Element selection for functional materials discovery by integrated machine learning of atomic contributions to properties

A. Vasylenko[1], D. Antypov[1], V. Gusev[1], M. W. Gaultois[1], M. S. Dyer[1], M. J. Rosseinsky[1*]

[1] Department of Chemistry, University of Liverpool, Crown Street L697ZD, UK

*corresponding author

Abstract

At the high level, the fundamental differences between materials originate from the unique nature of the constituent chemical elements. Before specific differences emerge according to the precise ratios of elements (composition) in a given crystal structure (phase), the material can be represented by its phase field defined simply as the set of the constituent chemical elements. Classification of the materials at the level of their phase fields can accelerate materials discovery by selecting the elemental combinations that are likely to produce desirable functional properties in synthetically accessible materials. Here, we demonstrate that classification of the materials' phase field with respect to the maximum expected value of a target functional property can be combined with the ranking of the materials' synthetic accessibility.

This end-to-end machine learning approach (PhaseSelect) first derives the atomic characteristics from the compositional environments in all computationally and experimentally explored materials, and then employs these characteristics to classify the phase field by their merit. PhaseSelect can quantify the materials' potential at the level of the periodic table, which we demonstrate with significant accuracy for three avenues of materials' applications: high-temperature superconducting, high-temperature magnetic and targetted energy band gap materials.

Introduction

Conceptualization of novel materials begins at the level of the periodic table with selection of chemical elements for synthetic investigation. There is a variety of possible ratios or compositions that can be formed from a set of chemical elements leading to different materials (phases); the field of these potential realizations can be defined as a material's phase field. The choice of a phase field to investigate ultimately determines the outcome of the synthetic work and the functional properties of the prospective materials.

The fundamental differences between atoms result in the variance in the materials' properties in thousands of compositions accumulated in materials databases[1–3]. Harvesting these statistical data, there has been a surge of machine learning (ML) methods aiming to predict the materials' properties from the knowledge of their structures and compositions[4,5]. Ranging from formation enthalpy[6] to energy band gap[7] to superconducting transition temperature[8], ML predictions enable fast screening of functional inorganic materials at scale, overcoming the otherwise forbidding combinatorial challenge for precise, but significantly more resource-demanding high-throughput quantum-mechanical calculations. At the same time, most of these high-performance ML models are based on the deep learning[9] or ensembles[10] methods that lack interpretability[11], hence they are not readily adopted by experimental teams. Improvement of interpretability of ML approaches without compromises on performance could bridge powerful ML methods with experimental workflows to form trusted ML-expert systems in material sciences.

Codification of the materials for statistical treatment involves description of the constituent chemical elements, often represented as vectors of their chemical and physical characteristics, that are combined linearly to describe a compound[6]. This approach relies on the expert selection of a number of exploited chemical characteristics as well as the relevance of these characteristics and the corresponding weights for the atomic descriptions in materials representations. This selection determines the quality of the model[12]. The composition-based models are predisposed to data leakage between training and

validation datasets via compositionally close datapoints, that impedes the extrapolation of patterns in materials-properties relationships onto unexplored materials that have distinct chemistries from those in the training set[13].

In this work, our goal is to assess the attractiveness of candidate functional materials at the high level of the periodic table by identifying the most promising phase fields that are likely to contain these candidates. This circumvents the combinatorial challenge of individual assessment of all possible compositions built from the chosen elements and aligns with the experimental challenge of identifying new functional materials from previously uncharted chemistry. We demonstrate that unsupervised learning of chemical elements combined with the attention technique for learning elemental contributions can be used for the accurate classification of the materials' functional performance at the level of the phase fields, while improving interpretability of the ML reasoning. This end-to-end integrated machine learning (PhaseSelect) of the materials databases can prioritise the materials with respect to both probability of a merit (maximum achievable value of the target property) and synthetic accessibility of the phase fields, while the existing vast chemical knowledge is learnt each time in the context of the specific target material function.

In our approach, the machine learns all atomic elements and their specific characteristics responsible for materials formation. This is achieved by exploring possible compositional combinations in all theoretically and experimentally studied materials[14], similarly to the concept in reference[15]. For each atom, a machine learns a vector, that encodes atomic characteristics learnt from the co-existence of atoms within some compositional environments and the absence of such co-existence with others. Thus built atomic vectors are then combined linearly to form a phase field representation, whereas attention mechanism[16] is trained to derive the weights to the atomic vectors that magnify the most prominent atomic contributions specific to a particular property. This offers a statistically-derived alternative to the expert knowledge-based manual selection of relevant chemical characteristics and their contributions, and enables the high-level ranking and classification of materials for functional

applications. Furthermore, by aggregating compositions into the phase fields in the input data, this high-level approach eliminates concerns of data leakage at the compositional level as all compositions within a phase field represent a single data entry.

We demonstrate a significant accuracy of PhaseSelect in classification of the materials with respect to three different properties: superconducting transition temperature, Curie temperature, and energy band gap, when learning the relevant property from SuperCon[3] and Materials Platform for Data Science (MPDS)[1] databases. Within these training and test sets, each phase field is labelled according to the maximum reported value of all materials within it. This maximum value is compared to the chosen thresholds (10K, 300K, 4.5eV) that reflect practical interests in high-temperature superconducting, magnetic materials and dielectrics respectively, and a class label is allocated accordingly.

In these applications, PhaseSelect demonstrates 80.4, 86.2, 75.6 % accuracy and 72.9, 84.2, 75.3% F1 score respectively. Furthermore, the phase field representations derived during properties classification are exploited to recognise patterns in elemental combinations that afford stable compositions in material databases and produce the ranking of synthetic accessibility for unexplored phase fields. The arising metrics of the phase fields – the merit probability (probability of achieving a high value of a property) and synthetic uncertainty (accessibility ranking) – can be orthogonally applied to any combination of elements at scale, creating a map of potentially attractive phase fields that can provide guidance to human researchers in the consequential and costly choice of phase fields for investigations and discovery of functional materials.

Results and discussion

PhaseSelect model architecture

At the level of the phase fields, relationships between elemental combinations and their synthetic accessibility have been studied with unsupervised machine learning and validated experimentally[12]. Here, we employ an integrated statistical description of atomic elements and their combinations to

learn what elemental combinations have high probabilities of both synthetic realization and high values of target properties. The architecture of the model is illustrated in Figure 1.

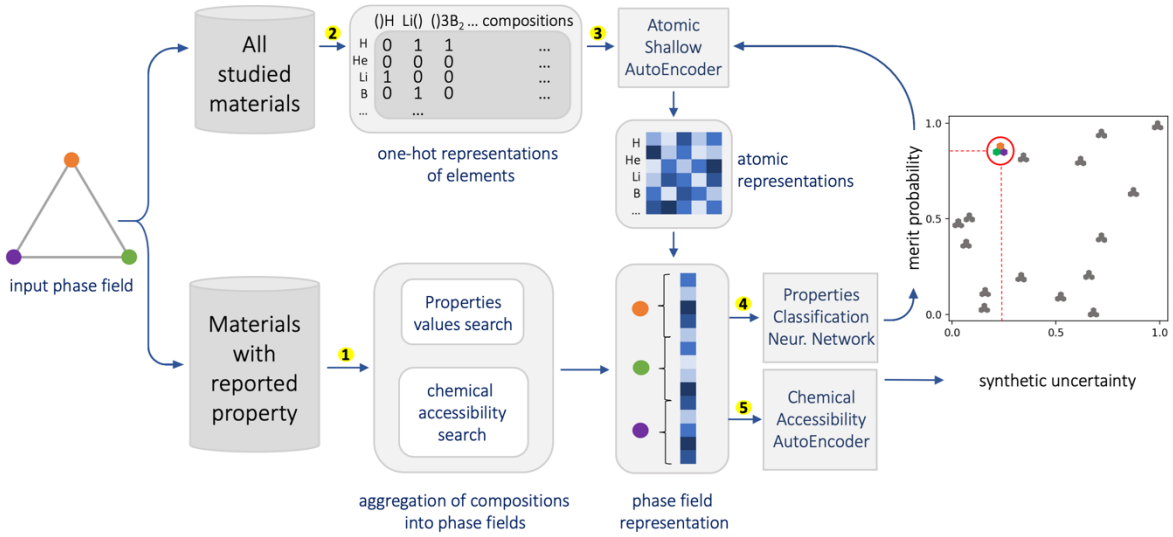


Figure 1. PhaseSelect predicts properties and chemical accessibility of phase fields. Model architecture. Arrows show the information flow between the various components described in this paper: 1) experimentally confirmed compositions are aggregated into the phase fields; the maximum values of the properties in the phase fields are selected; 2) compositional environments (elemental co-occurrence in materials) are aggregated from all theoretically and experimentally studied materials; 3) unsupervised learning of atomic representation from data collected in 2); 4) supervised classification of phase fields by maximum achievable values of the properties; the predicted probability of entering the high-value class is used as a merit probability; 5) unsupervised ranking of the phase fields by synthetic uncertainty; metrics derived in 4) and 5) result in a map of the phase fields' likelihood to form stable compounds with desired properties. The model is trained end-to-end so the losses of learning the atomic representation (3) and classification (4) are minimised simultaneously.

PhaseSelect consists of several connected modules (depicted as the sharp-corner rectangles in Figure 1) that pass information from the databases, while transforming the data (different data representations are depicted as the rounded-corner windows in Figure 1) and are trained simultaneously, while minimising the compound loss. We describe the data processing and the mechanisms of these modules in the following sections.

