Prediction of superconducting properties of materials based on machine learning models

Yongquan Jiang
Southwest Jiaotong University
Hu Jie (506338128@qq.com)
Southwest Jiaotong University
Yang Yan
Southwest Jiaotong University
Yu Siqi
Southwest Jiaotong University
Zuo Houchen
Southwest Jiaotong University

Research Article

Keywords: Superconducting materials, Critical temperature, Fermi energy level, Forbidden band width, Sub-network, Deep learning

Posted Date: August 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1891220/v1

Abstract

Superconducting materials have extremely high application value in our life. At present, the discovery of new superconducting materials relies on the experience of experts through a large number of "trial and error" experiments. Obtaining the properties of superconductors also requires a large number of experiments. In this paper, we propose to use the XGBoost model to identify superconductors, reaching 0.986 in accuracy. We apply the deep forest algorithm to predict the critical temperature of superconductors and the coefficient of determination reaches 0.944. We propose to apply the same algorithm to predict the forbidden band width of materials and the coefficient of determination reaches 0.917. A new sub-network structure is constructed to predict the Fermi level of materials and the coefficient of determination reaches 0.984. All of these algorithms have state-of-the-art performance. Finally, the model is tested with a publicly available dataset to identify 50 candidate superconducting materials with a critical temperature greater than 90K.

1 Introduction

In recent years, density generalized function theory in first principle calculations has received increasing attention with its widespread use. With the rise of machine learning, it has been increasingly applied in computing material properties [1–2]. Machine learning not only allows researchers to obtain the properties of compounds in a shorter period of time, but also reduces experimental costs. Since the discovery of superconductivity, scientists have continued to study superconductors because of its promising application in electricity transmission, strong current, electronics, and antimagnetism. A number of superconductors with high critical temperatures are discovered one after another. The critical temperatures of thiocyanide [3] and lanthanide [4] are over 200 K under the pressure of more than 150 GPa. Copper-based [5] and iron-based materials [6] are found to have very high critical temperature under room temperature and pressure. However, the strong electronic dependence of these materials makes it difficult to calculate the electronic structure and to predict the critical temperature using the first principle [7–8]. For properties such as band gap and Fermi energy levels, there is a systematic underestimation in standard DFT calculations [9].

With the great success of machine learning in image processing[10], natural language processing, etc., the application of machine learning in material science is gradually increasing. The implementation of the "Material Genome Project" means that machine learning will be widely used in Material science. On the one hand, material science in the 21st century continues to produce massive amount of data and a large-scale database has been systematically established [11] at the same time, allowing researchers to easily obtain data. Machine learning has better generalization ability coping with big data, which can learn the potential features in the data, search for the implicit relationship between the data and establish an accurate machine learning model. On the other hand, the computational simulation methods currently used are mainly finite element simulation, molecular dynamics, and first principle calculation, which not only has large limitations but also acquire high costs. Pilania et al [12]. used a machine learning approach in quantum mechanical computing combined with the concept of chemical similarity. They proposed a decision rule that creates a mapping between properties that can quickly and accurately predict properties based on fingerprints of chemical structures or electron charge density distributions. This method can significantly accelerate the discovery of specific materials. Jennings [13] et al. accelerated the discovery of nanoalloy catalysts using machine learning accelerated genetic algorithms, combining the robustness of genetic algorithms with the rapidity of machine learning.

