JOM
https://doi.org/10.1007/s11837-021-04695-x

Porous Metal Properties Analysis: A Machine Learning Approach

EDGAR AVALOS-GAUNA,[1,3] YUYUAN ZHAO,[2] LEON PALAFOX,[1] and PEDRO ORTIZ-MONASTERIO-MARTÍNEZ[1]

01.—Facultad de Ingeniería, Universidad Panamericana, Campus México, Augusto Rodin 498, 03920 Mexico City, CDMX, México. 02.—School of Engineering, The University of Liverpool, Liverpool L69 3GH, UK. 3.–e-mail: eavalos@up.edu.mx

In the past few decades, computer-aided techniques (i.e., numerical simulation) have complemented the research and development process (R&D) in material sciences. This approach is usually paired to experimental testing. Yet, both techniques have shown cost-efficiency disadvantages and are time consuming. Optimization algorithms like the ones used in machine learning have proven to be an alternative tool when dealing with lots of data and finding a solution. While the use of machine learning is a well-established technique in other research fields, its application in material science is relatively new. Material informatics provide a new approach to analyse materials such as porous metals by employing previous data sets. This article aims to study reliability to predict permeability and Forchheimers coefficient of lost carbonate sintering open-cell porous metal. The key features selected as predictors are porosity, pore size, and coordination number. A comparison among multiple linear regression, polynomial regression, random forest regressor and artificial neural network is revised.

INTRODUCTION

Studies of materials science and materials design have been linked to previous experiences and trial-error analysis since the beginning[1]. From these, empirical correlations had been obtained between processing conditions and desired performances[2]. By doing so, the main purpose was to obtain and enhance the material by modifying its known properties. This typical material selection approach is widely accepted to address product-level performance requirements. Many industries are enriched by new or improved multifunctional materials. Although they might have good properties, this does not necessarily mean that the materials have the right balance of properties needed for specific engineering applications[1]. For instance, in high-performance alloys and composite materials, maximum performance is often achieved within a specific range of properties with a small window of variance[3].

More recently, computational techniques have been included in evaluating the performance metrics necessary to support materials design[4]. This approach is known as numerical simulation (e.g., finite element models) and has been used during the past few decades to analyse different material properties. Yet, the principal issue in obtaining relevant properties from a numerical model is the correct selection of the main microstructure measures that have a higher influence on them. To solve this predicament, the selection is generally made based on experience[5]. Moreover, the time and expense required in any of these two approaches often limit the development of new materials[6]. Additionally, modern advances in experimental and computational sciences have generated vast amounts of data[7]. Consequently, new fast and reliable approaches are needed for data analysis of materials. One of the most promising and novel fields to deal with large amounts of data is materials informatics.

Materials informatics can be used for material analysis and developments[8]. This is due to the big data" generated by experimentation and simula-

(Received September 17, 2020; accepted April 15, 2021)

Published online: 10 June 2021

tions. Materials informatics offers unprecedented opportunities for new materials discovery with improved properties with the implementation of machine learning techniques[9]. Machine learning algorithms use statistical models and optimization algorithms to reveal patterns within data. Insights can be obtained through predictions or classifications depending on the chosen algorithm. The machine learning process is shown in Fig. 1. The advantage over traditional experimentation and simulations is that computers can often handle much larger and higher dimensional data in a more efficient manner[10].

Some applications of different algorithms implemented on material informatics can be found in the literature. For instance, Çeçen et al.[11] analysed polymeric fuelled cell microstructure properties. Principal component analysis and a multiple linear regression model were used to establish structure-property correlations. Correlations for diffusivity, tortuosity, and porosity exhibited higher accuracy than traditional models found in the literature. Tapia et al.[12] used a Gaussian model for porosity prediction on a metal-based additive manufacturing process. Porosity is a common defect that has been reported to occur in selective laser melting. In their study, the Gaussian model developed made use of laser power and scanning speed as input parameters to predict porosity. It was reported that low porosity can be predicted using the Gaussian model. Moreover, Khanzadeh et al.[13] implemented a different machine learning approach for porosity prediction on additive manufacturing pieces. In their study, direct laser deposition was considered as manufacturing technique. Their study focused on the classification of the melting pool thermal image streams for porosity obtention. Different classification techniques were tested such as support vector machines, principal component analysis, and K-nearest neighbours. It is reported that K-nearest neighbours (KNN) exhibited the best performance in classification of ill-structured melt pools for achieving low porosity in the final manufactured parts. Pardakhti et al.[14] studied the structural and chemical features of metallic organic frameworks (MOFs) and their correlation with methane absorption. The revised machine learning techniques were decision tree, Poisson regression, support vector machine (SVM), and random forest. Given the large dataset they had available, their training set only consisted of 8% of the total data. This was equal to 130,398 records. The random forest exhibited great accuracy for prediction (98%) and a lesser computational time than current techniques for absorption measurement.

Artificial neural networks (ANNs) are another viable machine learning technique. They have been widely used for supervised learning predictions. Moreover, Skinner and Broughton[15] described their wide applications for material sciences since early in this century. Dudsik and Strek[16], for instance, implemented ANNs when studying the strength properties of open-cell aluminium porous metal during compression. In their study, different ANN architectures were analysed in approximated stress-strain relations. To assess ANNs 'performance, different statistical metrics were used. The results showed that their approach provides an approximation of the mean absolute relative error (MARE) between 7 and 10%. Similarly, Altarazi et al.[17] tested an ANN to evaluate polyvinylchloride (PVC) composite properties. In their analysis, dif-

