526\text{-wtd\_std\_ElectronAffinity}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0153485
$B_{23}$	$h(\text{wtd\_std\_ElectronAffinity}-60.1526) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0071685
$B_{24}$	$h(8.6244\text{-mean\_FusionHeat}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0617544
$B_{25}$	$h(\text{mean\_FusionHeat}-8.6244) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	-0.0069835
$B_{26}$	$h(0.534908\text{-wtd\_entropy\_ThermalConductivity}) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	0.0865304
$B_{27}$	$h(\text{wtd\_entropy\_ThermalConductivity}-0.534908) \times h(\text{wtd\_std\_ThermalConductivity}-85.1085)$	1.1084331
$B_{28}$	$h(\text{wtd\_std\_ThermalConductivity}-85.1085) \times h(\text{wtd\_mean\_Valence}-2.38385)$	-0.1650073
$B_{29}$	$h(\text{wtd\_std\_ThermalConductivity}-85.1085) \times h(2.38385\text{-wtd\_mean\_Valence})$	-0.5945354
$B_{30}$	$h(\text{wtd\_std\_ThermalConductivity}-85.1085) \times (\text{std\_Valence}-0.433013)$	0.2547686
$B_{31}$	$h(\text{wtd\_std\_ThermalConductivity}-85.1085) \times h(0.433013\text{-std\_Valence})$	-0.8776123
$B_{32}$	$h(\text{wtd\_std\_ThermalConductivity}-85.1085) \times h(0.515047\text{-wtd\_std\_Valence})$	1.3096580



**Fig. 3** Representation of the first-order terms of the three more important input independent variables for the dependent superconducting critical temperature  $T_c$  variable: **a**  $T_c$  vs. Range atomic radius; **b**  $T_c$  vs. Mean density; and **c**  $T_c$  vs. Weighted standard deviation Thermal Conductivity



**Fig. 4** Representation of the second-order terms of the more important input independent variables for the dependent critical temperature  $T_c$  variable

with high allowances to assess the critical temperature  $T_c$  by means of the test dataset. Additionally, the Ridge, Lasso, and Elastic-net regression models were also built for the  $T_c$  output variable in order to predict the superconducting critical temperature of the superconductor state for different types of materials. Table 6 shows the determination and correlation coefficients ( $R^2$  and  $r$ ), root-mean-square error ( $RMSE$ ), and mean absolute error ( $MAE$ ) over

the testing set for the WOA/MARS, Ridge, Lasso, and Elastic-net models for the dependent  $T_c$  variable.

### 3.1 Significance of variables

Another important result of the current study is the relevance of the independent input variables in order to predict the superconducting critical temperature  $T_c$  for this non-linear complex problem (see Table 7 and Fig. 5).

**Table 6** Coefficients of determination ( $R^2$ ), correlation coefficients ( $r$ ), root-mean-square deviation (RMSE) and mean absolute error (MAE) over the testing set for the models fitted (WOA/MARS, Ridge, Lasso and Elastic-net) in this study using the training set

Error measure	WOA/MARS	Ridge	Lasso	Elastic-net
$R^2$	0.8005	0.6936	0.7295	0.7291
$r$	0.8950	0.8334	0.8541	0.8539
RMSE	15.14	18.77	17.64	17.65
MAE	10.75	14.50	13.43	13.44

**Table 7** Relative importance of the input physico-chemicals variables involved in the best fitted WOA/MARS-based model for the superconducting critical temperature  $T_c$  prediction according to criteria Nsubsets, GCV, and RSS

Input variable	Nsubsets	GCV	RSS
wtd_std_ThermalConductivity	31	100.0	100.0
std_atomic_mass	30	57.9	58.0
range_atomic_radius	30	57.9	58.0
wtd_mean_Valence	30	57.9	58.0
gmean_Density	29	42.0	42.2
wtd_entropy_ThermalConductivity	28	32.8	33.0
wtd_std_ElectronAffinity	27	27.2	27.5
mean_Density	26	25.4	25.6
wtd_range_ElectronAffinity	25	23.2	23.5
std_Valence	24	22.1	22.4
wtd_gmean_ThermalConductivity	23	21.0	21.3
wtd_std_Valence	20	17.9	18.1
wtd_std_atomic_mass	17	14.4	14.7
range_fie	14	11.9	12.2
wtd_gmean_Density	12	9.9	10.2
mean_FusionHeat	11	8.7	9.0