Aggregation of compositions into phase fields

For the classification and accessibility ranking of the phase fields (See bottom stream in Figure 1) we process the materials databases, where experimentally verified values of the target property are reported for a large number of compositions[1,3]. Materials built from the same constituent elements are aggregated into one phase field, with the associated property value corresponding to the maximum reported property value among all reported materials within this phase field. For example, in the SuperCon database, there are many compositions reported in Y-Ba-Cu-O phase field with a high critical temperature, including $YBa_2Cu_3O_7$ (T_c = 93 K) and $Y_3Ba_5Cu_8O_{18}$ (T_c = 100.1 K) – the highest reported temperature in Y-Ba-Cu-O. Hence, Y-Ba-Cu-O enters the data for training our classification model for superconductors with 100.1 K as the corresponding maximum value. Aggregation of materials with reported superconducting transition temperature, Curie temperature and energy band gap forms three datasets with 4826, 4753 and 40452 phase fields respectively. Division of the datasets into two classes by the threshold values for the corresponding properties – 10 K, 300 K and 4.5 eV for superconducting transition temperature, Curie temperature and energy band gap, respectively – forms reasonably balanced data classes with 3311:1515, 2726:2027 and 20910:19690 phase fields, respectively, with data distributions illustrated in Figure 2a-c. Furthermore, the remaining imbalances are taken into account by class-weighting in the corresponding classification models[17]. The rapidly decreasing number of explored phase fields with reported superconducting properties at temperatures above 10 K (See Figure 2b) proves development of reliable models for classification with respect to temperatures higher than 10 K challenging (See Supplementary Fig. 1)[8]. Nevertheless, despite the broad aggregation of high-temperature superconducting materials into a single class (with T_c > 10 K), accurate classification of unexplored materials into the two classes divided by the chosen threshold value would allow fast screening for novel high-temperature superconductors. Similarly, a binary classification enables fast screening of novel materials for applications as high-temperature magnetic materials and targetted band gap materials.

Across the three property datasets, the phase fields are formed from up to 12 constituent elements, with the majority of data represented by ternary, quaternary and quinary phase fields (See Figure 2d). The abundance of chemical elements among the explored materials in the databases is illustrated in Figure 2e. All datasets have similar trends with peaks for materials containing, e.g., carbon, oxygen, sulphur, with an especially pronounced match between elemental distribution in datasets with materials for superconducting and magnetic applications (See inset in Figure 2e). The data distributions across different chemical elements observed in Figure 2e, reflect the biases in the input data: e.g., magnetism is associated with Fe predominantly, while superconductivity with Cu, etc.

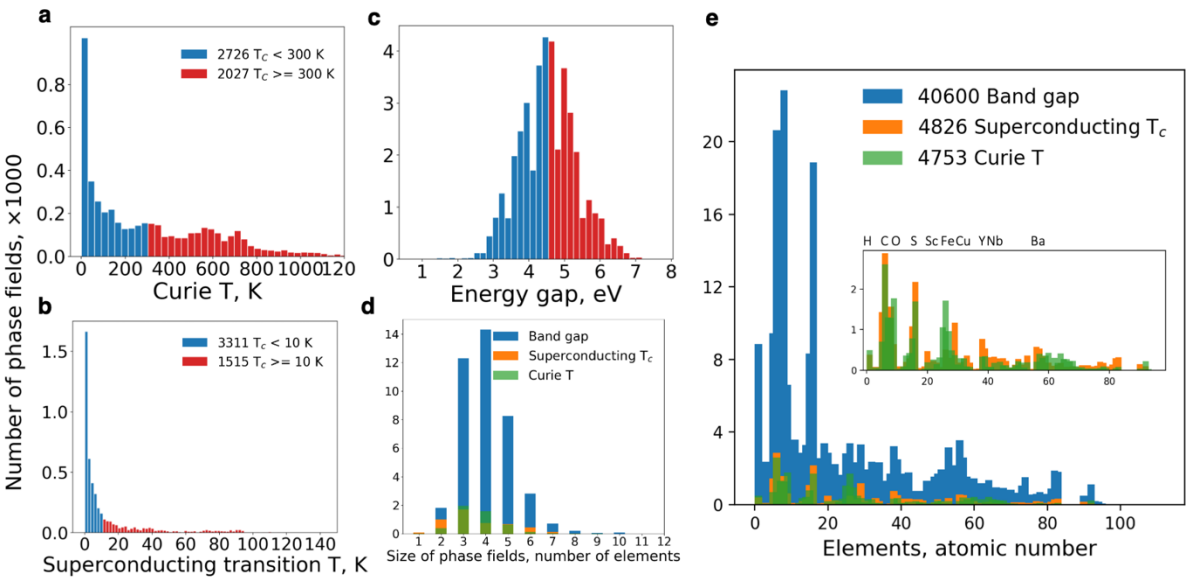


Figure 2. Aggregation of compositions into phase fields. a Distribution of phase fields of magnetic materials in MPDS[1] with respect to the maximum associated Curie temperature T_C. The materials' classes "low-temperature" and "high-temperature" magnets are divided at T_C = 300 K as 2726:2027 phase fields. **b** Distribution of phase fields of superconducting materials (joined datasets from SuperCon[3] and MPDS) with respect to the maximum associated superconducting transition temperature T_c. The materials' classes "low-temperature" and "high-temperature" superconductors are divided around T_c = 10 K as 3311:1515 phase fields. **c** Distribution of phase fields of materials with reported value of energy gap in MPDS with respect to the maximum associated band gap. The materials' classes "small-gap" and "large-gap" are divided around E = 4.5 eV as 20910:19690 phase fields. **d** Distributions of materials with respect to the number of constituent elements are similar for all datasets: the majority of the reported

compositions belong to ternary, quaternary and quinary phase fields. **e** Content of individual chemical elements among the explored materials in the databases; the total numbers of phase fields in the corresponding datasets are given in the legend. All datasets have similar trends with pronounced peaks for materials containing, e.g. carbon, oxygen, silicon. The inset illustrates overlap in trends for elemental distribution in explored materials for superconducting and magnetic applications.

Description of atoms by means of their compositional environments, which are shared by chemically similar elements, should mitigate the biases in the data accumulated over time due to the focused studies of particular families of materials.

Atomic representation and phase field representation

To learn atomic characteristics from the compositional environments – explored chemical compositions, where the atoms are found to form the variety of stable and metastable materials – we build a module for atomic representation based on a large materials database that includes both experimental and theoretical materials[14,15]. For each chemical element one can build a one-hot encoding vector from its instances in the database. The database is expanded into a table similarly to the approach proposed in reference[15] (depicted as a matrix of coexisting elements and compositional environments in the materials in Figure 1, 2)). The rows of the table correspond to the chemical elements, the columns are the remainders of the compositional formulas of the reported compounds, which we define here as compositional environments. For example, from stability of Li_3PO_4 we can learn about its constituent elements, Li, P, O and their compositional environments, "()3PO4", "()Li3O4" and "()4Li3P" respectively. In this notation, empty parentheses denote an element that by combining with the compositional environment forms a composition. Similarly, all alkali metals form the tri-"element" phosphates with "()3PO4", while trivalent elements do not, as they form the one-"element" phosphates with "()PO4" instead. In the proposed matrix representation[15], the intersections of the rows for elements with the columns for compositional environments are filled with ones if the resulting composition is reported in [14] and with zeros otherwise. The resulting sparse matrix represents coexistence of the chemical elements and compositional environments in the materials. We then

employ a shallow autoencoder neural network – an unsupervised ML technique – to reduce the dimensionality of this matrix, and to condense the information into the rich latent space of dimensionality k, in which similar atomic vectors (of length k) are grouped close to each other. We study the effects of the size of dimensionality k of thus derived atomic vectors on the classification accuracy to select the most efficient atomic description (Supplementary Fig. 1). We use the vectors of the most efficient latent space as atomic representations to build up the phase fields descriptions (Figure 3a).

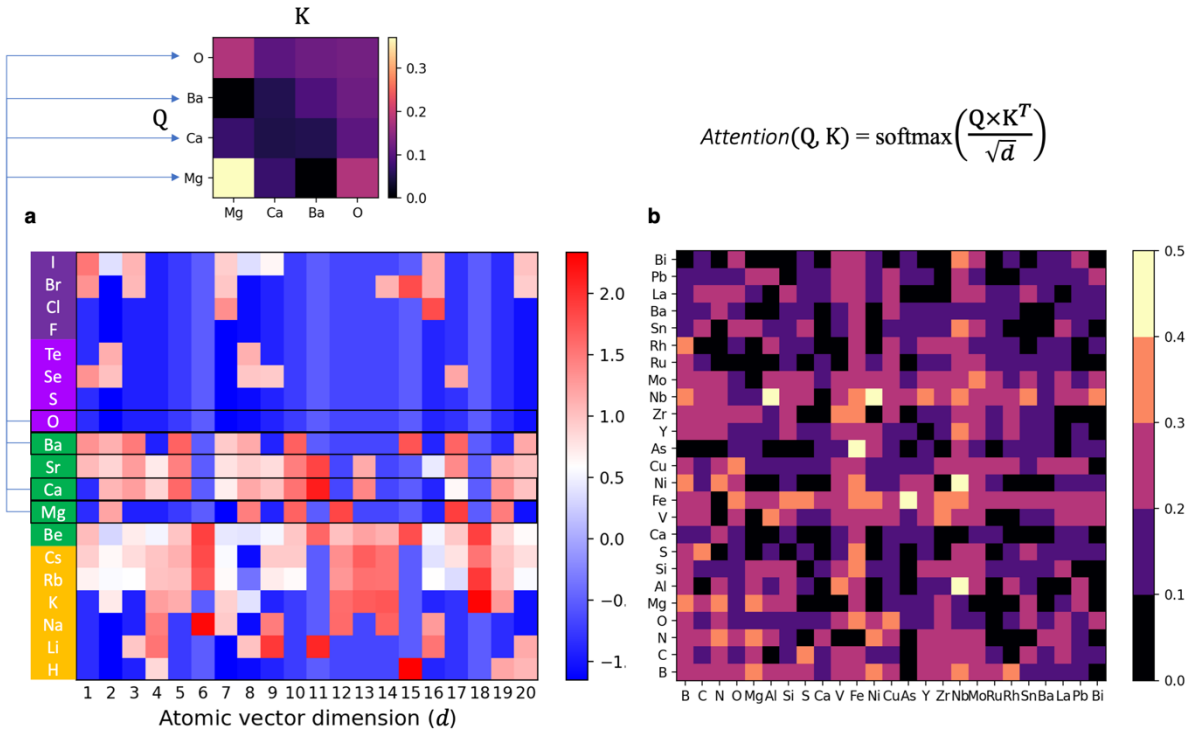


Figure 3. Atomic representations and their contributions to the phase fields' properties. a Atomic representation vectors in $k = 20$ dimensions for the 1st, 2nd, 16th and 17th atomic groups of the periodic table. The values (corresponding colour) illustrate differences and correlations between derived atomic features (vectors' components) in the neighbouring atoms and groups. The full stack of atomic vectors for the whole periodic table is extracted by PhaseSelect's atomic autoencoder shallow neural network, from the sparse matrix of chemical elements and compositional environments built for the Materials Project database[14,15]; for an example unexplored quaternary phase field, O-Ba-Ca-Mg, the corresponding contributions of the atomic elements to the likelihood of high-temperature superconductivity of this combination are calculated as the attention scores[16] (Supplementary Fig. 2-6). **b** Attention

scores are trained during the fitting of the model for phase fields classification by the target property. Here, attention to the atomic contributions to superconducting behaviour is visualised: combinations with e.g., Fe, Nb, Cu, Ni, Mo receive high attention in prediction of high-temperature superconductivity.