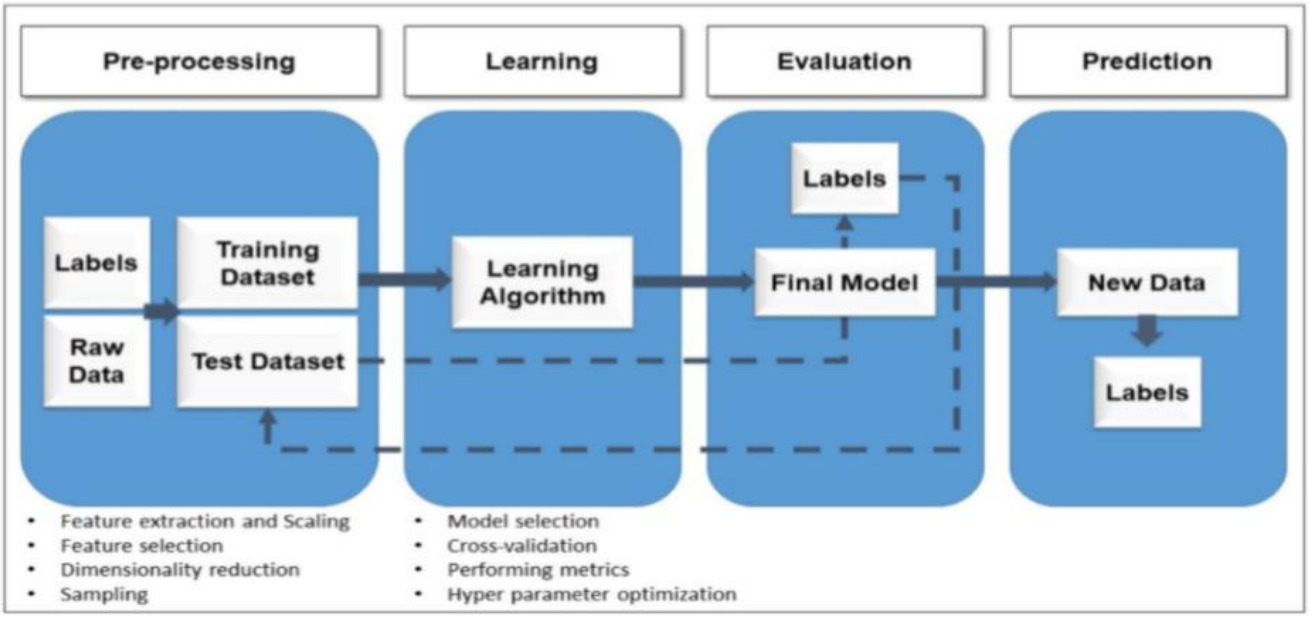

Fig. 1. Machine learning process.

ferent combinations of activation functions, network layers, and amounts of perceptrons per layer were used. The Levenberg-Marquardt back-propagation and radial basis training algorithms/activation functions were found to be the most appropriate for tensile strength, ductility, and density of extruded PVC composite predictions. In their study, the model accounts for weight percentages of virgin PVC, recycled PVC, CaCO3 filler, CaCO3 particle ductility, and density of extruded PVC.

METHODOLOGY

One of the key aspects of materials informatics is that it lowers the costs of R&D as it does not have to use many experimental tests and reduces the computational usage. In addition, the implementation time is faster than in the traditional way of doing research[4]. Therefore, the main purpose of this article is to demonstrate that data science tools are ideally suited to tackling the fluid flow problem of a porous copper geometry. The experimental results used in the current study can be found in Z. Xiao and Y. Zhao[18] and J. M. Baloyo and Y. Zhao[19]. The numerical analysis can be found in Edgar Avalos-Gauna[20]. The current analysis expands on the results obtained by numerical analysis and published by E. Avalos Gauna and Y. Zhao[21]. For the current work, the experimental and numerical results were coupled for machine learning purposes. To start with the machine-learning model, the following elements were implemented:

(1) Data Preparation

The first step is to generate a dataset that comprises the features (independent variables) and the targets or descriptors (dependant variables)[22]. This could come from experimental data results, numerical simulation, or, in this case, both entries (810 numerical results and 46 experimental tests). The reason for using two different sources of information is that the numerical records employed can represent the fluid flow behaviour observed during the experimental tests as stated in[20].The input variables and the target outcomes must be cleaned before any machine learning algorithm is implemented. The cleaning process includes looking for and dealing with missing data, eliminating non-relevant features, and an exploratory data analysis in most of the cases when using secondary data. This process was performed in both datasets. For the experimental data, average values were used where range values were available. The final dataset has 856 records for permeability and 826 for Forchheimers coefficient. Once all data are collected, there are two common approaches to bring different features onto the same scale: normalization and standardization. Normalization refers to rescaling all quantitative features to a range of (0, 1). Standardization on the other hand refers to rescaling all quantitative features to have a mean of 0 and a standard deviation of 1[23]. The procedure of normalization is expressed in Eq. 1:

$$x_{\text{norm}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where x_i represents a data value of the quantitative feature vector to be normalized. $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of this vector. This process can be done by the MinMaxScaler method from the python library sklearn, class preprocessing. For the standardization procedure, it can be achieved by implementing Eq. 2[24]:

$$x_{\text{std}} = \frac{x_i - \mu_x}{\sigma_x} \tag{2}$$

where μ_x and σ_x are the mean value and standard deviation of the quantitative feature vector to be standardized. This process can be done by the scale method, which can also be found in sklearn.preprocessing.

(2) Descriptor Selector

The next step is to define which material properties have to be selected as the key features that describe the performance of the material on a given aspect[25]. Descriptor or target selection consists of finding the most influential features by implementing different data extraction techniques and dimensionality reduction methods such as principal component analysis[6]. Depending on the number of elements in the material, the dimensionality of the input may differ. Hence, it is not possible to apply the same model to materials with different number of elements. Thus, an exploratory data analysis was performed to find correlations amongst all parameters available on the dataset. These correlations were ploted on a correlation matrix (Fig. 2). Positive relationships are marked in red and negative relationships are marked in blue. Stronger relatioships are marked with darker colors and weak relationships are marked with faded colors. In the current analysis, the predictor variables to be employed are: potassium carbonate diameter (K_2CO_3) (micrometres), porosity (f), copper particle diameter (micrometres), and coordination number.

