

PAPER • OPEN ACCESS

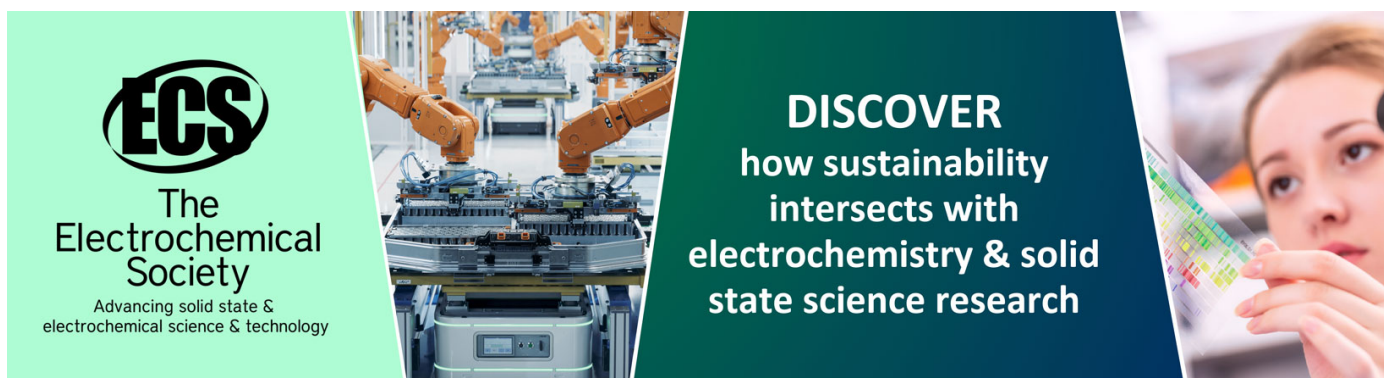
Magnetic microstructure machine learning analysis

To cite this article: Lukas Exl *et al* 2019 *J. Phys. Mater.* **2** 014001

View the [article online](#) for updates and enhancements.

You may also like

- [Stripe domains reorientation in ferromagnetic films with perpendicular magnetic anisotropy](#)
Louis-Charles Garnier, Massimiliano Marangolo, Mahmoud Eddrief *et al.*
- [Renormalized phonon spectrum in the Su–Schrieffer–Heeger model](#)
Stepan Fomichev and Mona Berciu
- [Machine learning \(ML\)-assisted surface tension and oscillation-induced elastic modulus studies of oxide-coated liquid metal \(LM\) alloys](#)
Kazi Zihan Hossain, Sharif Amit Kamran, Alireza Tavakkoli *et al.*



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research



PAPER

Magnetic microstructure machine learning analysis

OPEN ACCESS

RECEIVED
17 September 2018

REVISED
1 November 2018

ACCEPTED FOR PUBLICATION
20 November 2018

PUBLISHED
14 December 2018

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Lukas Exl¹ , Johann Fischbacher² , Alexander Kovacs² , Harald Oezelt² , Markus Gusenbauer² , Kazuya Yokota^{2,3,4}, Tetsuya Shoji^{3,4}, Gino Hrkac⁵ and Thomas Schrefl²

¹ WPI c/o Faculty of Mathematics, University of Vienna, A-1090 Vienna, Austria

² Department for Integrated Sensor Systems, Danube University Krems, Viktor Kaplan Str. 2/E, A-2700 Wiener Neustadt, Austria

³ Toyota Motor Corporation, Higashifuji Technical Center, Advanced Material Engineering Division 1200 Mishuku, Susono, Shizuoka, 410-1193, Japan

⁴ Technology Research Association of Magnetic Materials for High-efficiency Motors (MagHEM), Higashifuji-Branch, 1200 Mishuku, Susono Shizuoka 410-1193, Japan

⁵ College of Engineering, Mathematics and Physical Sciences, The University of Exeter, Exeter, EX4 4SB, United Kingdom

E-mail: tschrefl@gmail.com

Keywords: permanent magnets, machine learning, model order reduction, feature selection

Abstract

We use a machine learning approach to identify the importance of microstructure characteristics in causing magnetization reversal in ideally structured large-grained Nd₂Fe₁₄B permanent magnets. The *embedded Stoner–Wohlfarth method* is used as a reduced order model for determining local switching field maps which guide the data-driven learning procedure. The predictor model is a random forest classifier which we validate by comparing with full micromagnetic simulations in the case of small granular test structures. In the course of the machine learning microstructure analysis the most important features explaining magnetization reversal were found to be the misorientation and the position of the grain within the magnet. The lowest switching fields occur near the top and bottom edges of the magnet. While the dependence of the local switching field on the grain orientation is known from theory, the influence of the position of the grain on the local coercive field strength is less obvious. As a direct result of our findings of the machine learning analysis we show that edge hardening via Dy-diffusion leads to higher coercive fields.

1. Introduction

Permanent magnets are widely used in modern society. The high performance magnet market is dominated by Nd₂Fe₁₄B magnets. The six major application areas are acoustic transducers, air conditioning, electric bikes, wind turbines, hybrid and electric cars, and hard disk drives [1, 2]. Growing demands for permanent magnets are predicted for green technology applications such as sustainable energy production and eco-friendly transport. The generator of a direct drive wind mill requires high performance magnets of 400kg/MW power; and on average a hybrid and electric vehicle needs 1.25 kg of high end permanent magnets [3]. Another rapidly growing market is electric bikes. The global demand for rare-earth elements in permanent magnets will exceed 50 thousand tons per year in 2025 [3]. With the quest for rare-earth reduced or rare-earth free permanent magnets [4], an optimal control of the magnet's microstructure becomes increasingly important. In other fields of materials research, data driven machine learning approaches have been applied recently, in order to obtain a deeper understanding of the material's microstructure on its properties. Mangal and Holm [5] combined crystal plasticity based simulation with machine learning techniques for predicting stress hot-spots in polycrystalline metals. Using random-forest (RF) based machine learning they correlate the formation of grains with high stress by uniaxial tensile deformation with local microstructural features that describe crystallography, geometry, and connectivity. In another paper [6], they addressed the problem of feature selection for the classification of stress hot spots. They showed that a proper set of microstructural features is required, in order to find out what microstructural characteristics will cause high local stress during tensile deformation.