To emphasise the differences in the contributions of individual atoms to the phase field's properties, we employ the multi-head local attention[16] that calculates the attention scores – weights for the constituent atomic vectors contributing to the accuracy of the phase field classification for the target property. The attention scores are derived during the training and highlight the intermediate and interpretable results of the ML reasoning process well-aligned with the human understanding of chemistry of materials (See Figure 3b, Supplementary Fig. 2-6). When building a phase field representation for the downstream tasks of property classification and synthetic accessibility ranking, the phase field's atomic vectors are multiplied by their attention scores and then concatenated to form a $(n \times k)$-dimensional vector, where n is a number of constituent elements in a phase field, k is a chosen length of the atomic vector.

Classification by properties' values and ranking by synthetic accessibility

Classification in PhaseSelect is performed by a deep neural network (NN) that assigns the phase fields representation vectors to the corresponding classes of the properties' values. The phase fields in each dataset are divided into two classes (Figure 2a-c) that are labelled with '1' for the phase fields with associated property values above the chosen thresholds, and with '0' for the remaining phase fields. Three different classification models, one for each dataset - for superconducting materials and magnetic materials, and materials with a reported value of energy gap - are trained end-to-end with the architecture described in Fig. 1. Because the atomic characteristic and their relation to the materials properties are learnt from the reported chemistry, where the reports of the negatives (materials not possessing certain properties) are absent, the classification models are not trained to predict manifestation of target properties or their absence. Instead, for the phase fields that may contain

compositions with target properties, the classification models predict the probability of reaching high values of these properties within the phase fields. For example, in the training set for the materials with reported values of energy gap, none were reported with zero value (Fig. 2c). To verify the predictive power of the model trained on such data for the energy band gap classification, we have tested all 9816 intermetallic ternaries that do not have energy band gap values reported in MPDS (Supplementary discussion). 99.96% of the intermetallic ternary phase fields were classified as low energy gap materials (<4.5 eV) demonstrating the model's ability to extrapolate chemical patterns of atomic combinations – properties relationships, in absence of the zero-gap examples. On the other hand, this demonstrates vast generalisation of a model for the data regions where information is lacking.

The validation of the trained models is performed in the 5-fold cross-validation, where 5 models are trained on different 80% portions of the available data, with the remaining 20% used for testing. The average accuracy across the validation sets is 80.4, 86.2, 75.6 % for classification with respect to superconducting transition temperature, Curie temperature, and energy gap respectively. The validation datasets are used to tune the parameters of the NN models, such as dropout[18], learning rate, activisation[19], early stopping[17] and stochastic optimisation algorithm[20]. For the predictive models, we adopt all available data in the three datasets for training. Noting the stochastic nature of the machine learning NN, we employ averaging of the predicted probabilities over the ensemble of 300 models, this minimises the differences in training processes and derived models' parameters (Supplementary Fig. 10). The ensemble with the minimised variance in predictions enables assessment of the materials' properties not only by the assigned binary classes, that are threshold-dependent (Figure 4d, Supplementary Fig. 9, Supplementary Table 1), but also by the continuous values of probabilities as a measure of likelihood of achieving a desired property value. The latter helps to prioritise the materials for synthesis and further investigation.

In parallel to the classification module, a deep AutoEncoder neural network learns patterns of chemical accessibility from the experimentally verified materials data. Similarly to the procedure in [12], an

unsupervised de-noising AutoEncoder learns the patterns of similarity in data while reducing dimensionality of the phase fields representations. The training consists of two parts: encoding into a reduced dimensionality latent space, where phase fields representations are reorganised, so the similar phase fields are aligned, and decoding from the latent representation into the reconstructed images of original vectors. This reorganisation via the AutoEncoder enables ranking of the phase fields by their reconstruction errors, that reflect differences of individual entries from general patterns in data. Hence, elemental combinations that are unlikely to manifest conventional bonding chemistry nor to form synthetically accessible compositions exhibit high reconstruction errors[12]. We also find that predicted reconstruction errors converge to their average values when an ensemble of models is trained (See Supplementary Fig. 10b).

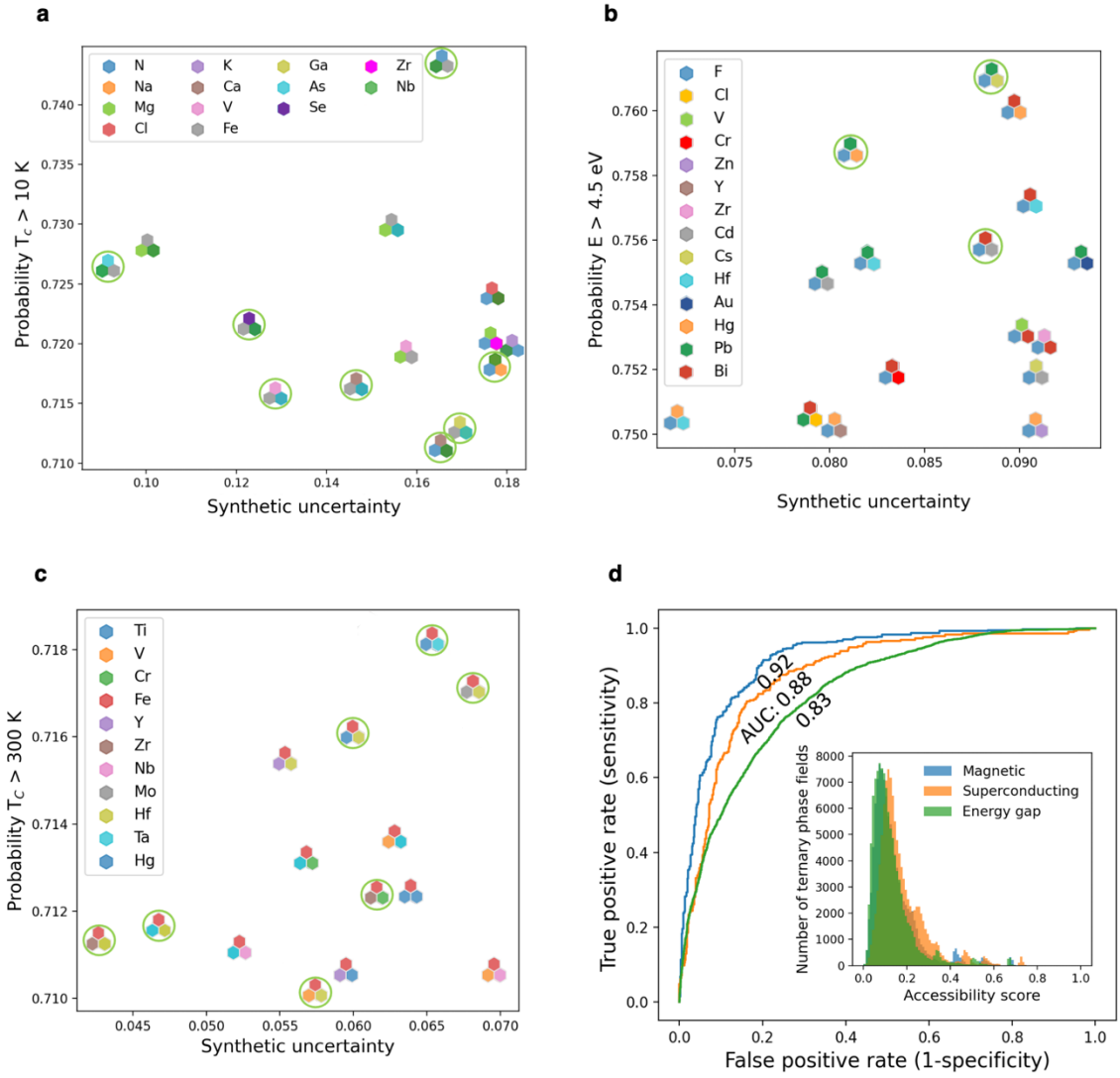


Figure 4. Probability of high-values properties and synthetic accessibility for unexplored ternary phase fields. Materials reported in ICSD[2], for which property values are not in SuperCon-v2018[3,8] or MPDS[1] are circled. a Unexplored ternary phase fields that are classified to exhibit superconductivity at T > 10 K with more than 70% probability and that have high likelihood of forming stable compounds with synthetic uncertainty (accessibility ranking) < 0.2, demonstrate trends in constituent elements: most of the top 50 phase fields are predicted to contain Mg, Fe, Nb and N. **b** Unexplored ternary phase fields that are classified to exhibit an energy band gap > 4.5 eV with more than 75% probability and that have high likelihood of forming stable compounds (with synthetic accessibility score < 0.1) demonstrate trends in distribution by constituent elements: different combinations of Hg-, F-, Bi-, Hf- and Pb-based phase fields have the highest probabilities. **c** Unexplored ternary phase fields that are classified to

exhibit magnetic properties at Curie T > 300 K with more than 71% probability and that have high likelihood of forming stable compounds (with synthetic accessibility score < 0.1) demonstrate trends in constituent elements: all top-ranked phase fields are Fe-based, with many phase fields containing Co and Y. **d** Receiver operating characteristics (ROC) of the classification models demonstrate high sensitivity and specificity of classifications for the range of thresholds of probabilities. The corresponding areas under the curves (AUC) demonstrate overall excellent performance of the model for magnetic materials, and good performance for both superconducting transition temperature and energy gap classifications. The inset illustrates close match of the distributions of 105995 ternary phase fields with respect to their synthetic accessibility scores for all three datasets.