(3) Model Selection and Evaluation

Extracting and validating the model, one must bear in mind that it must fit the data set and the type of data each feature has, without over- or under-fitting. Various regression methods are implemented. A thorough knowledge of the physics representing the problem is critical in selecting the most accurate machine learning technique[26]. Therefore, it is vital to understand the key features of a porous metal. Porous metals are a novel type of material with hollow spaces deliberately included into the material during its manufacturing process[27]. These hollow spaces or pores provide a unique set of properties to the material[28].

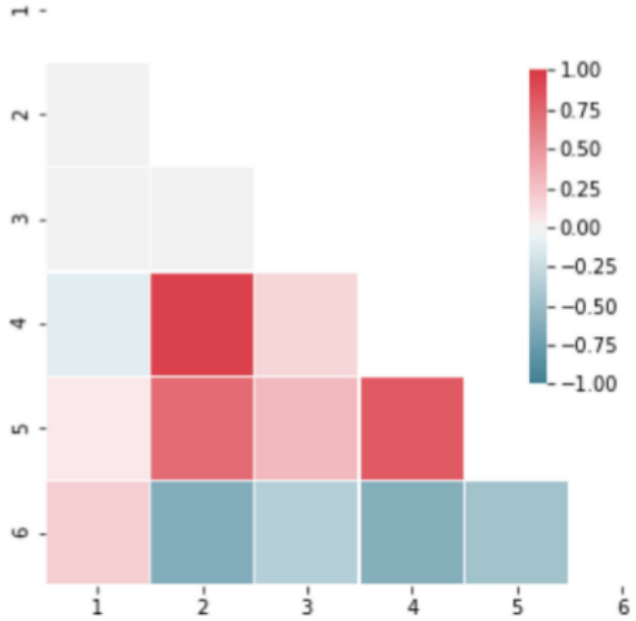

Fig. 2. Correlation matrix amongst the structural parameters: (1) K_2CO_3 diameter, (2) porosity, (3) Cu diameter, (4) coordination number, (5) permeability and (6) Forchheimer's coefficient..

For active cooling applications, the porous metal is expected to have an open-cell structure. The cooling system is composed of the porous metal medium and the fluid is used as a coolant flowing through the material. The literature indicates that the pressure drop across the sample is strongly affected by the pore structure[29], [30]. For the porous structure, many studies have analysed the flow problem within a porous media[31]. Fluid flow through porous metals can be considered turbulent if the Reynolds number increases to a critical value[32]. Thus, a nonlinear relationship is used to calculate the pressure drop due to the porous media. This relationship is known as Forchheimers equation[33]:

$$\frac{\Delta P}{\Delta L} = \frac{\mu v}{K} + \rho \geq v^2 \tag{3}$$

where ΔP is the pressure drop between the inlet and outlet of the porous media, ΔL is the length of the porous media, μ is the viscosity of the fluid, v is the Darcian velocity of the fluid (i.e., flow rate divided by the cross-sectional area), ρ is the density of the fluid (water in this case), K is the permeability of the porous media, and C is Forchheimers coefficient or form drag coefficient. K and C have often been the focus of investigations on porous media.

For the machine learning model selection, multiple linear regression, polynomial regression, random forest, and an ANN were tested to establish relationships between the input variables and the target outcomes. To do so, 80% of the data were set as the training subset and the remaining 20% was used as the testing subset. This process was achieved using the model_selection.train_test_split method from the scikit learn library. Once all data had been split, each model was trained using the first split of the data set. Each model was trained considering different algorithm parameters to achieve a better correlation coefficient. All models with their best chosen parameters were evaluated using ten-fold cross validation during the training part of the analysis.

(4) Model Implementation

As stated before, the most relevant features are obtained and then are used to make predictions of the target properties. This can be done by executing the trained model using new data (testing subset). During this phase, insights can be drawn from the predictions. This will derive from the development of specific types of materials or even tailored materials. For materials informatics, the literature designates five descriptor categories: constitutional, topological, physicochemical, structural, and quantum-chemical[25]. For this article, only the topological properties were used. Lastly, after the model has been implemented, it is crucial to evaluate its performance. This can be done by using statistical metrics such as Pearsons coefficient (R^2) or root mean square error (RMSE).

MACHINE LEARNING MODELS

For the current study, four different supervised machine learning techniques were employed. Multiple linear regression, polynomial regression, random forest regressor and ANN were tested to predict permeability and Forchheimers coefficient. As multiple linear regression is a particular case of the polynomial regression, only the later three will be explained.

(1) Polynomial Regression

This method has the advantage of being simple to understand, fast and straightforward to implement; given its simplicity, it is not adequate for complex problems[26]. The dependent variable (i.e.y) is assumed to be obtainable by evaluating a linear function of the explanatory variables (i.e. x). The linear regression is represented as[34]:

$$y = x^T \beta + \varepsilon \tag{4}$$

where x^T is the transpose vector of all explanatory variables, β is a vector of weights, which we must estimate, and ε is a vector with normally distributed zero mean, which represents random effects. Equation 4 is useful when the relationship between the predictors and target values is somehow linear. However, depending on the number (m) of elements on x^T, there are $2^m - 1$ models to test. Therefore, only those variables that contribute the most to predicting the target variable are considered.

Nonetheless, when the observed trend does not adjust well with this approach, it is possible to add

another term to the equation to account for the non-linear behaviour:

$$y = x_0^n \beta_0 + x_i^T \beta_i + \varepsilon \quad (5)$$

where x_i^T and $\boldsymbol{\beta}_i$ are still vectors, and $x_0^n \beta_0$ is the selected feature to display the nonlinear behaviour. In all cases of polynomial regression, the main objective is to find the parameters that minimise the loss function using Eq. (4):

$$\frac{\partial}{\partial x}\left(\frac{1}{n}\sum_{i}^{n}(y(x_i) - y_i)^2\right) = 0 \quad (6)$$

For the current study, 15 different combinations of the four numerical descriptors were tested, going from simple linear to combinations of two, three and all four descriptors. Results are shown in the following section.

