Compositional optimization of hard-magnetic phases with machine-learning models

Johannes J. Möller, Wolfgang Körner, Georg Krugel, Daniel F. Urban and Christian Elsässer

1 Fraunhofer Institute for Mechanics of Materials IWM, Wöhlerstr. 11, 79108 Freiburg, Germany
2 University of Freiburg, Freiburg Materials Research Center, Stefan-Meier-Str. 21, 79104 Freiburg, Germany

Machine Learning (ML) plays an increasingly important role in the discovery and design of new materials. In this paper, we demonstrate the potential of ML for materials research using hard-magnetic phases as an illustrative case. We build kernel-based ML models to predict optimal chemical compositions for new permanent magnets, which are key components in many green-energy technologies. The magnetic-property data used for training and testing the ML models are obtained from a combinatorial high-throughput screening based on density-functional theory calculations. Our straightforward choice of describing the different configurations enables the subsequent use of the ML models for compositional optimization and thereby the prediction of promising substitutes of state-of-the-art magnetic materials like Nd$_2$Fe$_{14}$B with similar intrinsic hard-magnetic properties but a lower amount of critical rare-earth elements.

1. INTRODUCTION

Machine Learning (ML) is currently establishing itself as an important tool in materials research. ML models have been successfully used for predicting grain boundary energies [1, 2] and mobilities [3] in pure metals, ferroelectric Curie temperatures and classes of perovskites [4–6], and electronic [7, 8] as well as magnetic properties [9–11]. The main applications of ML have, however, been limited to learning properties that are directly related to total/atomic energies or atomic forces [12–26]. We believe that this limitation originates mainly from the availability of high-accuracy density-functional theory (DFT) data for the energies of millions of structures and compositions in internet databases, such as the Materials Project [27], the Open Quantum Materials Database (OQMD) [28, 29], or the Novel Materials Discovery (NOMAD) repository [30]. The potential of ML to accurately predict and to optimize other materials properties, which are less well documented in internet databases, needs thus to be further explored.

The search for new hard-magnetic phases is an ideal application case for demonstrating the benefits of ML in materials research. On the one hand, many hard-magnetic materials are intermetallic phases that contain a multitude of atoms per unit cell. The unit cell of the prominent Nd$_2$Fe$_{14}$B phase, for instance, contains 68 atoms. This makes it infeasible even for combinatorial high-throughput screening (HTS) to cover the entire variety of possible phase compositions. For a reasonable variety of screened elements, the number of possible compositions and atomic arrangements within the unit cell easily adds up to billions or even trillions. ML overcomes this limitation by building a continuous model for the learned properties based on the training data. This enables a reliable interpolation between the training data and an accurate prediction of properties for unknown compositions. Today’s best hard-magnetic materials, which are key components in many green-energy technologies, contain substantial amounts of supply-critical rare-earth (RE) elements, such as Dy, Nd, or Sm. There is thus a considerable industrial demand for new materials with good hard-magnetic properties but less RE contents.

The objective of this paper is therefore to construct ML models for hard-magnetic properties and to subsequently use them for compositional optimization in order to identify those chemical compositions that exhibit good magnetic properties, but contain only few RE elements. For training the ML models, we select the ThMn$_{12}$-type crystal structure [31] from our materials database of hard-magnetic phases [32] as a promising substitute for the RE-rich state-of-the-art materials Nd$_2$Fe$_{14}$B and SmCo$_5$. Instead of using all the data in our database for training, we use a subset with equally-distributed and equally-spaced supporting points where only one alloying element (besides Fe) is considered. We believe that such a scenario where one holds a data set with not much more than 200 different configurations is representative for a typical systematic density-functional-theory (DFT) study of a certain non-standard property (which in our case is the magneto-crystalline anisotropy).

The paper is organized as follows: In section 2 we describe our computational approach, i.e., how we train the ML models to the hard-magnetic properties of the intermetallic phases in our database. In section 3 we present the results for validating, testing, and optimizing the ML models. These results are discussed in section 4 with focus on the accuracy and reliability of the ML predictions and on the potential of the optimized compositions. The discussion also includes comments on the achievements and limitations of our approach, on possible pathways for future extensions, and on discrepancies between the DFT training data and experimental results. The paper is summarized in section 5.

* Corresponding author: johannes.moeller@iwm.fraunhofer.de
2. METHODOLOGY

2.1. Machine Learning

The list of machine-learning (ML) models that have been used for predicting properties and behaviors of materials includes a variety of methods such as artificial neural networks [34–37], genetic algorithms [33, 34], Gaussian processes [17, 37], decision-tree and random-forest models [5, 36], and kernel-based methods [13, 14, 26, 36]. Which of the different methods is best suited for a certain analysis usually depends on the type (e.g., scalar values or images) and amount (hundreds, thousands or millions of samples) of the available training data and the targeted application of the ML model (value prediction/regression or classification).