By applying the trained ensembles of models to 105995 ternary phase fields (Supplementary discussion) and focusing on the unexplored materials that do not have any related compositions with reported properties in MPDS or SuperCon-v2018, we classify new elemental combinations with respect to the threshold values of superconducting transition temperature, Curie temperature and energy band gap and orthogonally rank candidate phase fields by their synthetic accessibility - degree of similarity with experimentally synthesized materials that are reported to exhibit these properties. We also highlight the phase fields, where compositions were synthesized and reported in ICSD, but for which there are no information about the properties discussed here in Supercon or MPDS, hence these phases fields did not enter the data for training. The large number of such phase fields among the top-performing candidates with respect to the measure of synthetic accessibility provides verification of the developed models and demonstrates that highly ranked candidates are likely to produce thermodynamically stable materials observed experimentally (See Figure 4a-c). We report the full list of likely candidates for novel superconducting materials among the phase fields that have been reported to form stable compounds in ICSD, but were not investigated from the perspectives of superconducting applications in [21] and its excerpt in Supplementary Table 7.

The top-performing phase fields according to both probability of exhibiting high values of properties and synthetic accessibility rank demonstrate trends produced by the constituent chemical elements:

Mg, Fe, Nb are predicted to constitute most of the top 50 phase fields that would yield stable compositions with superconducting transition temperatures above 10 K; similarly the top 50 magnetic ternary materials are Fe-based; while different combinations of Bi, Hf, Hg, Pb and F are predicted as most likely phase fields to contain stable compounds with energy gap of more than 4.5 eV, what can be expected from the simple bonding considerations as the majority of the latter are fluorides. While these predictions may align well with the human experts' understanding of chemistry, hence emphasizing the models' ability to infer complex atomic characteristics and phase fields-properties relationship from historical data, the models can also be used to identify unconventional and rare prospective elemental combinations as well as to rank the attractive candidate materials for experimental investigations.

Conclusions

Selection of elements is the cornerstone of the materials design. Quantitative assessment of the potential properties of the prospective materials at the level of their constituent elements mitigates the high risk of the consequential decisions in elaborate research of materials discovery. Classification of the materials for functional applications agglomerated into phase fields is also a route to the several orders of magnitude reduction of the combinatorial space. The end-to-end integrated architecture of PhaseSelect has demonstrated this capability of rendering the materials' phase fields in two orthogonal and equally challenging dimensions: merit probability and synthetic uncertainty. By employing PhaseSelect at the stage of conceptualization of the materials synthesis, human researchers can make use of numerical guidance in the selection of chemical elements that are most likely to produce new stable compounds with high probability of superior functional properties, combining this statistically derived quantitative information with the expert knowledge and understanding. The attention mechanism of PhaseSelect presents a route to interpretation of the machine learning for materials science and allows extrapolation of the knowledge of materials databases to the large number of

unexplored phase fields. These include multi-elemental materials, with prospective performance that could not be computationally assessed at scale with the methods developed to date.

Acknowledgements

We thank the UK Engineering and Physical Sciences Research Council (EPSRC) for funding through grants number EP/N004884 and EP/V026887. M.W.G. and V.G. thank the Leverhulme Trust for funding via the Leverhulme Research Centre for Functional Materials Design.

Data availability

The raw data used in this study is available at https://www.github.com/lrcfmd/PhaseSelect. The distribution of the phase fields' rankings, computed phase field's probability data generated in this study are available via University of Liverpool data repository at https://doi.org/10.17638/datacat.liverpool.ac.uk/1613

Software availability

The software developed for this study is available at https://www.github.com/lrcfmd/PhaseSelect.

Competing Interests Statement

The authors declare there are no competing interests.

Supporting Information

Machine Learning methodology and models, Training set for the Classification, Model validation, Atomic attention scores, Tools and libraries.

References

1. Villars, P., Cenzula, K., Savysyuk, I. & Caputo, R. Materials project for data science, https://mpds.io. (2021).

2. Zagorac, D., Müller, H., Ruehl, S., Zagorac, J. & Rehme, S. Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. *J. Appl. Cryst.* **52**, 918–925 (2019).

3. National Institute of Materials Science, Materials Information Station, SuperCon, http://supercon.nims.go.jp/index_en.html. (2011).

4. 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).

5. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).

6. Jha, D. *et al.* ElemNet : Deep Learning the Chemistry of Materials From Only Elemental Composition. *Sci. Rep.* **8**, 1–13 (2018).

7. 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).

8. Stanev, V. *et al.* Machine learning modeling of superconducting critical temperature. *NPJ Comput. Mater.* **4**, 1–14 (2018).

9. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

10. Polikar, R. Ensemble Learning. in *Ensemble Machine Learning: Methods and Applications* (eds. Zhang, C. & Ma, Y.) 1–34 (Springer US, 2012).

11. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **23**, 18 (2021).

12. Vasylenko, A. *et al.* Element selection for crystalline inorganic solid discovery guided by unsupervised machine learning of experimentally explored chemistry. *Nat. Commun.* **12**, 5561 (2021).

13. Meredig, B. *et al.* Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Mol. Syst. Des. Eng.* **3**, 819–825 (2018).

14. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002 (2013).

15. Zhou, Q. *et al.* Learning atoms for materials discovery. *PNAS* **115**, E6411–E6417 (2018).

16. Vaswani, A. *et al.* Attention Is All You Need. *arXiv:1706.03762 [cs]* (2017).

17. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. (2015).

18. 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).

19. Agarap, A. F. Deep Learning using Rectified Linear Units (ReLU). *arXiv:1803.08375 [cs, stat]* (2019).

20. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]* (2017).

21. Vasylenko, A. PhaseSelect: Element selection for functional materials discovery by integrated machine learning of atomic contributions to properties, https://github.com/lrcfmd/PhaseSelect. (2021).

Supplementary Information

Element selection for functional materials discovery by integrated machine learning of atomic contributions to properties

A. Vasylenko[1], D. Antypov[1], V. Gusev[1], M. W. Gaultois[1], M. S. Dyer[1], M. J. Rosseinsky[1,*]

[1]Department of Chemistry, University of Liverpool, L697ZD Liverpool, UK

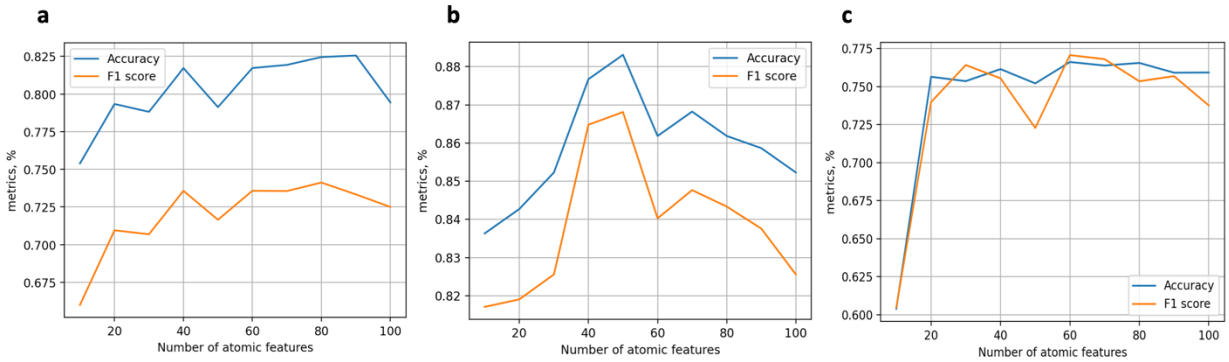
[*]corresponding author

Contents

Atomic features encoding

For unsupervised learning of atomic features from the materials database[1], we employ an approach similar to reference[2], in which we substitute single value decomposition with a shallow autoencoder. A shallow autoencoder is a 3-layer neural network, in which the input and output layers have a large number of neurons that corresponds to the size of the input vectors – sparse one-hot encoding