(2) Random Forest (RF)

A widely used and effective technique in machine learning involves the use of learning models known as ensembles. An ensemble takes multiple individual learning models and combines them to produce an aggregate model that is more powerful than any of its individual learning models alone[35]. Random forest is a supervised machine learning ensemble that encompasses several decision trees for prediction or classification problems[9]. A decision tree represents a procedure where data are analysed based on their attributes[13]. It is a supervised learning algorithm that is mostly used for classification problems. It works for both discrete and continuous dependent variables. It involves a selection process that can be described as a sequence of binary selections[36]. By fitting several decision trees, a random forest uses various subsamples of the dataset and uses averaging to improve the predictive accuracy and control overfitting of the final prediction. The number of trees on any random forest model usually is around tens or hundreds of trees depending on data complexity. For the current analysis different numbers of trees were used, ranging from 50 to 2000. Additionally, a ten-fold test was performed.

(3) Artificial Neural Network (ANN)

ANNs are a novelty approach widely employed in recent years in applications such as pattern recognition and material science[26]. ANNs are inspired by how the nervous system, neural networks and neurons are composed. In reality, biological neurons are interconnected in a network structure, with each neuron having an electrical signal as input and later transmitting a response to a neighbouring neuron when a certain activation threshold is surpassed[37]. In ANNs, neurons are called perceptrons, and the back propagation method is often used to train the networks weights and biases for each perceptron with the sigmoid function as the most common activation function used[38]:

$$\theta(u) = \frac{1}{1+e^{-u}} \quad (7)$$

where u represents a linear function within a perceptron. An ANN is formed by layers of several numbers of perceptrons. Each perceptron on one layer is interconnected with the perceptrons from the next layer. These processing units are made up of one layer of inputs, a group of hidden layers, and one output layer. The final goal of the ANN is to learn about the information presented to produce one output report[39]. Thus, during the construction of network architecture, an optimal combination of these parameters is sought. With the inclusion of more perceptrons, more layers, or different activation functions, it is possible to increase the models accuracy; however, it might incur data overfitting. Therefore, the model should be complex enough to have good predictions and yet somehow simple enough to avoid overfitting.

For this process, an extensive grid search was carried out. Different combinations of the activation function (identity, ReLU, sigmoid), type of solver (lbfgs, sgd, adam), maximum number of iterations (2000, 2500), hidden layers (1 to 3), and neurons per layer (1 to 20) were tested. In total, 6174 different networks were analysed using a ten-fold analysis. The final ANN architecture used for the current study comprised five layers as shown in Fig. 3. The first layer is the input layer for all predictor variables (four features). Then, the following three layers are hidden layers for data processing with their corresponding activation function. The last layer corresponds to the output layer for the target prediction.

RESULTS AND DISCUSSION

A numerical dataset was employed to train different machine learning models to predict permeability and Forchheimers coefficient. The actual behaviour of both variables is shown in Fig. 4. The dataset was transformed in a way that can be processed by all machine learning techniques proposed in the current study. All records were analysed using normalized and standardized values. However, only standardized results are shown as the difference in accuracy between normalized and standardized values was negligible. Then, the dataset was divided into 80% for training and 20% for testing. Later, ten-fold cross validation was used on each technique considering the best chosen parameters to avoid overfitting.

Finally, the trained algorithms were compared against the testing set to measure accuracy. The target features for the current analysis are permeability and Forchheimers coefficient. Therefore, they were computed in two separate analyses.

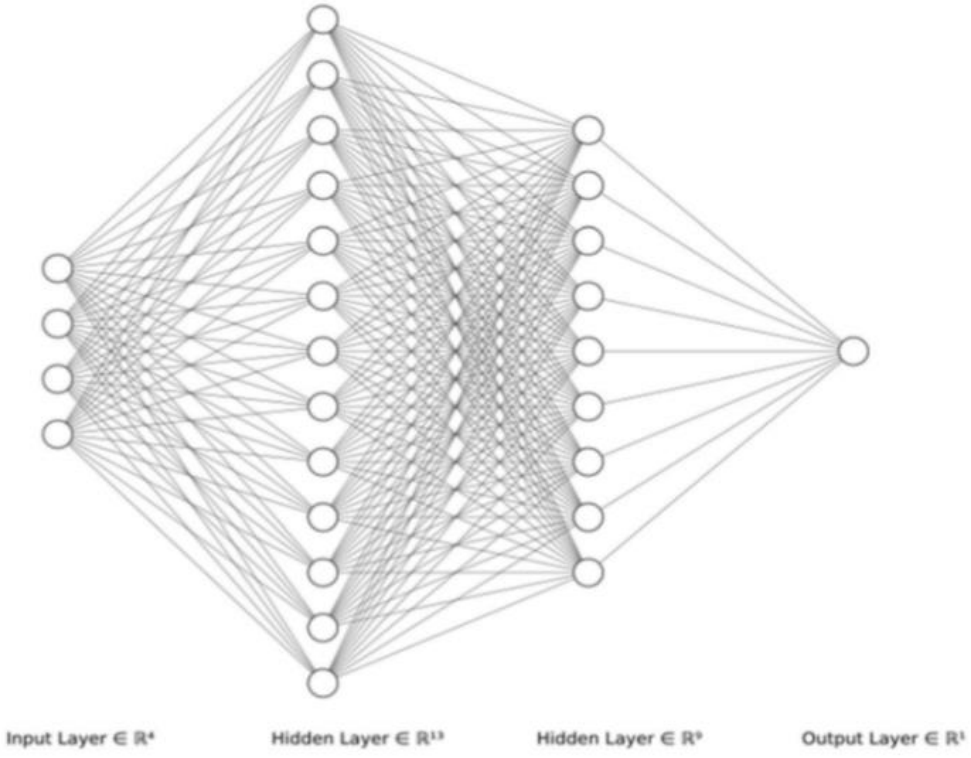

Fig. 3. Artificial Neural Network architecture used for the current study..