Modern Nd₂Fe₁₄B permanent magnets show a granular structure. Ideally, the grains are separated by a nonmagnetic grain boundary phase [7]. In order to improve the isolation of the grains by a nonmagnetic Nd-rich grain boundary phase, a high Nd content and a dopant such as Al [7] or Ga [8] are required. In this work we investigate the influence of the microstructure on the local coercivity of permanent magnets with ideal structure. We assume grains that are completely separated by a nonmagnetic phase, and we do not introduce any soft magnetic defects. Using machine learning techniques we identify the microstructural characteristics that may cause weak grains, which are defined as the grains that will reverse first when an increasing opposite field is applied to the magnet. By neglecting defects and ferromagnetic grain boundaries we focus on the effects of key structural features that are common to any polycrystalline material such as grain size, grain shape, grain sphericity, and crystallographic orientation. We use machine learning techniques to study microstructural features only and choose as a reference material Nd₂Fe₁₄B as the nowadays most important permanent magnets material. The anisotropy field of Nd₂Fe₁₄B is $\mu_0 H_A = 7.65$ T, whereby it is considered to be the maximum possible coercive field [9]. Since we do not include any soft magnetic defects, we expect nucleation fields in the range of 3–7.65 T for the investigated ideal grain structures.

2. Methods

2.1. Dataset generation

We investigate magnetic multigrain structures in view of their switching field distribution aiming at predicting grains with low switching field (weak grains) and those with high switching field (strong grains), respectively. More precisely, we take the 20th percentile of the lowest switching fields in the distribution as a threshold for labeling weak grains. We generate synthetic microstructures consisting of polyhedral grains using the software Neper [10, 11]. We use the default grain growth parameter which gives a wider grain size distribution and higher grain sphericities than a standard Voronoi tessellation. The grain size normalized by the average grain size, $D/\langle D \rangle$, follows a lognormal distribution with a standard deviation of 0.35. The sphericity s is a metric for the shape of the grains [11]. It is defined as the ratio of the surface area of a sphere with equivalent volume to the surface area of the grain. The quantity $1 - s$ follows a lognormal distribution with a mean of 0.145 and a standard deviation of 0.03. We investigate three scenarios depending on the standard deviation of the misorientation angle of the anisotropy direction: $\sigma_\theta = 0^\circ$, $\sigma_\theta = 5^\circ$, and $\sigma_\theta = 15^\circ$. For each scenario 10 synthetic microstructures with 1000 grains each were generated. Seven structures were randomly selected to form the so-called *training set*, hence containing 7×1000 grains. The remaining three structures build the so-called *test set*, which is used in the very end to measure the performance of the predictor model. Figure 1(a) shows a typical microstructure. Figures 1(b)–(d) show the distributions of some features in the training set: the misorientation angle of the anisotropy axes, the distance of the grain from the magnet's center, and the grain size.

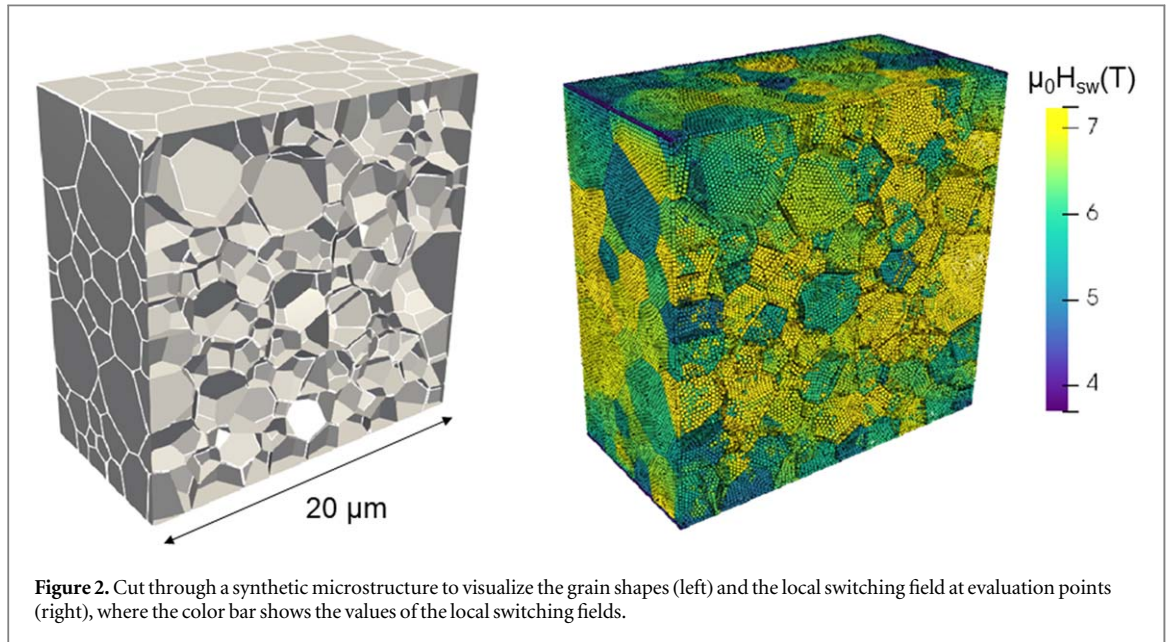
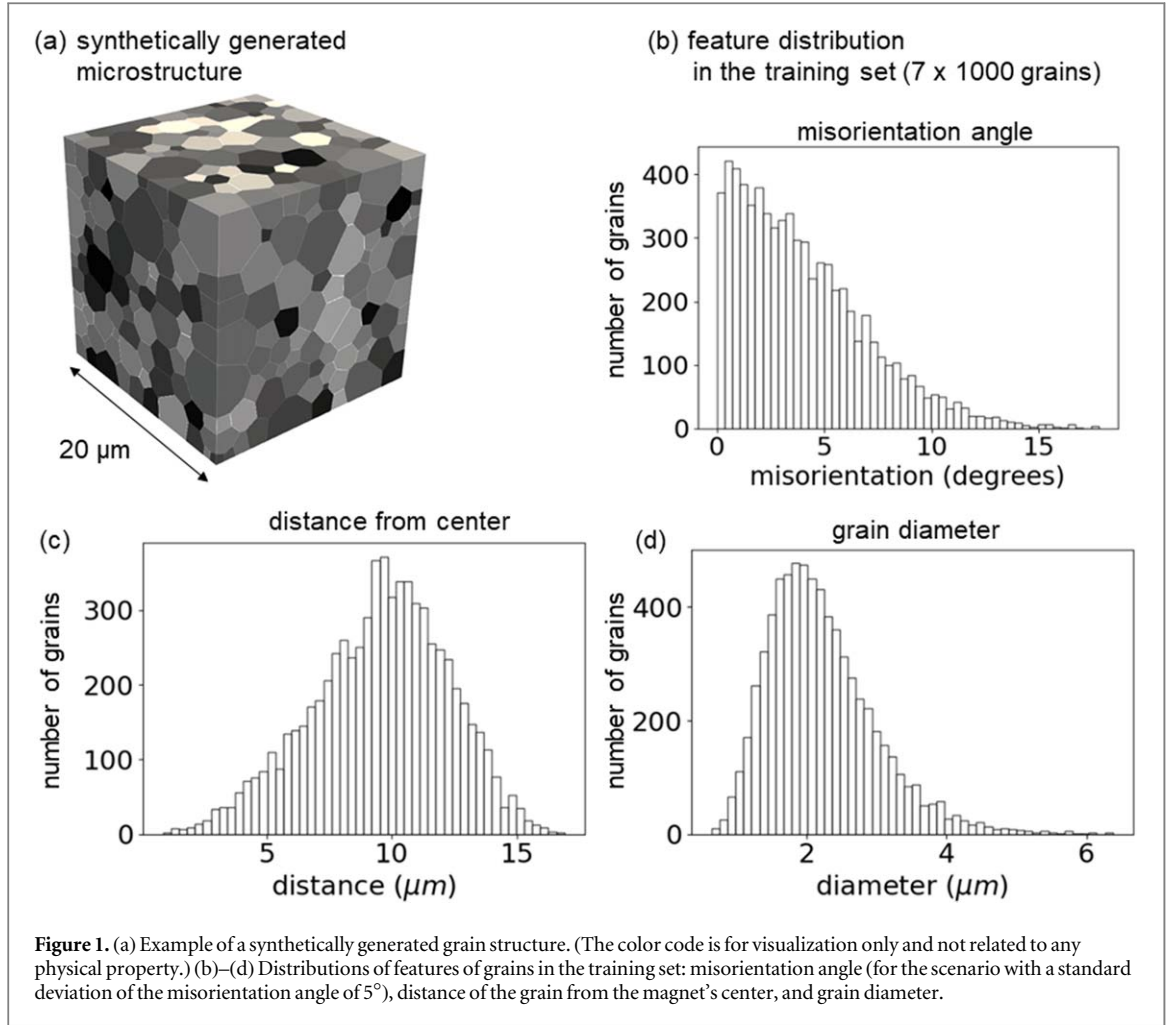
Switching field values are calculated near the surface of the grains which serve as underlying datasets for the *microstructure machine learning analysis*. Figure 2 shows a cut through the grain structure, the locations of the field-evaluation points, and the calculated switching fields. Since there are no pinning sites for domain walls within a grain a reversed domain will expand through the grain once it is nucleated. Therefore, the minimum value of the switching fields within a grain defines its reversal field which is used for machine learning. For the simulations we use the material properties of Nd₂Fe₁₄B (anisotropy constant $K_1 = 4.9$ MJ m⁻³, spontaneous magnetic polarization $\mu_0 M_s = 1.61$ T, and exchange constant $A = 8$ pJ m⁻¹ [12]) and a mean grain size of 2 μ m. Here μ_0 is the permeability of vacuum.