With the increase of the amount of data on superconducting materials, the application of machine learning in superconductivity has shown a spurt in growth. Hutcheon[14] et al. applied machine learning methods in an effort to identify superconducting hydrides. They identified and filtered stable candidate objects through structure search and then performed electron-phonon calculations to obtain the critical temperature of the candidate objects, obtaining critical temperatures up to 115 K for RbH at 50 GPa and 90 K for CsH at 100 GPa. Zhang and Xu[15] et al. developed a Gaussian process regression (GPR) model to reveal the relationship between the process parameters and the superconducting transition temperature of the superconductor BiPbSrCaCuOF, which has a promising future with strong stability and high accuracy. Owolabi[16] et al. developed a computational intelligent model (CIM) with lattice parameters as descriptors using the support vector regression method(SVR) to estimate the Tc of 31 different YBCO superconductors. The estimated superconducting transition temperatures are in good agreement with the experimental values. CIM allows fast and accurate estimation of the temperature of any processed YBCO superconductor without any complex equipment. Alizade and Mohammadizadeh[17] used Debye temperature and critical transition temperature as descriptors to build a machine learning model to predict the electron-phonon coupling constant of 28 elements with an accuracy of 0.88 using cross-validation techniques. Lee and Lee [18] et al. used supervised learning algorithms and convolutional neural networks and successfully identified the signature of the topological superconductivity-*Majorana zero mode* and demonstrated that neural network model has the best performance. Roter and Dordevic [19] obtained a coefficient of determination of 0.93 using chemical descriptors of superconductors as input of the machine learning model. They also obtained several promising new superconductors based on this model. Zhang Y and Xu X [20] et al. developed a Gaussian process regression model based on structural and topological parameters (including lattice constants, volume and bonding parameters topological index H31) to predict the critical temperature of doped iron-based superconductors. Stable and accurate, the model helps to estimate the critical temperature quickly. Konno [21] et al. proposed a method called "periodic table reading", which represents the periodic table in a deep learning way. It learns to read the periodic table and the elemental laws to discover new superconductors beyond the training data. The model obtained a R^2 value of 0.92 for the target material Tc in the superconductor database. Stanev [22] et al. modeled the critical temperature (Tc) of more than 12,000 known superconductors in the SuperCon database. They first classified the materials into two categories based on their Tc values comparing to 10 K and then trained a classification model to predict this label. The model only uses coarse-grained features based on chemical composition. It reaches an accuracy of 92%. Afterwards, seperate regression models were built to predict Tc values for Cu-based, Fe-based and low Tc compounds. These models also have good performance with the learned predictors providing potential insights into the mechanisms behind superconductivity in different material families. It combined classification and regression models into one single integrated pipeline and searched the entire Inorganic Crystal Structure Database (ICSD) for potential novel superconductors. It succeeded in identifying 30 non-Cu-based and non-Fe-based oxide candidates. Zeng [23] et al. developed the atomic table convolutional neural network (ATCNN) that requires only elemental composition to learn relevant features from its own structure and the accuracy of the model exceeds the results of standard DFT calculations. More than 20 potential compounds with high

critical temperatures were screened from the existing database. Jha, D[24] et al. proposed a deep learning model Elemnet. The model automatically captures the physical and chemical interactions and similarities between different elements, thus enabling the prediction of material properties with higher accuracy and speed. The application of machine learning in predicting material properties has a promising future.

In this manuscript, we construct machine learning models to predict whether a material is superconducting, the critical temperature of superconductivity, the forbidden band width and the Fermi energy level. The forbidden band width represents the energy state interval between the valence band and the conduction band where energy density is zero. The size of the forbidden band width is used to determine whether the material has semiconductor or insulator properties. The Fermi energy level is the highest energy level of electrons at absolute zero. The Fermi energy level is a very important physical parameter in semiconductor physics. As its value is known, the statistical distribution of electrons in each quantum state at a certain temperature can be known. In order to accurately identify superconductors, this manuscript uses the XGBoost model. We apply the deep forest model to predict the critical temperature and forbidden band width of superconducting materials for the first time. We propose a novel sub-network structure applied to predict the Fermi energy level. The sub-network model uses atomic frequency coding to extract features and then manually adds atomic number, outermost electron number, electron affinity, first ionization energy, and atomic radius features based on the features. All the features are used as inputs of the model. All other machine learning models use atomic frequencies as input.

2 Theoretical Approach

Previously, researchers would spend a lot of time constructing the suitable feature vector to build a machine learning model with good performance. After the emergence of deep learning, models try to learn features automatically from raw data and give target attributes as output directly. This approach, called end-to-end, has become the mainstream deep learning method nowadays. However, in materials science, good machine learning models are still obtained by constructing suitable feature engineering. Constructing features manually based on experience is called non-end-to-end learning and is usually time-consuming and the results are subject to chance.

In this manuscript, the critical temperature data of superconducting materials are obtained from the SuperCon public dataset [25], containing 21263 records. In this manuscript, we extracted 9399 high-energy stable insulators with DFT band gaps greater than 0.1 eV from the Materials Project database as a non-superconductor dataset. We also extracted 3896 Fermi energy level (Ef) data from the Open Quantum Database (OQMD) [26], as well as 5886 compounds with forbidden band width (Eg) from previous literature [27].