In this work, we make use of the kernel-based Support Vector Regression (SVR) method to construct numerical ML models for $K_1$, $\mu_0M$ and $E_t$ (together denoted as target properties $y$ in the following). SVR is a non-linear regression analysis that makes use of the so-called kernel trick. In kernelized ML methods, the kernel function maps the input space, in which the target property $y$ is usually not a linear function of the feature vector $x$, into a higher dimensional space where such a linear relationship may exist. SVR is based on the concept of Support Vector Machines (SVMs), which were originally developed in the 1960s [38] for classification purposes. In SVR models, the predicted value $y^m$ (superscript 'm' for model) for a feature vector $x$ is determined as

$$y^m(x) = \sum_i w_i k(x_i, x) + b,$$

where $w_i$ are the individual weights for each training vector $x_i$, $k(x_i, x)$ is the kernel function, and $b$ is the constant intercept ('$b$' stands for bias).

The kernel $k(x_i, x)$ is typically of linear, polynomial or Gaussian (radial basis function, 'rbf') type [39]:

$$k_{\text{linear}}(x_i, x) = x_i \cdot x,$$

$$k_{\text{poly}}(x_i, x) = [\gamma(x_i \cdot x)]^d,$$

$$k_{\text{rbf}}(x_i, x) = \exp \left[ -\gamma \| x_i - x \|^2 \right].$$

Here, the subscript '2' indicates the Euclidean ($L_2$) distance and '.' the inner product, $d$ is the degree of the polynomial kernel, and $\gamma$ represents the width of the respective kernel function.

SVR uses an $\varepsilon$-insensitive loss function [40, 41], which means that optimizing a SVR model involves minimizing the weights $w_i$. This procedure leads to a flat evolution of $y(x)$ and inherently reduces the risk of over-fitting [42] by simultaneously allowing for some larger deviations (up to a value of $\varepsilon$) between $y$ and $y^m$ for individual samples ('outliers'). In this paper, we use $\varepsilon = 0.1$ T, $1.0$ MJ/m$^3$, and $0.01$ eV/atom for the ML models for $\mu_0M$, $K_1$, and $E_t$, respectively. This choice can be considered as the targeted accuracy of the constructed models.

Besides the choice of the kernel function (and its width $\gamma$), SVR models have an additional regularization parameter $C$, which determines the trade-off between the flatness of $y^m(x)$ and the amount up to which deviations larger than $\varepsilon$ are tolerated [41] (a higher $C$ allows higher possible values for the weights $w_i$). As mentioned before, $\gamma$ determines the width of the kernel and therefore the contribution of neighboring feature vectors to the prediction (a smaller $\gamma$ resulting in a higher contribution of samples close by). For a given kernel function, both hyperparameters control the complexity of the model and need to be optimized in order to prevent under- and over-fitting. To yield a model that is as universal as possible, both $C$ and $\gamma$ should be as small as possible. Note that $\gamma$ has no meaning for a linear kernel.

Extensive descriptions of the SVR method are given in Refs. [40, 41]. The SVR algorithm is used in its implementations in the Python ML package scikit-learn [39, 43]. To assess the potential benefit of ML over classical fitting methods, we also parameterized linear regression (LR) functions ($y^m = \sum_j w_j x_j + b, j = 0 \ldots N_{\text{features}}$) for our data. Note that in this case, contrary to the SVR method described above, the weights $w_j$ are determined for each component of $x$, i.e., for the feature $x_j$, not for each training sample $x_i$.

2.2. Material Database

The material database used for training and validating the ML models originates from our previous study on rare-earth-lean intermetallic ReA$_{12}$X compounds [32]. Here, Re is either Ce or Nd, and A can be one of the magnetic transition-metal elements Mn, Fe, Co, and Ni, the non-magnetic elements Ti, V, Cr, Cu, Zn, Al, Si, and P, or a mixture of them. The element X is one of the small interstitial elements B, C, or N.

The ReA$_{12}$X structure is based on the ThMn$_{12}$ type, which was first reported by de Mooij et al. [31]. The symmetry-equivalent Wyckoff sites of this body-centered tetragonal structure (space group #139, I4/mmm symmetry) are the 2a (occupied by Re elements), 8i, 8j, 8f (occupied by A elements), and 2b (occupied by X elements, if any), see Fig. 1 for illustration. Note that the number of atoms per formula unit (14) is only half of all atoms per conventional tetragonal unit cell (28). The tetragonal lattice parameters $a = 8.566$ Å and $c = 4.802$ Å, as well as the crystal coordinates of the Wyckoff sites were taken from the work of Isnard et al. [44].