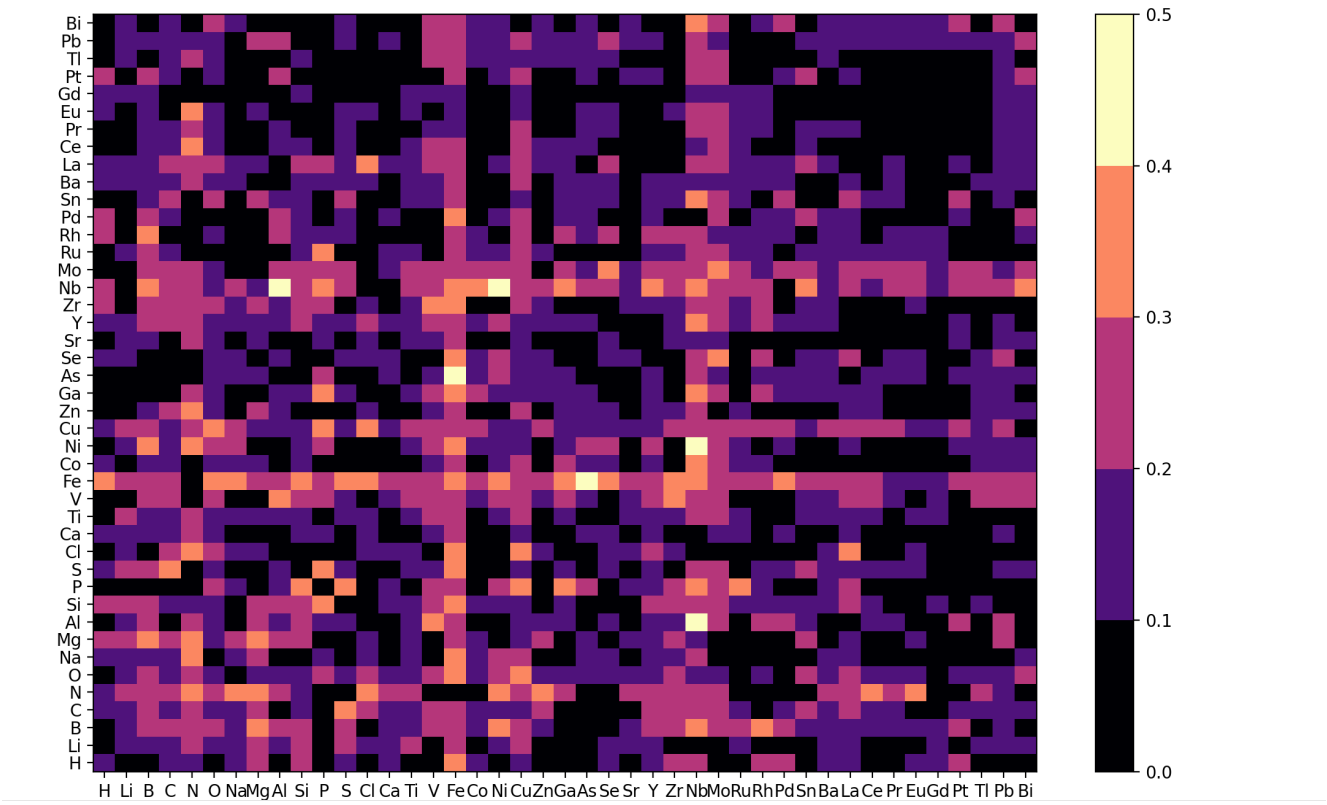
representations of atoms in the database. A single latent layer in between the input and output is a bottleneck aiming to extract the essential patterns in the data, while decreasing its dimensionality and filtering out the less representative and noisy information. One can further use thus trained representations as the atomic features. To maximise the quality and the descriptive power of the extracted atomic features, we study the effect of the size of the latent layer on the metrics of the downstream classifications. In this work, we train the shallow autoencoder simultaneously with the classification neural network in the end-to-end fashion. When trained separately for classification of superconducting, magnetic materials, and materials with a reported band gap, the end-to-end models based on the different sizes of atomic vectors have the metrics depicted in Supplementary Figure 1 a, b, c respectively. Although the best performance for classification of different properties is achieved at different numbers of atomic features in each of the three cases, there is similar trend for these dependencies. This trend suggests that a small number (< 40) of features cannot fully capture the variation in data, and a large set of features (> 80) contains too much noise, hence there is an optimal number of atomic descriptors for each model.



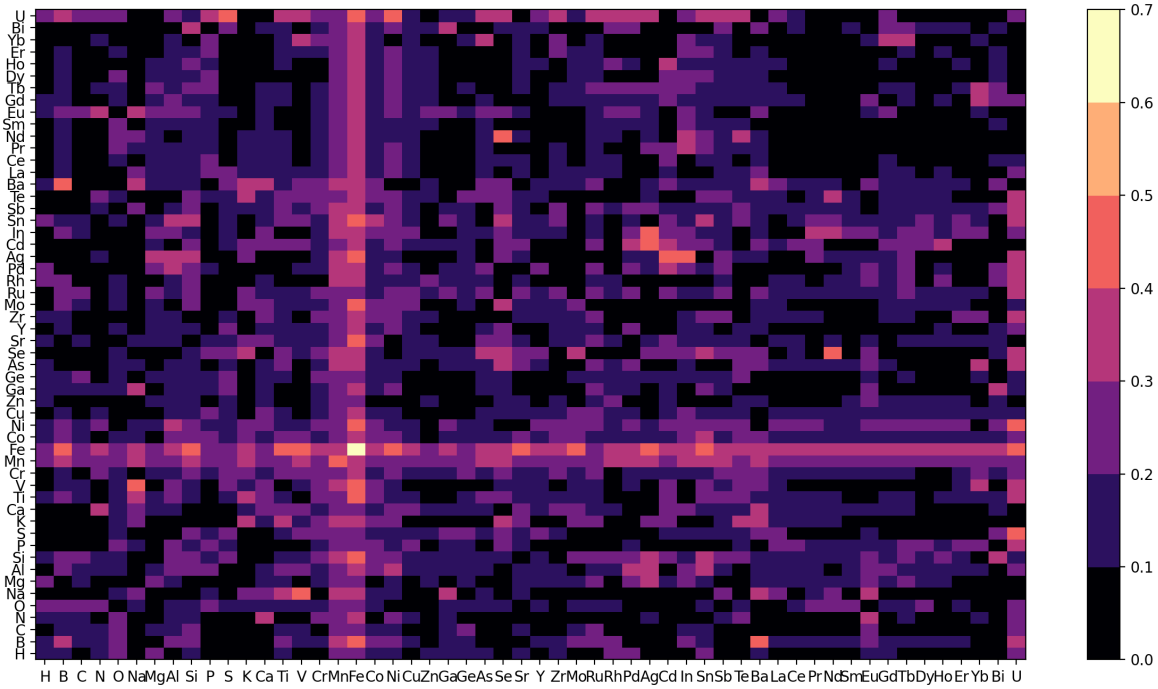
Supplementary Figure 1. Changes in classification metrics for models with different number of atomic features. **a** Accuracy and F1 score for classification of materials with respect to the maximum of superconducting transition temperature threshold 10 K: the best performing model has 80 atomic features; **b** Accuracy and F1 score for classification of materials with respect to the maximum of Curie transition temperature threshold 300 K: the best performing model has 50 atomic features; **c** Accuracy and F1 score for classification of materials with respect to the maximum of energy band gap threshold 4.5 eV: the best performing model has 60 atomic features.

Attention to atomic contributions maximizing the properties

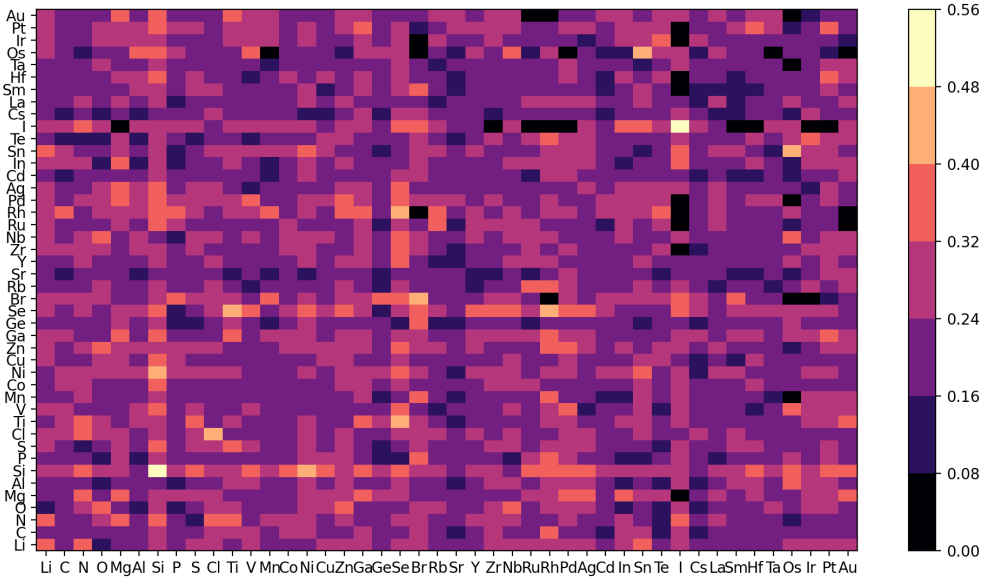
In the end-to-end classification models, we employ an attention mechanism[3] to emphasize those atomic contributions that minimize the combined loss, and hence maximize classification metrics. To incorporate information about atomic bonding interplay from all available data, the variance in size of the phase fields is alleviated by zero-padding in the phase fields representation module that further allows extrapolation of the patterns derived from the explored materials onto the candidate phase fields of arbitrary number of elements. We extract the attention scores obtained during the training of the models that illustrate atomic contributions to the properties manifested by the phase fields (Supplementary Figures 2-6). For visualisation, the attention scores are averaged across the attention heads and across all instances of the atomic pairs in the corresponding datasets. In Supplementary Figure 11, distributions of the averaged attention scores are plotted for the atoms that contribute the most to identify phase fields that manifest particular properties.