Ultimately, the most relevant input variable according to WOA/MARS approach in the  $T_c$  forecasting is Weighted Standard Deviation Thermal Conductivity. The second most significant input variable is Standard Deviation Atomic Mass, followed by: Range Atomic Mass, Weighted Mean Valence, Geometric Mean Density, Weighted Entropy Thermal Conductivity, Weighted Standard Electron Affinity, Mean Density, Weighted Range Electron Affinity, Standard Valence, Weighted Geometric Mean Thermal Conductivity, Weighted Standard Valence, Weighted Standard Atomic Mass, Range First Ionization Energy, Weighted Geometric Mean Density and Mean Fusion Heat.

We found that the most influential attributes were related to thermal conductivity. This is to be expected as both superconductivity and thermal conductivity are driven by lattice phonons and electrons transitions [8]. Also, the influence of ionic properties (related to the first ionization

energy and electron affinity) could likely reflect the capability of superconductors to form ions, which is related to the movement through the crystalline lattice. This interpretation aligns well with BCS theory of superconductivity [2]. The knowledge of the physico-chemical features that are more directly related to the critical temperature can facilitate the study of superconducting materials.

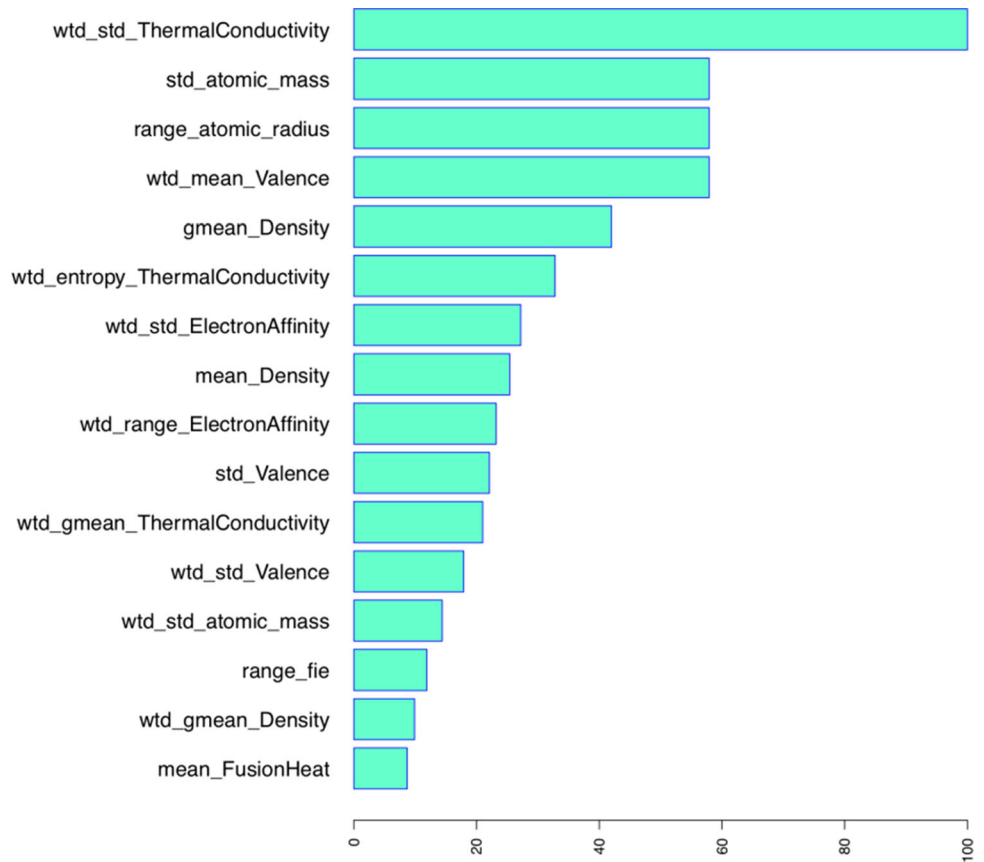
Overall, the MARS-based technique has demonstrated itself to be an extremely accurate and highly satisfactory tool to indirectly assess the superconducting critical temperature  $T_c$  (dependent variable), conforming to the real observed data in this study, as a function of some main measured physico-chemical parameters. Specifically, Fig. 6 indicates the comparison between the experimental and predicted  $T_c$  values employing the WOA/MARS, Ridge, Lasso, and Elastic-net regression models for the test dataset. Thus, it is essential to combine the MARS methodology with the WOA optimizer to overcome this nonlinear regression problem through a novel hybrid approach that is significantly more robust and more effective than the three remaining regression models. In particular, the modeled and measured  $T_c$  values were found to be highly correlated. Table 8 shows the  $T_c$  observed and predicted for the first materials in Fig. 6.