To create the magnetic-properties database, we employ a HTS approach with DFT calculations using the tight-binding linear-muffin-tin-orbital atomic-sphere-approximation method (TB-LMTO-ASA) [45] based on the work of Fähnle et al. [46, 47]. The several thousand phases were calculated with the HTS setup described by Drebov et al. [48] which allows a fully automated generation of new phases by combinatorial substitution of sets of
In order to convert our zero-temperature single-crystal, single-domain DFT-HTS results for \( K_1 \) and \( H_a \) to room-temperature estimates, which can serve as guidelines for experimental efforts, the calculated values need to be divided by four for Nd. For Ce, a division by 35 provides a conservative estimate. These heuristic factors have been derived from benchmark calculations of well-known hard-magnetic materials (see Table II and discussion in Ref. [32]).

For the optimization of the hyperparameters of our ML model (see Secs. 2.1 and 3.1) we used the ReFe\(_{12}\)\(_{0.4}\)A\(_{4}\)X \((z = 0...3)\) subset of all the different Re(Fe,A)\(_{12}\)X configurations in our database. In this data set, each A position \((8i, 8j, 8f)\) which initially has been occupied by Fe atoms, was subsequently fully replaced by other alloying elements A. For each choice of the \( N_X = 3 \) choices for X, there exists one configuration without alloying elements (i.e. \( z = 0 \)) and \( N_{sub} = 7 \) possible configurations for each of the \( N_A = 11 \) alloying elements A, namely three for both \( z = 1 \) and \( z = 2 \) and one for \( z = 3 \).

For \( \text{Re} = \text{Nd} \), the data set therefore consists of 234 \((= N_X \cdot (1 + N_{sub} \times N_A) = 3 \cdot (1 + 7 \times 11)\) configurations. For \( \text{Re} = \text{Ce} \), two compounds could not be converged in the given crystal structure, thus the data set contains only 232 configurations.

In total, our database contains 3,080 entries. Besides the 466 configurations used for training the ML models, the data includes 2,614 compounds with multiple substitutional alloying elements and partially chemically heterogeneous occupations of Wyckoff sites. For example the 8i position can be occupied by two Co atoms and six Fe atoms. This remaining data set was used for the subsequent testing (see Sec. 3.2) of our model with the previously optimized hyperparameters. Note that the test data are inherently difficult to be predicted with our realistic choice of training data as they contain lots of situations about which the trained model was not informed at all (i.e., two different alloying elements in the same configuration). The objective of our work is not to create the best possible model (for which we should obviously use the whole data set for training) but to reflect a realistic scenario and therefore use the typical outcome of a systematic DFT study in terms of number and distribution of the data points.

2.3. Material Representation

For a straightforward decoding of the structural and chemical information stored in the feature vectors \( x \), the construction of \( x \) needs to be bijective. This means that a reversible mapping exists between \( x \) and the represented atomistic configuration. We introduce here a descriptor that sufficiently fulfills this requirement. It contains the number of atoms \( n_j^{(i)} \) of each chemical species \( i \) resolved...
FIG. 2. Illustration of descriptors for selected Re(Fe,A)_{12}X compounds. Note that the first seven configurations are included in the training set (ReFe_{12-4z}A_{4z}X, z = 0 ... 4) whereas the last four compositions belong to the test set (see text for details).

on each of the Wyckoff positions \(j\),

\[
\mathbf{x} = (n_1^{(1)}, \ldots, n_P^{(1)}, n_1^{(2)}, \ldots, n_P^{(2)}, \ldots, n_1^{(N)}, \ldots, n_P^{(N)}),
\]

where \(N\) is the number of different elements in the data set and \(P\) is the number of different Wyckoff positions in the crystal structure. When certain positions are occupied only by certain elements (e.g., the 2a site of the ReA_{12}X structure is occupied only by Re atoms) the general form of \(\mathbf{x}\) given in Eq. 7 can be simplified by leaving out those entries for which \(n_j^{(i)}\) is zero throughout the whole data set.

The descriptors for a data set of the ReA_{12}X structure with, for instance, two different Re elements on the 2a position (\(j = 1\)), three different A elements on the 8i, 8j, and 8f positions (\(j = 2, 3, 4\)), and two different interstitial elements on the 2b position (\(j = 5\)) are then written as

\[
\mathbf{x} = (n_1^{(1)}, n_2^{(1)}, n_3^{(3)}, n_4^{(3)}, n_5^{(4)}, \ldots, n_{4}^{(N-2)}, n_{5}^{(N-1)}, n_{5}^{(N)}),
\]

and have 13 components in total [instead of \(N \cdot P = 7 \cdot 5 = 35\) with Eq. (7)]. Fig. 2 illustrates the descriptors constructed in this way for selected Re(Fe,A)_{12}X compounds.