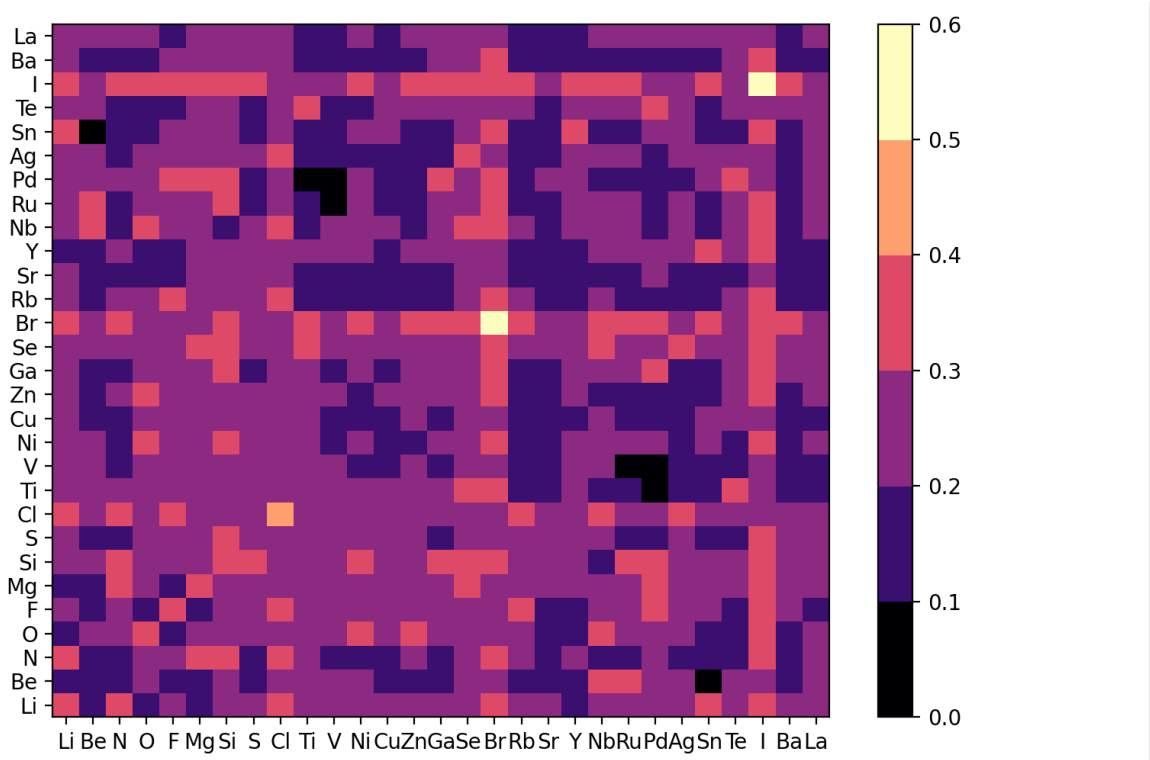
Supplementary Figure 2. Attention to atomic pairs that maximize accuracy of classification of high-temperature superconducting materials. Attention scores vary from 0 to 1. By focusing on the atomic pairs with the highest scores, when describing the phase field, accuracy of classification of these phase fields is maximized. This suggests the atomic pairs with the most prominent contributions allowing high-temperature superconductivity, e.g. Nb-Al, Nb-Ni, Cu-O and Fe-As.



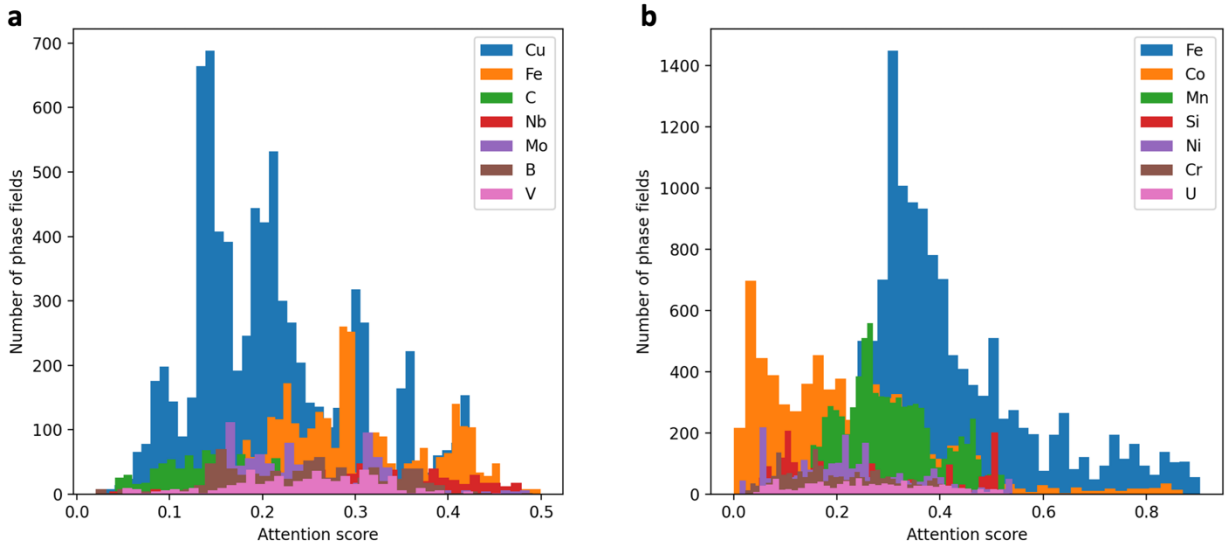
Supplementary Figure 3. Attention to atomic pairs that maximize the accuracy of classification of high-temperature magnetic materials. Attention scores vary from 0 to 1. By focusing on the atomic pairs with the highest scores, when describing the phase field, accuracy of classification of these phase fields is maximized. This suggests the atomic pairs with the most prominent contributions allowing high-temperature magnetic behaviour, with Mn, Fe and Co included in the majority of such pairs.



Supplementary Figure 4. Attention to atomic pairs that maximize accuracy of classification of materials with energy gap <4.5 eV. Attention scores vary from 0 to 1. By focusing on the atomic pairs with the highest scores, when describing the phase field, accuracy of classification of these phase fields is maximized. The majority of the atoms in the phase fields have 0.3-0.5 attention score, and contribute equally to identification of low energy gaps.



Supplementary Figure 5. Attention to atomic pairs that maximize accuracy of classification of materials with energy gap > 4.5 eV. Attention scores vary from 0 to 1. By focusing on the atomic pairs with the highest scores, when describing the phase field, accuracy of classification of these phase fields is maximized. This suggests atoms and atomic pairs with the most prominent contributions to the materials with energy gap > 4.5 eV, e.g. I, Br, Se, Cl, Si.



Supplementary Figure 6. Distribution of attention scores for the most contributing atoms to the functional materials a High-temperature superconducting materials; **b** high-temperature magnetic materials.

The atomic contributions weights are also used for building a model for an arbitrary number of elements in a phase field. For this, we create all phase fields representations vectors of an equal size *l*,

corresponding to the largest phase field in a database, and pad the smaller phase field vectors, of size *s*, with *l* - *s* zeros, that will have zero attention weights, but will further enable formation of a neural network layer for processing of all input data in a single model. The described construction of a phase field representation with local attention weights also makes the model insensitive to the order in which atomic elements are listed in a phase field, without the need to take into account all possible permutation of the elements.

Models' training and validation

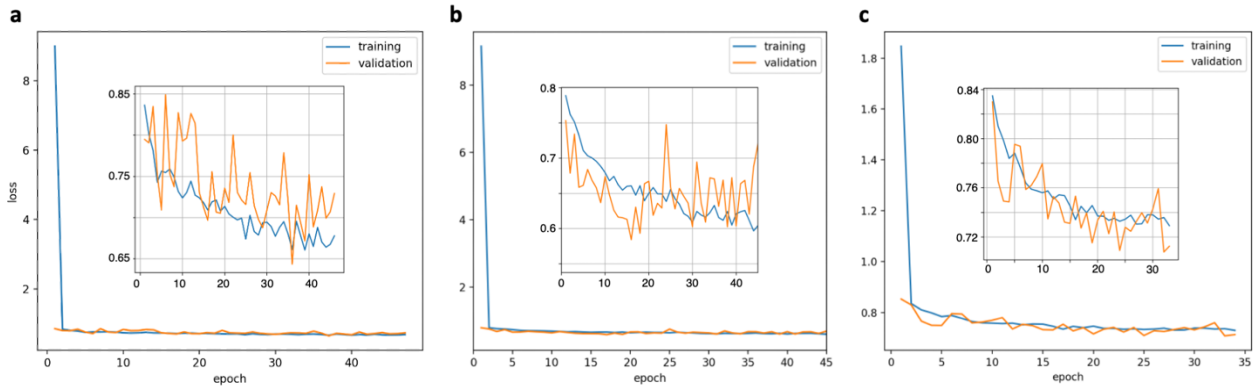
To validate the models' performance we employ 5-fold cross-validation for each dataset: phase fields with reported values of superconducting transition temperature, phase fields with reported values of Curie transition temperature, phase fields with reported values of energy gap. In 5-fold cross-validation, the data is divided into the training and test sets (80% and 20% of data respectively) in 5 different ways so 5 different models are examined with respect to the ability of the models' chosen architecture to generalise and extrapolate the information learnt from 5 different subsets of data onto the unseen areas. The accuracy and F1 scores of the classification models are presented in Supplementary Table 1.

Supplementary Table 1. Accuracy and F1 scores for classification models in 5-fold cross-validation

	Superconducting T_c=10K		**Magnetic T_c=300K**		**Energy gap 4.5 eV**	
test data subset	Accuracy,%	F1 score,%	Accuracy,%	F1 score,%	Accuracy,%	F1 score,%
0-20%	80.9	73.3	86.8	84.5	75.5	75.6
21-40%	83.6	77.1	86.7	85.7	75.2	74.8
41-60%	78.7	71.7	85.9	82.1	75.9	75.6
61-80%	79.7	71.3	85.5	84.1	76.0	75.1
81-100%	79.2	71.0	86.0	84.4	75.7	75.5
Average:	**80.4**	**72.9**	**86.2**	**84.2**	**75.6**	**75.3**

The performance metrics from the 5 models for each dataset are then averaged to describe a general ability of the models' architecture to learn from the available data. During the training of the end-to-end classification models, the weights and biases of the autoencoder and classifier neural networks are trained simultaneously, while the corresponding losses – reconstruction error and binary cross-entropy,

respectively – are minimized as a combined loss during back propagation with Adam optimization[4]. The typical training of the classification models for the superconducting, magnetic and energy band gap datasets are converged under 50 epochs as illustrated in the Supplementary Figure 7.