The extracted data on the critical temperature of superconducting materials are preprocessed. Materials with the same composition usually have different Tc values due to the different experimental conditions under. For example, H_2S, which is not a superconductor under ambient conditions, has a particularly high Tc under high pressure[3]. There are two different Tc values for H_2S in the database, 185 K and 60 K, respectively. To avoid confusion, for such compounds with multiple Tc values, if the maximum value exceeds twice the minimum value, the data is deleted. Otherwise, the average value is taken as the Tc value of the material and the duplicates are deleted. Unproven superconducting materials, such as HWO_3, are deleted as well. Data with elemental coefficients greater than 50 in the chemical formula, such as Hg_{1234} O_{10+z} and data with uncertain oxygen content Determined data such as $Yb_{16}BaCuO_{27}$ are excluded. In the dataset, compounds Hg, MgB_2, FeSe and $YBa_2Cu_3O_7$ are typical representatives of elemental superconductors, conventional BCS superconductors, iron-based superconductors and copper-based superconductors, respectively. They are used to test the generalization ability of the ATCNN model. Before segmenting the training and test sets, these compounds, including Hg, Mg_xB_y, Fe_xSe_y and $YBa_2Cu_xO_y$ (where x and y denote the content of the corresponding elements) are removed from the cleaned-up dataset. The collated dataset has a total of 13598 superconductors. The data of non-superconductor materials as well as forbidden band widths and Fermi energy levels are not further processed because they do not contain duplicate data and were measured in the same environment.

For the Fermi energy level of the material, the neural network structure named sub-network is shown in Fig. 3. In this manuscript, there are a total of six attributes, namely, atomic frequency, electronegativity, electron affinity, first ionization energy, electron affinity energy, and number of outermost electrons, all of which are commonly and easily obtained features. Each feature is extracted with one network as shown in the left part of Fig. 1. After repeated experiments, we finally determined the number of layers of each network structure and the number of neurons. There are 6 fully-connected layers and the number of neurons in each layer is 60, 100, 200, 100, 50, 16, respectively, which is the blue part in Fig. 1. The fully-connected layer is followed by the Dropout layer in orange. Dropout can then dormant some neurons, reduce the number of model parameters, thus speed up the model training and also improve the generalization performance of the model. After the Dropout layer, the batch normalization layer is added, which is the yellow part in Fig. 1. The main role of this layer is to ensure that the distribution of the data in each layer of the network does not change significantly with the change of parameters, thus avoiding the phenomenon of internal covariate shift and ensuring a more stable training process. The last fully-connected layer uses the Linear activation function as the mapping function, which is the green part in Fig. 1. The output of the six networks are connected to a new network. The network mainly consists of fully-connected layer, Dropout layer and batch normalization layer, 4 cells in total. After repeated experiments, the number of neurons in each cell is finally determined to be 80, 64, 32, and 8 respectively, and the activation function of each layer is the Tanh function. The final mapping function uses Linear to map the matrix to a specific value, thus achieving the regression of the whole sub-network.

To avoid the occurrence of overfitting, Dropout is used with the dropout rate set to 0.1. The training process is terminated after 400 epoches, because the loss hardly decreases after that. The loss graph is shown in Fig. 2(c).

The deep forest[28] method used in this manuscript was proposed by Prof. Zhihua Zhou in 2017.

3 Results And Discussion

Each dataset is randomly divided into a training set (80%) and a test set (20%), and the experimental results of all models below are obtained using the same dataset. For the dataset identifying superconductors and non-superconductors, the output label of superconductors is set to 1 and the output label of 9399 insulators is set to 0. After training with the XGBoost classification model, the results are shown in Figure 1, and the AUC and TPR in the test set are 0.98 and 98.48% respectively. The results compared with other models are shown in Table I, which shows that the AUC and TPR of the XGBoost model are better than those of the ATCNN model [23], the only similar literature to our knowledge.

Table I Experimental results of different classification models

Models	AUC	TPR (%)
GBDT	0.96	96.47
ATCNN[23]	0.97	93.78
SVR	0.97	91.9
DecisionTree	**0.98**	97.84
RandomForest	**0.98**	98.23
XGBoost	**0.98**	**98.48**

For the collated superconducting critical temperature dataset, this manuscript proposes to use the atomic frequency as the input of the deep learning model and the deep learning model we chose is deep forest. The experimental results are shown in Figure 2(a), and in the test set, the value of R^2, MAE and RMSE are 0.944, 4.04, and 7.51 respectively. Using the same training set and test set, the comparison results with similar literature are shown in Table II with the bolded indicating the best result in that column. In this manuscript, all algorithms in Table II are implemented and tested in the same hardware and software environment.