The multiple linear regression and polynomial regression analyses were performed using combinations of all four predictor variables. For that, an algorithm was created to generate different linear models that were tested to measure the correlation coefficient. The algorithm considered all possible combinations with the predictor variables. Each predictor variable and target outcome were labelled accordingly as shown in Eq. (8).

$$X = \begin{cases} x_0 : \text{Pore size} \\ x_1 : \text{Porosity} \\ x_2 : \text{Copper size} \\ \text{x}_3 : \text{Coordination number} \end{cases} \qquad Y = \begin{cases} y_0 : \text{Permeability} \\ y_1 : \text{Forchheimers coeff} \end{cases} \tag{8}$$

From all possible combinations, only the top three combination results are shown in Table I. The results showed that the multiple linear regression considering all four predictors obtained the highest correlation coefficients. For the permeability regression, the highest correlation achieved was 0.761 and for the Forchheimers coefficient was 0.625. These results agree well with the nonlinear behaviour of both variables as shown in Fig. 4 and Eq. (3). For the polynomial algorithm, different grees were employed to improve accuracy. One important aspect to keep in mind is to check for overfitting[36]. For the current study, the starting point was a degree 2 polynomial. It was observed that overfitting started to occur at a polynomial of degree 4 for both target values. Thus, it was decided to stop at a degree 3 polynomial.

Table I. Pearson coefficients for multiple linear regression combinations

Y	Combinations of X	R^2
y_0	x_0, x_1, x_2, x_3	0.761
y_0	x_0, x_1, x_3	0.756
y_0	x_0, x_2, x_3	0.723
y_1	$x_0 x_1, x_2, x_3$	0.625
y_1	x_0, x_1, x_2	0.572
y_1	x_1, x_2, x_3	0.547

The next step was to implement other models that could account for the nonlinearities. One of the chosen models was random forest. The only parameter that was modified to assess a better prediction accuracy was the number of estimators. By increasing the number of estimators (from 50 up to 2000), the correlation coefficient and the differences between folds stabilize as shown in Fig. 5. Yet, the final number of estimators was set to 175 as this provides good prediction accuracy (99.1%) and the difference between the ten folds was the smallest (1.61%).

For the ANN, several combinations of parameters were tested. The results for the top 10 combinations

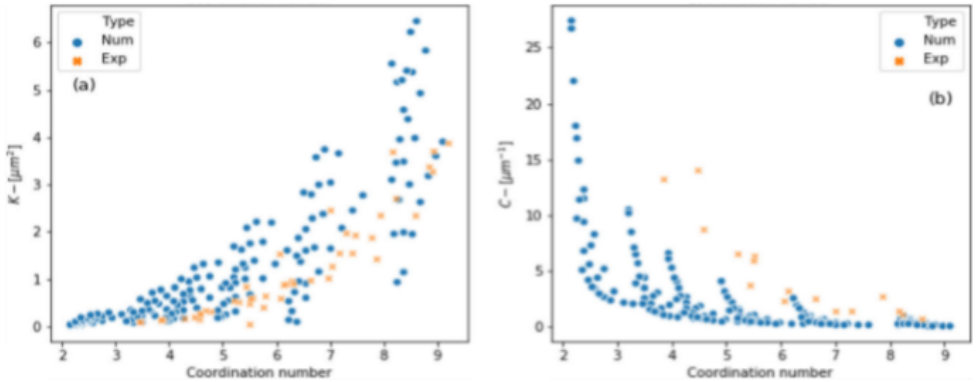

Fig. 4. Relationship between (a) permeability with the coordination number and (b) Forchheimers coefficient and coordination number..

Table II. ANN architecture results during the grid search process

Rank	Activation	Hidden_layer_sizes	Max_iter	Mean_test_score	Std_test_score	Hidden_layer_index
1	Logistic	(7, 12, 12)	2000	0.995879	0.00425	130
2	Logistic	(16,)	2000	0.995643	0.0034	245
3	Logistic	(7, 9)	2500	0.995469	0.00359	119
4	Logistic	(16, 3, 12)	2000	0.995389	0.00399	256
5	Logistic	(10, 15, 9)	2500	0.995368	0.00364	185
6	Logistic	(7, 15, 18)	2500	0.995315	0.00384	139
7	Logistic	(16,)	2000	0.995309	0.00363	245
8	ReLU	(13, 9, 9)	2500	0.995305	0.00401	220
9	ReLU	(16, 6, 12)	2000	0.995283	0.0045	263
10	Logistic	(16,)	2500	0.995268	0.00356	245

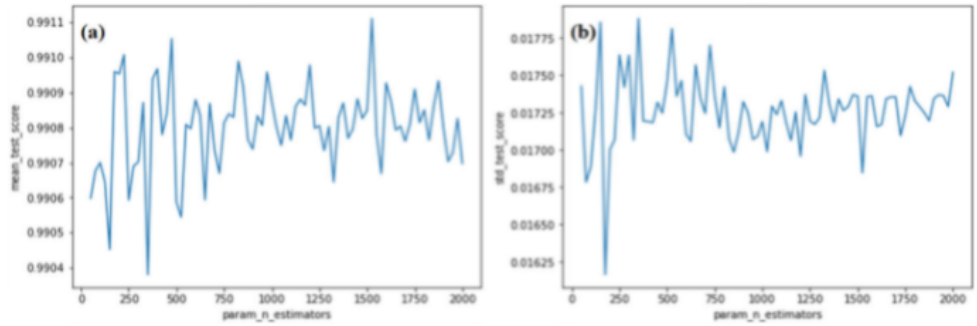

Fig. 5. Random forest (a) mean score prediction and (b) standard deviation of the ten-fold test vs. number of estimators..

from 6174 possible architecture combinations are shown in Table II. No further search was carried out as these achieved results are already over the 99% correlation score. The final selection for the ANN was ranked number 8. It was observed that ReLU activation function converged faster than the sigmoid. Additionally, the difference in correlation coefficient score compared to the other architectures was minimal.