## 4 Conclusion

Based on the abovementioned results, several core discoveries of this study can be drawn:

- Existing analytical models to predict the superconducting critical temperature  $T_c$  from the observed values are not accurate enough as they make too many simplifications of a highly nonlinear and complex problem. Consequently, the use of machine learning methods such as the novel hybrid WOA/MARS-based approach employed in this study offer the best option for making accurate estimations of the  $T_c$  from experimental samplings.
- The hypothesis that the identification of  $T_c$  can be determined with precision by employing a hybrid

**Fig. 5** Relative importance of the input physico-chemical variables to predict the critical temperature  $T_c$  for the optimized fitted WOA/MARS-relied model according to the *GCV* parameter criterion



WOA/MARS-based approach in a wide variety of superconductors has been successfully validated here.

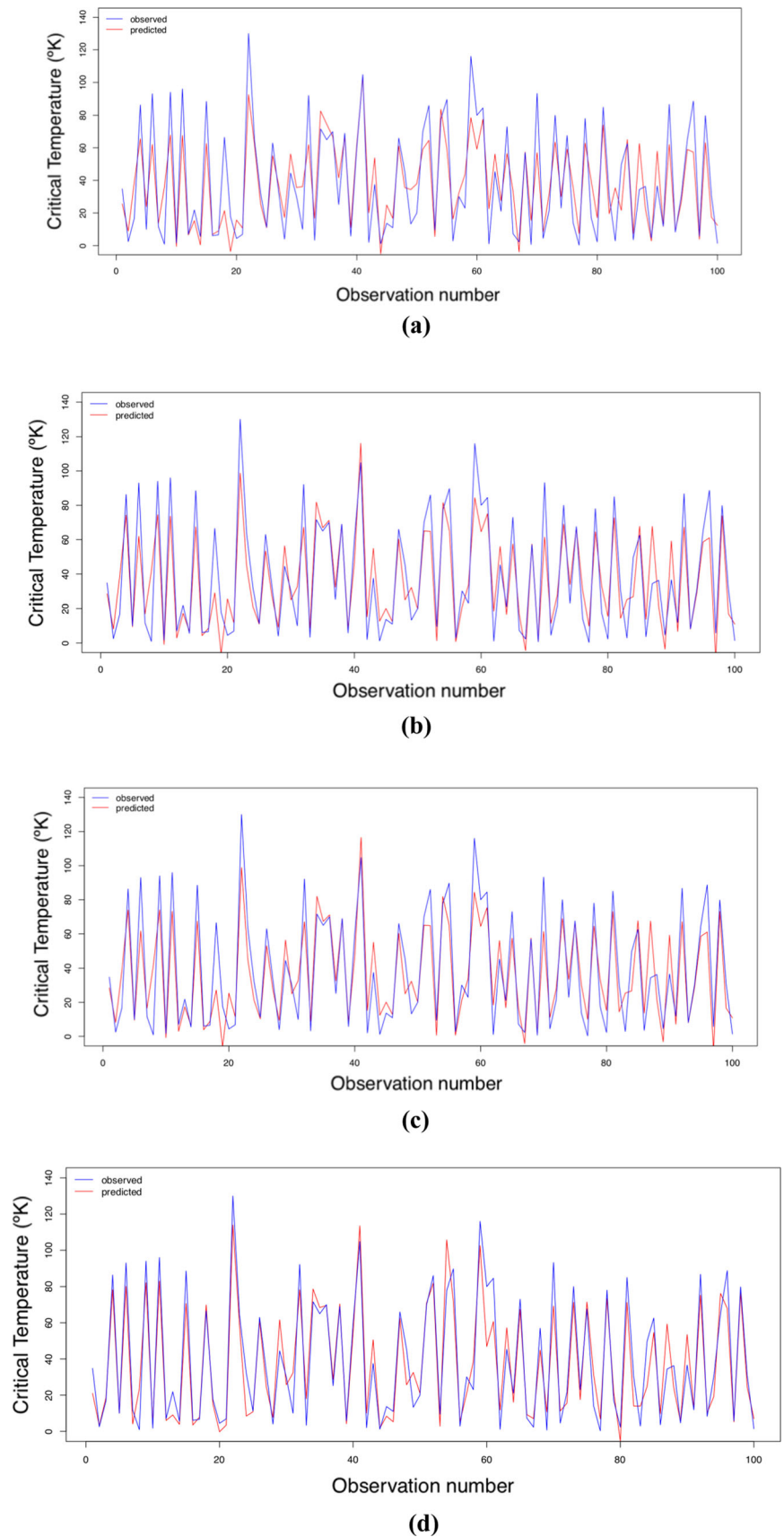
- The application of this MARS-based methodology to the complete experimental dataset belonging to the  $T_c$  resulted in a satisfactory coefficient of determination and correlation coefficient whose values were 0.8005 and 0.8950, respectively.
- The ranking according to the order of importance of the input variables entailed in the estimation of the  $T_c$  from experimental samplings in different superconductors has been established. Specifically, Weighted Standard Thermal Conductivity has been identified as the single most important factor in predicting critical temperature  $T_c$ . It is also important to note the successive order of importance, which is as follows: the Standard Atomic Mass, Atomic Range Radius, Weighted Mean Valence, Geometric Mean Density, Weighted Entropy Thermal Conductivity, Weighted Standard Electron Affinity, Mean Density, Weighted Range Electron Affinity, Standard Valence, Weighted Geometric Mean Thermal Conductivity, Weighted Standard Valence, Weighted