Table II Critical temperature prediction results on the test set

Network Model	R^2	MAE	RMSE
SVM[22]	0.714	11.36	17.06
K Nearby	0.891	5.88	10.40
DecisionTree	0.888	5.40	10.50
GBDT	0.852	8.33	12.16
ExtraTree	0.892	5.32	10.40
1DCNN	0.891	6.25	10.54
Artificial Neural Networks	0.896	6.15	10.64
XGBoost	0.916	5.71	9.34
Bagging	0.918	4.80	9.07
ATCNN[23]	0.891	5.88	10.40
Sub-network	0.920	5.41	8.78
Random Forest[3]	0.928	4.58	8.53
Deep Forest	0.**945**	**4.04**	**7.51**

Table II shows that the value of R^2, MAE, and RMSE of the test set of the deep forest model are 0.945, 4.04, and 7.51 respectively, which are optimal in the published literature.

The scatter plot of the material Fermi energy level is shown in Figure 2(b) and the final results obtained using different models are shown in Table III with the bolded indicating the best value in that column. From Table III, it can be seen that the sub-network model can achieve more accurate predictions of the Fermi energy levels. For the Fermi energy level dataset, the MAE of ExtraTree and the sub-network model are both 0.1, which is lower than the MAE of DFT in literature [29] and the 0.15 of ElemNet model in literature [24]. This also means that the accuracy of the model used in this manuscript exceeds the DFT calculation. The R^2, MAE, and RMSE value of the sub-network model can also be seen in Table III at 0.984, 0.10, and 0.14 respectively.

Table III Fermi energy level prediction results on test set

Network Model	R^2	MAE	RMSE
1DCNN	0.952	0.17	0.25
KNN	0.867	0.26	0.40
SVM[16]	0.887	0.24	0.38
ExtraTree	0.966	**0.10**	0.20
XGBoost	0.938	0.20	0.27
ElemNet[24]	-	0.15	-
ATCNN[23]	0.965	0.13	0.21
Random Forest[22]	0.955	0.13	0.23
DFT[29]	-	0.136~0.81	-
Deep Forest	0.977	0.11	0.17
Sub-network	**0.984**	**0.10**	**0.14**

In this manuscript, the atomic frequency number is used as the input for the deep learning model. The deep forest model is chosen to predict the forbidden band width of the material. In the test set, the MAE, RMSE, and R^2 of the deep forest model are 0.27, 0.44, and 0.917 respectively. The corresponding scatter plot is shown in Figure 2(d). The results are shown in Table IV with the bolded indicating the best value of the column.

Table IV The forbidden band width prediction results on test set

Network Model	R^2	MAE	RMSE
SVM[16]	0.555	0.59	0.56
ExtraTree	0.617	0.44	0.83
1DCNN	0.649	0.52	0.82
KNN	0.697	0.45	0.79
ATCNN[23]	0.814	0.35	0.63
Random Forest[22]	0.811	0.34	0.64
XGBoost	0.818	0.39	0.64
Bagging	0.825	0.34	0.64
Sub-network	0.866	0.34	0.55
CGCNN[30]	-	0.388	-
DFT[29]	-	0.6	-
Deep Forest	**0.917**	**0.27**	**0.44**

As shown in Table IV, the MAE, RMSE, and R^2 of the deep forest model all outperform other machine learning models such as ATCNN, CGCNN, and SVM. It also outperforms the DFT calculation method, which shows that the accuracy of the deep forest model has surpassed the accuracy of the DFT calculation method.

In order to verify the effectiveness of the depth forest model in predicting the critical temperature of superconductivity, four materials were extracted from [21], all of which have measured critical temperature values. The values predicted using the depth forest model in this manuscript were compared with literature [21] and the results are shown in Table V, with the bolded indicating the best value in the row. Table V shows that the prediction error of the deep forest model for $CaBi_2$ is much lower than that of literature [21] and the prediction errors of the remaining materials are within 3%.

Table V Critical temperature verification results

Materials	Actual Tc value（K）	The literature [21] predicted Tc values（K）	Deep forest model predicted Tc values（K）
$CaBi_2$	2	14.85（642%）	**5.91**（195%）
HfV_4Zr	10	10.17（1.7%）	**9.92**（0.8%）
$Au_{0.5}Nb_3Pt_{0.5}$	10	**10.13**（1.3%）	10.30（3%）
$Hf_{0.5}Nb_{0.2}V_2Zr_{0.3}$	10	**10.11**（1.1%）	9.83（1.7%）

To further verify the practicality of the model, in this manuscript, for the 100,000 materials in the COD database[31], each material is first identified using the XGBoost model. If it is a superconducting material, the critical temperature value is then predicted using the deep forest model. Finally, the candidate superconducting materials with a critical temperature greater than 90 K are screened with a total of 50 materials satisfying the conditions collected. The detailed results are shown in the Appendix.