A comparison of all four models with their selected parameters during the ten-fold training phase is shown in Fig. 6. The results showed variations during permeability calculations amongst all models. The largest variation was observed for the multiple linear regression whereas the minimum variation was obtained with the random forest regressor. Polynomial regression and ANN showed some outliers that could be related to the imbalance between numerical data and experimental data. Some data balancing techniques could be implemented to improve these results. For the Forchheimer coefficient, the training phase showed more consistency for all methods except for multiple linear regression. Later, all four methods were analysed using the testing subset.

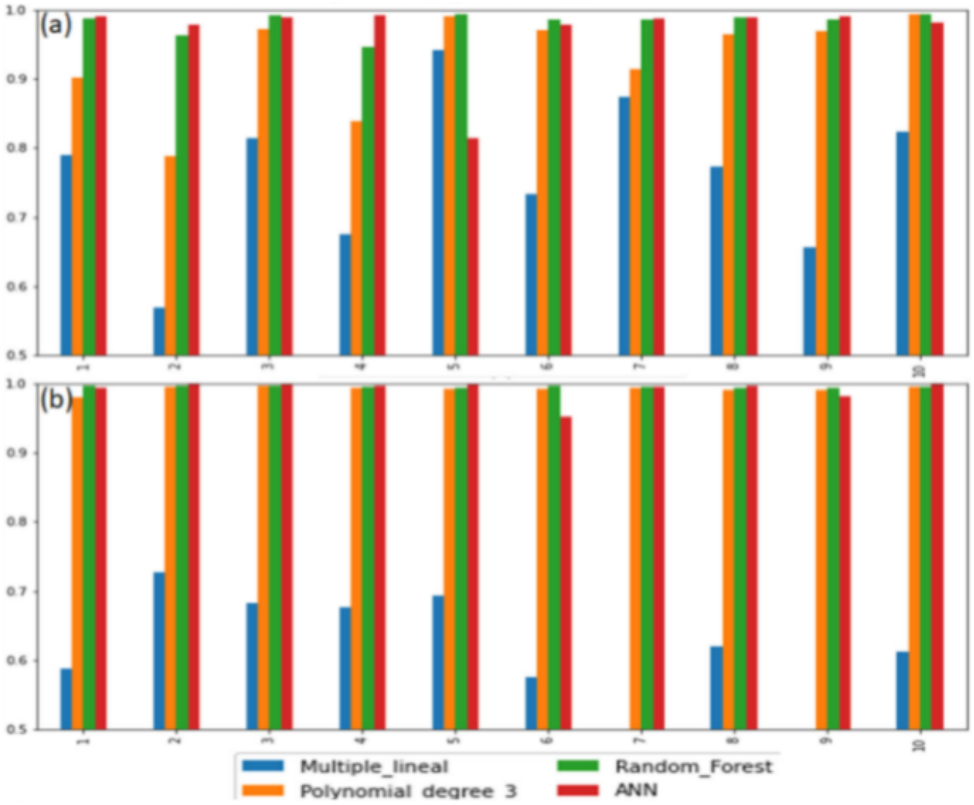

Fig. 6. Ten-fold analysis results during the training phase for (a) permeability and (b) Forchheimer's coefficient..

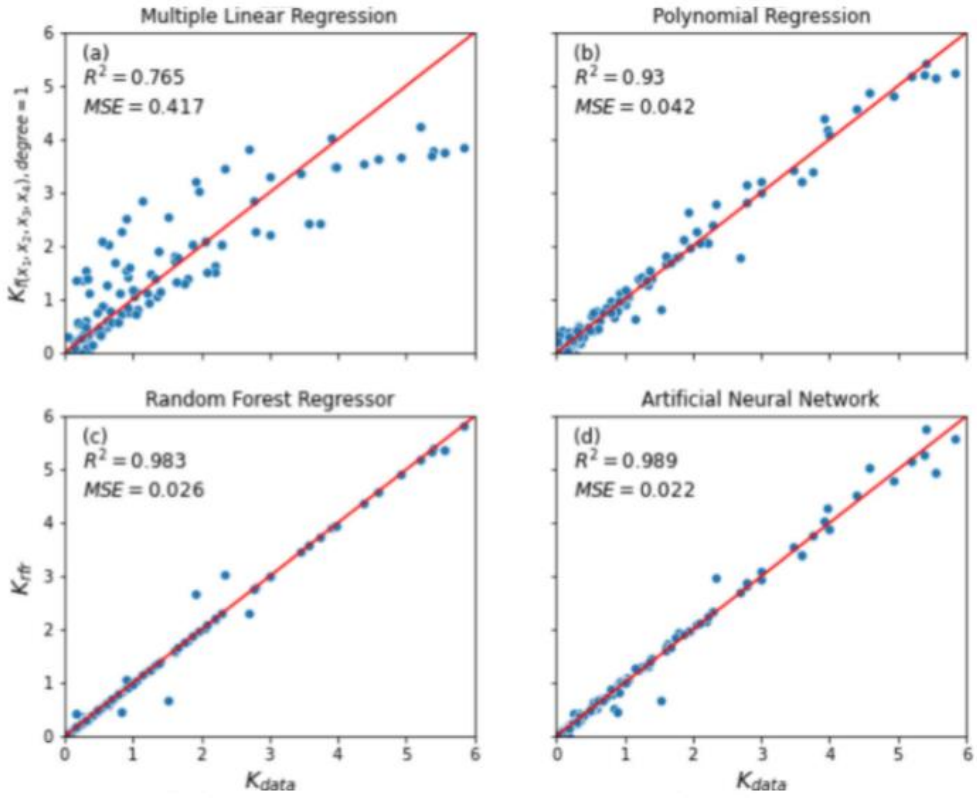

Fig. 7. Q-Q plot for permeability and predicted values using different machine learning techniques: (a) multiple linear regression, (b) polynomial regression, (c) random forest regressor, (d) artificial neural network..