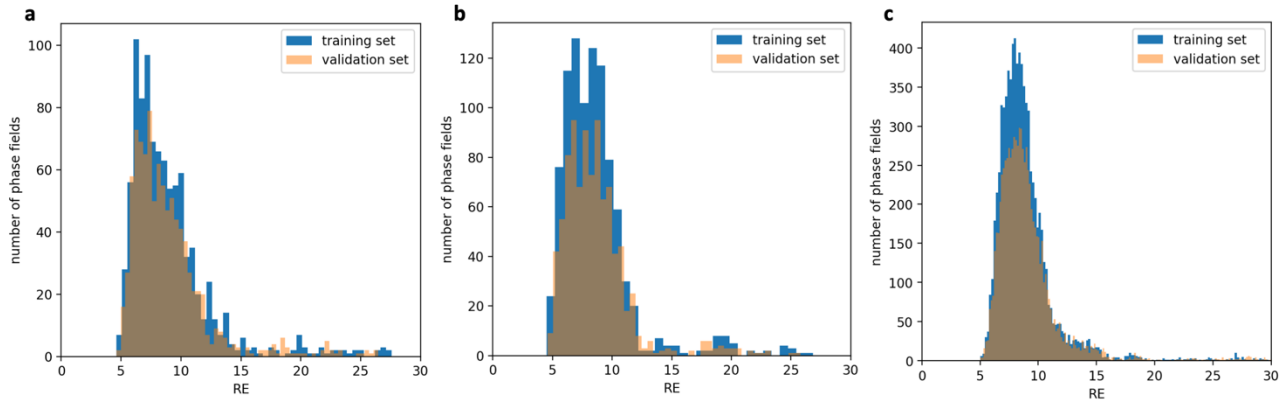
Supplementary Figure 7. Training progress of the end-to-end classification models. a Classification of the superconducting materials, training on 4826 phase fields; **b** Classification of the magnetic materials, training on 4753 phase fields; **c** Classification of the materials' energy band gap, training on 40452 phase fields.

For validation of the unsupervised models for the phase fields ranking with respect to synthetic accessibility, we employ an approach developed in [5]. We perform 5-fold cross validation, in which the validation error is defined as the percentage of entries in the test set that evaluated with normalized reconstruction errors in the 20% of the maximum Supplementary Table 2. Additionally, we compare the predicted reconstruction errors for the validation sets with the ground truth reconstruction errors obtained for the same entries in unsupervised training, when the entries are included in the training data (Supplementary Fig. 8) and calculate the mutual information score adjusted against chance[6] (Supplementary Table 2). The typical training process of the ranking autoencoder neural network for different datasets are depicted in Supplementary Fig. 9.

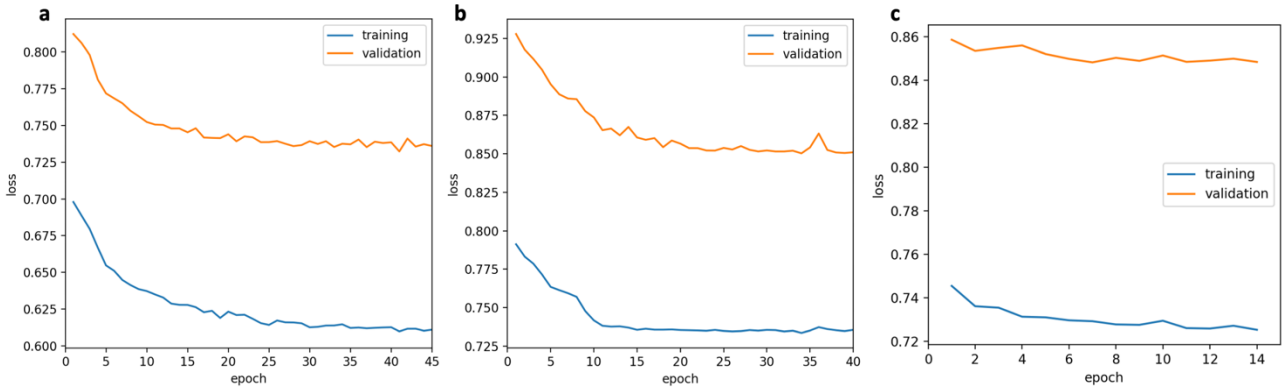
Supplementary Table 2. Accuracy and Adjusted Mutual Information Score (AMIS) for ranking autoencoder models in 5-fold cross-validation

	Superconducting materials		**Magnetic materials**		**Energy gap materials**	
test data subset	Accuracy,%	AMIS	Accuracy,%	AMIS	Accuracy,%	AMIS
0-20%	96.1	0.69	94.7	0.64	97.2	0.77

21-40%	97.4	0.76	95.3	0.66	98.6	0.78
41-60%	97.7	0.68	93.5	0.66	97.7	0.75
61-80%	95.1	0.79	94.9	0.64	98.7	0.75
81-100%	96.6	0.72	93.9	0.68	97.8	0.81
Average:	**96.6**	**0.73**	**94.5**	**0.66**	**98.0**	**0.77**



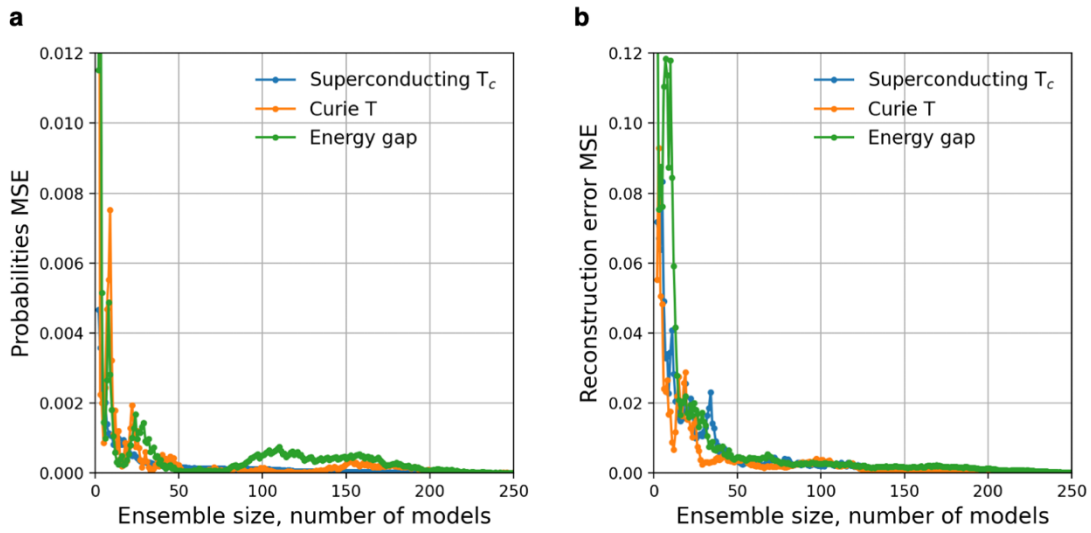
Supplementary Figure 8. Distribution of reconstructions errors (RE) for the phase fields. RE for the same phase fields are calculated in two approaches: 1) in unsupervised learning, as a part of a training set – used as ground truth RE for AMIS calculation in Supplementary Table 2; 2) predicted by the model trained on 80% of the remaining data – as a validation set. **a** Superconducting materials; **b** magnetic materials; **c** materials with reported energy gap.



Supplementary Figure 9. Training progress of the ranking autoencoder models. a ranking of the superconducting materials, training on 4826 phase fields; **b** ranking of the magnetic materials, training on 4753 phase fields; **c** ranking of the materials with the reported energy band gap, training on 40452 phase fields.

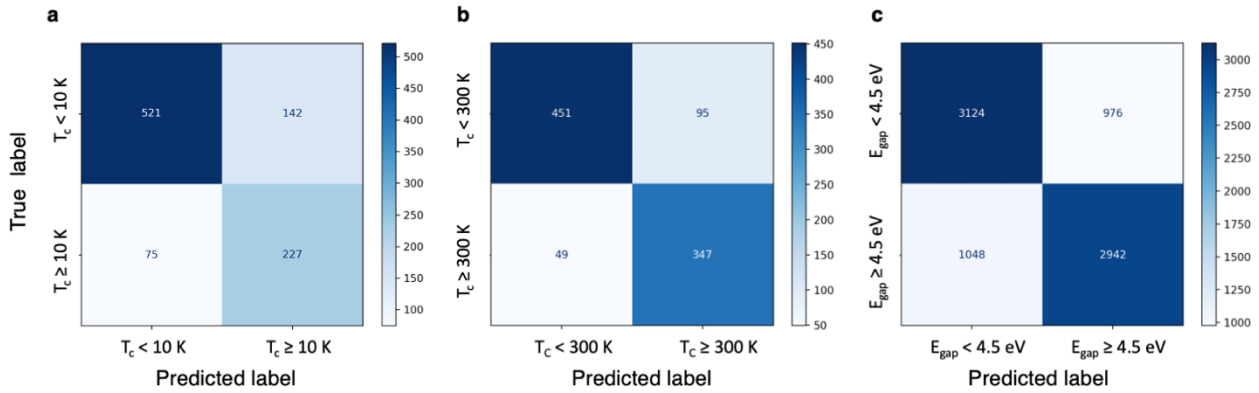
To take into account statistical variance in both supervised and unsupervised results from the neural networks trained at different instances, we average the results across the ensemble of 250 neural networks. Convergence of deviations of results in terms of the mean square errors from the running

average values is illustrated in Supplementary Figure 10. For all datasets, for both supervised classifying neural network and ranking autoencoders, the average values converge when more than 200 models are considered.