4 Conclusion

(1) Random Forest, XGBoost and Decision Tree models can all be used as models for identifying superconducting materials, among which XGBoost has the highest accuracy. This model outperforms the classification models in ATCNN [23].

(2) For the prediction of superconducting critical temperature and forbidden band width of the material, the deep forest model has the best determination coefficient R^2, mean absolute error and root mean square error compared with similar literatures.

(3) For the Fermi energy level of the material, the determination coefficient R^2, the mean absolute error, and the root mean square error of the sub-network model we proposed are optimal compared with similar literatures.

Declarations

Funding

The research work was supported by the National Natural Science Foundation of China (No.61976247) and the Fundamental Research Funds for the Central U

Author Contribution

Author 1 (First Author): Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing - Original Draft；

Author 2 (Corresponding Author): Conceptualization, Software, Funding Acquisition, Resources, Supervision, Writing - Review & Editing.；

Author 3: Visualization, Investigation；

Author 4: Resources, Supervision；

Author 5: Software, Validation；

Conflict of Interest

All authors disclosed no relevant relationships.

References

1. Paul Raccuglia, Katherine C. Elbert, Philip D. F. Adler, Casey Falk, Malia B. Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A. Friedler, Joshua Schrier and Alexander J. Norquist. Machine-learning-assisted materials discovery using failed experiments[J]. Nature, 2016, 533: 73–76.
2. Patrick Avery, Xiaoyu Wang, Davide M. Proserpio, Cormac Toher, Corey Oses, Eric Gossett, Stefano Curtarolo, Eva Zurek. Predicting superhard materials via a machine learning informed evolutionary structure search[J]. npj Computational Materials, 2019, 5(1): 1–11.
3. Drozdov A P, Eremets M I, Troyan I A, Ksenofontov V, Shylin S I. Conventional superconductivity at 203 Kelvin at high pressures in the sulfur hydride system[J]. Nature, 2015, 525(7567): 73–76.
4. Somayazulu M, Ahart M, Mishra A K, Geballe Z, Balnidi, Maria, Meng, Yue, Struzhkin, Viktor V, Hemley RJ. Evidence for Superconductivity above 260 K in Lanthanum Superhydride at Megabar Pressures[J]. Physical Review Letters, 2019, 122(2): 1–6.
5. Bednorz J G, Mueller K A Z. Possible high Tc superconductivity in the Ba-La-Cu-O system[J]. Zeitschrift für Physik B Condensed Matter, 1986, 64(2):189–193..
6. Yoichi Kamihara, Takumi Watanabe, Masahiro Hirano, Hideo Hosono. Iron-based layered superconductor La[O1-xFx]FeAs (x = 0.05–0.12) with Tc = 26 K[J].Journal of the American Chemical Society, 2008, 30(11):3296–3297.
7. Kohn W, Sham L J. Self-Consistent Equations Including Exchange and Correlation Effects[J]. Physical Review, 1965, 140(4A):1133–1138.