Standard Atomic Mass, Range First Ionization Energy, Weighted Geometric Mean Density and Mean Fusion Heat in the obtained  $T_c$  outcome.

- The principal role of the accurate hyperparameter determination in the MARS-based methodology in relation to the regression performance carried out for critical temperature  $T_c$  has been established using the WOA optimizer.

In conclusion, this procedure can be applied to successfully predict the superconducting critical temperature  $T_c$  of a variety of superconductors; however, it remains essential to consider the different physico-chemical features of each superconductor and/or experiment. Hence, the WOA/MARS-based method proves to be an extremely robust and useful answer to the nonlinear problem of the estimation of the  $T_c$  from experimental samplings in different superconductors. Researchers interested in finding high temperature superconductors may use the model to narrow their search. As a future extension of this work, we intend to apply the presented methodology to a more extensive database [43]. For instance, researchers could use

**Fig. 6** Observed vs. predicted superconducting critical temperature  $T_c$  values using 100 samples from the testing dataset for four different models: **a** Ridge regression model ( $R^2 = 0.6936$  and  $r = 0.8334$ ); **b** Lasso regression model ( $R^2 = 0.8005$  and  $r = 0.8541$ ); **c** Elastic-net regression model ( $R^2 = 0.7291$  and  $r = 0.8531$ ); and **d** WOA/MARS-relied model ( $R^2 = 0.8005$  and  $r = 0.8950$ )