This description has the advantage that structural and chemical information are encoded separately. We briefly note that this representation is somewhat similar to the descriptor which has been used very recently by Faber et al. for predicting the formation energies of elpasolite crystals [14].

The construction of this descriptor involves a space group analysis of the crystal structure, for which we use the FINDSYM tool of Stokes & Hatch as implemented in the ISOTROPY Software Suite [50, 51].

2.4. Optimization of Compositions

To efficiently search for compositions that exhibit optimal combinations of \(\mu_0 M\) and \(K_1\), we optimize the model functions \(y^m(\mathbf{x})\) (with \(y\) being \(\mu_0 M\) or \(K_1\)) using the stochastic basin-hopping optimization (BHO) algorithm as implemented in the Python package scipy [52]. Details of the optimization procedure are given in the A. BHO has the advantage of allowing both bounded and constrained components of \(\mathbf{x}\). The components of the Wyckoff descriptor have to be bounded because each of them has a lower limit (0) and an upper limit (maximal number of atoms per Wyckoff position, e.g., eight atoms for the 8i sites). At the same time, the components of \(\mathbf{x}\) are constrained because each non-zero component decreases the possible maximal value of all other components representing the same Wyckoff position.

After optimization, the descriptor \(\mathbf{x}'\) that fulfills the search criteria, i.e., \(y^m(\mathbf{x}') > y_t\) (subscript ‘t’ for target), can be directly mapped to the corresponding chemical composition and distribution of the elements on the Wyckoff positions. This fast and efficient procedure directly benefits from the bijective nature of the Wyckoff descriptor, which allows a reversible mapping between \(\mathbf{x}\) and the respective material.

In the optimization, we effectively seek descriptors that maximize the anisotropy coefficient \(K_1\) (and thereby the anisotropy field \(H_\text{a}\)) by using a relatively conservative target value for \(\mu_0 M_\text{t} = 1.4\) T since it is well known that the highest values for \(\mu_0 M\) are given by compositions containing mainly Fe with (small) contents of Co (see Slater-Pauling curve [53–55]). Note that we are not aiming at minimizing \(E_\text{f}\) as a consequence of limitations in the input data since our reference database for determining \(E_\text{f}\) contains only elemental crystal phases to compare with. For this reason, we only use the criterion \(E^\text{ft}_\text{t} < E^\text{ft}_\text{t} + \text{moderate target value } E^\text{ft}_\text{t} = 0.1\) eV/atom for the optimized compositions (because \(\pm 0.1\) eV/atom is about the predictive power of the TB-LMTO-ASA method for compound formation energies).

3. RESULTS

Since the hyperparameters for the SVR algorithm (kernel function \(k\), regularization parameter \(C\), and width \(\gamma\)) are not known beforehand, we first determine the optimal set of hyperparameters that maximizes the predictive power of the model and minimizes its tendency for overfitting. The performance of the optimized ML models is then tested for unseen compositions from our database. Finally, we use an optimization procedure to find the material composition \(\mathbf{y}\)—more precisely—its descriptor, for which our two figures of merit are within the desired range of values: \(\mu_0 M > 1.4\) T and \(K_1\) maximal.
3.1. Training the ML Models

To find the optimal set of hyperparameters for the ML models, we varied the regularization parameter $C$ for the SVR as well as the kernel width $\gamma$ for various kernel functions. $C$ was varied between 0.1 and 1000 and $\gamma$ between $10^{-7}$ and 1.0, with the kernels being linear, polynomial (degree $d = 2$), and rbf functions. The numeric hyperparameters $C$ and $\gamma$ were varied on a logarithmic grid.

To validate the ML models built with a specific set of hyperparameters, we determine the Pearson correlation coefficient $\rho$ and the mean absolute error (MAE) \cite{42,56},

\[
\rho(y, y^m) = \frac{\text{cov}(y, y^m)}{\sigma_y \sigma_{y^m}},
\]

\[
\text{MAE}(y, y^m) = \frac{\sum_{i=1}^{N} \| y^m_i - y_i \|}{N_{\text{samples}}},
\]

where cov is the covariance function, $\sigma$ denotes the standard deviation, and $N_{\text{samples}}$ is the number of samples. The correlation coefficient $\rho$ measures the linear relationship between two (normally distributed) datasets. It varies between $-1$ and $+1$, where values close to zero indicate that no correlation exists. The limiting values $\rho = +1 (-1)$ imply a perfectly direct (indirect) linear proportionality between $y^m$ and $y$. The MAE (in units of the predicted property) indicates the average difference between $y^m$ and $y$ and should therefore be as small as possible.

