Materials Today: Proceedings 66 (2022) 648–652

Contents lists available at ScienceDirect

Materials Today: Proceedings

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

Random Forest Regressor based superconductivity materials investigation for critical temperature prediction

G. Revathy [a,*], V. Rajendran [a], B. Rashmika [b], P. Sathish Kumar [c], P. Parkavi [a], J. Shynisha [a]

[a] Vels Institute of Science Technology & Advanced Studies, Pallavaram, Chennai 600 117, India
[b] Srisivasubramaniya Nadar College of Engineering, Kalavakkam 603110, India
[c] Bharath Institute of Science and Technology, Chennai-600073, India

ARTICLE INFO

Article history:
Available online 15 April 2022

Keywords:
Regressor
Superconductor
Critical temperature

ABSTRACT

Ever since its invention over past hundred years, superconductivity has been the subject of intense investigation. However, numerous aspects of this unusual phenomenon stay unknown, the most notable of which being the relationship among superconductivity, compound/structural assets of materials as well. Every superconductor materials transition temperature that lies in between 1 Kelvin and 10 Kelvin. Based on critical temperature of materials, superconductivity materials classified into two namely less than 10 Kelvin, greater than 10 Kelvin. Several regression models are developed here to analyze the critical temperatures of more than 12,000 known superconductors accessible through Super Con metadata, in order to sustain. After studying and implementing the aforementioned techniques, Random Forest Regressor stood out and gave the best results in terms of R^2 score metrics initial value as 91.2% and after normalizing features in superconductivity metadata, R^2 score value reaches 92.79% in predicting the temperature values of superconductors.

Selection and peer-review under responsibility of the scientific committee of the International Conference on Thermal Analysis and Energy Systems 2021.

1. Introduction

When some materials are cooled below a particular temperature, known as the superconducting critical temperature, Tc, they exhibit superconductivity. It became a focus of research after the detection of superconductors over a century in the past. The expected quantum occurrence of superconductivity is caused by the restricted magnetism among paired electrons. As mentioned in Fig. 1, two features on superconductivity materials described (a) No resistivity, and (b) ideal dia-magnetism.

There are few other characteristic properties such as resistance, impurity of materials, pressures, stress, and temperature, effects of isotopes [13], and magnetic fields that uniquely distinguish superconductors from other materials. Revathy et.al [1] proposed several regression models in predicting the critical temperature of superconducting materials with the dataset and delivered the comparisons of accuracies for each model. The work extended by performing exploratory data analysis with the dataset and by using them to predict the critical temperature. Secondly, implementing four regression models (with increased accuracy) for prediction based on certain properties like atomic mass, mean entropy, range density, thermal conductivity.

Based on these following intentions, the authors analyzed superconductivity dataset.

i. To forecast the critical temperature utilising statistical analysis of machine learning techniques, particularly regression models, based on variables derived from learning phase.
ii. Performance of novel approach shows better in foreseeing critical temperature should be possible during testing stage.
iii. Metrics such as RMSE, MSE and R^2 score value are estimated to find the performance in predicting critical temperature of superconductivity substances.

2. Background

Several authors investigated superconductivity materials using XGBoost approach [2], hybrid combination of CNN, LSTM by [3] attained R2 value of 89.9%, Bayesian Neural Network [4] achieves 92%, machine learning approach [5], regression methods [6] with 92% R2 score for predicting Tc. The authors of [7 9] established

* Corresponding author.
E-mail address: grevathy19@gmail.com (G. Revathy).

https://doi.org/10.1016/j.matpr.2022.03.515
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the International Conference on Thermal Analysis and Energy Systems 2021.