**Table 8**  $T_c$  observed and predicted for some of the first materials in Fig. 6 for WOA/MARS model

Material	Observed $T_c$ (°K)	Predicted $T_c$ (°K)
Mg <sub>1</sub> B <sub>1.94</sub> C <sub>0.06</sub>	34.8	21
Y <sub>1</sub>	2.5	3
Sm <sub>1.25</sub> Gd <sub>0.6</sub> Ce <sub>0.15</sub> Cu <sub>1</sub> O <sub>3.97</sub>	16.8	18.6
Bi <sub>2</sub> Sr <sub>2</sub> Ca <sub>1</sub> Cu <sub>2</sub> O	86.3	78.2
Li <sub>1.4</sub> Zr <sub>1</sub> Ni <sub>1</sub> Cl <sub>1</sub>	10	10
Gd <sub>1.1</sub> Ba <sub>1.9</sub> Cu <sub>3</sub> O <sub>7</sub>	93	80.1
Pb <sub>1</sub> Mo <sub>6</sub> S <sub>6</sub> O <sub>2</sub>	11.7	4.1
Zr <sub>41.2</sub> Ti <sub>13.8</sub> Cu <sub>12.5</sub> Ni <sub>10</sub> Be <sub>22.5</sub>	0.9	23.6
Hg <sub>1</sub> Ba <sub>2</sub> Cu <sub>0.99</sub> Mg <sub>0.01</sub> O <sub>4</sub>	94	82
Pd <sub>0.95</sub> Pt <sub>0.05</sub> Te <sub>2</sub>	1.71	4.4
Hg <sub>0.9</sub> Ba <sub>2</sub> Cu <sub>1.05</sub> O <sub>4</sub>	96	82.9
Sn <sub>0.05</sub> Fe <sub>1</sub> Se <sub>0.93</sub>	7	6
Sr <sub>1</sub> Fe <sub>1.75</sub> Rh <sub>0.25</sub> As <sub>2</sub>	21.9	9.1
Nb <sub>0.29</sub> Re <sub>0.71</sub>	5.6	3.8
Y <sub>1</sub> Ba <sub>1.6</sub> Sr <sub>0.4</sub> Cu <sub>3</sub> O <sub>6.8</sub>	88.5	70.7
Mo <sub>0.865</sub> Re <sub>0.135</sub>	6.1	3.5
Lu <sub>4</sub> Sc <sub>1</sub> Ir <sub>4</sub> Si <sub>10</sub>	6.64	7.8
Y <sub>0.8</sub> Pr <sub>0.2</sub> Ba <sub>23</sub> O <sub>7.6</sub>	66.5	69.8
Na <sub>1</sub> Fe <sub>0.99</sub> Co <sub>0.01</sub> As <sub>1</sub>	17.8	15.3

this dataset along with new data (such as pressure or crystal structure) to make better models.

**Acknowledgements** The authors gratefully recognize the computational help supplied by the Department of Mathematics at the University of Oviedo as well as financial assistance from the Research Projects PGC2018-098459-B-I00 and FC-GRUPIN-IDI/2018/000221, both of which are partially financed by European Funds (FEDER). Likewise, the authors would like to express their gratitude to Anthony Ashworth for his English revision of this research paper.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

