ARTICLE OPEN

Atom table convolutional neural networks for an accurate prediction of compounds properties

Shuming Zeng[1,2], Yinchang Zhao[3], Geng Li[1,2], Ruirui Wang[1,2], Xinming Wang[1,2] and Jun Ni[1,2]

Machine learning techniques are widely used in materials science. However, most of the machine learning models require a lot of prior knowledge to manually construct feature vectors. Here, we develop an atom table convolutional neural networks that only requires the component information to directly learn the experimental properties from the features constructed by itself. For band gap and formation energy prediction, the accuracy of our model exceeds the standard DFT calculations. Besides, through data-enhanced technology, our model not only accurately predicts superconducting transition temperatures, but also distinguishes superconductors and non-superconductors. Utilizing the trained model, we have screened 20 compounds that are potential superconductors with high superconducting transition temperature from the existing database. In addition, from the learned features, we extract the properties of the elements and reproduce the chemical trends. This framework is valuable for high throughput screening and helpful to understand the underlying physics.

INTRODUCTION

Machine learning (ML) techniques are gaining popularity in estimating the properties of materials,[1–8] inspired by the high-throughput first-principles calculations based on density functional theory (DFT).[9–11] This new tool enables researchers to explore compounds in the vast chemical space in a reasonable amount of time. The key problem for ML methods applied in materials science is to construct descriptors, which map the structure and composition to a fixed length vector. To ensure that the feature vector of arbitrary crystal systems have the same size, early researches usually consider the overall properties of the compounds,[12] while recent advances have treated materials as crystal graph with atomic properties encoded.[4,5,7] Despite the differences in approaches, they all show good performances in predicting crystals or/and molecules properties, such as formation energy and elastic moduli.

However, for properties like band gap (E_g), there is a systematic underestimation in standard DFT calculations compared with experiments.[13] To bridge the gap between theoretical calculations and the experimental values, ML models based on eXtreme Gradient Boosting (XGBoost), random forest (RF), and support vector machine (SVM) have been applied to directly estimate the experimental E_g[14] and superconducting critical temperature (T_c),[15,16] recently. Nevertheless, in order to obtain the best predictive performance, well-constructed features are required. However, the choice of the manually constructed descriptors is quite arbitrary, and it is hard to explain why those features are chosen rather than others. Here, we treat compounds as atomic table (AT) and propose a generic framework called atom table convolutional neural networks (ATCNN) to predict the compounds properties obtained experimentally. In this framework, the descriptors are learned by itself, and no additional prior knowledge, such as atomic properties and the underlying physics are

utilized except for the component information. Although the detailed structure parameters are ignored, for a specific composition, the experimental structure is often determined and unique. In other words, the composition is entangled to the crystal structure. Therefore, our approach also has the advantage of overcoming the difficulty of obtaining accurate atomic position information through experiments.

Under the ATCNN framework, we have constructed ML models to predict the experimental T_c, formation energy (E_f), and E_g. The performance of these models has been greatly improved compared to the previous models. For the T_c prediction, to avoid system bias, we propose a data-enhancement method that enables the model to distinguish superconductors and non-superconductors. Utilizing the well-trained model, we have screened out dozens of compounds that are potential high T_c materials from the existing materials database. For the prediction of E_g, the accuracy of our model exceeds hybrid functional calculations that is considered to be accurate in calculating E_g,[17] which means the ATCNN model has great application prospects in the semiconductor industry.

RESULTS

Before the emergence of deep learning, researchers spent a lot of time constructing appropriate feature vectors to obtain ML model with superior performance. Deep learning algorithms try to learn high-level features from raw data, and directly output the target properties. This end-to-end approach has achieved great success in image recognition,[18,19] speech recognition[20,21] and machine translation.[22] However, in the field of materials science, feature engineering is still one of the most important aspects of building ML models.[4,12] The method of manually constructing features and then predicting the properties according to the features is called

[1]State Key Laboratory of Low-Dimensional Quantum Physics, Department of Physics, Tsinghua University, 100084 Beijing, P.R. China; [2]Collaborative Innovation Center of Quantum Matter, 100084 Beijing, P.R. China and [3]Department of Physics, Yantai University, 264005 Yantai, P.R. China
Correspondence: Jun Ni (junni@mail.tsinghua.edu.cn)
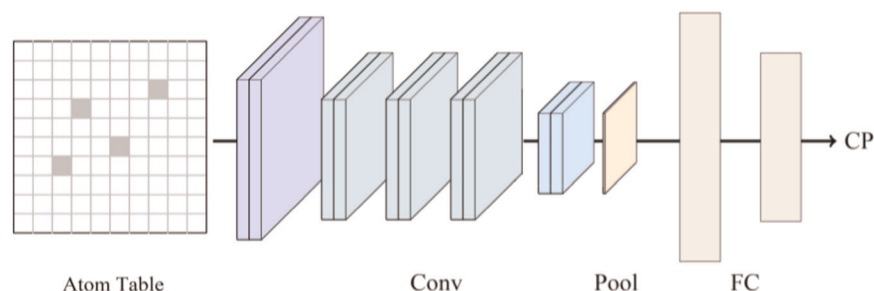
Fig. 1 Schematic diagram of the ATCNN model for T_c prediction. A typical ATCNN model contains one input layer (atom table), one output layer (compound property, CP), several convolutional layers (Conv), one pooling layer (Pool)", and several fully connected layers (FC). The size of blue, teal, and cyan colored kernels of Conv are 5×5, 3×3, and 2×2, respectively. For pool layer, the max-pooling method is used and the size is 2×2. The detailed hyperparameters are presented in Table S1

non-end-to-end learning, which is not only inefficient, but usually does not achieve optimal performance[21] due to the improperly constructed features. Here, we develop an end-to-end framework ATCNN to directly predict the target properties. In this approach, no other prior knowledge is used except for the components. The detailed construction of the ATCNN model is presented in the "Methods" section and a typical ATCNN model is shown in Fig. 1.

The experimental data for T_c, E_g, and E_f are extracted from the SuperCon database,[24] previous literature,[14] and the Open Quantum Materials Database (OQMD),[25] respectively. The atomic number of the elements involved is within 86, covering the first six period elements. For the E_g and E_f data, a total of 3896 and 5886 compounds are included, respectively. These data are used without further screening, and each data set is randomly divided into training set (80%) and test set (20%) (see Table S4 in the Supplementary Information). However, in the SuperCon database, materials of the same composition often have different T_c values, which are obtained from different experiments. For example, at ambient pressure, H_2S is not a superconductor, but under high pressure, it has a very high T_c.[26] Because of this, there are two very different T_c values of H_2S in the database, namely 185 and 60 K, respectively. To avoid causing confusion, for a compound with multiple T_c values, if the maximum exceeds the minimum twice, this material is removed, otherwise the average is taken as the T_c value of this material and the duplicate data is thrown away. In addition, we have also removed unreasonable data, including the unconfirmed 2D high-T_c superconductor HWO_3,[27] and those with the coefficient of an element in the chemical formula >50 like "Hg1234O10+z", or with uncertain oxygen content such as "Yb16Ba1Cu2Oz".

In the cleaned dataset of T_c, Hg, MgB_2, FeSe, and $YBa_2Cu_3O_7$ are retained to test the generalization ability of the ATCNN model, which are typical representatives of elemental superconductors, conventional BCS superconductors, iron-based superconductors, and copper-based superconductors. Before splitting training set and test set, the related compounds including Hg, Mg_xB_y, Fe_xSe_y, and $YBa_2Cu_xO_y$ are removed from the cleaned dataset, where x and y denote the content of corresponding elements. Therefore, compounds like MgB_2, $Mg_{0.31}B_{0.69}$, and $Mg_{0.9}B_2$ are all removed. The cleaned data set contains 13,598 superconductors, which is divided into a training set (80%) and a test set (20%). When determining model hyperparameters, 20% of the training data is used to validate the model, and each hyperparameter is determined by multiple tests. In the validation set, the model with five Conv layers performs best, as shown in Fig. S2. Besides, the number of the kernels in each Conv layer, the size of kernel, the number of the FC layers and the number of neurons in each FC layer are also tested. After hyperparametric optimization, the structure of the ATCNN-I model is determined, as shown in Fig. 1. In ATCNN-I, each Conv layer contains 64 kernels, and the size of hidden layers are 200 and 100 (see Table S1 in the Supplementary

Table 1. Statistical summary of the prediction performance of ATCNN-I and ATCNN-II on superconductors

Model	Compounds	Data size	Test result		
		Test set	MAE (K)	RMSE (K)	r^2
ATCNN-I	Total	2720	4.27	8.14	0.97
	I	1122	6.98	10.89	0.94
	II	287	4.90	7.64	0.83
	III	69	9.77	13.98	0.85
	IV	1242	1.69	3.66	0.87
ATCNN-II	Total	2720	4.12	8.19	0.97
	I	1122	7.25	11.02	0.94
	II	287	4.62	7.57	0.83
	III	69	9.49	12.91	0.88
	IV	1242	1.71	3.84	0.86

Information), respectively. To avoid overfitting, the dropout method[28] is used and the dropout rate of the FC layers is set to 0.2. The training process is terminated after 500 epochs, because the error of the model on the validation set decreases before 500 epochs, and then the error almost remains unchanged, or even tends to increase, as shown in Fig. S1. The test results are summarized in Table 1 and shown in Fig. 2c.

In the test set, the MAE, RMSE, and the coefficient of determination (r^2) are 4.12, 8.14 K and 0.97, respectively. The overall performance of the ATCNN-I model is much better than the previous fine tuned RF model (which has an r^2 of nearly 0.88)[15] and XGBoost model (which has an r^2 of 0.92 and a RMSE of 9.5 K).[16] In addition, except for Hg, the predicted values of T_c in the independent dataset (Hg, MgB_2, FeSe, and $YBa_2Cu_3O_7$) are nearly the same compared with the experimental results (see Table 2), showing that the ATCNN-I model has strong generalization ability. For Hg, its superconducting property can be learned from compounds that contain Hg element in the training set, such as $HgSr_2Cu_{10}O_4$, $Hg_{0.76}Tl_{0.76}BaCuO_{4.5}$, and $HgBa_2CuO_{4.19}$.

However, like the previous models, ATCNN-I has difficulty distinguishing between superconductors and non-superconductors, because the training data only contains superconductors. To fix this problem, we add 9399 energetic stable insulators with the DFT band gap larger than 0.1 eV to the 13,598 superconductors data set, and 80% of them are mixed up to the training set, while 20% of them are added to the test set. These insulators are extracted from Materials Project repository[9] and they are treated as non-superconductors. The full training and test data set are listed in the Table S5. After retraining, we get the ATCNN-II model. As shown in Fig. 2, Tables 1 and 2, ATCNN-II has similar performance on superconductors with ATCNN-I, and the MAE, RMSE, r^2 are 4.12, 8.19 K and 0.97, respectively. Both models have
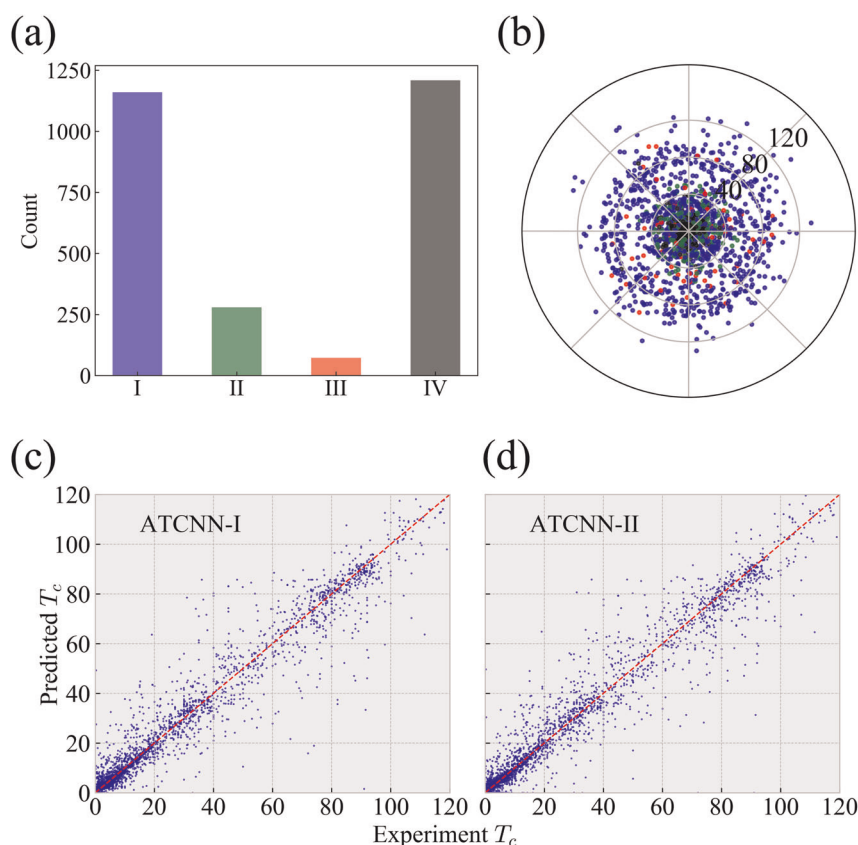
Fig. 2 Detailed results of ATCNN-I and ATCNN-II models on superconductors. The test set only contains superconductors. **a** Statistics of the compounds in the test set. The blue, red, green, and gray bars indicate the number of compounds containing Cu, Fe, Cu–Fe, and others, respectively. **b** The distributions of T_c of the compounds in **a**. **c** Comparison of the ATCNN-I model predicted T_c against the experimental T_c in the test set. **d** Comparison of the ATCNN-II model predicted T_c against the experimental T_c in the test set

Table 2. Comparison of experimentally measured T_c with the values predicted by ATCNN-I and ATCNN-II models

Compound unit	Experiment (K)	ATCNN-I (K)	ATCNN-II (K)
Hg	4.12[29]	1.4	2.6
MgB$_2$	39[30]	38.0	38.7
α-FeSe	8[31]	9.8	8.1
YBa$_2$Cu$_3$O$_7$	91[32]	91.5	90.6
YBa$_2$Cu$_3$O$_{6.6}$	60[32]	63.7	67.2

The related materials are removed from the data set before training and testing these two models

capability of capturing the change of T_c when the composition changes slightly, such as YBa$_2$Cu$_3$O$_7$ and YBa$_2$Cu$_3$O$_{6.6}$. The biggest difference between ATCNN-I and ATCNN-II is in predicting the non-superconductors. In Figs. S3, S4 and Table S5, the predictive performances on superconductors and non-superconductors of these two models are presented. In the full test data set which contains 2720 superconductors and 1880 non-superconductors, the MAE and RMSE of the ATCNN-I model increase to 8.76 and 10.45 K, while that of the ATCNN-II model decrease to 3.17 and 6.91 K. The reason is that ATCNN-I treats compounds as superconductors and predicts non-zero T_c values, even if they are insulators, but ATCNN-II gives absolute zero T_c values for most insulators and non-superconducting metal such as alkali metal. If the predicted $T_c > 0$ and $T_c = 0$ are classified to superconductors

and non-superconductors, respectively, 8.9% of the superconductors and 2.2% of the non-superconductors are misclassified by ATCNN-II, while 2.9% of the superconductors and 99.6% of the non-superconductors are misclassified by ATCNN-I in the full test set, as shown in the confusion matrices in Fig. S3. The ability to distinguish between superconductors and non-superconductors can greatly increase the efficiency of searching for new superconductors. For example, from 20,574 energetic stable materials in Materials Project database, we screen out 20 materials with large predicted T_c (see Table S6). These selected compounds are potential high T_c materials.

To further analyze the model, the 2720 superconductors in the test set are divided into four groups: the first group (I) has 1122 materials, all of which contain Cu; the second group (II) has 287 materials, all of which contain Fe; the third group (III) has 69 materials, all of which contain both Cu and Fe elements; the rest are classified as the fourth group (IV), with a total of 1242 materials. The first, second, and fourth categories roughly represent copper-based superconductors, iron-based superconductors, and conventional BCS superconductors, respectively. Their statistical distributions are shown in Fig. 2a, b. The predicted results for each group are summarized in Table 1. It can be seen from the statistical results that both the two ATCNN models perform better in group I and group IV, but worse in group II and group III. The possible reason is that the number of iron-based superconductors in the data set is much smaller than that of conventional BCS superconductors and copper-based superconductors. Therefore, it should be cautious when predicting the T_c of iron-based superconductors.
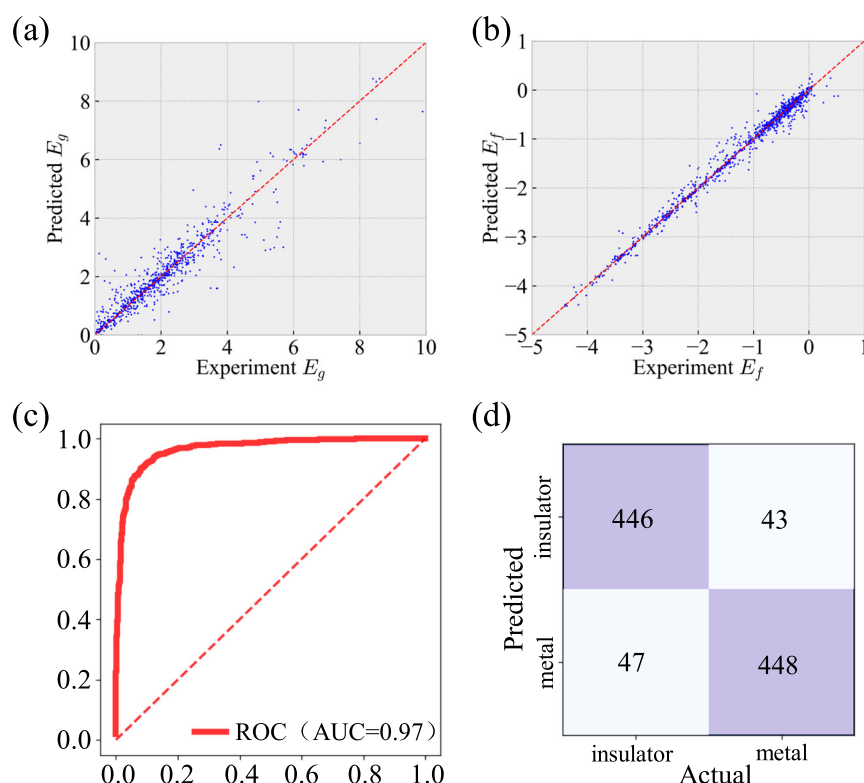
Fig. 3 a Comparison of the ATCNN model predicted E_g against the experimental E_g. **b** Comparison of the ATCNN model predicted E_f against the experimental E_f. **c** ROC curve for the metal/insulator classification model with an AUC of 0.97. **d** Confusion matrix of the metal/insulator classification model

In the prediction of crystal properties, electronic structures are often used to test the performance of the ML model.[4–7,33,34] As a general framework, the ATCNN model can also be applied to other experimental data, such as E_g and E_f. Due to the small amount of data, only one (two) Conv, one Pool, and two hidden layers are used for the prediction of E_f (E_g). The detailed network hyperparameters are listed in Tables S2 and S3. Figure 3a, b show the comparison of the experimental E_g and E_f with the prediction of the ATCNN model. It is clear that the ATCNN model achieves excellent agreement between the experimental data and predicted values in the test set. For E_f prediction, the MAE and r^2 are 0.078 eV/atom and 0.99, while the MAE of the DFT calculation with respect to the experimental measurement is 0.81–0.136 ev/atom,[11] which means that the accuracy of the ATCNN model has exceeded the DFT calculations. In addition, from the data point of view, for the same training data size (4708) of E_f, the performance of the ATCNN model is much better than the structure free ElemNet model (MAE of about 0.15 eV/atom).[33] We attribute the better performance of the ATCNN model to its unique network structure, since different network structures lead to different solutions. Due to the fully connected network structure, the ElemNet model treats the relationship between elements as equivalent, and it requires a lot of data to learn the unique relationship between elements. But for the ATCNN model, because of the convolution network structure, the relationship between elements is naturally different. Besides, for the E_f prediction, the size of ATCNN model is much small than ElemNet model and does not easily cause overfitting. Therefore, the ATCNN model performs better than the ElemNet model when the dataset is small.

For E_g prediction, the MAE of the ATCNN model is 0.307 eV, while the MAE of the CGCNN model is 0.388 eV,[5] and the MAE between standard DFT calculations and experimentally measured values is 0.6 eV,[35] demonstrating the superior performance of our

model. Compared with E_f, the prediction of E_g seems to be not so accurate, and the r^2 is 0.94. Nevertheless, the performance of the ATCNN model is better than the gradient-boosting decision tree model based on the property-labeled materials fragment descriptors, which has an r^2 of 0.90,[4] and the SVR model using the same experimental data, which achieves an r^2 of 0.90.[14] To quantify the capabilities of the ATCNN model, we apply the trained model to a set of specific compounds which have been studied intensely by different levels of theory, and are often used to benchmark new methods.[36–38] The predicted results on the selected compounds are listed in Table 3. Among these methods, the PBE calculated band gaps ($E_{g,PBE}$) differ greatly from the experimental value, and are all underestimated. The GW approximation method[36] is the most accurate when calculating E_g, which resulted in a MAE of 0.22 eV and an RMSE of 0.33 eV. However, the GW-type calculation is not currently amenable to high-throughput calculations due to the expensive computational cost. The most effective way for high-throughput screening is to predict E_g by the ATCNN model, because it is both accurate (with a MAE of 0.25 eV and an RMSE of 0.58 eV, which are only worse than the GW calculation) and fast. In addition, the structure of the E_g prediction model can be used for metal/insulator classification. For the sake of comparison, the same dataset which contains 2458 unique insulators and 2458 metals as the SVC mdoel[14] is used to train and test the classification model. The performance of the ATCNN classification model is characterized by the receiver operating characteristic (ROC) curve and the confusion matrix, as shown in Fig. 3c, d. The area under the ROC curve (AUC) is 0.97, the same as the SVC model. From the confusion matrix, it can be seen that ATCNN model is more accurate in classifying metals (91.2% vs. 88.8%), but slightly worse in classifying semiconductors than SVC model (90.5% vs. 95.2%).

DISCUSSION

For any ML algorithm applied in materials science, model interpretability is desirable. Here, we take the E_f prediction model as an example to illustrate how to extract knowledge from the ATCNN model. Generally, visualizing the element representations is a good way to examine the learned features. In Fig. 4a, the 50 dimensional representations (for each element, the first FC layer outputs 50 values) of main-group elements are shown. However, these features are intertwined with each other and it is difficult to visually see the relationship between elements. To better understand the high-dimensional feature vectors, all features are first decoupled by the principal component analysis method (PCA) and then projected onto the space spaned by the first two principal axes. PCA is a method of feature extraction and dimensionality reduction, and it is widely used in previous studies to visualize the high-dimensional features.[39–41] As seen in Fig. 4b, alkali metals (group I), alkali earth metals (group II), halogens (group VII), and

rare gases (group VIII) are clustered in different regions. Besides, the elements O, N, and S are close to halogens, showing strong non-metallicity, while the element In is close to alkali earth metal, showing metallicity. Since the E_f of elemental crystals are all 0, the features are not learned through the elements, but are learned through the compounds. The PCA results reflect the periodic law of the elements, and confirm that the ATCNN model indeed learns the properties of elements. Besides, the E_f reflects the stability of the compounds, and the stability of the compounds is related to the arrangement of electrons outside the nucleus,[25,42] that is, to the periodic law of elements. Thus, the PCA results indicate that the ATCNN model has captured the underlying physics of E_f.

In summary, we treat compounds as atom tables and propose a universal ML framework called ATCNN to predict the experimental measured properties. The ATCNN model automatically learns the features needed and directly predict the properties without having to manually construct feature vectors. Under this framework, we construct ATCNN-I and ATCNN-II model to predict the superconducting transition temperature. The ATCNN-I is accurate to predict the T_c of superconductors, while the ATCNN-II is not only accurate, but also able to distinguish between superconductors and non-superconductors. Using the ATCNN-II model, we have screened dozens of unexplored compounds that are potential high T_c materials. In addition, for experimental measured E_f and E_g, the accuracy of the ATCNN model exceeds the standard DFT calculations. Furthermore, we use PCA method to analyze the learned features of main-group elements, and find that the ATCNN model indeed learns the properties of elements and reproduces the chemical trends which reflects the underlying physics of E_f.

Table 3. Comparison of experimentally measured band gap (E_g), with values calculated by PBE functional ($E_{g,PBE}$), hybrid functional ($E_{g,HSE}$), the GW approach ($E_{g,GW}$), the SVR model that relies on manually constructed feature descriptors ($E_{g,SVR}$), and the ATCNN model ($E_{g,ATCNN}$)

Compound	E_g (eV)	$E_{g,PBE}$ (eV)	$E_{g,HSE}$ (eV)	$E_{g,GW}$ (eV)	$E_{g,SVR}$ (eV)	$E_{g,ATCNN}$ (eV)
GaN	3.2	1.62	3.14	3.32	4.45	3.45
CdTe	1.6	0.62	1.52	1.76	1.43	1.58
LiF	14.2	9.2	11.47	15.1	9.87	12.39
TiO$_2$	3.42	2.13	3.67	3.73	3.99	3.44
CuSbS$_2$	1.38	0.9	1.69	1.1	1.39	1.40
ZnS	3.91	2.07	3.49	4.15	3.12	3.86
Cu$_2$ZnSnS$_4$	1.6	0.28	0.09	1.64	1.75	1.75
PbTe	0.19	0	0.19	0.26	0.2	0.31
CaAs	1.52	0.19	1.12	1.52	1.28	1.46
ZnO	3.44	0.67	2.49	3.2	3.41	3.46
MAE		1.65	0.67	0.22	0.75	0.25
RMSE		2.1	1.0	0.33	1.46	0.58

The experimental data and the calculated values are extracted from previous literature[14]
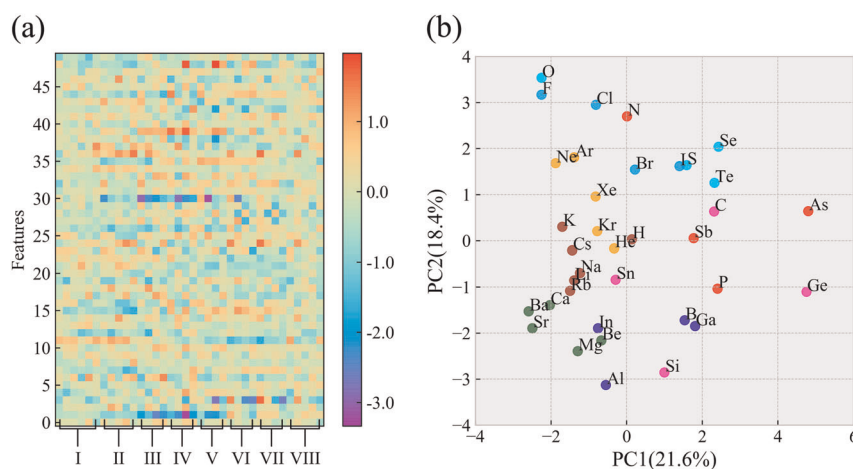
METHODS

In the framework of ATCNN, a compound is treated as a 10×10 pixels image that is called AT. Each pixel of AT represents an element, and its value is the proportion of this element in the compound. Therefore, a 10×10 pixels AT can represent any materials within 100 elements. Because all compounds involve only the first 86 elements of the periodic table, the size of the AT is sufficient. The elements represented by each pixel can be specified randomly, but must be unique and deterministic. For convenience, we specify them in the order of the periodic table of elements. That is, the first pixel represents the proportion of the H element, and the last pixel represents the proportion of the Fm element. We do not use the actual structure of the periodic table as AT because we try to build the ML models with as little prior knowledge as possible, and it is difficult to represent Actinides and Lanthanides in the periodic table. AT represents

Fig. 4 Features analysis of the E_f prediction model. **a** Illustration of the feature vectors of 36 main-group elements in vector space. Group I represents the first main-group elements, group II represents the second main-group elements, and so on. **b** Projection of the feature vectors of 36 main-group elements onto the plane spanned by the first and second principal axes (PC1 and PC2). The percentage represents the ratio of the variance on the principal axe direction. Elements are colored according to the their elemental groups

only the compound, and its feature vector will be learned through convolutional neural networks (CNN).

A typical CNN contains two major components: convolutional layers (Conv) and pooling layers (Pool). In the first Conv, the convolution is performed on the AT with the use of the filters or kernels to then produce feature maps. Each convolution kernel produces a feature map of the original map size,

$$\Lambda_l = \text{Conv}(\text{AT}, K_{l, m \times m}, \phi), \tag{1}$$

where Λ_l is the lth feature map produced by the lth kernel $K_{l,m \times m}$ with the size of $m \times m$. The kernels K are learned automatically by the network. ϕ is the non-linear activation function called Rectified Linear Unit (Relu)[23] which is defined as $\phi(x) = \max(0, x)$. The $n + 1$th Conv takes the feature maps Λ^n produced by the nth Conv as input and outputs new feature maps Λ^{n+1},

$$\Lambda_l^{n+1} = \text{Conv}(\Lambda^n, K_{l, m \times m}, \phi). \tag{2}$$

After several Convs, the final feature maps Λ^f are obtained. Then the max-pooling method is used to reduce the size of Λ^f and produce feature vector v

$$\Lambda^r = \text{Pool}(\Lambda^f, P_{h \times h}), \tag{3}$$

$$v = \sum_{l, ij} \oplus \Lambda_{l, ij}^r, \tag{4}$$

where $\oplus$ is the concatenation operator and $\Lambda_{l, ij}^r$ is the entry of the lth feature map of Λ^r. $P_{h \times h}$ denotes the pool with the size of $h \times h$. The role of the Pool layer is to reduce the size of the feature map, but in our model, the size of the Atom Table is just 10×10, so we do not use the Conv layer and the Pool layer alternately like the common CNN. In fact, using one Pool layer after the final Conv layer is best in our models. Finally, several FC hidden layers are used to capture the complex relation between feature vector and target property. The parameters of the entire network are learned by minimizing the loss function. For predictions of T_c, E_g, and E_f, the loss functions are root mean square error (RMSE), mean absolute error (MAE), respectively. To ensure that the results are non-negative, we add the Relu activation function in the output layer in the predictions of T_c and E_g. Thus, the predicted T_c and E_g are positive and consistent with the actual situation.

DATA AVAILABILITY
The data sets used to generate the results in this work are available at https://github.com/xinyu1905/ATCNN.

CODE AVAILABILITY
The codes that use the ATCNN models are available at https://github.com/xinyu1905/ATCNN.

AUTHOR CONTRIBUTIONS
J.N. and S.Z. designed the research. S.Z. worked on the model. S.Z., Y.Z. and G.L. collected the data. S.Z., X.W. and J.N. wrote the text of the manuscript. All authors discussed the results and commented on the manuscript.

ADDITIONAL INFORMATION
Supplementary Information accompanies the paper on the *npj Computational Materials* website (https://doi.org/10.1038/s41524-019-0223-y).

Competing interests: The authors declare no competing interests.

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

REFERENCES
1. Pilania, G., Wang, C., Jiang, X., Rajasekaran, S. & Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **3**, 2810 (2013).
2. Schütt, K. T. et al. How to represent crystal structures for machine learning: towards fast prediction of electronic properties. *Phys. Rev. B* **89**, 205118 (2014).
3. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 16028 (2016).
4. Olexandr Isayev et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 15679 (2017).
5. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018a).
6. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet—a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
7. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
8. Tehrani, A. M. et al. Machine learning directed search for ultraincompressible, superhard materials. *J. Am. Chem. Soc.* **140**, 9844–9853 (2018).
9. Jain, A. et al. Commentary: the Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
10. Curtarolo, S. et al. Aflow: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
11. Kirklin, S. et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **1**, 15010 (2015).
12. De Jong, M. et al. A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. *Sci. Rep.* **6**, 34256 (2016).
13. Chan, M. K. Y. & Ceder, G. Efficient band gap prediction for solids. *Phys. Rev. Lett.* **105**, 196403 (2010).
14. Zhuo, Y., Tehrani, A. M. & Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **9**, 1668–1673 (2018).
15. Stanev, V. et al. Machine learning modeling of superconducting critical temperature. *npj Comput. Mater.* **4**, 29 (2018).
16. Hamidieh, K. A data-driven statistical model for predicting the critical temperature of a superconductor. *Comp. Mater. Sci.* **154**, 346–354 (2018).
17. Heyd, J. & Scuseria, G. E. Efficient hybrid density functional calculations in solids: assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional. *J. Chem. Phys.* **121**, 1187–1192 (2004).
18. Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (eds Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Red Hook, NY, USA, 2012).
19. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. 29th IEEE Conference on Computer Vision and Pattern Recognition* (eds Agapito, L., Berg, T., Kosecka, J. & Zelnik-Manor, L.) 770–778 (IEEE Computer Society, Los Alamitos, CA, USA, 2016).
20. Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Proc. Mag.* **29**, 82–97 (2012).
21. Song, W. & Cai, J. *End-to-end Deep Neural Network for Automatic Speech Recognition*. Standford CS224D Reports (2015).
22. Wu, Y. et al. Google's neural machine translation system: bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144 (2016).
23. Hahnloser, R. H. R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J. & Seung, H. S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* **405**, 947 (2000).
24. National Institute of Materials Science, Materials Information Station. SuperCon. https://supercon.nims.go.jp/index_en.html (2011).
25. 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). *JOM* **65**, 1501–1509 (2013).
26. 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. *Nature* **525**, 73 (2015).
27. Reich, S., Leitus, G., Popovitz-Biro, R., Goldbourt, A. & Vega, S. A possible 2D H$_x$WO$_3$ superconductor with a T_c of 120 K. *J. Supercond. Nov. Magn.* **22**, 343–346 (2009).
28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
29. Delft, D. V. & Kes, P. The discovery of superconductivity. *Phys. Today* **63**, 38–43 (2010).

30. Xu, M. et al. Single crystal MgB$_2$ with anisotropic superconducting properties. *Appl. Phys. Lett.* **79**, 2779 (2001).

31. Subedi, A., Zhang, L., Singh, D. J. & Du, M. H. Density functional study of FeS, FeSe, and FeTe: electronic structure, magnetism, phonons, and superconductivity. *Phys. Rev. B* **78**, 134514 (2008).

32. Cava, R. J. et al. Oxygen stoichiometry, superconductivity and normal-state properties of YBa$_2$Cu$_3$O$_{7-\delta}$. *Nature* **329**, 423 (1987).

33. Jha, D. et al. Elemnet: deep learning the chemistry of materials from only elemental composition. *Sci. Rep.* **8**, 17593 (2018).

34. He, Y., Cubuk, E. D., Allendorf, M. D. & Reed, E. J. Metallic metal-organic frameworks predicted by the combination of machine learning methods and ab initio calculations. *J. Phys. Chem. Lett.* **9**, 4562–4569 (2018).

35. Jain, A. et al. A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.* **50**, 2295–2310 (2011).

36. Shishkin, M., Marsman, M. & Kresse, G. Accurate quasiparticle spectra from self-consistent GW calculations with vertex corrections. *Phys. Rev. Lett.* **99**, 246403 (2007).

37. Clark, S. J. & Robertson, J. Screened exchange density functional applied to solids. *Phys. Rev. B* **82**, 085208 (2010).

38. Crowley, J. M., Tahir-Kheli, J. & William, A. Goddard III. Resolution of the band gap prediction problem for materials design. *J. Phys. Chem. Lett.* **7**, 1198–1203 (2016).

39. Xie, T. & Grossman, J. C. Hierarchical visualization of materials space with graph convolutional neural networks. *J. Chem. Phys.* **149**, 174111 (2018b).

40. Zhou, Q. et al. Learning atoms for materials discovery. *Proc. Natl Acad. Sci. USA* **115**, E6411–E6417 (2018).

41. Herr, J. E., Koh, K., Yao, K., & Parkhill, J. Compressing physical properties of atomic species for improving predictive chemistry. arXiv preprint arXiv:1811.00123 (2018).

42. Pyykkö, P. Understanding the eighteen-electron rule. *J. Organomet. Chem.* **691**, 4336–4340 (2006).