8. A. Jain, S. P. Ong, G. Hautier, W Chen, Kristin A.Persson.Commentary: The materials project: A materials genome approach to accelerating materials innovation[J]. APL Materials, 2013, 1(1): 0–11.
9. Chan, M. K. Y. & Ceder, G. Efficient band gap prediction for solids. Physical Review Letters, 2010, 105(19): 1–4.
10. Zhang W, Zhang F, Chen W, Jiang Y, Song D. **Fault State Recognition of Rolling Bearing Based Fully Convolutional Network[J]**. Computing in Science & Engineering, 2019, **21**(**5**): 55–63.
11. Jain A, Hautier G, Ong S P, Persson K. New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships[J]. Journal of Materials Research, 2016, 31(8): 977–994.
12. Pilania G, Wang C, Jiang X, Ramprasad R. Accelerating materials property predictions using machine learning[J]. Scientific Reports, 2013, 3(1):1–6.
13. Jennings P C, Lysgaard S, Hummelshj J S, Vegge T, Bligaard T. Genetic algorithms for computational materials discovery accelerated by machine learning[J]. Npj Computational Materials, 2019, 5(1):1–6.
14. Hutcheon M J, Shipley A M, Needs R J. Predicting novel superconducting hydrides using machine learning approaches[J]. Physical Review B, 2020, 101(14):1–9.
15. Zhang Y, Xu X. Machine Learning F-Doped Bi(Pb)-Sr-Ca-Cu-O Superconducting Transition Temperature[J]. Journal of Superconductivity and Novel Magnetism, 2021, 34(1): 63–73.
16. Owolabi T O, Akande K O, Olatunji S O. Application of computational I ntelligence technique for estimating superconducting transition temperature of YBCO superconductors[J]. Applied Soft Computing, 2016, 43: 143–149.
17. Alizadeh Z, Mohammadizadeh M R. Predicting Electron-Phononon Coupling Constants of Superconducting Elements by Machine Learning [J]. Physica C Superconductivity, 2018, 558:7–11.
18. Lee J H, Lee H C. Machine learning study of the deformed one-dimensional topological superconductor[J]. Journal of the Korean Physical Society, 2021, 79(2):173–184.
19. Roter B, Dordevic S V. Predicting new superconductors and their critical temperatures using unsupervised machine learning[J]. Physica C: Superconductivity and its Applications, 2020, 575:1–4.
20. Zhang Y, Xu X. Predicting doped Fe-based superconductor critical temperature from structural and topological parameters using machine learning[J]. International Journal of Materials Research, 2021, 112(1):2–9.
21. Konno T, Kurokawa H, Nabeshima F, Sakishita Y,Ogawa R, Hosako I,Maeda A. Deep learning model for finding new superconductors[J]. Physical Review B, 2021, 103(1):1–6.
22. Stanev V, Oses C, Kusne A G, Rodriguez E, Takeuchi I. Machine learning modeling of superconducting critical temperature[J].Npj Computational Materials, 2018(1):405–418.
23. Zeng S, Zhao Y, Li G, Wang R, Wang X, Ni J. Atom table convolutional neural networks for an accurate prediction of compounds properties[J].Npj Computational Materials, 2019, 5(1):387–393.
24. Jha D, Ward L, Paul A, Liao W, Choudhary A, Wolverton C, Agrawal A. Elemnet: Deep learning the chemistry of materials from only elemental composition[J]. Scientific reports, 2018, 8(1): 1–13.
25. National Institute of Materials Science, Materials Information Station. SuperCon. https://supercon.nims.go.jp/index_en.html (2011).
26. Saal J E, Kirklin S, Aykol M, Meredig B, Wolverton C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)[J]. JOM - Journal of the Minerals, Metals and Materials Society, 2013, 65:1501–1509.
27. Zhuo, Y, Tehrani, A. M, Brgoch, J. Predicting the band gaps of inorganic solids by machine learning[J]. Journal of Physical Chemistry Letters, 2018,9(7): 1668–1673.
28. Zhou Z H, Feng J. Deep Forest: Towards an Alternative to Deep Neural Networks[J]. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), 2017.
29. Kirklin, S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Rvhl S, Wolverton. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies[J]. Npj Computational Materials, 2015, 1(1):15–29.
30. Xie, T. Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties[J]. Physical Review Letters, 2018, 120(14):1–6.
31. Saulius Gražulis, Adriana Daškevič, Andrius Merkys, Daniel C, Luca L, Miguel Q, R S N, Peter M, T D B, Le BA. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration [J]. Nucleic Acids Research, 2012, 40(1): 420–427.