To estimate whether the model is over-fitted, i.e., it exactly matches the training samples but nothing else, we performed tenfold cross validation (CV) \cite{42}. The data set used for building and optimizing the ML model consists of 234 (232) samples of type $\text{ReFe}_{12.4x}\text{A}_{4x}X$ for $\text{Re} = \text{Nd}$ (Ce). It is randomly divided into ten subsets of which nine are used to train a new ML model and the tenth is used for validation. This procedure is repeated ten times and the average values of the correlation metrics $\rho$ and MAE for the ten validation sets are used as a measure of the performance of the model with respect to unseen test data.

### TABLE I. Optimal set of hyperparameters [kernel $k$, degree $d$ for polynomial ($p$) kernels, regularization parameter $C$, and kernel width $\gamma$] and correlation metrics ($\rho$ and MAE) from tenfold CV. $\mu_0 M$ varies between 0 and 2 T; $E_i$ varies between $-1$ and 1 eV/atom; $K_1$ varies between 10 and 90 MJ/m$^3$ for Nd and between 50 and 230 MJ/m$^3$ for Ce. For $\mu_0 M$, only configurations with a significant magnetization of $\mu_0 M > 0.25$ T have been used for training (181 for Nd, 179 for Ce).

| Property (unit) | Hyperparameters | $C$ (Nd) | $C$ (Ce) |
|-----------------|-----------------|---------|---------|
| $\mu_0 M$ (T)   | $k(d), C, \gamma$ | $\rho$ (MAE) | $\rho$ (MAE) | $\rho$ (MAE) |
| $K_1$ (MJ/m$^3$) | $p(1), 100, 1$ | 0.95 (0.09) | 0.91 (0.11) |
| $E_i$ (eV/atom) | $p(2), 0.1, 0.1$ | 0.95 (0.043) | 0.94 (0.04) |

The optimal set of hyperparameters for each combination of ML model and material property were determined by taking those hyperparameters for which $C$ and $\gamma$ were minimal and satisfying $\rho > 0.95 \rho_{\text{max}}$ ($\rho_{\text{max}}$: maximal correlation coefficient for each model/material property combination). This ensured that the models were both as accurate and as universal as possible at the same time. The optimal set of hyperparameters and the corresponding correlation metrics $\rho$ and MAE obtained with these hyperparameters are listed in Tab. I. The correlation coefficients obtained with CV are in the range of $\rho = 0.88 \ldots 0.95$ for the models with the optimal set of hyperparameters. The CV results are displayed in Fig. 3 for the Nd(Fe,A)$_{12}$X compounds.

3.2. Testing of the ML models with unseen compositions

For testing our ML models we use test sets that contain configurations, where the the alloying atoms fractionally occupy the 8i Wyckoff site. The resulting predictions for Nd(Fe,A)$_{12}$X are presented in Fig. 4(top) for SVR and Fig. 4(bottom) for LR. The determined correlation coefficients $\rho$ and MAEs are summarized in the last column of Tab. I. For $K_1$ and $E_i$, $\rho$ is somewhat lower than the CV results while the correlation for $\mu_0 M$ is similar to the one determined by CV.

Most of the outliers in the SVR predictions for $\mu_0 M$, see Fig. 4(top), originate from Mn-containing compounds. This can be attributed to the magnetic behavior of Mn, which can change from ferromagnetic to anti-ferromagnetic magnetism depending on the precise atomic arrangement \cite{57}. For $K_1$ and $E_i$, most of the outliers are attributed to phases where the non-ferromagnetic elements Al, P, Si, Ti, and Zn represent more than 66% of all A elements in the compound. In both cases, the evolution of the calculated properties with the elemental concentrations is less systematic and predictions are inherently difficult. It is important to note that the outliers are, however, uncritical for the final optimization because they do not lead to compositions that meet our search criteria.

The test results give us a good indication on how well the ML models will interpolate between the training data, or in other words, how well the models will predict the whole property space of the $\text{ReA}_{12}X$ crystal structure (for the trained $\text{Re}$, $\text{A}$, and $X$ elements). Note that the achieved MAEs are of the same magnitude as the desired accuracy $\varepsilon$ of the SVR models (see section 2.1).

3.3. Optimal Compositions

A selection of the obtained optimal compositions is presented in Tab. II along with their respective values for $\mu_0 M$ and $K_1$ and the expected values for $(BH)_{\text{max}}$ and $H_a$ estimated according to Eqs. (5-6).
FIG. 3. Results of the tenfold cross validation of the SVR models for $\mu_0 M$ (left), $K_1$ (center), and $E_f$ (right) for Nd-containing compounds. The colored circles indicate the predictions for each of the ten validation runs. The used (optimal) sets of hyperparameters are given in Tab. I. The MAEs are given in the units of the corresponding axes.