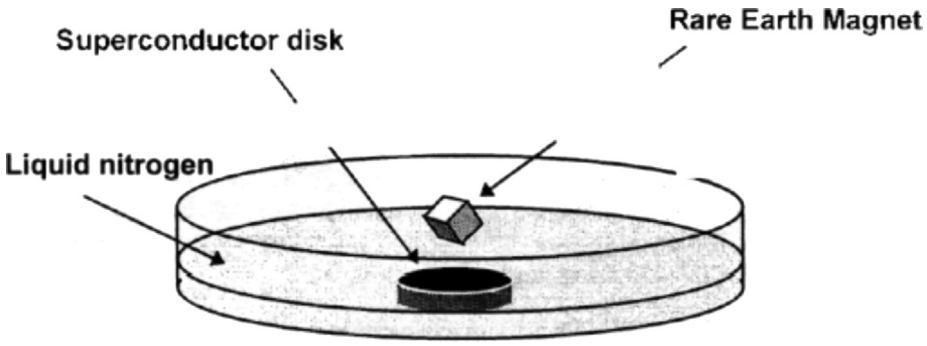

Fig. 1. Picture demonstrating Meissner effect.

novel superconductor materials of 3% which relied on occurrence and perception with authority proceeded in trial and error approach. The investigation done [17 18] achieved maximum efficacy in transferring electrical energy by superconductor materials. Many materials were condensed to suitable pressure greater than 1 million times atmospheric pressure hence such compound materials turn into superconducting at 250 Kelvin [20 21]. The related metadata gathered to build vectors, then developed machine learning approaches to predict Tc [15] [31] also deep learning techniques utilized by [8 10] for identifying actinic elements in materials. The periodic table permits deep method to discover the materials rule [24] that beyond learning phase. The cuprate materials [11] and iron metals [12] during ambient pressure anywhere exceptional superconductivity afar BCS structure was recognized. The combined method of Anderson, coherent based approximation [14] for effectively evaluates Tc with comprehensive mathematical computation. In early analysis, the model was learnt to detect less dimensional descriptors with less than 10 Kelvin [16], also reveal very effortless systematic appearance therefore attained enhances in the lead the Allen-Dynes fit. Both classification based method and regression approach identify materials via non-living crystallographic framework metadata [19]. The samples distinguished as both non cuprate with 30 samples and non-iron based materials [23 26].

3. Proposed Methodology

The main aim of our proposed work is to predict the T_c of the superconductors based on the features in the superconductivity dataset using machine learning [22] techniques. The basic workflow of this model is depicted in Fig. 2 illustrates that how machine based regression algorithms suitable in identifying the materials based on critical temperature.

3.1. Data acquisition

Initially, the dataset have taken from open source repositories comprises instance, 81 attributes extorted as of 21,263 occurrences. The goal here is to predict the critical temperature based on the features extracted. On cleaning the data, the samples were 21197. The important features such as fie, atomic mass, radius, density, electron affinity, fusion heat, valence electron, critical temperature and thermal conductivity are utilized in current work. The samples are splitted into 80% training phase as 19,657 samples and remaining 20% for testing as 4239 samples.

3.2. Data Pre-processing

In the data pre-processing phase, we analysed each instance of the dataset for duplicates, null and missing values. Each parameter's data type such as the atomic_mass, entropy, and conductivity was also observed for reporting misinformation. After removing the parameters that contained duplicate rows and null values, the total number of instances reduced to 21197.

3.3. Machine learning algorithms

Four different ML algorithms were chosen- the Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor and Simple Linear Regressor to analyse the features and predict the targeted T_c.

3.3.1. Random Forest Regressor

The first and foremost algorithm mainly used for predicting T_c is the Random Forest regressor. Tests are taken over and again to find the T_c from the training data with the features of the superconductors, so every instance point is having an equivalent likelihood of getting chosen, and every one of the 21,197 examples have a similar size as the training set. The model is then trained for each bootstrap sample, and the forecast is recorded for each sample.

3.3.2. Decision Tree Regressor

Decision tree builds regression models in the form of a tree structure, having branches and leaves. The leaf node (here, the T_c) represents a decision on the numerical target [25]. The strength based on critical field virtually temperature independent at very low temperatures, but as the temperature rises, the critical field strength decreases until it becomes zero on T_c. By $T < T_c$, a very little magnetic field is all that is needed to destroy superconductivity. The critical strength of superconductivity with temperature dependence is defined as equation (1).

$$\boldsymbol{B_c}(\boldsymbol{T}) = \boldsymbol{B_c}(0)\left[1 - \frac{\boldsymbol{T}^2}{\boldsymbol{T_c}}\right] \tag{1}$$

$\boldsymbol{B_c}(0)$ represents the critical field strength of superconductivity, Tc denotes the critical temperature.

3.3.3. Gradient Boosting Regressor

A Gradient Boosting Machine or GBM combines the predictions from multiple decision trees to generate the critical temperature of superconductors in the dataset [24]. In gradient boosted trees a lot of those weak learners from the obtained critical temperature results after prediction are built in a sequential way and each new weak learner comes to reduce the error of the previous combination of weak learner.

3.3.4. Linear Regressor

Based on independent variables, regression models a goal prediction value. It is mostly utilised in forecasting and determining

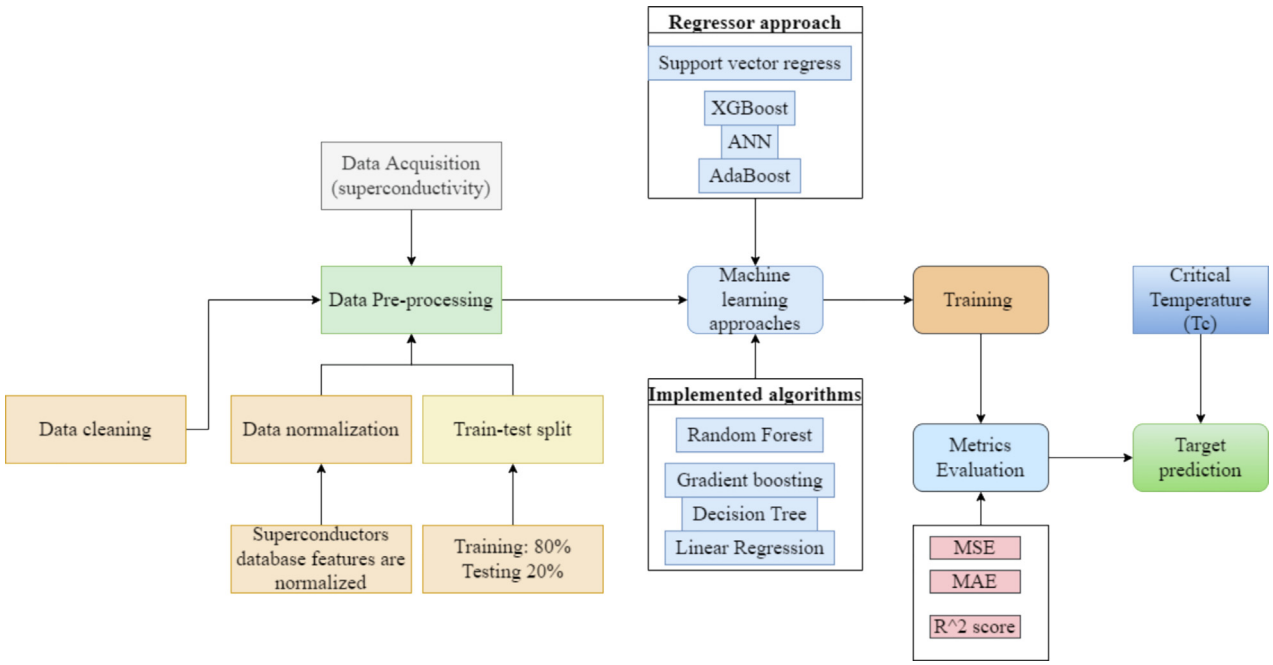

Fig. 2. Proposed workflow approach for Tc prediction via regression methods.

the relationship between variables. The various features of superconductors are used to predict the critical temperature. Linear regression is utilized to forecast the value of a dependent variable (y) based on the value of an independent variable (x).

3.4. Training

The split training data (16957) superconductors' data samples are fitted with the algorithms and the machine learns the correlation patterns and various associations present in the features of the superconductivity dataset.