Figures

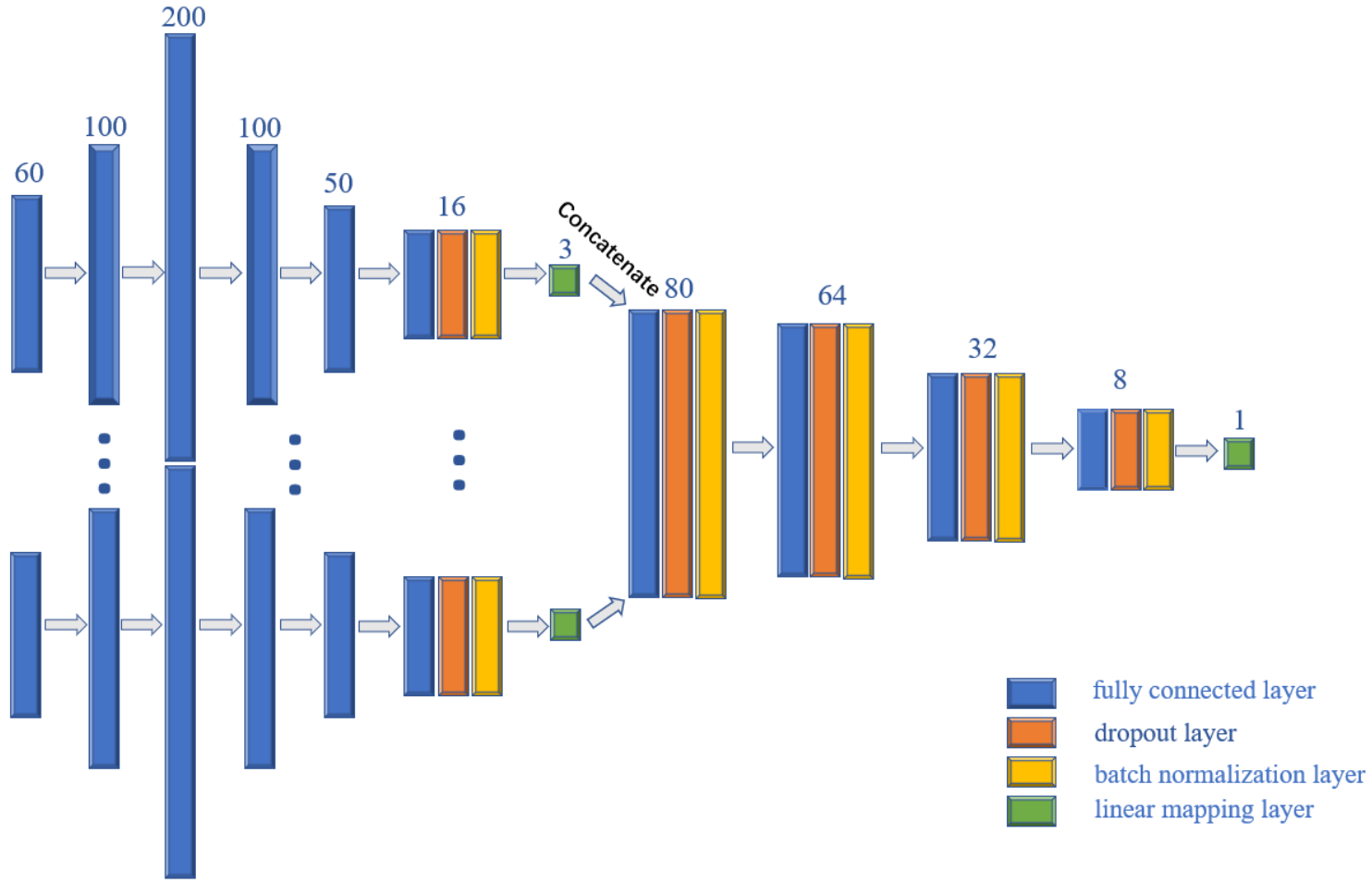

Figure 1

Schematic diagram of the sub-network

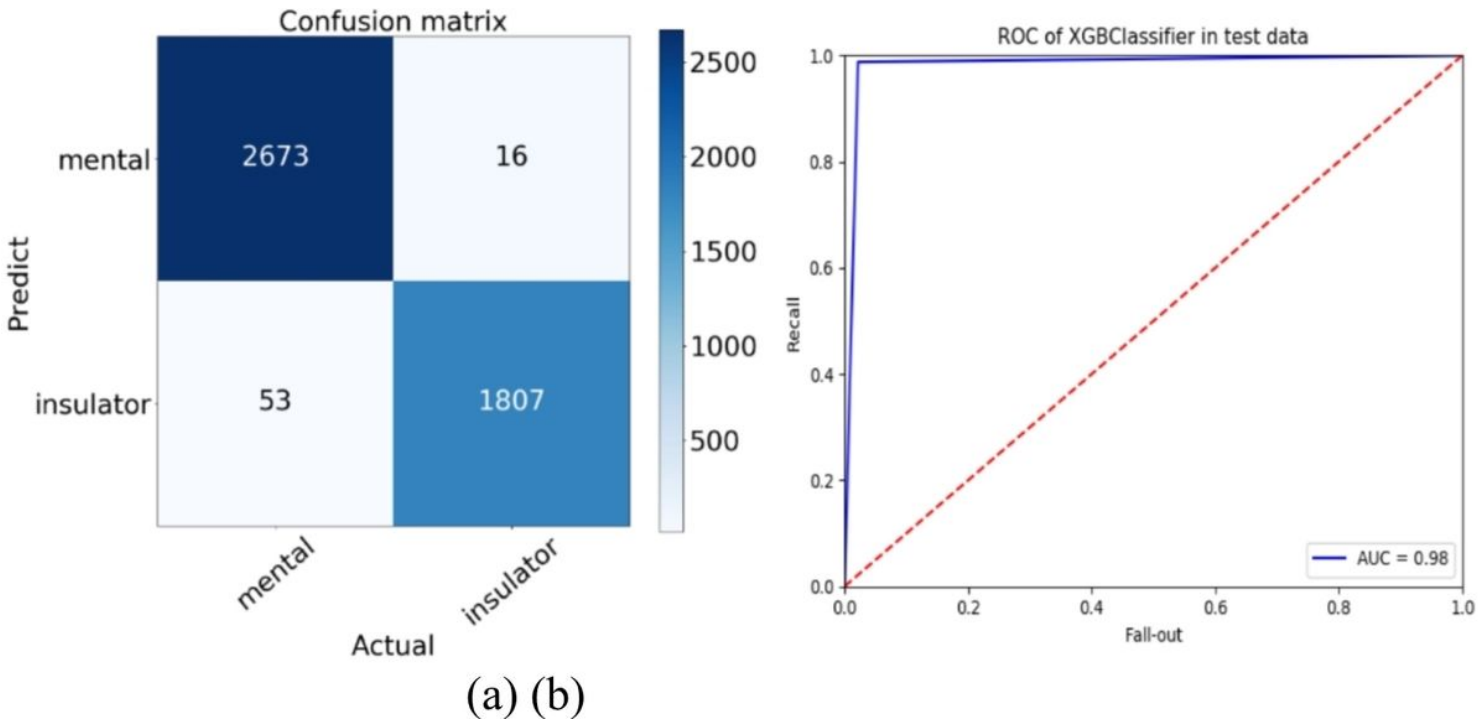

Figure 2