FIG. 4. Testing of the SVR models (top) and LR (bottom) for $\mu_0 M$ (left), $K_1$ (center), and $E_f$ (right) for Nd-containing compounds. The colored circles indicate the corresponding predictions for the 832 unseen compositions used as test samples. The grey squares are the predictions for the models trained on the whole data set (training and validation data). The used (optimal) sets of hyperparameters for the SVR model are given in Tab. I. The MAEs are given in the units of the corresponding axes.

Even though the test results (section 3.2) already indicate that the ML predictions should be accurate, we veri-
fied the ML-predicted values with \textit{a posteriori} performed TB-LMTO-ASA calculations. In addition to the optimal compositions, the magnetic properties of the most promising candidates drawn from our database [32] as well as the state-of-the-art benchmark materials SmCo$_5$ and Nd$_2$Fe$_{14}$B are presented for completeness. The predicted and calculated values are compared to experimental data whenever available. Note that the values for the formation energy $E_f$ are not given in the table explicitly, but they are negative for all configurations shown.

The optimal compositions given in Tab. II exemplify the versatility of our approach, i.e. the combination of an optimization algorithm with a bijective descriptor. Two compositions were, for instance, optimized under the constraint that they should contain Ti, which is known for its beneficial effect on the phase stability [61]. The predicted NdFe$_{8.5}$Cu$_{3.5}$N phase, as another example, avoids the elements Co and Ni, which are both more expensive than Cu.

4. DISCUSSION

In the discussion, we focus on the comparison between the ML predictions and DFT-determined values of the optimal compositions. In addition, we critically assess the potential of the proposed compositions to substitute the state-of-the-art materials NdFe$_{14}$B and SmCo$_5$. Finally, we demonstrate the potential benefit of using kernel-based ML methods over classical LR fitting.

4.1. Comparison between ML predictions and DFT

Although the ML models were trained only to samples containing one alloying element (in addition to Fe), it is remarkable that the predictions for the optimal compositions (which contain often at least two alloying elements) are in overall very good agreement with the DFT calculations performed \textit{a posteriori} (the average ML-DFT difference is below 4%). A plausible reason for this is the property of iron, other transition metals (Tm), and their alloys (whose electronic band structures are mainly formed by the Tm d-orbitals) that along the Tm series in the periodic systems the band structures and densities of states for a specific crystal structure are only scaled in width but hardly changed in shape with varying number of electrons occupying the bands (see, e.g. Refs. [62, 63]).

The apparent success of the machine to learn the underlying physics should not distract us from the fact that the interpolation behavior between training samples is mainly determined by the kernel function. This property of ML models is visualized in Fig. 5 where the predicted $K_1$ values for different contents of Co on the 8i site, $n_{8i}(Co) = 0 \ldots 8$ ($z$), in the NdFe$_{12-z/2}Co_{z/2}N$ phase space are plotted for different kernel functions in comparison with TB-LMTO-ASA results. Whereas with the polynomial kernel ($d = 2$), the nearly linear increase of $K_1$ is very well reproduced, the linear kernel fails already in predicting the training data at $z = 0$ and 8, respectively. Due to its higher flexibility, the ‘rbf’ kernel exactly reproduces the training data points. The drawback of this flexibility is that the deviation from the DFT results becomes very large in the range of $z$ values.

One of the recipes of success for ML of materials properties is therefore the combination of a descriptor, whose components have an almost linear relationship with the modelled properties, and a comparably inflexible kernel function. Since many physical properties have polynomial relationships of low degrees with the material composition, this approach is expected to work equally well for other materials properties.

At the same time, one limitation of our approach is that the employed descriptor encodes all chemical and structural information implicitly. ML Models built with this descriptor are thus not transferable to other crystal structures. Promising \textit{universal} descriptors, which may overcome this shortcoming, have already been proposed for molecular systems [26, 64] and their suitability for use needs to be tested for periodic crystals in the future [13].

4.2. Potential of the optimized compositions

The hard-magnetic properties of both benchmark materials, which contain comparably high amounts of RE elements, are currently known as the upper limits for $(BH)_{\text{max}}$ (NdFe$_{14}$B) and $H_a$ (SmCo$_5$). As to be expected, our candidates do not beat these values, but they represent alternatives with significantly reduced RE contents.