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

## References

- Ashcroft NW (2003) Solid state physics. Thomson Press Ltd, Delhi
- Tinkham M (2004) Introduction to superconductivity. Dover Publications, New York
- Kittel C (2005) Introduction to solid state physics. Wiley, New York
- Annett JF (2004) Superconductivity, superfluids, and condensates. Oxford University Press, Oxford
- Poole CP Jr, Prozorov R, Farach HA, Creswick RJ (2014) Superconductivity. Elsevier, Amsterdam
- Abrikosov AA (2017) Fundamentals of the theory of metals. Dover Publications, New York
- Hamidieh K (2018) A data-driven statistical model for predicting the critical temperature of a superconductor. *Comput Mat Sci* 154:346–354
- Huebener RP (2019) Conductors, semiconductors, superconductors: an introduction to solid-state physics. Springer, Berlin
- Matthias BT (1955) Empirical relation between superconductivity and the number of electrons per atom. *Phys Rev* 97:74–76
- Riaz M, Hashmi MR (2019) Linear diophantine fuzzy set and its applications towards multi-attribute decision-making problems. *J Intell Fuzzy Syst* 37:5417–5439
- Riaz M, Garg H, Farid HMA, Chinram R (2021) Multi-criteria decision making based on bipolar picture fuzzy operators and new distance measures. *Comput Model Eng Sci* 127(2):771–800
- Riaz M, Naeem K, Chinram R, Iampan A (2021) Pythagorean m-polar fuzzy weighted aggregation operators and algorithm for the investment strategic decision making. *J Math* 2021(ID6644994):1–19
- Riaz M, Hashmi MR, Pamucar D, Chu Y (2021) Spherical linear diophantine fuzzy sets with modeling uncertainties in MCDM. *Comput Model Eng Sci* 126:1125–1164
- Riaz M, Hamid T, Afzal D, Pamucar D, Chu Y (2021) Multi-criteria decision making in robotic agri-farming with q-rung orthopair m-polar fuzzy sets. *PLoS ONE* 16(2):e0246485
- Friedman JH (1991) Multivariate adaptive regression splines. *Ann Stat* 19:1–141
- Sekulic SS, Kowalski BR (1992) MARS: A tutorial. *J Chemometr* 6:199–216
- Friedman JH, Roosen CB (1995) An introduction to multivariate adaptive regression splines. *Stat Methods Med Res* 4:197–217
- Hastie T, Tibshirani R, Friedman JH (2003) The elements of statistical learning. Springer, New York
- Zhang WG, Goh ATC (2013) Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Comput Geotech* 48:82–95
- Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67
- Gharehchopogh FS, Gholizadeh H (2019) A comprehensive survey: Whale Optimization Algorithm and its applications. *Swarm Evol Comput* 48:1–24
- Ebrahimgol H, Aghaie M, Zolfaghari A, Naserbegi A (2020) A novel approach in exergy optimization of a WWER1000 nuclear power plant using whale optimization algorithm. *Ann Nucl Energy* 145:107540
- Yildirim H, Özkale MR (2019) The performance of ELM based ridge regression via the regularization parameters. *Expert Syst Appl* 134:225–233
- Moreno-Salinas D, Moreno R, Pereira A, Aranda J, de la Cruz JM (2019) Modelling of a surface marine vehicle with kernel ridge regression confidence machine. *Appl Soft Comput* 76:237–250
- Melkumova LE, Shatskikh SY (2017) Comparing Ridge and LASSO estimators for data analysis. *Procedia Eng* 201:746–755