3.5. Evaluating the model

In this stage, testing data (remaining 4239 samples) are used to check the score of the model in predicting the T_c. Test data instances are fed into the trained model and evaluate the output with actual data to know the accuracy level. The erroneous results were also studied such as MSE, MAE to learn and implement tuning techniques in future.

3.6. Predictions

In this phase, the critical temperature of superconductors found via applied machine learning regression model after training. Also, the model performances are evaluated using several metrics such as MSE, RMSE and MAE which produces the better outcomes in prediction of critical temperature in superconductors.

4. Metrics Evaluation

To assess if regression models are correct or deceptive, it is necessary to examine the various evaluation measures. We employ a variety of metrics such as the R^2 score, Mean Absolute Error and Mean Squared Error to assess the performance of machine learning models, particularly regression models.

4.1. Mse

The sum of inaccuracy among actual, predicted samples in metadata makes squaring calculated as the Mean Squared Error, or MSE. It's a loss function for algorithms that are fit or optimised using a regression problem's least squares framework. The units of the MSE are squared units is described as equation (2).

$$\text{(Mean Squared Error)},\ \text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{2}$$

Where N corresponds to number of samples y_i signifies the actual value and $\hat{y}_i$ represents the predicted value.

4.2. Mean Absolute Error, or MAE

Like root mean squared error, the error score's units correspond to the units of the anticipated target value. The average of the absolute error numbers is used to generate the MAE score evaluated using equation (3).

$$\text{(Mean Absolute Error)},\ \text{MAE} = \frac{\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right|}{N} \tag{3}$$

4.3. R^2 score

R2 score can range from 0 to 100 percent. The highest R^2 score was obtained from training the data with random forest regressor, which gave a value of 92%. All the proposed methods of regression, as well as their R2 scores and mean errors, are compared in this

Table 1
Summarization of metrics for proposed machine based regression methods.

Machine learning model/algorithm	MAE value	MSE value	R^2 score
Random Forest	5.30	9.70	91.99
Decision Tree	6.18	12.31	87.12
Gradient Boosting	8.70	12.90	85.85
Linear Regression	13.50	17.84	72.95

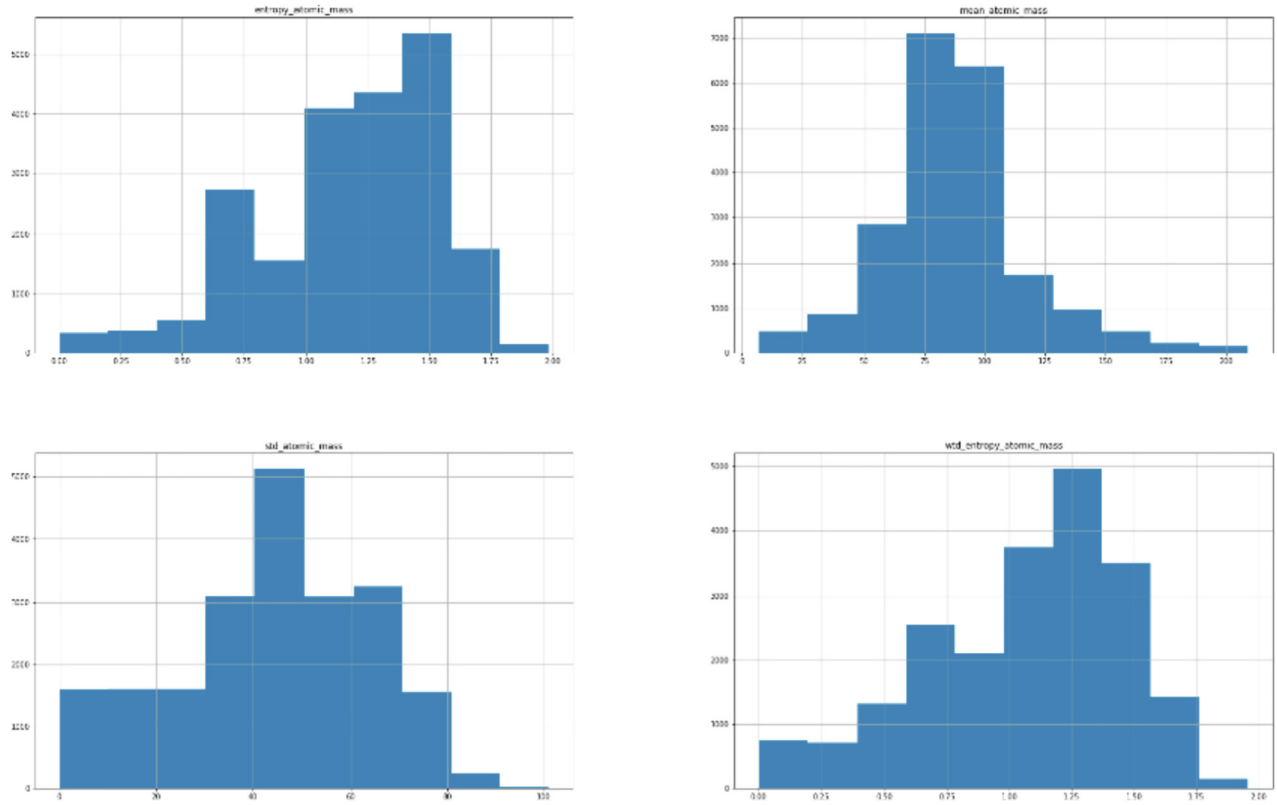

Fig. 3. Histogram plot for atomic mass (entropy, weighted entropy, mean, standard atomic mass).

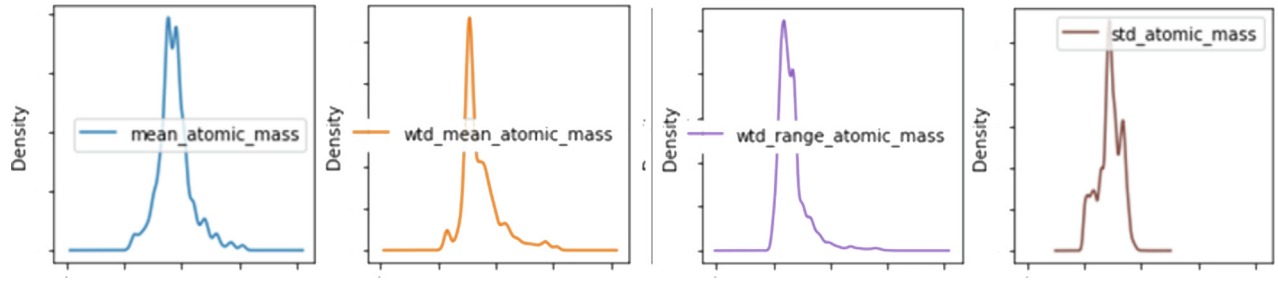

Fig. 4. Density plot for atomic mass comprises mean, weighted mean, range and weighted range.

study. The following Table 1. illustrates the metrics value of MAE, MSE and R^2 score.

Moreover, exploratory data analysis found several attributes like normal mean, mean weight, weight entropy, range, weighted range, and std dev, correlation maps, histogram, density plot and box plot were analyzed using regression methods.

5. Implementation of experimental outcomes

5.1. Histogram

Histograms divide data into bins and are the quickest way to see how each attribute in a dataset is distributed. The histograms used in the analysis of superconducting materials gives us a count of the number of observations of every features in each visualisation has been analyzed. The following is the histogram plot of various atomic masses comprises entropy, mean, weighted entropy, and standard atomic mass depicts in Fig. 3.

5.2. Density plot

Density plots are also like histograms but have a smooth curve drawn through the top of each bin. We can call them as abstracted histograms. The following is the density plots of various atomic masses are depicted from Fig. 4.

Table 2
Comparison on proposed R^2 values among regression techniques.