In comparison with the best results from the HTS
TABLE II. Selected compositions with high $K_1$ and $\mu_0 M$ above 1.4 T found by optimizing the ML models built for the REA$_{12}$X crystal structure. Predictions of the ML models are indicated with a superscript ‘m’ and compared to a posteriori performed DFT (TB-LMTO-ASA) calculations. The Wyckoff positions of Fe substitutes are specified as superscripts (e.g., Co$_{0.5}$ means that Co occupies one out of eight 8i positions, which is equivalent to 0.5 atoms in the formula unit). The best candidates from the HTS [32] are marked with an asterisk (*). The calculated $K_1$ values are scaled (after the slash) for better comparability with experiments (see section 2.2 for the heuristic scaling factors). The macroscopic values for $(BH)_{max}$ and $H_a$ are estimated according to Eqs. (5-6). The available experimental values for $\mu_0 M$, $K_1$, $(BH)_{max}$, and $H_a$ are given in parentheses. $E_f$ is negative for all listed compositions.

| Composition | Predictions of ML Models | TB-LMTO-ASA Calculations | Estimations of macroscopic values |
|-------------|--------------------------|---------------------------|----------------------------------|
|             | $\mu_0 M^m$ (T) $K_1^m$ (MJ/m$^3$) | $\mu_0 M$ (T) $K_1$ (MJ/m$^3$) | $(BH)_{max}^{EST}$ (kJ/m$^3$) $H_a$ (T) |
| CeFe$_8$Co$_4$/$^8/16$Cu$_4/16$Si$_1/16$N | 1.45 165 | 1.45 169 / 4.8 | 337 233 |
| CeFe$_7$Ni$_4$/$^8/16$Co$_2/8$N | 1.48 167 | 1.57 170 / 4.9 | 398 216 |
| CeFe$_8$Ni$_4$/$^8/16$N* | — — | 1.61 167 / 4.8 | 417 207 |
| NdFe$_8$Co$_2$/$^8/16$Co$_4/8$Si$_1/4$C | 1.42 57 | 1.46 60 / 15 | 346 83 |
| NdFe$_8$Co$_3$/$^8/16$Ni$_4$ | 1.48 57 | 1.57 55 / 14 | 397 70 |
| NdFe$_8$Ni$_4$/$^8/16$N* | — — | 1.68 57 / 14 | 457 67 |

| Benchmark Materials |
|---------------------|
| SmCo$_5$ | — — | 1.07 (1.07 [58]) 69 (26 [58]) | 184 (219 [59]) 129 (40.4 [58]) |
| Nd$_2$Fe$_{14}$B | — — | 1.87 (1.86 [60]) 19 (4.9 [58]) | 563 (516 [59]) 20 (6.6 [58]) |

Finally, intermetallic phases with hypothetically good hard-magnetic properties can exist in reality only if they have sufficient thermodynamic stability. As stated above, in our DFT database of Ref. [32] on which our present ML work is based, the formation energies of intermetallic phases are calculated with the TB-LMTO-ASA method and only with respect to their constituting single-component phases. The ASA is a well suited and justified approximation for calculating hard-magnetic key quantities $\mu_0 M$ and $K_1$ of REA$_{12}$-based phases with approximately close-packed crystal structure [32, 47]. For estimating formation-energy differences between competing multi-component phases more precisely, total-energy calculations with more accurate full-potential DFT methods are necessary. However, theoretically predicted phases with promising hard-magnetic properties are already valuable information for experimentalists who want to produce hard-magnetic materials by metallurgical means even without prior theoretical stability prediction. Future work on extending DFT databases to hard-magnetic phases with accurate phase-stability information is a desirable task. However, for the present ML work it is not an issue because the presented ML approach to predict chemical compositions of the specific REA$_{12}$ structure type with good hard-magnetic properties, is not limited by a limited precision of the thermodynamic-stability data. Once that the DFT database is extended with more accurate full-potential DFT data for formation energies of competing multi-component phases, the presented ML approach will be immediately applicable to such an improved database.

4.3. Comparison between ML and LR

As it can be seen from the testing results in Fig. 4, the $\mu_0 M$ correlation metrics for the SVR ML model (with a linear kernel function) are nearly identical to those of LR. This clearly illustrates that the benefits of ML come only into play if a non-linear kernel (as for $K_1$ and $E_f$) is used. In our work, this improvement is a factor of 2 for the MAE and around 0.2 for the correlation coefficient.