26. Spencer B, Alfandi O, Al-Obeidat F (2018) A refinement of Lasso regression applied to temperature forecasting. *Procedia Comput Sci* 130:728–735
27. Wang S, Ji B, Zhao J, Liu W, Xu T (2018) Predicting ship fuel consumption based on LASSO regression. *Transp Res D Transp Environ* 65:817–824
28. Al-Obeidat F, Spencer B, Alfandi O (2020) Consistently accurate forecasts of temperature within buildings from sensor data using ridge and lasso regression. *Future Gener Comput Syst* 110:382–392
29. Zhao H, Tang J, Zhu Q, He H, Li S, Jin L, Zhang X, Zhu L, Guo J, Zhang D, Luo Q, Chen G (2020) Associations of prenatal heavy metals exposure with placental characteristics and birth weight in Hangzhou Birth Cohort: Multi-pollutant models based on elastic net regression. *Sci Total Environ* 742:140613
30. Chou S-M, Lee S-M, Shao YE, Chen I-F (2004) Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. *Expert Syst Appl* 27:133–142
31. de Cos Juez FJ, Sánchez Lasheras F, García Nieto PJ, Suárez Suárez MA (2009) A new data mining methodology applied to the modelling of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women. *Int J Comput Math* 86:1878–1887
32. Álvarez Antón JC, García Nieto PJ, de Cos Juez FJ, Sánchez Lasheras F, Blanco Viejo C, Roqueñí Gutiérrez N (2013) Battery state-of-charge estimator using the MARS technique. *IEEE Trans Power Electron* 28:3798–3805
33. Chen M-Y, Cao M-T (2014) Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Appl Soft Comput* 22:178–188
34. Zhang W, Goh ATC, Zhang Y, Chen Y, Xiao Y (2015) Assessment of soil liquefaction based on capacity energy concept and multivariate adaptive regression splines. *Eng Geol* 188:29–37
35. Kisi O (2015) Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J Hydrol* 528:312–320
36. Vu DT, Tran X-L, Cao M-T, Tran TC, Hoang N-D (2020) Machine learning based soil erosion susceptibility prediction using social spider algorithm optimized multivariate adaptive regression spline. *Measurement* 164:108066
37. Kumar S, Rai B, Biswas R, Samui P, Kim D (2020) Prediction of rapid chloride permeability of self-compacting concrete using multivariate adaptive regression spline and minimax probability machine regression. *J Build Eng* 32:101490
38. Zheng G, Yang P, Zhou H, Zeng C, Yang X, He X, Yu X (2019) Evaluation of the earthquake induced uplift displacement of tunnels using multivariate adaptive regression splines. *Comput Geotech* 113:103099
39. Li DHW, Chen W, Li S, Lou S (2019) Estimation of hourly global solar radiation using multivariate adaptive regression spline (MARS)—a case study of Hong Kong. *Energy* 186:115857
40. García-Nieto PJ, García-Gonzalo E, Alonso Fernández JR, Díaz Muñoz C (2019) Modeling algal atypical proliferation using the hybrid DE-MARS-based approach and M5 model tree in La Barca reservoir: a case study in northern Spain. *Ecol Eng* 130:198–212
41. García-Nieto PJ, García-Gonzalo E, Bové J, Arbat G, Duran-Ros M, Puig-Bargues J (2017) Modeling pressure drop produced by different filtering media in microirrigation sand filters using the hybrid ABC-MARS-based approach, MLP neural network and M5 model tree. *Comput Electron Agr* 139:65–74
42. Wang T, Ma H, Liu J, Luo Q, Wang Q, Zhan Y (2021) Assessing frost heave susceptibility of gravelly soils based on multivariate adaptive regression splines model. *Cold Reg Sci Technol* 181:103182
43. Superconducting Material (SuperCon) Database (2021) National Institute for Materials Science (NIMS), Japan. <https://supercon.nims.go.jp/en>
44. Le TD, Noumeir R, Quach HL, Kim JH, Kim JH, Kim HM (2020) Critical temperature prediction for a superconductor: a variational Bayesian neural network approach. *IEEE T Appl Supercon* 30(4):1–5
45. Li S, Dan Y, Li X, Hu T, Dong R, Cao Z, Hu J (2020) Critical temperature prediction of superconductors based on atomic vectors and deep learning. *Symmetry* 12(262):1–13
46. Roter B, Dordevic SV (2020) Predicting new superconductors and their critical temperatures using machine learning. *Physica C* 575:1353689575
47. Dua D, Graff C (2019) UCI machine learning repository. School of Information and Computer Sciences, University of California, Irvine, CA, USA. <http://archive.ics.uci.edu/ml>
48. Freedman D, Pisani R, Purves R (2007) *Statistics*. W.W. Norton & Company, New York
49. Knaf GJ, Ding K (2016) *Adaptive regression for modeling nonlinear relationships*. Springer, Berlin
50. McClave JT, Sincich TT (2016) *Statistics*. Pearson, New York
51. Wasserman L (2003) *All of statistics: a concise course in statistical inference*. Springer, New York
52. Simon D (2013) *Evolutionary optimization algorithms*. Wiley, New York
53. Picard R, Cook D (1984) Cross-validation of regression models. *J Am Stat Assoc* 79:575–583
54. Milborrow S (2020) *Earth: multivariate adaptive regression spline models*, R Package, version 4.5.0, R Foundation for Statistical Computing, Vienna, Austria. <https://cran.r-project.org/web/packages/earth/index.html>. Accessed 11 Oct 2020
55. Friedman JH, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33:1–22

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