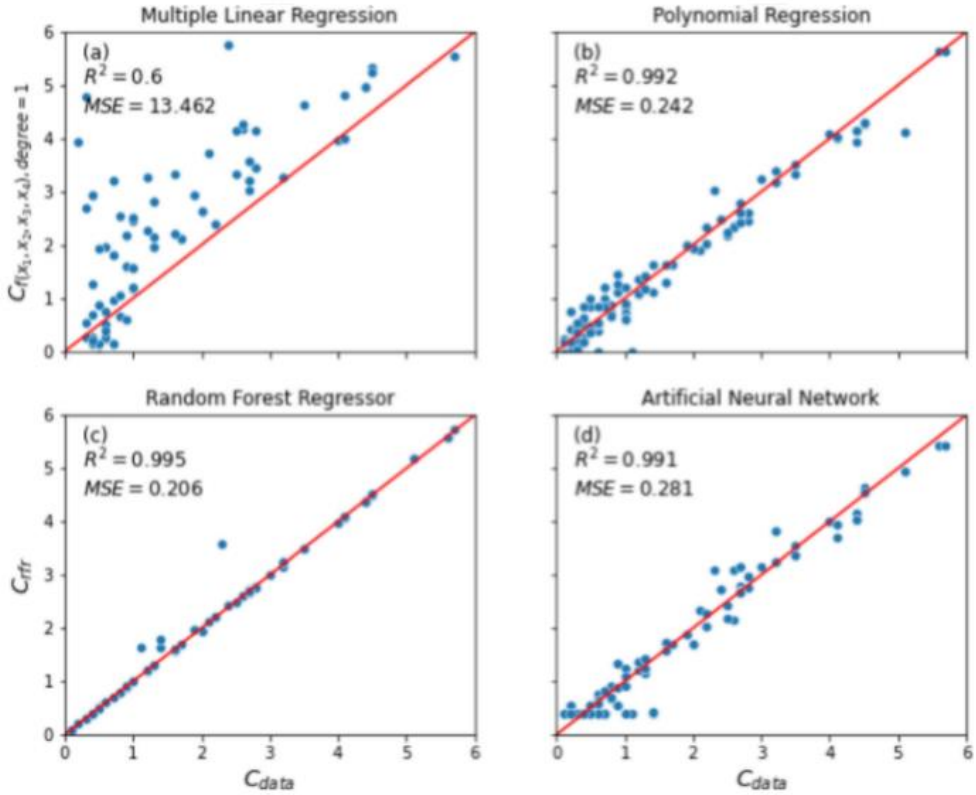

Fig. 8. Q-Q plot for Forchheimers coefficient and predicted values using different machine learning techniques: (a) multiple linear regression, (b) polynomial regression, (c) random forest regressor, (d) artificial neural network..

Q-Q plots were designed to show how well each model performed when trying to predict K and C during the testing phase. The X-axis corresponds to the real values and the Y-axis is the predicted value of each model. Figure 7a shows a multiple linear regression of degree 1 between the structural parameters and permeability. The Q-Q plot shows some correlation between the predictors and target outcome. Yet, the linear model only successfully predicted 76% of the expected values for permeability during the testing phase. Additionally, the observed mean squared error (MSE) was 0.417. By increasing the polynomial degree it was possible to achieve better average accuracy (93%) and an almost ten times lower MSE. However, during the training phase, the ten-fold test for polynomial regression exhibited a wide range between the less optimum and the top result of the analysis going from 78.7% to 99.3% respectively. The next step that followed was the analysis of the data with a random forest regressor, i.e. Fig. 7c. The random forest algorithm was set to use 175 decision trees with a minimum split of 2. The average testing accuracy obtained for this model was 98.3% with an MSE 0.026. The mean score and standard deviation during training for the ten-fold analysis for the random forest regressor were 98.253% and 1.548% respectively, meaning that all tests obtained a similar accuracy. Finally, the predictor variables were used on an ANN. Different activation functions were tested but, in the end, regular linear was used on each of the three hidden layers. The results for that analysis are given in Fig. 7d. This technique averaged the best testing accuracy to predict permeability amongst the other four techniques (98.9%). Finally, the MSE was slightly lower than the random forest regressor. Moreover, during the ten-fold training phase, ANNs exhibited an accuracy close to 96.91% and, overall, the standard deviation was 5.47%. Thus, random forest and ANN were suitable for permeability predictions. Yet, random forest tends to overfit the data (this might explain the consistency during training) while neural networks keep learning when new data become available.

For the Forchheimers coefficient prediction, similar results were observed. During the testing phase the multiple linear regression model only predicted 60% correctly. Fig. 8a shows how dispersed these results are. The MSE was also quite large, i.e. 13.462, meaning that the difference in prediction for some points was huge. This behaviour was consistent with what was observed during the ten-fold training phase (61.5% for the mean score with a standard deviation of 10.95%). After implementing the other machine learning techniques, an improvement in both metrics (accuracy and MSE) was observwed. All three cases showed accuracy of 99%, but the random forest regressor had the best

prediction accuracy (99.5%). For the MSE, the ANN (Fig. 8d) had more difference in errors with the predicted values compared to random forest and polynomial regression. However, the ten-fold training phase showed differences amongst machine learning techniques. For instance, the ANN minimum accuracy rate was 95.25%. This led to a standard deviation for the ten-fold training phase for ANN of 1.46%. This could imply that the ANN is not learning properly from the dataset and some parameter tuning might need to be used. Only the random forest and polynomial regression models had similar results throughout their corresponding ten-fold training phase: 99.3% and 0.47% averaged accuracy and standard deviation for the polynomial regression and 99.52% and 0.14% averaged accuracy and standard deviation for the random forest regressor. Overall, the results showed a big improvement from the multiple linear regression with all three algorithms for both target variables (K and C). Moreover, it was possible to correlate the structural parameters of the porous metal to target values that in the literature are often obtained mainly through experimental means.