2.1.1. Embedded Stoner–Wohlfarth (ESW) method

The micromagnetic calculation of switching fields in permanent magnet models relies on hysteresis computation usually using numerous successive total energy minimization steps for slightly varying external field strength. This is only feasible for models in the nanometer regime with a few grains. Since our data driven approach requires hundreds of grains our models are too large for conventional micromagnetic simulations. Hence we apply a reduced order model for the prediction of critical fields, called the *ESW method* [13]. The approach has its origin in the work of Schrefl and Fidler [14] and adapts the original Stoner–Wohlfarth model for small ferromagnetic particles in a way to additionally account for long-range interactions of uniformly magnetized grains. First the total field is calculated

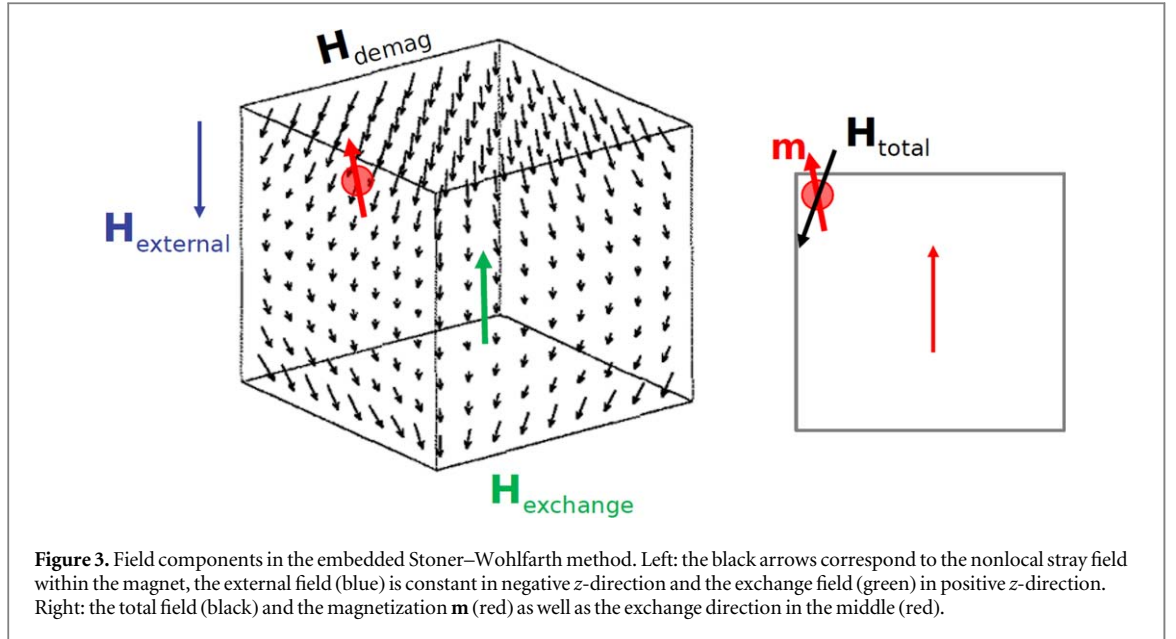
$$\mathbf{h}_{\text{tot}} = \mathbf{h}_{\text{ext}} + \mathbf{h}_{\text{demag}} + \mathbf{h}_x, \quad (1)$$

at evaluation points which are located at a distance d from the grain surface. The total field is the sum of the external field \mathbf{h}_{ext} , the demagnetizing field $\mathbf{h}_{\text{demag}}$, and the exchange field \mathbf{h}_x . To this end the stray field computations are accomplished by analytical formulas for polyhedral geometries [15] calculating



$$\mathbf{h}_{\text{demag}}(\mathbf{x}) = \frac{1}{4\pi} \sum_j \sum_k \int_{S_{jk}} \mathbf{n}_{jk}(\mathbf{x}') \cdot \mathbf{m}_j \frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3} dS'. \quad (2)$$

The demagnetizing field at point \mathbf{x} is the sum of surface integrals over the surfaces of all polyhedra S_{jk} , where the index j runs over all grains and the index k over the surfaces of a grain. The vector \mathbf{n}_{jk} is the outer normal to



surface k of grain j ; and \mathbf{m}_j is the magnetization vector of grain j . The perpendicular component of the demagnetizing field grows with no bound towards the edges of a polyhedron, which is compensated by the exchange field [16]. The exchange field comes from the Heisenberg exchange $\mathbf{h}_x(\mathbf{x}) = -\sum J_{xi} \mathbf{s}_i$ taking into account the exchange interactions of the spin at point \mathbf{x} with the neighboring spins \mathbf{s}_i . In a continuum approach the exchange integral J_{xi} is replaced by an expression containing the exchange constant A [17]. Here we assume that the spins within a grain remain parallel before switching. Therefore the exchange field acting at the evaluation point \mathbf{x} is parallel to the magnetization vector of the grain. In the ESW model we set its magnitude to $h_x = (1/(\mu_0 M_s))A/d^2$ [13]. Please note that in an ideal $\text{Nd}_2\text{Fe}_{14}\text{B}$ permanent magnet the grains are separated by a nonmagnetic grain boundary phase [7]. Therefore we do not take into account any exchange interactions with neighboring grains. Figure 3 shows the field components in a cubic particle. The distance d is $1.2L_{\text{ex}}$. The exchange length L_{ex} is $\sqrt{A/(\mu_0 M_s^2)}$.

According to Stoner–Wohlfarth [18], the switching field of a small uniformly magnetized particle can be given in terms of the angle ψ between the easy axis and the total field by the formula [19]

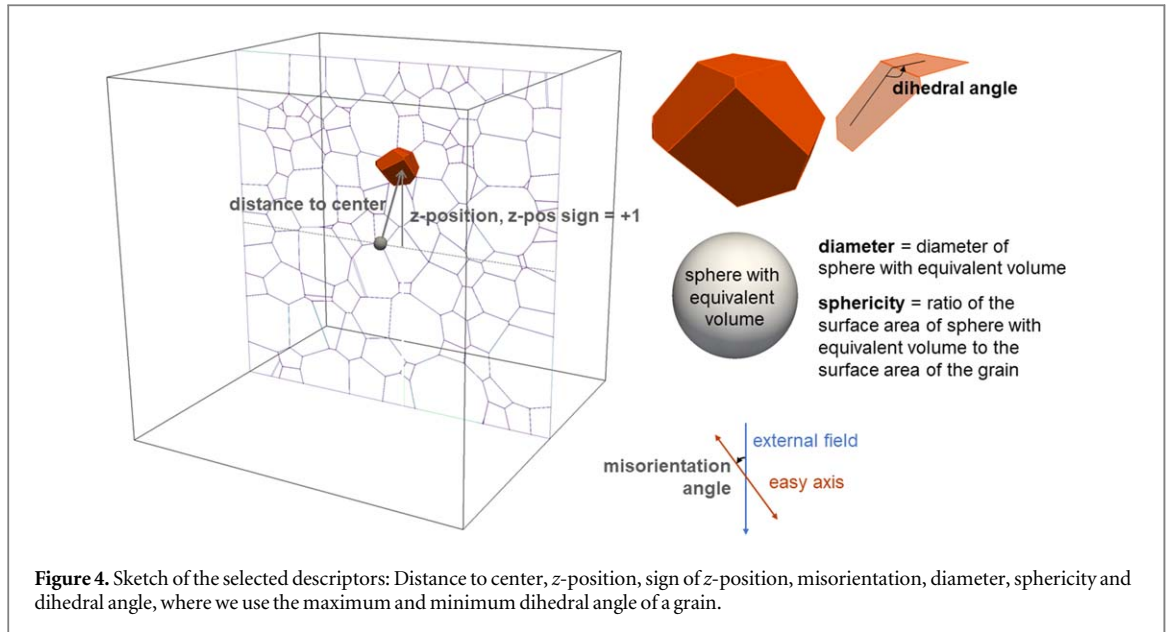
$$h_{\text{sw}} = f(\psi)h_N, f(\psi) = (\sin^{2/3}\psi + \cos^{2/3}\psi)^{-3/2}, \quad (3)$$

where h_N is the ideal nucleation field [20], $h_N = 2K_1/(\mu_0 M_s)$. In a hard magnetic particle the easy axis coincides with the magneto-crystalline anisotropy direction. The Stoner–Wohlfarth switching field (3) is evaluated for a varying external field at target points at a distance d away from the surface of the polyhedral grains [13, 21], where the angle between the anisotropy direction and the total field (1) is taken. Please note that in the remanent state the magnetization can be assumed to be approximately parallel to the anisotropy direction. The local switching field at a target point is the smallest value of $|\mathbf{h}_{\text{ext}}|$ which makes the total field greater than the value obtained from (3), that is $|\mathbf{h}_{\text{tot}}| > h_{\text{sw}}$. Then we compute the minimum switching field over all target points of a grain. This minimum value is the switching field of the grain, which is then used for labeling weak and strong grains in the subsequent machine learning task.

2.1.2. Microstructure attributes

Our main intuition is that weak points in permanent magnet grain structures can be well understood by their (mainly) geometrical microstructure attributes. The machine learning approach will assign these features to each grain together with the grain label (weak or strong grain) according to calculated switching field values using the ESW model as an effective reduced order model. The following geometrical attributes are assigned:

- The absolute value of the z -coordinate of the center of the polyhedral grain measured from the center of the cube ($z\text{-position}$).
- The sign of the z -position ($z\text{-pos sign}$).
- The distance to the center of the magnet (distance).
- The diameter of the polyhedron (diameter) defined as the diameter of a sphere with equivalent volume.
- The number of next neighbor grains (no of neighbors).



- The sphericity of the grain (*sphericity*).
- The absolute deviation of the current grain diameter from the average diameter of the next neighbors (*diam variation*).
- The maximum dihedral angle of the polyhedron (*max dihedral angle*).
- The minimum dihedral angle of the polyhedron (*min dihedral angle*).