Supplementary Figure 10. Convergence of the mean square errors (MSE) of the average predicted scores with the number of models in the ensemble. a Probabilities of phase fields belonging to a binary class are averaged over an ensemble of models. MSE of the average scores decrease below 0.001 for ensembles larger than 200 models for all datasets. **b** MSE of the average reconstruction errors, used as synthetic accessibility scores of phase fields decrease below 0.005 for ensembles larger than 200 models.

The ensembles of the trained models for each dataset are then used to classify the phase fields with respect to the corresponding properties. For randomly selected 20% of the phase fields from each dataset, the classification predictions are illustrated with the confusion matrices in Supplementary Figure 11.



Supplementary Figure 11. Confusion matrices for binary classification models with threshold probability 0.5. a Superconducting materials classification of 20% of the collected data from MPDS[7] and SuperCon[8] with respect to transition temperature 10 K; **b** magnetic materials classification of 20% of the collected data from MPDS, with respect to Curie temperature 300 K; **c** classification materials with reported values of energy band gap with respect to energy gap value 4.5 eV, test set is 20% of randomly selected data collected from MPDS.

The corresponding average accuracy, F1 score and the Matthews' correlation coefficients (MCC) are presented for the three models in Supplementary Table 3.

Supplementary Table 3. Average binary classifications metrics of the maximum values of exhibited properties in the phase field.

Metrics	Superconducting Tc >10 K	Magnetic Tc > 300 K	Energy gap > 4.5 eV
Accuracy, %	80.4	86.2	75.6
F1 score, %	72.9	84.2	75.3
MCC	0.608	0.711	0.523

Combination of probabilities of high-values properties (merit probability) and synthetic uncertainties

We combine the outcomes of the classifying neural network and autoencoder to rank unexplored ternary combinations of elements. For the unexplored ternary combinations we consider all possible combinations of 87 atoms, that exclude rare and toxic elements and have sufficient data in Materials Project to be reasonably well learnt with the proposed unsupervised approach described above. The total number of ternary combinations, therefore, is $87 \times 86 \times 85 / 3! = 105995$, among them 12297 have a reported value of energy band gap in MPDS (and in a peer-reviewed literature), 1953 are reported to have magnetic properties and a corresponding Curie temperature in MPDS, and 1716 are reported to have superconducting properties and a corresponding critical temperature in a combined data from SuperCon and MPDS.

The best ranking combinations, illustrated in Figure 4 in the main text are presented in the Supplementary Tables 4-6. Among the considered phase field there are entries that have been

synthesized and reported in ICSD-v2021[9], but do not have records in MPDS and SuperCon concerning the properties studied here. These entries did not enter the training datasets and are highlighted in bold in the Supplementary Tables 4-6. These entries have been predicted to have low synthetic uncertainty, that provides experimental verification of the proposed method for ML assessment of synthetic accessibility. The full list of the predicted scores for the yet experimentally unexplored ternary phase fields can be found along with the PhaseSelect software[10].

Supplementary Table 4. Predicted probabilities of the best unexplored ternary phase fields to manifest superconducting T_c > 10 K and their synthetic uncertainty scores. The phase fields, in which compounds are synthesized[9] but were not included into the training data[7,8] are highlighted in bold.

Phase fields	Probability Tc > 10 K	Synthetic uncertainty
N Fe Nb	**0.7433**	**0.1643**
Mg Fe As	0.7295	0.1557
Mg Fe Nb	0.7278	0.1016
Fe As Nb	**0.7261**	**0.0903**
N Cl Nb	0.7238	0.1781
Fe Se Nb	**0.7212**	**0.124**
N Mg Zr	0.72	0.1776
N K Nb	0.7194	0.1798
Mg V Fe	0.7189	0.1589
N Na Nb	**0.7187**	**0.1774**
Ca Fe As	**0.7162**	**0.1477**
V Fe As	**0.7154**	**0.1299**
Fe Ga As	**0.7126**	**0.1709**
N Ca Nb	**0.7111**	**0.1665**

Supplementary Table 5. Predicted probabilities of the best unexplored ternary phase fields to manifest Curie T_c > 300 K and their synthetic uncertainty scores. The phase fields, in which compounds are synthesized[9] but were not included into the training data[7] are highlighted in bold.

Phase fields	Probability T_c > 300 K	Synthetic uncertainty
Ti Fe Ta	**0.7181**	**0.0658**
Fe Mo Hf	**0.717**	**0.0685**
Ti Fe Hf	**0.716**	**0.0603**
Fe Y Nb	0.7159	0.0681
Fe Y Hf	0.7154	0.0557
V Fe Ta	0.7136	0.0631
Cr Fe Ta	0.7131	0.057
Ti Fe Hg	0.7123	0.0642
Cr Fe Zr	**0.7123**	**0.0619**
Fe Hf Ta	**0.7117**	**0.0467**
Fe Zr Hf	**0.7113**	**0.0426**
Fe Nb Ta	0.7111	0.0522
Fe Y Hg	0.7106	0.0598
V Fe Nb	0.7106	0.0699
V Fe Hf	**0.7101**	**0.0575**

Supplementary Table 6. Predicted probabilities of the best unexplored ternary phase fields to manifest energy band gap > 4.5 eV and their synthetic uncertainty scores. The phase fields, in which compounds are synthesized[9] but were not included into the training data[7] are highlighted in bold.

Phase fields	Probability E_g > 4.5 eV	Synthetic uncertainty
Cs F Pb	**0.7613**	**0.8852**
F Hg Bi	0.7603	0.0897
F Hg Pb	**0.759**	**0.0811**
F Te Hf	0.7575	0.0975
F Y Bi	0.7575	0.0984
F Hf Bi	0.7574	0.0906
F As Hf	0.7572	0.0984

Cl I Hf	0.7561	0.0999
F Cd Bi	**0.7561**	**0.0882**
F Au Pb	0.7556	0.0932
F Hf Pb	0.7556	0.082
F Cd Pb	0.755	0.0796
F V Bi	0.7531	0.0904

Prediction of superconducting behaviour for reported phase fields in ICSD-v2021

We apply PhaseSelect ensembles of classification models to identify likely candidates for novel superconducting materials among the phase fields that have been reported to form stable compounds in ICSD-v2021, but were not investigated from the perspectives of superconducting applications and reported in MPDS and SuperCon (hence were not included into the training dataset). The excerpt of these predictions is presented in Supplementary Table 7; classification of all binary, ternary and quaternary phase field in ICSD with respect to the maximum accessible value of superconducting critical temperature is uploaded in[10].

Supplementary Table 7. Predicted probabilities of superconducting behaviour at T_c > 10 K for the best ternary phase fields reported to form stable structures in ICSD. (Excerpt for p > 0.7. The full list is in[10]).

Phase fields	Probability T_c > 10 K	Phase fields	Probability T_c > 10 K
Fe N Nb	0.7466	Mo N Nb	0.7142
Fe Li N	0.7391	C Li N	0.7131
Fe Ga N	0.7383	As Fe Nb	0.7113
C Fe N	0.7352	Al N Nb	0.7111
Fe Mo N	0.7343	C Ga N	0.7102
Ba Fe N	0.733	Ga N V	0.7101
Fe N Se	0.7324	N Nb V	0.7096
Ca Fe N	0.7322	Ca Fe O	0.709
C Mg N	0.7305	C N V	0.709
Fe Mg O	0.7302	Ba Fe O	0.7087
Li Mg N	0.7291	B Mg N	0.7077
Ga N Nb	0.7262	Fe Nb Se	0.7075
Ga Mg N	0.7261	C K N	0.7075
C N Nb	0.726	Ca Mg N	0.7065

Fe N Zr	0.725	As Ca Fe	0.7058
Mg Mo N	0.7229	C Cl N	0.7056
Fe Ga Nb	0.7227	N Na Nb	0.7056
Fe N Sr	0.7226	As Fe V	0.7055
Cl Mg N	0.7218	N Nb Zr	0.7054
Cu Fe O	0.7197	Ba N Nb	0.7037
Fe N Sn	0.7194	As Fe Ga	0.7034
Fe Mn N	0.7192	Fe N Pt	0.7023
Fe N O	0.7186	C N Na	0.702
As Fe O	0.7175	Ca N Nb	0.7019
Fe N Y	0.7173	As Ba Fe	0.7017
C Mo N	0.717	C Ca N	0.7014
Fe Ga V	0.7157	As Fe K	0.7007
Li N Nb	0.7143		

Tools and Libraries

PhaseSelect[10] has been built using Python 3.7.4, Tensorflow 2.4.1, Scikit-learn 0.24.0, Numpy 0.19.2, Pandas 1.1.4. The figures in the main text and Supplementary figures are created using Matplotlib 3.3.4.

Supplementary References

1. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002 (2013).

2. Zhou, Q. *et al.* Learning atoms for materials discovery. *PNAS* **115**, E6411–E6417 (2018).

3. Vaswani, A. *et al.* Attention Is All You Need. *arXiv:1706.03762 [cs]* (2017).

4. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]* (2017).

5. Vasylenko, A. *et al.* Element selection for crystalline inorganic solid discovery guided by unsupervised machine learning of experimentally explored chemistry. *Nat. Commun.* **12**, 5561 (2021).

6. Vinh, N. X., Epps, J. & Bailey, J. Information theoretic measures for clusterings comparison: is a correction for chance necessary? in *Proceedings of the 26th Annual International Conference on Machine Learning* 1073–1080 (Association for Computing Machinery, 2009). doi:10.1145/1553374.1553511.

7. Villars, P., Cenzula, K., Savysyuk, I. & Caputo, R. Materials project for data science, https://mpds.io. (2021).

8. National Institute of Materials Science, Materials Information Station, SuperCon, http://supercon.nims.go.jp/index_en.html. (2011).

9. Zagorac, D., Müller, H., Ruehl, S., Zagorac, J. & Rehme, S. Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. *J. Appl. Cryst.* **52**, 918–925 (2019).

10. Vasylenko, A. PhaseSelect: Element selection for functional materials discovery by integrated machine learning of atomic contributions to properties, https://github.com/lrcfmd/PhaseSelect (2021).