CONCLUSION

This article presented a way to calculate the permeability and Forchheimer coefficient from data generated experimentally and numerically. The model implemented a correlation matrix to establish the main features to use for analysis, which were pore size, porosity, metal particle size and coordination number. With the implementation of multiple linear regression, it was possible to obtain the best correlation factor for the target properties by using a combination of four predictor features. The multiple linear regression was later used to assess the effectiveness of the machine learning approaches. With this combination of features, three different machine learning techniques were tested to improve the accuracy of prediction. The results showed that polynomial regression provided a big improvement over the multiple linear regression. Moreover, it accounted for most of the nonlinearities in the data. Additionally, the results suggest that the addition of a third-degree term improves the calculation of permeability and Forchheimers coefficient. For the other two machine learning techniques, random forest regressor performed best for both analyses. Although ANN had better results in some areas, it showed a lack of consistency in the ten-fold training phase. Further parameter tuning is suggested to compensate for the differences. In summary, this article successfully presented a novel approach for porous metal material property calculation. Random forest regressor (99.5%) and polynomial regression (99.2%) were the best adequate algorithms for Forchheimers coefficient whereas ANN (98.9%) had better results for predicting permeability.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. B.D. Conduit, N.G. Jones, H.J. Stone, and G.J. Conduit, *Scr. Mater.* 146, 82. (2018).
2. K. Matouš, M.G.D. Geers, V.G. Kouznetsova, and A. Gillman, *J. Comput. Phys.* 330, 192. (2017).
3. T. Lookman, P.V. Balachandran, D. Xue, J. Hogden, and J. Theiler, *Curr. Opin. Solid State Mater. Sci.* 21, 121. (2017).
4. J.H. Panchal, S.R. Kalidindi, D.L. McDowell, and C.A.D. Comput, *Aided Des.* 45, 4. (2013).
5. A. Gupta, A. Cecen, S. Goyal, A.K. Singh, and S.R. Kalidindi, *Acta Mater.* 91, 239. (2015).
6. K. Rajan, *Mater. Today* 8, 38 (2005).
7. J.M. Rondinelli, N.A. Benedek, D.E. Freedman, A. Kavner, E.E. Rodriguez, E.S. Toberer, and L. W. Martin, *Am. Ceram. Soc. Bull.* 92, 14. (2013).
8. B. Meredig, *Curr. Opin. Solid State Mater. Sci.* 21, 159. (2017).
9. A. Agrawal and A. Choudhary, *APL Mater.* 4, 053208. (2016).
10. L. Himanen, A. Geurts, A. S. Foster, and P. Rinke, *Adv. Sci.* 6, 1900808. (2019).
11. A. Çeçen, T. Fast, E.C. Kumbur, and S.R. Kalidindi, *J. Power Sources* 245, 144. (2014).
12. G. Tapia, A.H. Elwany, and H. Sang, *Addit. Manuf.* 12, 282. (2016).
13. M. Khanzadeh, S. Chowdhury, M. Marufuzzaman, M.A. Tschopp, and L. Bian, *J. Manuf. Syst.* 47, 69. (2018).
14. M. Pardakhti, E. Moharreri, D. Wanik, S.L. Suib, R. Srivastava, and A.C.S. Comb, *Sci.* 19, 640. (2017).
15. A.J. Skinner, and J.Q. Broughton, *Model. Simul. Mater. Sci. Eng.* 3, 371. (1995).
16. M. Dudzik and A.M. Stręk, *Math. Probl. Eng.* 2020, 1. (2020).
17. S. Altarazi, M. Ammouri, and A. Hijazi, *Comput. Mater. Sci.* 153, 1. (2018).
18. Z. Xiao, and Y. Zhao, *J. Mater. Res.* 28, 2545. (2013).
19. J.M. Baloyo and Y. Zhao, *MRS Proc.* 1779, 39. (2015).
20. E. Avalos-Gauna, Numerical Simulation of Active Cooling Using Porous Metals, PhD Thesis, (University of Liverpool, 2017), pp. 126-187.
21. E. Avalos Gauna and Y. Zhao, Metall. *Mater. Trans.* B (2017).
22. W. Lu, R. Xiao, J. Yang, H. Li, and W. Zhang, *J. Mater.* 3, 191. (2017).
23. S. Raschka, *Python machine learning*, 1st edn. (Packt Publishing, Birmingham, 2015), pp 110–111.
24. S.S. Skiena, *The data science design manual*, 1st edn. (Springer, 2017), pp 103–104.
25. H. Ohno, *Appl. Soft Comput. J.* 46, 17. (2016).
26. E. Bélisle, Z. Huang, S. Le Digabel, and A.E. Gheribi, *Comput. Mater. Sci.* 98, 170. (2015).
27. Y. Zhao, *J. Powder Metall. Min.* 02, 2. (2013).
28. J. Banhart, *Prog. Mater. Sci.* 46, 559. (2001).
29. Ö. Bağci, M. Özdemir, and N. Dukhan, *Transp. Porous Media* 104, 501. (2014).
30. J. M. Baloyo, Porous Metals with Novel Structures for Optimum Heat Exchange Performance, PhD Thesis, (University of Liverpool, 2016), pp. 21–33.
31. M. Faiz, Comparison of Porous Media Permeability: Experimental, Analytical and Numerical Methods, PhD Thesis, (Loughborough University, 2014), pp. 7–34.
32. J. Despois, and A. Mortensen, *Acta Mater.* 53, 1381. (2005).
33. C. Xin, Z. Rao, X. You, Z. Song, and D. Han, *Energy Convers. Manag.* 78, 1. (2014).
34. D. Forsyth, *Probability and statistics for computer science*, 1st edn. (Springer International Publishing, UK, 2018), pp 308–313.

35. M. Bowles, *Machine learning in python, essential techniques for predictive analysis*, 10th edn. (John Wiley & Sons Inc, Indianapolis, IN, 2015), pp 211–253.
36. C.M. Bishop, *Pattern recognition and machine learning*, 9th edn. (Springer, New York, 2006), pp 663–665.
37. N. Artrith, and A. Urban, *Comput. Mater. Sci.* 114, 135. (2016).
38. T.M. Ahmed, P.L. Green, and H.A. Khalid, *Road Mater. Pavement Des.* 18, 141. (2017).
39. H. Chiroma, S. Abdulkareem, and A. Gital, *Proc. Int.* I, 10. (2014).