In permanent magnets the magnetocrystalline anisotropy energy is expressed by $K_1 \sin^2(\varphi - \theta)$ where φ is the angle between the magnetization and the saturation direction and θ is the angle between the c -axis of the tetragonal crystal and the saturation direction. In the ESW model the orientation dependence of the switching field expressed by (3) describes the reduction of the anisotropy field by a factor that depends on the angle ψ between the easy axis and the total field (1). Hence, additionally to geometrical features we assign the orientation of the easy axis for each grain:

- The orientation angle θ of the grains (*misorientation*).

Figure 4 shows a sketch of some of the descriptors.

The contribution of each of the above attributes in predicting weak and strong grains is studied statistically by the machine learning approach. These features represent an already preselected and rather uncorrelated subset of a larger possible set of attributes. For instance, attributes like the surface area, the volume and the diameter of the grains exhibit correlation coefficients above 0.95. Pearson's correlation coefficient [22] is a measure of the tendency of the features to increase or decrease together. Therefore, we only took one 'representative' when the correlation coefficient between a pair of features was greater than 0.76. For example, we only take the grain diameter and drop surface area and volume. The correlation matrix for the selected descriptors is shown in figure 5 which in addition includes the local switching field attribute and misorientation corresponding to polar angles with standard deviation of 5° and zero mean. In section 3.2 we will study the significance of the features in explaining local switching as indicated by feature importance measures based on different predictor models.

2.2. Machine learning methods

Machine learning is a statistical approach that aims at automating analytical model fitting for data analysis, for instance finding clusters/structures in data or generating data-based predictive decision tools. For very comprehensive introductions to machine learning the reader is referred to [23, 24]. We use so-called *supervised learning*, where the training data also includes the true solutions. In our case, the training data consist of grains together with their microstructure features and labels (switching fields). We aim at classifying weak grains, that is, predicting those feature classes which exhibit a switching field below a certain threshold (class of weak grains)

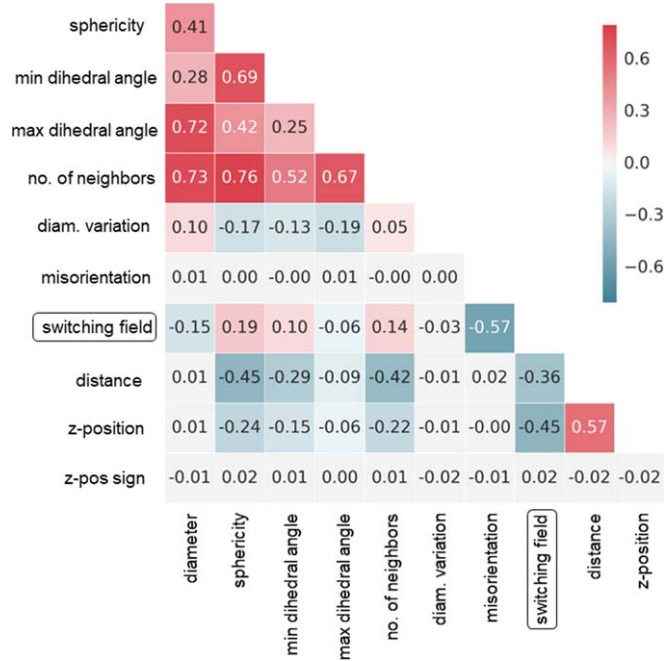


Figure 5. Correlation matrix of the selected descriptors including the local switching field. Misorientation corresponding to polar angles with standard deviation of 5° and zero mean. All correlation coefficients are smaller or equal 0.76.

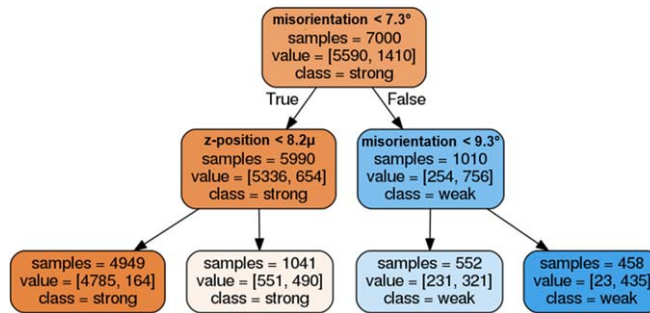


Figure 6. Example of individual decision tree. First, the training set consisting of 7000 grains is split into 5990 strong grains and 1010 weak grains depending on the crystallographic orientation. The two nodes in the second level are split depending on the z-position and on the misorientation, respectively.

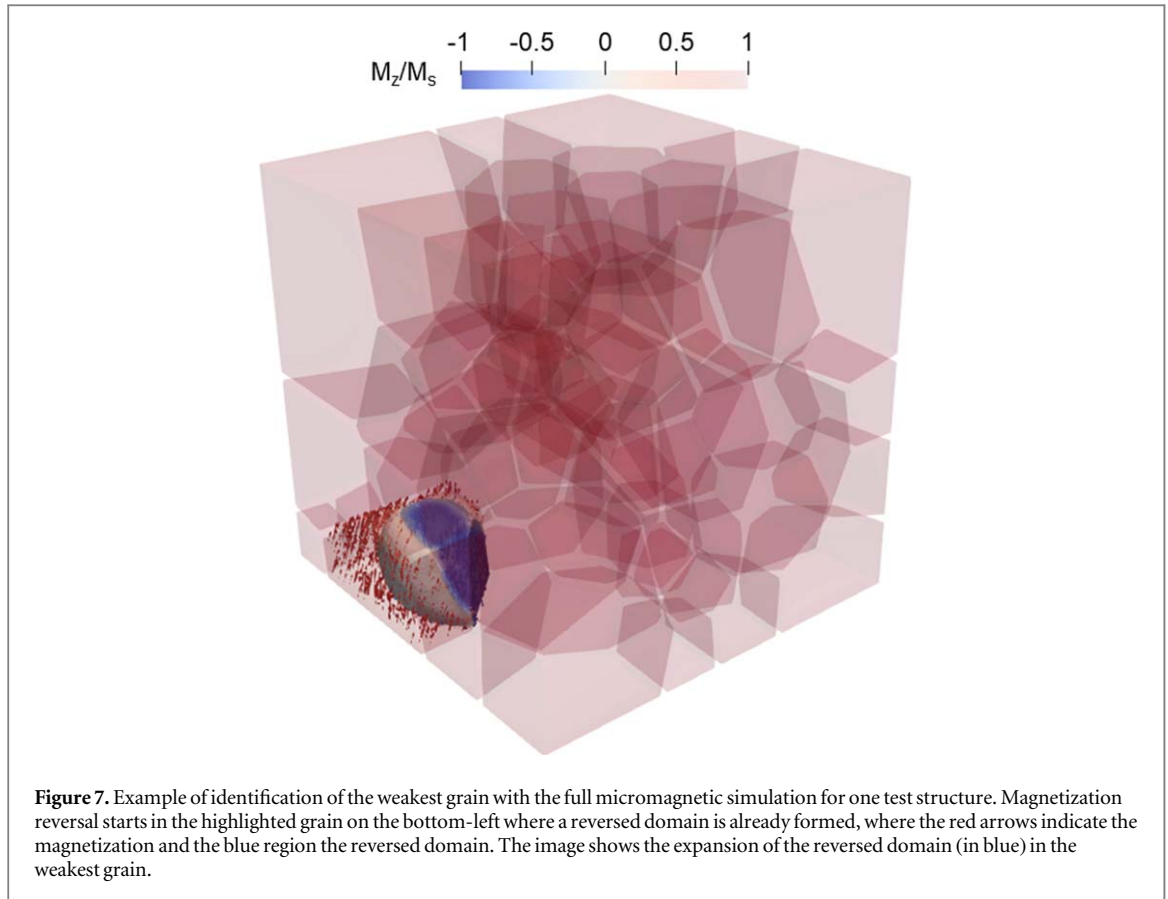
and above it (class of strong grains), respectively. This refers to as *binary classification*. The learning algorithm produces a function that maps a sample's feature vector to the class of weak grains or to the class of strong grains. Beside *classification* a second common supervised learning task is *regression*, which would try to predict values instead of classes, that is, a function that maps a given feature vector to predictions for switching field values. We will compare *logistic regression* with ℓ^1 - and ℓ^2 -regularization and RF [25]. We also get insight into the *feature importance* causing weak grains. However, similar as in [26] we observe best results for the RF algorithm. RF algorithms are *bagging methods* (short for *bootstrap aggregation*) built up by combining predictions of individual *decision trees* trained over randomly generated sub-training samples with replacement (bootstrap sample). At any instance an average of the individual estimators is taken to generate the ensemble model. An example of one decision tree with depth two is given in figure 6.

An important and nontrivial task is the performance measure of a classifier. The *accuracy* of a model is the amount of correctly predicted instances relative to all instances. Depending on the tightness of the threshold of the switching field value (=decision threshold) used for classifying weak grains any accuracy could be achieved. For instance, if the smallest 10% of all grains are labeled as weak, a classifier which invariably predicts strong grains will have a 90% accuracy. A way out is to determine the *confusion matrix* of a binary classifier, that is to count the number of times instances of one class (strong or weak grain) are classified correctly (true weak or strong) or incorrectly (false weak or strong), respectively. The ratio of the number of true weak grains and all grains classified as weak is called *precision*, the accuracy of positive predictions (weak grains). A high precision

means that few strong grains are erroneously classified as weak, where possibly many weak instances can still be erroneously classified as strong. Instead, the so-called *recall* (also *true positive rate*) is the ratio of the number of true weak grains and the sum of true weak and false strong instances. A high recall means that few weak grains are erroneously classified as strong, where possibly many strong instances can still be erroneously classified as weak. Obviously there is a trade-off between precision and recall. The harmonic mean of precision and recall is the *f1-score* of the binary classifier. Machine learning models depend on various hyperparameters. Hyperparameters refer to underlying model settings such as tree depth in a random forest or polynomial degree in a regression model. Such parameters, which control the model capacity, can not be tuned by learning on the training set since this would lead to overfitting to the training set, that is, the maximum possible model capacity is chosen (like highest possible polynomial degree in regression). Such an over-fitted model would correctly predict every sample of the training set but lacks predicting the general patterns of the data set leading to poor generalization error on new data. The traditional way to overcome this is to split the training set into two disjoint subsets, the validation set and the actual training data set [23]. The validation set is therefore separated from the training (and test) set and thus never observed by the training algorithm. The validation set is used to estimate the generalization error (model fit) in an unbiased manner while calibrating the hyperparameters. In this sense the hyperparameters are ‘trained’ on the validation set. However, if the dataset is too small alternative procedures need to be applied to avoid reducing the available training data by completely setting aside a validation set. One way is to use resampling techniques for hyperparameter tuning, such as *k*-fold cross validation. This procedure is based on repeated training and testing of models on different randomly chosen disjoint and roughly equally sized subsets of the original (training) dataset. It can be seen as an iterated version of the traditional training/validation set splitting but leveraging the complete training set. The test error (e.g. *f1-score*) of each model is estimated by taking the average error across *k* different trials for which always one held-out subset serves as the test set and the rest for training. This averaging strategy has a beneficial effect on the error of each model as it is a close approximation of really unseen data [5]. The choice of *k* is usually 5 or 10, however this is not a formal rule. As *k* gets larger, the difference in size between the (original) training set and the resampling subsets gets smaller and thus also the difference between the estimated performance error and the true error gets smaller. This latter difference is referred to as the bias of the technique [1]. By balancing a model’s complexity one achieves an optimal trade-off between bias and variance of a model [24, 27]. In fact, the authors of [27] make a recommendation for small sample sizes of using *k*-fold cross validation because of the good variance and bias properties for only minor additional computational costs due to the rather small sample sizes. A slight variant of this method is to select the *k* partitions in a way that makes the folds balanced with respect to the distribution of the outcomes [28], which is then referred to as stratified *k*-fold cross validation. To avoid over-fitting, which refers to small bias but high variance we choose stratified five-fold cross validation, i.e. a smaller *k* to achieve decent small bias. In addition, Breiman [25] showed that bagging reduces the variance of the overall ensemble relative to any individual learner in the ensemble.

We then maximized the *f1-score* by searching optimal values for the tree depth, the number of trees, and the number of features to consider when splitting a node, see section 3.2. We calculated the confusion matrix with respect to the test set where we used 50% probability for the class membership threshold in the forthcoming analysis. Another performance measure is the *receiver operating characteristic* (ROC) curve which plots the true positive rate versus the false positive rate. The area under the ROC curve (AUC) is a common evaluation metric where values close to 1 indicate a good classifier.

In a decision tree important features are likely to appear closer to the root of the tree, whereas unimportant features are found near the leaves or not at all. Estimates of a feature’s importance in an RF classifier can be calculated by the average depth at which it appears across all trees. Another approach to determine feature importance is a model-agnostic version called *model reliance*, where feature importance is indicated by the amount of increase of model error, for example measured by the AUC or any other performance measure, by fitting a model after permuting the features [29, 30]. We will use the model-agnostic approach [24] as implemented in *Skater* [31]. *Skater* measures the mean absolute value of the change in predictions given a perturbation of a certain feature. The idea behind is the following: the algorithm works through all features in the test set and replaces the values corresponding to a single feature by randomly chosen feature values from the training set and calculates new predictions. The more important a specific feature, the more the predictions will change as a function of perturbing the feature. This approach works for any predictor model. Further, we will compare with features’ importance measure based on the magnitude of coefficients when using logistic regression models.



3. Results

3.1. Micromagnetic validation

In the case of structures consisting of very few grains we can validate our approach with full micromagnetic computations including the conventional determination of the magnetostatic field via Maxwell's equations. The question is whether a trained RF model can predict where magnetization reversal will start. We create 100 granular structures consisting of only 64 grains each with a mean grain size of 50 nm. We split the data structures into 80 training structures and 20 test structures. For each structure the grain orientations with respect to the z -direction are set randomly according to a zero-mean normal distribution with a standard deviation of 5° for the polar angle and a uniform distribution for the azimuthal angle. We first label the grains as 'weak' or 'strong' according to the switching fields computed by the ESW model. Then we train an RF model on the training set using the Python library *Scikit-Learn* [32]. In order to validate the model, we perform full micromagnetic simulations using the finite element method [33]. Following the demagnetization curve we compute the grain and corresponding switching fields where magnetization reversal starts. This identifies the true weakest grains in the test set (see figure 7). In 16 out of the 20 test cases the RF prediction of the weakest grain coincides with the results from full micromagnetic simulations.

We also estimated the model error of the ESW model. In 18 out of 20 cases the weakest grains according to the ESW and full micromagnetic switching fields coincide. This discrepancy reflects the model error mainly corresponding to the simplified stray field calculation in the ESW model which does not take into account reversible magnetization rotations before switching.

Considering both, the model error of the ESW model and the performance measure of the RF model (see table 3) gives an overall accuracy of 80% in accordance with the above validation result.

3.2. Microstructure machine learning analysis

We use ten multigrain models with 1000 grains each, where we randomly put aside three models for the validation (this is the test data set). For the grains in each model we determine the feature values and calculate the true labels by the ESW method in order to supervise the subsequent learning process. Note that the generation of the dataset is very expensive, in fact, both the creation of the microstructure as well as the calculation of the labels limit the usable size of data. The anisotropy directions are set randomly according to a uniform distribution for the azimuthal angle and a zero-mean normal distribution with a standard deviation of $\sigma_\theta = 0^\circ$, $\sigma_\theta = 5^\circ$, or

Table 1. Confusion matrices (with threshold 50% for classification) for the ℓ^1 -logistic regression model for 0° , 5° , and 15° standard deviation of the misorientation angle. The model performance metrics include accuracy, precision, recall, f1-score and AUC.

$\sigma_\theta = 0^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2334	74	0.9430	0.9423	0.9630	0.9426	0.99
	1	97	495					
$\sigma_\theta = 5^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2288	121	0.8903	0.8858	0.8903	0.8869	0.94
	1	208	383					
$\sigma_\theta = 15^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2314	95	0.8917	0.8864	0.8917	0.8862	0.92
	1	230	361					

Table 2. Confusion matrices (with threshold 50% for classification) for the ℓ^2 -logistic regression model for 0° , 5° , and 15° standard deviation of the misorientation angle. The model performance metrics include accuracy, precision, recall, f1-score and AUC.

$\sigma_\theta = 0^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2333	75	0.9427	0.9420	0.9427	0.9423	0.99
	1	97	495					
$\sigma_\theta = 5^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2288	121	0.8903	0.8858	0.8903	0.8869	0.94
	1	208	383					
$\sigma_\theta = 15^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2313	96	0.8913	0.8860	0.8913	0.8859	0.92
	1	230	361					

$\sigma_\theta = 15^\circ$ for the polar angle. This determines three different scenarios. For each scenario we label grains with a switching field smaller than the 20th percentile of the switching field distribution as ‘weak’ and use the records of the training set to train logistic regression classifier with ℓ^1 - and ℓ^2 -regularization and a random forest model applying the Python library `Scikit-Learn` [32]. The optimal value for the regularization strength, $\alpha = 1/C$, for logistic regression was determined by a grid search-based hyperparameter tuning using five-fold cross-validation as described above. The values turned out to be identical for ℓ^1 - and ℓ^2 -regularization: In the $\sigma_\theta = 0^\circ$ case $C = 1$, in the $\sigma_\theta = 5^\circ$ case $C = 7.74$ and in the $\sigma_\theta = 15^\circ$ case $C = 1$. The optimal hyperparameter values in the random forest model for the three scenarios were found to be as follows:

The $\sigma_\theta = 0^\circ$ case: 20 for the maximum tree depth, 500 for the number of trees in the forest and the square root of all features for the optimal number of features considered for splitting a node. The $\sigma_\theta = 5^\circ$ case: 10 for the maximum tree depth, 500 for the number of trees in the forest and all features for the optimal number of features considered for splitting a node. The $\sigma_\theta = 15^\circ$ case: 10 for the maximum tree depth, 200 for the number of trees in the forest and all features for the optimal number of features considered for splitting a node.

Table 1 shows the confusion matrices together with several model performance metrics for the ℓ^1 -logistic regression classifier model also with 50% threshold for class membership probability. Table 2 depicts the case for the ℓ^2 -regression model. Table 3 shows the confusion matrices with 50% threshold for class membership probability and several model performance metrics for the RF model for $\sigma_\theta = 0^\circ$, $\sigma_\theta = 5^\circ$, and $\sigma_\theta = 15^\circ$. Confusion matrices and performance measures give roughly the same results for the two logistic models, while they perform slightly worse than the random forest predictor especially in the smaller σ_θ -regime.

Table 4 compares all models, in detail it gives for all σ_θ -cases the Pearson correlation of the features to the local switching field as well as all determined feature importance for the logistic regression and random forest models. We remark that only the random forest model is able to weight the distance as an important feature in the cases with misorientation, which explains, together with the decisive role of the z-position, the weak edges of

Table 3. Confusion matrices (with threshold 50% for classification) for the random forest model for 0° , 5° , and 15° standard deviation of the misorientation angle. The model performance metrics include accuracy, precision, recall, f1-score and AUC.