XGBoostClassifier

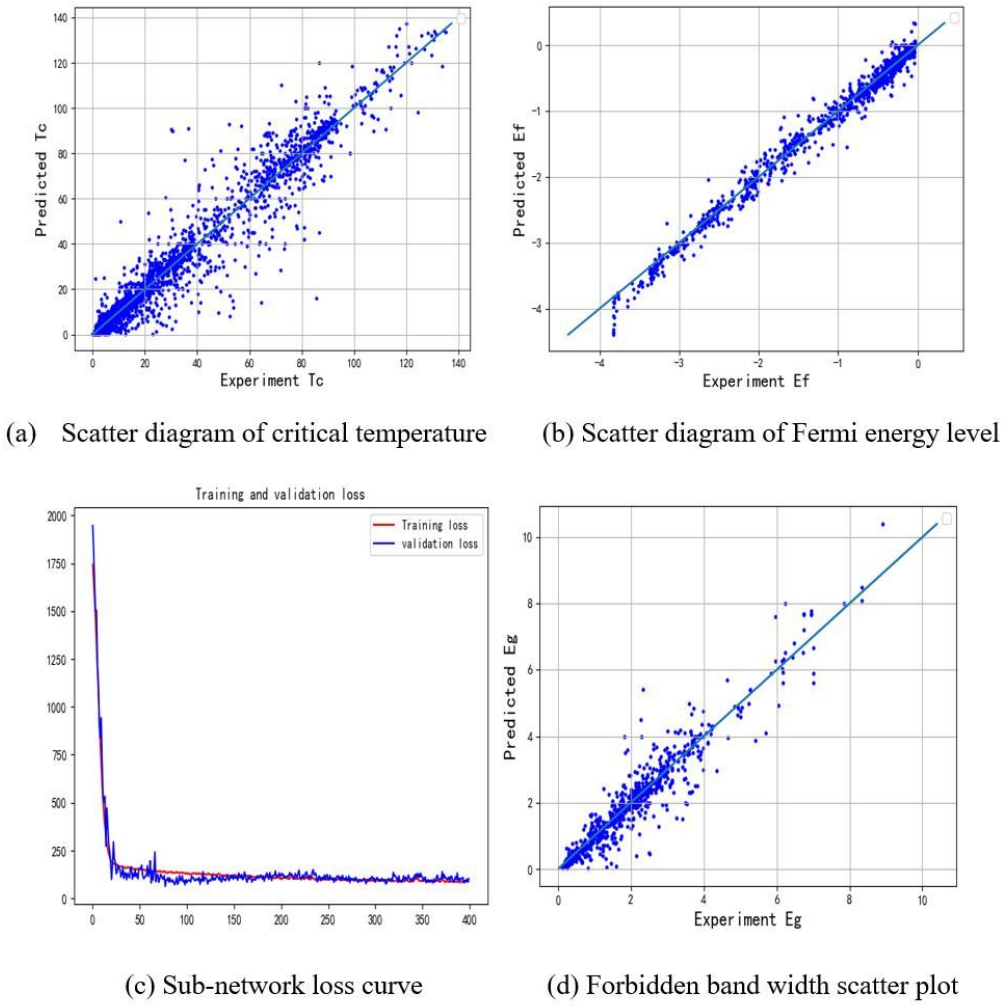

Figure 3

Graph of prediction results

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx