scientific reports

OPEN

DFT-aided machine learning-based discovery of magnetism in Fe-based bimetallic chalcogenides

Dharmendra Pant[1], Suresh Pokharel[2], Subhasish Mandal[3], Dukka B. KC[2] & Ranjit Pati[1,4✉]

With the technological advancement in recent years and the widespread use of magnetism in every sector of the current technology, a search for a low-cost magnetic material has been more important than ever. The discovery of magnetism in alternate materials such as metal chalcogenides with abundant atomic constituents would be a milestone in such a scenario. However, considering the multitude of possible chalcogenide configurations, predictive computational modeling or experimental synthesis is an open challenge. Here, we recourse to a stacked generalization machine learning model to predict magnetic moment (μB) in hexagonal Fe-based bimetallic chalcogenides, Fe_xA_yB; A represents Ni, Co, Cr, or Mn, and B represents S, Se, or Te, and x and y represent the concentration of respective atoms. The stacked generalization model is trained on the dataset obtained using first-principles density functional theory. The model achieves MSE, MAE, and R^2 values of 1.655 $(\mu B)^2$, 0.546 (μB), and 0.922 respectively on an independent test set, indicating that our model predicts the compositional dependent magnetism in bimetallic chalcogenides with a high degree of accuracy. A generalized algorithm is also developed to test the universality of our proposed model for any concentration of Ni, Co, Cr, or Mn up to 62.5% in bimetallic chalcogenides.

Abbreviations

DFT   Density function theory
MSE   Mean squared error
MAE   Mean absolute error
ML   Machine learning
DT   Decision tree
RF   Random forest
SVM   Support vector machine
SVR   Support vector regressor
ANN   Artificial neural network
KNN   K-nearest neighbour
XGB   Extreme gradient boosting
LR   Linear regression
SD   Standard deviation
API   Application program interface
PBE   Perdew–Burke–Ernzerhof
PAW   Projector augmented wave
VASP   Vienna ab initio simulation package

Permanent magnets have been playing a great role in the development of science since their discovery[1]. They are used almost in every sector of current technology[2]. The growing awareness of green earth and renewable energy sources has also boosted the use of permanent magnets in energy sectors such as hydro-energy, wind, and wave energy as well as in electric vehicles. With the increased demand for permanent magnets, the rare-earth elements

[1]Department of Physics, Michigan Technological University, Houghton, MI 49931, USA. [2]Department of Computer Science, Michigan Technological University, Houghton, MI 49931, USA. [3]Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA. [4]Henes Center for Quantum Phenomena, Michigan Technological University, Houghton, MI 49931, USA. ✉email: patir@mtu.edu

used in manufacturing strong permanent magnets are in a critical state of running out. Numerous experimental and theoretical research works have been done to develop new magnetic materials[3–12].

For example, iron-based chalcogenides have been extensively studied, both theoretically and experimentally for their intriguing magnetic behavior[13–20]; ferromagnetism, ferrimagnetism, and antiferromagnetism are reported for different chalcogens[14,15,18]. Varying the elemental composition of the transition metal chalcogenides by changing both the metal elements as well as the chalcogens may reveal a new form of magnetic material. However, investigating all possible compositions of chalcogenides is an open challenge. The experimental investigation involves the synthesis and characterization of these materials, which are prohibitively expensive and time-consuming. Predictive calculation based on the first-principles density functional theory (DFT)[21–23] that explicitly includes electron–electron interactions within an effective single-particle picture is also numerically challenging considering a multitude of compositional configurations of chalcogenides. In a situation like this, a well-established data-driven approach could offer a faster and computationally cost-effective alternative to those expensive and time-consuming experimental or computational methods. In recent years, much data-driven research has been performed to study magnetic properties[24–29], band gaps[30,31], as well as chemical properties[32,33] of materials using machine learning models trained on DFT and experimental data. The catalytic activity of the complicated chemical system has been investigated using machine learning methods[32]. Also, the accurate predictions of band gaps in functionalized MXene exhibit the credibility of the machine learning approach[30]. Additionally, complex phenomena such as magnetic ordering, and magnetic moment have been successfully studied in 2d materials using a data-driven approach[24]. DFT-aided autonomous material search system has been designed to identify magnetic alloys[25]. Furthermore, the properties of rare-earth lean magnets are studied using DFT-aided machine learning[34]. In particular, the growing interest in studying the magnetic properties of materials using DFT-based machine learning models has highlighted the importance of DFT in the field of data-driven material science[22,25,29–32].

In this work, considering the recent advances of artificial intelligence in the multidisciplinary field of science and technology[35–38], we attempt to apply machine learning methods to develop a predictive tool that learns meaningful patterns from data and predicts the compositional dependent magnetism in Fe-based bimetallic chalcogenides Fe_xA_yB; A represents Ni, Co, Cr, or Mn, and B represents S, Se, or Te, and x and y are the concentration of respective elements. In order to develop a machine learning-based approach for identifying magnetism in Fe-based bimetallic chalcogenides, we generate a dataset of structures representing 4348 compositional configurations of Fe-based bimetallic chalcogenides Fe_xA_yB using density functional theory (DFT) calculations. We obtain magnetization of each compositional configuration using spin-polarized DFT calculation. This dataset is subsequently used to train the various ML algorithms such as Linear Regression, Support Vector Regressor[39,40], Random Forest[41], Decision Trees[42], K-nearest neighbors[43], Extreme Gradient Boosting[44,45], and Artificial Neural Network[46,47]. Based on a tenfold cross-validation[48] score, we selected the six best machine learning algorithms to develop an ensemble model based on stacked generalization for predicting magnetism in bimetallic chalcogenides. We obtained MSE, MAE, and R^2 values of 1.655 $(\mu B)^2$, 0.546 (μB), and 0.922 when we tested the final stacked model on an independent DFT test data.

Materials and methods

Our approach to discovering magnetism in Fe-based bimetallic chalcogenides is based on a supervised machine-learning approach. Initially, we generated a dataset of 4348 structures representing various compositional configurations and then performed DFT[21] calculations and obtained magnetization in the unit cell. Then, we performed feature engineering where we employed a set of descriptors (features) that are suitable for describing magnetization in the chalcogenides. We divided the dataset into training and test sets. Subsequently, we trained the model using cross-validation and grid search methods to determine the best-performing model. Finally, we tested the performance of our proposed model on the (independent) test set. Each of these four stages is briefly discussed in the following subsections.

Dataset. To create the dataset, we employed the first principles DFT[21] calculations. We started with constructing a primitive cell of hexagonal (space group p63/mmc) Iron-Sulfide (FeS) consisting of two Fe atoms and two S atoms. It has been reported that the chalcogenide can be easily synthesized in the hexagonal form as compared to the tetragonal structure[49]. Vienna ab initio simulation package (VASP)[50,51] is used for the DFT calculations; a plane wave basis with a cutoff energy of 720 eV is used. The atomic structure in the unit cell is optimized without symmetry constraint until the residual force on each atom is smaller than 0.001 eV/ Å. The convergence criterion for the total energy is set at 10^{-10} eV. The exchange and correlation are approximated using a gradient-corrected Perdew–Burke–Ernzerhof (PBE)[52] exchange–correlation functional and the electron–ion interactions are treated with the Projector Augmented Wave (PAW)[53] potential. A Monkhrost-Pack scheme with a $3 \times 3 \times 3$ K-point grid is used to sample the first Brillouin zone in the reciprocal lattice. Using the optimized lattice parameters, we expanded the primitive cell of hexagonal FeS to a bigger unit cell which consists of 16 Fe-atoms and 16 S-atoms as shown in Fig. 1.

Subsequently, we used a substitution technique to create bimetallic chalcogenides (Fe_xA_yB) of different atomic compositions; x and y represent the respective concentration of elements. Substitution technique[24,54] is a common practice in material science to create a defect as well as new material. The Fe atoms in the structures were substituted by A (Ni, Co, Cr, or Mn) and S atoms by B (Se or Te). To describe the local geometry of the structure, we designated four atomic sites by S1, S2, S3, and S4 as shown in Fig. 1. A number was assigned to these sites S1, S2, S3, and S4 depending upon how many Fe-atoms are replaced at those atomic sites keeping the chalcogen concentration unchanged. For example, when we substitute two Fe-atoms at site S1, one Fe atom at site S2, and no substitution at S3 and S4, then S1 = 2, S2 = 1, and S3 = S4 = 0. This leads to x = (16–3)/16 and y = 3/16. Based
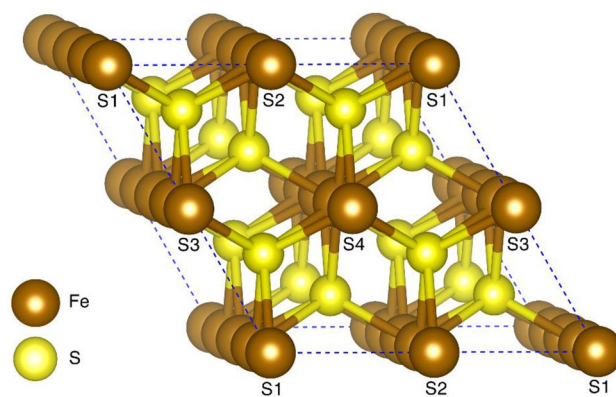
Figure 1. Unit cell of hexagonal FeS. S1, S2, S3, and S4 represent the substitutional sites for the transition metal elements.

on the number and site of substitutions, we generated 4348 bimetallic chalcogenide structures of different compositions. Subsequently, spin-polarized DFT calculations were performed to calculate magnetization for each of these compositions to obtain the data set to develop the machine learning models.

Once the dataset is created, the next step is to identify/define the descriptors or features for the problem as the choice of descriptors is one of the most important aspects of any machine learning-based approach. We choose 12 descriptors for our problem where the 8 descriptors describe the concentration of metal elements such as metal Fe, Ni, Co, Cr, Mn, and chalcogens elements S, Se, and Te. For example, in $Fe_{0.6875}Ni_{0.3125}Se$, the eight descriptors are 0.6875, 0.3125, 0, 0, 0, 1, 0, 0 representing the concentration of each element for possible bimetallic chalcogenide configurations (please see the data table in the provided GitHub link). Four descriptors describe the location of substitutional sites S1, S2, S3, and S4. The choice of these descriptors was motivated by the fact that magnetic ordering in the substituted chalcogenides is dependent upon the substitutional sites too. The calculated magnetic moment of the unit cell obtained from the DFT is taken as the target variable and these 12 descriptors are the input for the supervised machine learning framework. To understand the correlations between the features, a correlation matrix is generated (see Fig. S2 in Supplementary Information). We observed a low level of correlation between the features.

The data were randomly split into a training set and a test set in the proportion of 85:15. Subsequently, both the training and test data set are normalized separately to avoid information leakage from the test set to the training set. The size of the training and test dataset is given in Table 1.

Algorithm training and model selection. We trained seven different machine learning algorithms: Linear Regression (LR)[55], Support Vector Regressor (SVR)[39,40], Random Forests (RF)[41], Decision Trees (DT)[43], K-Nearest Neighbours (KNN)[43], Extreme Gradient Boosting (XGBoost)[56], and Artificial Neural Network (ANN). We used scikit-learn[57] and TensorFlow Keras API[58] to implement these models. To find the optimal hyperparameters for our model, we extensively performed tenfold cross-validation on the training set and grid search on different combinations of hyperparameters (See Table S1 in Supplementary Material). The description of algorithms and cross-validation techniques are briefly described in the following subsections.

We started with LR[55], a popular machine-learning model that provides the best linear fit to the data points by finding a linear relationship between features and target output by minimizing the distance between the target value and the predicted value that lies on the best-fit line. The basic LR model takes the form: $y = W^T X$, where, y is the target, $X = (1, x_1, x_2, \ldots, x_n)$ is the input feature vector, $W = (w_0, w_1, w_2, \ldots, w_n)$ is the weight vector.

For nonlinear relationships between features and the target, other algorithms such as DT, RF, KNN, SVR, XGBoost, and ANN offer better performance. DT[42] splits the training examples into a tree-like structure based on the significant splitter in the input features. The splitting results in various leaf nodes, each of which represents a different prediction. RF[41,59], which is also known for its capability in solving nonlinear problems, uses an ensemble learning approach that relies on the output of multiple decision trees. Thus, RF is a more powerful estimator as compared to DT and is less prone to overfitting and bias. The K-Nearest Neighbor (KNN)[43] is a supervised algorithm that estimates the association between features and target variables based on the average output of the other nearest K data points. In our experiment, we set the number of nearest neighbors to K = 5.

Dataset	Number of samples
Training data	3695
Independent testing data	653
Total data	4348

Table 1. Number of DFT-data points used for training and testing the machine learning models.

3

SVM[39] uses the kernel functions that transform the low dimensional data into a higher dimensional feature space such that it can find a separation hyperplane that maximizes the margin between different classes. For regression problems, SVR[40] fits the best hyperplane on training data to predict the discrete values. We have used the XGBoost[56] which is a gradient-boosting-based decision tree algorithm. It uses a gradient descent approach to minimize the loss and combines different models using an ensemble approach. XGBoost and RF have nearly similar model representations with different training algorithms. XGBoost is based on serialized base learners, whereas RF is based on parallelized base learners. We have also implemented an artificial neural network (ANN) with two hidden layers using a simple feed-forward neural network architecture that learns by comparing initial outputs with the provided target by adjusting weights and biases through backpropagation. The architecture of the ANN-based model is shown in Fig. 2. After hyperparameter tuning, we found the best-performing neural network consists of two hidden layers having 256 and 64 neurons respectively. The details of hyperparameters are given in Table S1 (Supplementary Material). We have also analyzed how the features are contributing to a model prediction. Based on these calculations, which are performed using the random forest model, the concentrations of elements such as Mn, Cr, Te, and S are found to be the most dominant features as shown in Fig. S3 (Supplementary Information).

K-fold cross validation. To search for the best hyperparameters and to compare the performance of different models, we implemented a K-fold cross-validation algorithm on the training data. The K-fold cross-validation relies on a data partitioning scheme to ensure that the model can generalize the pattern on a diverse dataset. In this method, the dataset is randomly divided into K different sets. Following this, the model is trained using K-1 sets of the dataset and tested against the remaining set. The process is repeated K times and the results are statistically analyzed to choose the best-performing model. In this work, each model is trained and fine-tuned using K = 10 through a grid search process.

Stacked generalization. After examining the individual models, we combined the best-performing models using a stacked generalization[60] algorithm to improve the predictive performance. Stacked generalization is an ensemble approach that combines two or more pre-trained models (base learners) followed by a second-level regression model (meta learner). In this method, we stacked six base learners (DT, RF, SVR, XGB, KNN, and ANN) followed by a meta learner (RF) as shown in Fig. 3. It is noteworthy to point out that LR was omitted from the stacked generalization as its performance was not satisfactory.

Despite being a powerful technique that relies on the strength of multiple models, the stacked generalization approach is more prone to data leakage while performing cross-validation because the same dataset is used to train both the base models and meta-regression models resulting in model overfitting. The data leakage and overfitting in the cross-validation stage may mislead the model selection process. Hence, in this work, we implemented a stacking algorithm with cross-validation proposed by Wolpert[60] to prevent data leakage and overfitting. In this technique, initially, the data is randomly divided into K sets. In the first stage, the base models are trained using K-1 sets and score-level features are extracted from those base models using the remaining set of training data. The process is repeated K times and each time a new dataset is prepared using the score-level features of the base models. In the second stage, a meta-regression model is trained using the data constructed from the first stage. Finally, each base model is trained using the entire training dataset and stacked together, which is subsequently connected to the previously trained meta-regression model (RF) to form the final model. It should
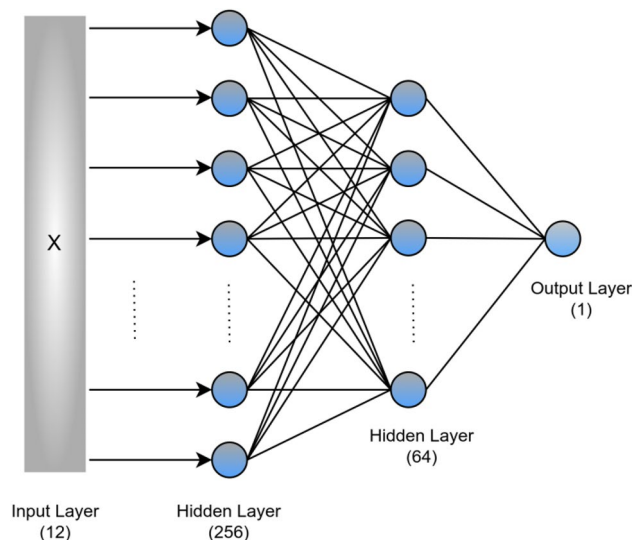

Figure 2. The architecture of an artificial neural network with two hidden layers with 256 and 64 neurons respectively.
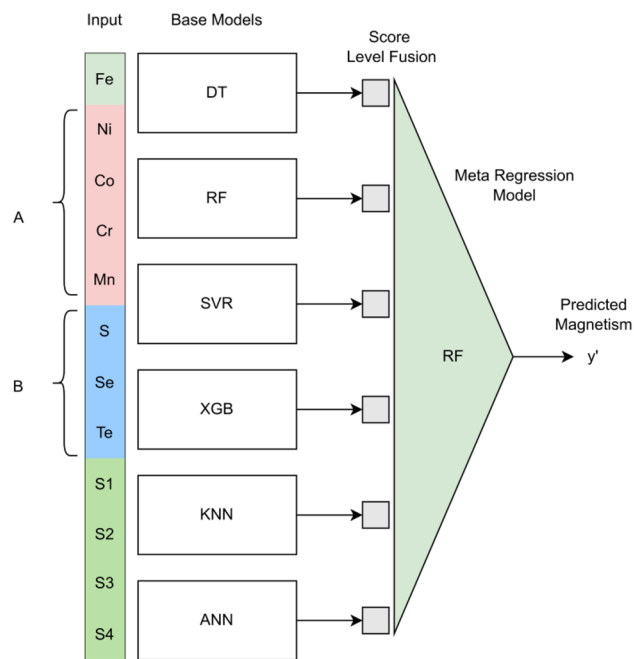
Figure 3. Block diagram for stack generalization.

be noted that in this work, each base learner and meta learner are trained and evaluated independently using the tenfold cross-validation method.

Performance evaluation. To evaluate the performance of different machine learning regression models, we use three evaluation metrics: mean square error (MSE), mean absolute error (MAE), and coefficient of determination (R^2), which are defined below.

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

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

$$\text{Coefficient of determination } \left(R^2 \right) = 1 - \frac{\sum_{i}^{n} \left(y_i - \hat{y}_i \right)^2}{\sum_{i}^{n} \left(y_i - \overline{y}_i \right)^2}$$

where, y_i, $\hat{y}_i$ and $\overline{y}$ are the true, predicted, and the average value of y respectively.

MSE measures the average of the squares of residuals, while MAE measures the average of the residuals. They both have positive values; a smaller value indicates less error and better performance. MSE penalizes the model with larger errors than the MAE and hence is more sensitive to the outliers in the data. The lower the values of MSE and MAE are, the better the predictive performance of a model. The R^2 score, also known as the coefficient of determination, is also a statistical measure in a regression model that represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s). Its value lies between 0 and 1. Since R^2 alone does not measure the accuracy of the predictions[61,62], we have used this metric in conjunction with MSE and MAE to measure the performance of the regression models used in our study.

Results and discussion

First, we performed some exploratory data analysis on the DFT dataset. It reveals that the bimetallic chalcogenides containing S and Se are found to have higher magnetic moments compared to those containing Te as shown in Fig. 4. This fact is supported by earlier research[49] that revealed FeS and FeSe exhibiting stronger magnetization than FeTe. Furthermore, we found that Fe-chalcogenides containing Cr and S have higher magnetic moments than those containing other transition metal elements (Ni, Co, or Mn) and chalcogen elements (Se and Te) as shown in Fig. 4. An increase in the magnetic moment is also noticeable when Cr or Mn concentration increases in chalcogenides containing S or Se (see Fig. S4 in Supplementary Information).

Further, the substitutional sites of transition metal elements in the chalcogenides are found to influence the target value (magnetic moment).
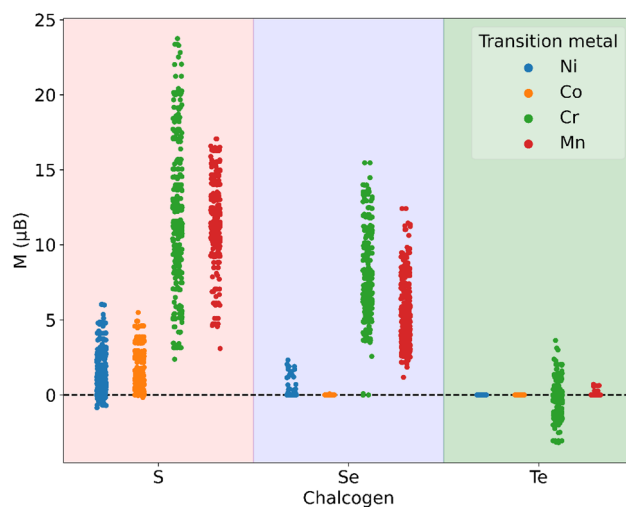
Figure 4. Dot plot showing the magnetic moment of Fe-based bimetallic chalcogenides; Fe_xA_yB where A represents Ni, Co, Cr, or Mn, and B represents S, Se, or Te, and x and y represent the concentration of respective atoms. Three shaded regions differentiate the magnetic moments of the chalcogenides containing S, Se, and Te respectively. Blue, orange, green, and red color dots correspond to the value of magnetic moments of the chalcogenides containing transition metal Ni, Co, Cr, and Mn respectively. Bimetallic chalcogenide with Cr and S exhibits a higher range for the magnetic moments.

Tenfold cross-validation results of base learners and meta learners. Table 2 shows the performance comparison of various machine learning models on the training dataset mentioned in Table 1. The RF model is found to perform well based on the comparison of mean MSE, mean MAE, and mean R^2. The detailed calculation of performance measures is provided in Tables S2, S3 and S4 (Supplementary Information). Except for the LR approach, all models perform reasonably well, which could be understood from the plausible non-linear relationship between the target variable and the features in our dataset. The well-performing models are subsequently used to develop the final stacked model.

Next, to find the best meta-regression model for the stacked generalization approach, we used the output from the individual base models to train LR, RF, and XGB and recalculated mean MSE, mean MAE, and mean R^2. The results are presented in Table 3. RF model is found to be the best meta-learner.

Evaluation of the final stacked model on an independent (DFT) test dataset. Finally, we trained the models using the entire training dataset and tested them on the independent test dataset. Upon testing against

Model name	Mean MSE(SD)	Mean MAE(SD)	Mean R²(SD)
Linear regression (LR)	7.00 (0.48)	2.16 (0.07)	0.69(0.02)
Decision tree (DT)	1.67 (0.32)	**0.54**(0.05)	0.93(0.01)
Random forest (RF)	**1.44** (0.21)	0.55(0.04)	0.93(0.01)
Support vector regressor (SVR)	1.62 (0.18)	0.57(0.03)	**0.92**(0.01)
Extreme gradient boosting (XGB)	1.58 (0.21)	0.68(0.04)	0.93(0.01)
K-nearest neighbour (KNN)	1.74 (0.30)	0.63(0.04)	**0.92**(0.01)
Artificial neural network (ANN)	1.45 (0.19)	0.59(0.04)	0.93(0.01)

Table 2. 10-Fold cross-validation results of different machine learning models. The best value in each column is highlighted in bold.

Model name	Mean MSE(SD)	Mean MAE(SD)	Mean R²(SD)
Linear regression (LR)	1.40 (0.19)	0.56 (0.04)	0.94(0.01)
Extreme gradient boosting (XGB)	1.36 (0.27)	0.54(0.04)	0.94(0.01)
Random forest (RF)	**1.29** (0.28)	**0.50** (0.05)	**0.94** (0.01)

Table 3. 10-Fold cross-validation results of various meta-learners. Significant values are in bold.

the independent 653 DFT-test data points (Table 1), the MSE, MAE, and R^2 of the stacked model are found to be 1.655 $(\mu B)^2$, 0.546 μB, and 0.922 respectively compared to tenfold cross-validation values of 1.29 $(\mu B)^2$, 0.50 μB and 0.94. These results show that our final model performs equally well on the test data as it did during the validation indicating the generalizability of the approach.

Figure 5 shows the comparison between the true magnetization and predicted magnetization for each data point on the independent DFT test dataset. We have sorted the test data points in ascending order based on the value of the magnetic moment obtained from the DFT (true value). One can notice that the predictive performance of the model is much better for M < 14 μB. This is expected as only 3% of the entire training dataset is available to train the machine learning model for M > 14 μB. Though we noticed over and under prediction of magnetic moments in some instances, our model identifies non-magnetic and magnetic chalcogenides with a high degree of certainty. A deeper analysis of the under and over-predicted region reveal that a limited number of data points having similar target value are used during the training of the model, which explains the variation in the predicted target value (M) in those regions. Nevertheless, our study shows that the complex electronic interactions involved in the DFT calculations are well captured by the purposed model to predict magnetism in bimetallic chalcogenides.

Expanding the applicability of the proposed model.   In this work, a unit cell of FeS having 16 Fe and 16 S atoms is used to generate the DFT dataset. As a result, we have predefined fixed values of y (0.0625, 0.125, 0.1875, 0.25, 0.3125, 0.375, 0.4375, 0.5, 0.5625, 0.625) and x (= 1 − y) as multiples of 1/16. Furthermore, the descriptors S1, S2, S3, and S4 can each take integer values from 0 to 4. To overcome this limitation and increase the flexibility of our model, we have developed a generalized algorithm that can take the concentration y (or x) in percentages with a constraint of $0 < y < 0.625$. It also allows the user to choose the concentration of substituted atoms on the atomic sites S1, S2, S3, and S4 in percentages. Then, the algorithm calculates the input features y, x, S1, S2, S3, and S4 in the suitable form required to feed the ML model. The detailed procedure is explained in Algorithm 1. The implementation of the algorithm is available in the provided GitHub link.
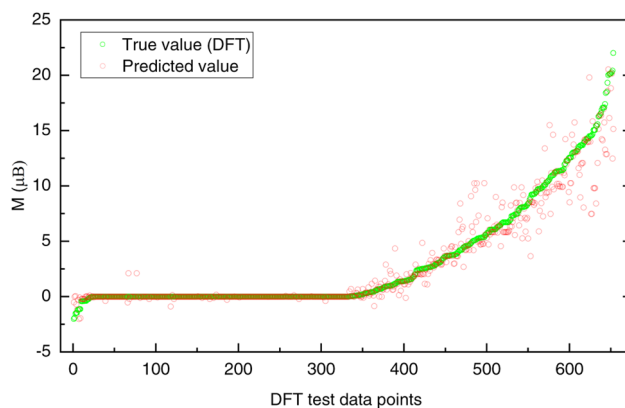

Figure 5. Scatter plots showing true (green circle) versus predicted (red circle) magnetic moments (M) in bimetallic chalcogenides. Data points from the independent test dataset, sorted in ascending order based on the value of the magnetic moment (M) obtained from DFT (true value).

Algorithm 1: Generalization Algorithm for a wide range of inputs

---

Inputs:

Read y

Where y is the percentage of A in Fe_xA_yB.

Constraints:

  i. $A \in \{Ni, Cr, Co, Mn\}, B \in \{S, Se, Te\}$

  ii. $x+y \leftarrow 100$, where x is the percentage of Fe in Fe_xA_yB.

  iii. $0 < y \leq 62.5$.

Read S_i for $i \leftarrow 1$ to 4

Where S_i is the percentage of element A on i^{th} site of the structure Fe_xA_yB.

Constraints:

  i. $S_i \leq \frac{1}{4*y} \times 100 \times 100$

  ii. $\sum_{i \leftarrow 1}^{4} S_i = 100$

Step 1:

Convert percentages into real numbers to fit our pre-trained model

$$y' \leftarrow y \times \frac{16}{100}$$

$$x' \leftarrow 16 - y'$$

Step 2 (Reduction step):

Convert percentage values to the original range of 16

for $i \leftarrow 1$ to 4

$$S_i' \leftarrow y' \times \frac{S_i}{100}$$

Step 3:

Input $\frac{x'}{16}, \frac{y'}{16}$, and S_i' as the concentration of Fe, the concentration of A, and the number of atoms A on sites i=1 to 4.

Conclusions

The quest for new magnetic materials that are cheaper than the rare-earth element-based magnets is of significant interest in current years for their application that ranges from data storage to automotive vehicles to biomedical fields to the green energy sector. Experimental and computational investigation of possible alternative magnetic materials is expensive and time-consuming. In this work, we have used a data-driven framework that would accelerate the discovery of new magnetic materials. We used an optimized FeS structure and employed a substitution technique to design new bimetallic chalcogenides of different compositions. The first principle DFT is used to generate training and test data. After training and evaluating several machine learning models, we have developed a stacked machine learning model combining several best-performing base models as the final predictive tool. The final model has shown a high degree of accuracy on the independent DFT-test data with MSE, MAE, and R^2 values of 1.655 $(\mu B)^2$, 0.546 μB, and 0.922 respectively. Additionally, we have developed a generalized algorithm to expand the applicability of our model to a wide range of inputs. The predicted data reveal the Fe-based bimetallic chalcogenides containing chalcogen element S and a higher concentration of transition metal Cr yielding higher magnetic moments than the other configurations, which is consistent with the DFT data. This work presents a strategy for discovering a new magnetic material made from less expensive and more abundant elements that would eventually replace the costly existing magnetic materials made out of rare-earth metals.

Data availability

References

1. Coey, J. M. D. *Magnetism and Magnetic Materials* (Cambridge University Press, 2001). https://doi.org/10.1017/CBO9780511845000.
2. Thompson, D. A. & Best, J. S. The future of magnetic data storage techology. *IBM J. Res. Dev.* **44**, 311–322 (2000).
3. Balasubramanian, B., Das, B., Skomski, R., Zhang, W. Y. & Sellmyer, D. J. Novel nanostructured rare-earth-free magnetic materials with high energy products. *Adv. Mater.* **25**, 6090–6093 (2013).
4. Kuz'min, M. D., Skokov, K. P., Jian, H., Radulov, I. & Gutfleisch, O. Towards high-performance permanent magnets without rare earths. *J. Phys. Condens. Matter* **26**, 064205 (2014).
5. Lottini, E. *et al.* Strongly exchange coupled core|shell nanoparticles with high magnetic anisotropy: A strategy toward rare-earth-free permanent magnets. *Chem. Mater.* **28**, 4214–4222 (2016).
6. Anagnostopoulou, E. *et al.* Dense arrays of cobalt nanorods as rare-earth free permanent magnets. *Nanoscale* **8**, 4020–4029 (2016).
7. Balamurugan, B., Das, B., Zhang, W. Y., Skomski, R. & Sellmyer, D. J. Hf–Co and Zr–Co alloys for rare-earth-free permanent magnets. *J. Phys. Condens. Matter* **26**, 064204 (2014).
8. Gao, T. R. *et al.* Combinatorial exploration of rare-earth-free permanent magnets: Magnetic and microstructural properties of Fe–Co–W thin films. *Appl. Phys. Lett.* **102**, 022419 (2013).
9. Pant, D., Aryal, S., Mandal, S. & Pati, R. Emergence of ferromagnetism due to spontaneous symmetry breaking in a twisted bilayer graphene nanoflex. *Nano Lett.* **21**, 7548–7554 (2021).
10. Vishina, A. *et al.* High-throughput and data-mining approach to predict new rare-earth free permanent magnets. *Phys. Rev. B* **101**, 094407 (2020).
11. Gao, Q., Opahle, I., Gutfleisch, O. & Zhang, H. Designing rare-earth free permanent magnets in Heusler alloys via interstitial doping. *Acta Mater.* **186**, 355–362 (2020).
12. Pant, D. & Pati, R. Phase transition from a nonmagnetic to a ferromagnetic state in a twisted bilayer graphene nanoflake: The role of electronic pressure on the magic-twist. *Nanoscale* **14**, 11945–11952 (2022).
13. Liu, A., Chen, X., Zhang, Z., Jiang, Y. & Shi, C. Selective synthesis and magnetic properties of $FeSe_2$ and $FeTe_2$ nanocrystallites obtained through a hydrothermal co-reduction route. *Solid State Commun.* **138**, 538–541 (2006).
14. Kang, L. *et al.* Phase-controllable growth of ultrathin 2D magnetic FeTe crystals. *Nat. Commun.* **11**, 3729 (2020).
15. Deng, Y. *et al.* Gate-tunable room-temperature ferromagnetism in two-dimensional Fe_3GeTe_2. *Nature* **563**, 94–99 (2018).
16. Oyler, K. D., Ke, X., Sines, I. T., Schiffer, P. & Schaak, R. E. Chemical synthesis of two-dimensional iron chalcogenide nanosheets: FeSe, FeTe, Fe(Se, Te), and $FeTe_2$. *Chem. Mater.* **21**, 3655–3661 (2009).
17. Matthews, P. D., Akhtar, M., Malik, M. A., Revaprasadu, N. & O'Brien, P. Synthetic routes to iron chalcogenide nanoparticles and thin films. *Dalton Trans.* **45**, 18803–18812 (2016).
18. Nakamura, K., Ito, T. & Freeman, A. J. Half-metallic ferrimagnetism in zincblende Mn-doped transition metal chalcogenides. *Phys. Rev. B* **72**, 064449 (2005).
19. Dong, C. *et al.* Revised phase diagram for the $FeTe_{1-x}Se_x$ system with fewer excess Fe atoms. *Phys. Rev. B* **84**, 224506 (2011).
20. Böhmer, A. E. *et al.* Origin of the tetragonal-to-orthorhombic phase transition in FeSe: A combined thermodynamic and NMR study of nematicity. *Phys. Rev. Lett.* **114**, 027001 (2015).
21. Parr, R. G. & Yang, W. *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1994).
22. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
23. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
24. Rhone, T. D. *et al.* Data-driven studies of magnetic two-dimensional materials. *Sci. Rep.* **10**, 15795 (2020).
25. Iwasaki, Y. *et al.* Machine learning autonomous identification of magnetic alloys beyond the Slater-Pauling limit. *Commun. Mater.* **2**, 31 (2021).
26. Kusne, A. G. *et al.* On-the-fly machine-learning for high-throughput experiments: Search for rare-earth-free permanent magnets. *Sci. Rep.* **4**, 6367 (2015).
27. Halder, A., Ghosh, A. & Dasgupta, T. S. Machine-learning-assisted prediction of magnetic double perovskites. *Phys. Rev. Mater.* **3**, 084418 (2019).
28. Möller, J. J., Körner, W., Krugel, G., Urban, D. F. & Elsässer, C. Compositional optimization of hard-magnetic phases with machine-learning models. *Acta Mater.* **153**, 53–61 (2018).
29. Merker, H. A. *et al.* Machine learning magnetism classifiers from atomic coordinates. *iScience* **25**, 105192 (2022).
30. Rajan, A. C. *et al.* Machine-learning-assisted accurate band gap predictions of functionalized MXene. *Chem. Mater.* **30**, 4031–4038 (2018).
31. Zhuo, Y., Mansouri Tehrani, A. & Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **9**, 1668–1673 (2018).
32. Wan, X., Zhang, Z., Yu, W. & Guo, Y. A density-functional-theory-based and machine-learning-accelerated hybrid method for intricate system catalysis. *Mater. Rep. Energy* **1**, 100046 (2021).
33. Jinnouchi, R. & Asahi, R. Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning algorithm. *J. Phys. Chem. Lett.* **8**(17), 4279–4283. https://doi.org/10.1021/acs.jpclett.7b02010 (2017).
34. Halder, A., Rom, S., Ghosh, A. & Saha-Dasgupta, T. Prediction of the properties of the rare-earth magnets $Ce_2 Fe_{17-x} Co_x CN$: A combined machine-learning and Ab Initio study. *Phys. Rev. Appl.* **14**, 034024 (2020).
35. Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M. & Fazzio, A. From DFT to machine learning: Recent approaches to materials science–a review. *J. Phys. Mater.* **2**, 032001 (2019).
36. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
37. The rise of data-driven modeling. *Nat. Rev. Phys.* **3**, 383–383 (2021).
38. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5**, 83 (2019).
39. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
40. Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **14**, 199–222 (2004).
41. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
42. Breiman, L. *Classification and Regression Trees* 1st edn. (Routledge, 1984). https://doi.org/10.1201/9781315139470.
43. Bentley, J. L. Multidimensional binary search trees used for associative searching. *Commun. ACM* **18**, 509–517 (1975).
44. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *JSTOR* **29**, 1189–1232 (2001).

45. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2013).
46. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).
47. Specht, D. F. A general regression neural network. *IEEE Trans. Neural Netw.* **2**, 568–576 (1991).
48. Gareth, J., Daniela, W., Trevor, H. & Robert, T. *An Introduction to Statistical Learning* (Springer, 2013).
49. Parker, D. S. Strong 3D and 1D magnetism in hexagonal Fe-chalcogenides FeS and FeSe vs. weak magnetism in hexagonal FeTe. *Sci. Rep.* **7**, 3388 (2017).
50. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
51. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
52. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
53. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
54. Picard, C. J. & Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter* **23**, 053201. https://doi.org/10.1088/0953-8984/23/5/053201 (2011).
55. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer, 2016).
56. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, New York, USA https://doi.org/10.1145/2939672.2939785 (2016).
57. Pedregosa, F., Varoquaux, G., Gramfort, A. & Michel, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
58. TensorFlow Developers. *TensorFlow*. 10.5281/ZENODO.4724125 (2022).
59. Geurts, P., Ernst, D. & Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **63**, 3–42 (2006).
60. Wolpert, D. H. Stacked generalization. *Neural Netw.* **5**, 241–259 (1992).
61. Spanos, A. *Probability Theory and Statistical Inference: Empirical Modeling with Observational Data* (Cambridge University Press, 2019).
62. Hagquist, C. & Stenbeck, M. Goodness of fit in regression analysis: R2 and G2 reconsidered. *Qual. Quant.* **32**, 229–245 (1998).

Acknowledgements

Author contributions

R.P. designed and supervised the research. D.P. designed bimetallic chalcogenides and performed DFT calculations. S.M. also performed DFT calculations. D.P. collected DFT data, performed data preprocessing, data exploration, designed and developed base ML models. D. P. and S.P. designed and developed stacked ML models and generalized algorithm. D. K. oversaw the ML aspects of the whole project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-30438-w.

Correspondence and requests for materials should be addressed to R.P.

Reprints and permissions information is available at www.nature.com/reprints.

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