5. SUMMARY

In this work, we successfully applied a ML approach to a common materials science problem, namely the search for a material composition that optimizes a certain physical property. Based on our HTS database, we trained ML models that reliably predict the hard-magnetic key properties of RE(Fe$_{A}$)$_{12}$X compounds with up to three different non-ferrous alloying elements $A$ based on input data with compounds containing only one alloying element. By optimizing the ML model functions we identified optimal compositions for the REA$_{12}$X structure that exhibit an increase in magnetocrystalline anisotropy field $H_a$ by more than 10 % as compared to our previous HTS results. Although in both $(BH)_{max}^{EST}$ and $H_a$ the predicted
optimal compositions do not excel the current state-of-the-art materials SmCo$_5$ and Nd$_2$Fe$_{14}$B at the same time, they can still be valuable for bridging the gap between these [comparably low $(BH)_{\text{max}}$ and high $H_k$ for SmCo$_5$ and vice versa for Nd$_2$Fe$_{14}$B] while having significantly lower RE contents and being therefore more potential cost efficient and less supply critical.

Our results clearly highlight the potential of ML methods for materials discovery and design. With an appropriate choice of model hyperparameters and material representation it is possible to accurately predict the materials properties in the entire compound space and to identify compositions that optimize the learned property. In the end, the applicability of ML is only limited by the availability and accuracy of training data. For the future, we believe that ML of materials property. In the end, the applicability of ML is only limited by the availability and accuracy of training data. For the future, we believe that ML of materials properties will be especially helpful, when the underlying structure-composition-property relationships are not yet completely clarified, as e.g., in the case of $K_1$ for novel RE-lean magnets.

6. ACKNOWLEDGMENTS

Financial support for this work was provided by the Fraunhofer Lighthouse Project Critical Rare Earths.

Appendix A: Optimization details

Using basin-hopping optimization, we intend to determine the descriptor $x$ that minimizes the following function:

$$F(x) = -\Xi_1(E_t^m, E_t)\Xi_2(\mu_0 M^m, \mu_0 M_t)\Xi_2(K_1^m, K_{1t})$$

where the subscript 't' denotes the desired target values. Note that $\mu_0 M^m$, $K_1^m$, and $E_t^m$ are functions of $x$, which is omitted in Eq. (A1) for better readability. The smooth threshold function $\Xi_1$ (Eq. (A2)) becomes one ($\mu_0 M^m$, $K_1^m$, and $E_t^m$) below (above) the target value $E_{t0}$ ($\mu_0 M_t$, $K_{1t}$). Both functions converge to zero otherwise. The functions $\Xi_1 (E_t)$ and $\Xi_2 (\mu_0 M$ for $K_1)$ are defined as follows:

$$\Xi_1(z) = \frac{1}{2} \text{erfc} \left[ \frac{5(z - 1/2)}{2} \right]$$

$$\Xi_2(z)/y^m = 1 - \frac{1}{4} \{ \text{tanh} [-5(z - 0.95)] + 1 \}$$

with $z = y^m/y_t$ and $y$ being either $\mu_0 M$, $K_1$, or $E_t$, erf is the error function and erfc is its complement. Obviously, the complementary error function in $\Xi_1 (z)$ appears as being a logical choice for a function that should penalize values which become higher than certain target values.

The functional form of $\Xi_2 (z)$, on the other hand, needs some more explanation. It was chosen such that values smaller than the target value are less strongly penalized than values above it. By comparing $\Xi_2$ with its two main point-symmetric constituents, namely the tanh and erf functions, this behavior is visually exemplified in Fig. 6. The somewhat unorthodox choice of $\Xi_2 (z)$ has the objective to 'guide' the optimization algorithm in the correct direction (increasing $\mu_0 M$ and $K_1$ values). Above the thresholds, however, all values contribute equally to $F(x)$, i.e., only by their magnitudes.

---

[1] Kiyohara, S., Oda, H., Miyata, T., Mizoguchi, T., Prediction of interface structures and energies via virtual screening. Sci Adv 2016;2(11).

[2] Kiyohara, S., Oda, H., Tsuda, K., Mizoguchi, T., Acceleration of stable interface structure searching using a kriging approach. Jpn J Appl Phys 2016;55(4):044502. URL: http://stacks.iop.org/1347-4065/55/i=4/a=044502?key=crossref.6edc2feef01670642598a371ff15639a. doi:10.7567/JJAP.55.044502

[3] Rosenbrock, C.W., Homer, E.R., Csányi, G., Hart, G.L.W.. Discovering the Building Blocks of Atomic Systems using Machine Learning 2017:URL: http://arxiv.org/abs/1703.06236. arXiv:1703.06236.

[4] Balachandran, P.V., Xue, D., Lookman, T.. Structure-Curie temperature relationships in BaTiO3 -based ferroelectric perovskites: Anomalous behavior of (Ba,Cd)/TiO3 from DFT, statistical inference, and experiments. Phys Rev B 2016;93(14):1–12. doi:10.1103/PhysRevB.93.144111