S. NO	Regression technique	R^2 score values
1.	Random Forest Regressor	**92.79**
2.	XGBoost Regressor	91.66
3.	Artificial Neural Networks	88.03
4.	Support Vector Regressor	75.78
5.	Decision Tree Regressor	88.44
6.	Gradient Boosting Regressor	85.49
7.	AdaBoost Regressor	67.39
8.	Simple Linear Regressor	72.51

Now, this work is extended with normalizing the superconductivity material features during processing approach. Using normalization method, the datas related with superconductors are structured in sequence, to modify the numeric variables present in columns into another forms with no deformation. To obtain better r^2 score value while predicting critical temperature, additional regression techniques are applied after normalizing the features in superconductivity dataset in which random forest Regressor attained maximum R^2 score as 92.79% in predicting the critical temperature available in superconductor materials illustrated in Table 2.

6. Conclusions

In conclusion, using a basic yet robust machine learning method, specifically a regression algorithm, this study was able to predict the critical temperatures of superconducting materials using a variety of key characteristics. Several models were developed using superconductor information, with the best suited model being the Random Forest Regressor, which had R2 = 92.79 percent and RMSE = 9.7 K. Prior to the prediction, exploratory data analysis was performed to determine the association between the factors, and it was discovered that the atomic mass had a strong relationship with the critical temperature. The next work will be improved by picking superconductor material attributes and implementing layers in Convolutional Neural Networks, a deep based technique that generates superior predictions through hyper-parameter adjustment, particularly batch size and learning rate.

CRediT authorship contribution statement

G. Revathy: Conceptualization, Methodology, Software, Data curation, Writing – original draft. **V. Rajendran:** Conceptualization, Methodology, Software, Data curation, Writing – original draft. **B. Rashmika:** Visualization, Investigation. **P. Sathish Kumar:** . **P. Parkavi:** . **J. Shynisha:** Software, Validation, Writing – review & editing.

Declaration of Competing Interest

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

References

[1] G. Revathy, V. Rajendran, P. Sathish Kumar, Prediction study on critical temperature (C) of different atomic numbers superconductors (both gaseous/solid elements) using machine learning techniques, Mater. Today: Proc. 44 (5) (2021) 3627–3632, https://doi.org/10.1016/j.matpr.2020.10.091.
[2] Kam Hamidaih, A Data-Driven Statistical Model for Predicting the Critical Temperature of a Superconductor, Comput. Mater. Sci. 154 (2018) 346–354, https://doi.org/10.1016/j.commatsci.2018.07.052.
[3] Shaobo Li, Yabo Dan, Xiang Li, Hu Tiantian, Rongzhi Dong, Zhuo Cao Jianjun Hu, Critical Temperature Prediction of Superconductors Based on Atomic Vectors and Deep Learning, Special Issue, Mater. Sci.: Synthesis, Structure, Properties 12 (2) (2020) 262, https://doi.org/10.3390/sym12020262.
[4] T. D. Le, R. Noumeir, H. L. Quach, J. H. Kim, J. H. Kim and H. M. Kim, "Critical Temperature Prediction for a Superconductor: A Variational Bayesian Neural Network Approach," in IEEE Transactions on Applied Superconductivity, vol. 30, no. 4, pp. 1-5, June 2020, Art no. 8600105, 10.1109/TASC.2020.2971456.
[5] B. Roter, S.V. Dordevic, Predicting new superconductors and their critical temperatures using unsupervised machine learning, Physica C (Amsterdam, Neth.) 575 (2020), https://doi.org/10.1016/j.physc.2020.1353689.
[6] V. Stanev, C. Oses, A.G. Kusne, et al., Machine learning modeling of superconducting critical temperature, Npj Comput. Mater. 4 (2018) 29, https://doi.org/10.1038/s41524-018-0085-8.
[7] K. Hamidieh, A data-driven statistical model for predicting the critical temperature of a superconductor, Comput. Mater. Sci. 154 (2018) 346–354.
[8] S. Zeng, Y. Zhao, G. Li, R. Wang, X. Wang, J. Ni, NPJ Comput. Mater. 5 (84) (2019).
[9] H. Hosono, K. Tanabe, E. Takayama-Muromachi, H. Kageyama, S. Yamanaka, H. Kumakura, M. Nohara, H. Hiramatsu, S. Fujitsu, Exploration of new superconductors and functional materials, and fabrication of superconducting tapes and wires of iron pnictides, Sci. Technol. Adv. Mater. 16 (3) (2015), https://doi.org/10.1088/1468-6996/16/3/033503.
[10] T. Konno, Hodaka Kurokawa, Fuyuki Nabeshima, Yuki Sakishita, Ryo Ogawa, Iwao Hosako, Atsutaka Maeda, Deep Learning Model for Finding New Superconductors, APS Phys. Phys. Rev. B 103 (2021) 014509.
[11] J.G. Bednorz, K.A. Müller, ZeitschriftfürPhysik B Condensed Matter 64 (1986) 189.
[12] Y. Kamihara, T. Watanabe, M. Hirano, H. Hosono, J. Am. Chem. Soc. 130 (2008) 3296.
[13] E. Maxwell, Isotope Effect in the Superconductivity of Mercury, Phys. Rev. 78 (4) (1950) 477, https://doi.org/10.1103/PhysRev.78.477.
[14] Hidetoshi Fukuyama, Yasumasa Hasegawa, Kei Yosida, Critical Temperature of Superconductivity Caused by Strong Correlations, SpringerLink Novel Superconductivity (1987) 407–410, https://doi.org/10.1007/978-1-4613-1937-5_45.
[15] S.R. Xie, G.R. Stewart, J.J. Hamlin, P.J. Hirschfeld, R.G. Hennig, Functional form of the superconducting critical temperature from machine learning, APS Phys. Phys. Rev. B 100 (2019) 174513.
[16] P.B. Allen, R.C. Dynes, Transition temperature of strong-coupled superconductors reanalyzed, APS Phys. Phys. Rev. B 12 (1975) 905.
[17] Anton Matasov, Varvara Krasavina, Visualization of superconducting materials, SpringerLink SN Appl. Sci. 2 (1463) (2020).
[18] J.J. Hamlin, Superconductivity near room temperature, Nature 569 (2019) 491–492, https://doi.org/10.1038/d41586-019-01583-y.
[19] S.C. Wimbush, N.M. Strickland, A public database of high temperature superconductor critical current data, IEEE Trans. Appl. Supercond. 27 (4) (2016) 1–5.
[20] N. Wagner, J.M. Rondinelli, Theory-guided machine learning in materials science, Front. Mater. 3 (2016) 28.
[21] D.M. Dimiduk et al., Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering, Integrating Mater. Manuf. Innovation (2018) 1–16.
[22] S. Baskar, D. Sendil Kumar, R. Dhinakaran, A. Prabhakaran, B. Arun, Mohanraj Shanmugam, Experimental Studies on Mechanical and Morphological Property of the Natural and SBR/BR Hybrid Rubber, Mater. Today Proc. (4 Jul 2020).
[23] Hao Li, Kecheng Wang, Yujia Sun, Christina T. Lollar, Jialuo Li, Hong-Cai Zhou, Recent advances in gas storage and separation using metal–organic frameworks ISSN 1369-7021 Mater. Today 21 (2) (2018) 108–121, https://doi.org/10.1016/j.mattod.2017.07.006.
[24] E. Hoffer, I. Hubara, D. Soudry, Train longer, generalize better: closing the generalization gap in large batch training of neural networks, in: in *Advances in Neural Information Processing Systems*, 2017, pp. 1731–1741.
[25] R.S. Chandel, R. Kumar, J. Kapoor, Sustainability aspects of machining operations: A summary of concepts, Mater. Today: Proc. (2021), https://doi.org/10.1016/j.matpr.2021.04.624.
[26] G. Revathy, S. Zuhair Affan, M. Suriya, P. Sathish Kumar, V. Rajendran, Optimization study on competence of power plant using gas/steam fluid material parameters by machine learning techniques, Mater. Today: Proc. (2020), https://doi.org/10.1016/j.matpr.2020.07.245.