$\sigma_\theta = 0^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2374	34	0.9727	0.9725	0.9727	0.9725	0.99
	1	48	544					
$\sigma_\theta = 5^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2310	99	0.9243	0.9231	0.9243	0.9236	0.97
	1	128	463					
$\sigma_\theta = 15^\circ$		Predicted:		Performance	Metrics			
		0	1	Accuracy	Precision	Recall	f1-score	AUC
Actual:	0	2308	101	0.9027	0.8989	0.9027	0.8995	0.93
	1	191	400					

Table 4. The first column describes the features, the second column gives the Pearson correlation of the features to the local switching field and all further columns give the feature importance for the different models: LogReg (ℓ^1) and LogReg (ℓ^2) denote logistic regression with ℓ^1 - and ℓ^2 -regularization, respectively, and RF indicates the random forest model. In the cases of the logistic regression models we also give feature importance based on the determined coefficients' magnitude. In all other cases we also report the feature importance computed with *Skater*. All this is given for the three scenarios $\sigma_\theta = 0^\circ$, 5° and 15° .

Feature	Pearson correlation	LogReg (ℓ^1) coefficients	LogReg (ℓ^1) Skater	LogReg (ℓ^2) coefficients	LogReg (ℓ^2) Skater	RF Skater
$\sigma_\theta = 0^\circ$						
misorientation	0.00	0.01	0.00	0.01	0.00	0.02
z-position	-0.68	0.47	0.46	0.46	0.47	0.40
distance	-0.63	0.36	0.36	0.36	0.35	0.30
diameter	-0.18	0.12	0.13	0.13	0.14	0.09
no of neighbors	0.27	0.02	0.02	0.01	0.01	0.05
sphericity	0.35	0.00	0.00	0.00	0.00	0.04
diamvariation	-0.04	0.00	0.00	0.00	0.00	0.03
maxdihedral angle	-0.03	0.00	0.00	0.00	0.00	0.03
mindihedral angle	0.21	0.01	0.01	0.01	0.01	0.03
z-pos sign	0.02	0.00	0.00	0.00	0.00	0.00
$\sigma_\theta = 5^\circ$						
misorientation	-0.57	0.35	0.38	0.35	0.38	0.41
z-position	-0.45	0.23	0.22	0.23	0.22	0.27
distance	-0.36	0.02	0.02	0.02	0.02	0.12
diameter	-0.15	0.15	0.14	0.15	0.14	0.06
no of neighbors	0.15	0.15	0.14	0.15	0.14	0.03
sphericity	0.19	0.02	0.01	0.02	0.02	0.05
diamvariation	-0.03	0.02	0.02	0.02	0.02	0.02
maxdihedral angle	-0.06	0.04	0.04	0.04	0.04	0.02
mindihedral angle	0.10	0.03	0.02	0.03	0.02	0.02
z-pos sign	0.02	0.00	0.00	0.00	0.00	0.00
$\sigma_\theta = 15^\circ$						
misorientation	-0.77	0.40	0.45	0.40	0.45	0.45
z-position	-0.22	0.17	0.16	0.17	0.16	0.16
distance	-0.08	0.05	0.05	0.05	0.05	0.09
diameter	-0.14	0.16	0.15	0.16	0.15	0.07
no of neighbors	-0.06	0.13	0.11	0.13	0.11	0.06
sphericity	-0.02	0.02	0.02	0.02	0.02	0.05
diamvariation	0.00	0.00	0.00	0.00	0.00	0.03
maxdihedral angle	-0.10	0.01	0.00	0.01	0.01	0.04
mindihedral angle	-0.03	0.04	0.04	0.04	0.04	0.04
z-pos sign	-0.04	0.02	0.02	0.02	0.02	0.00

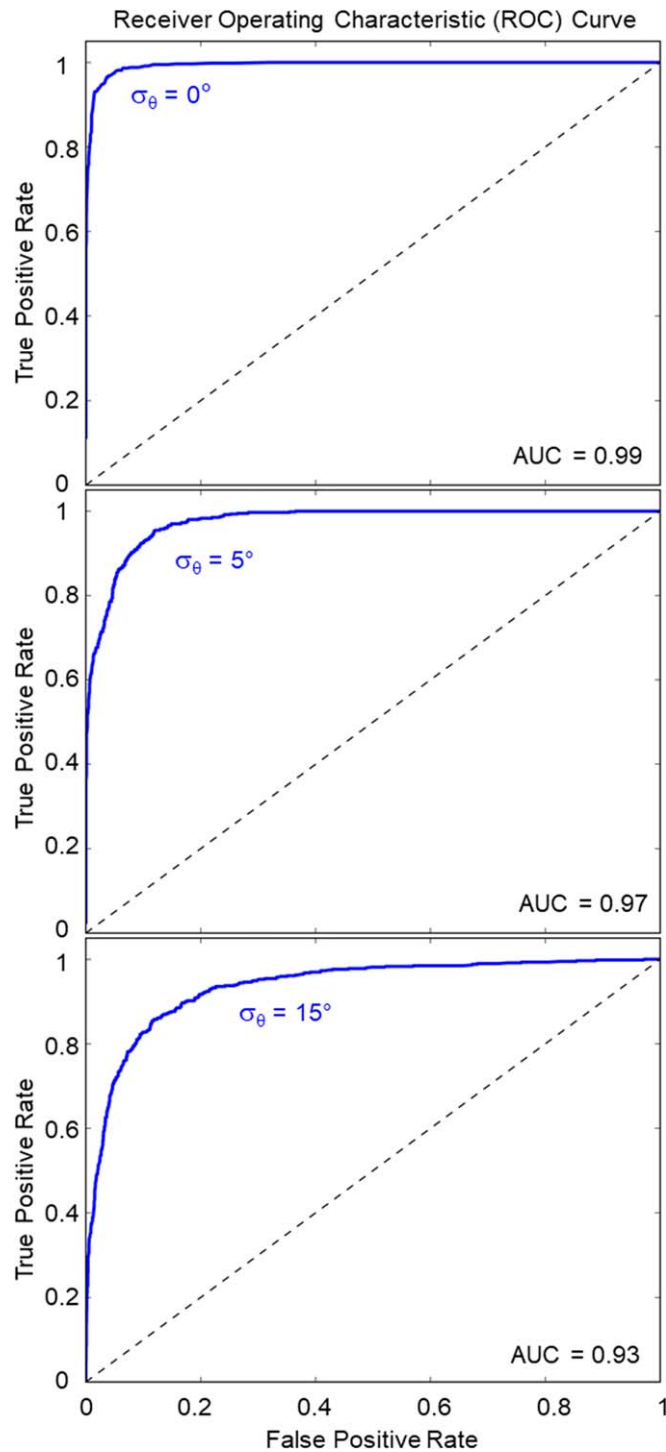
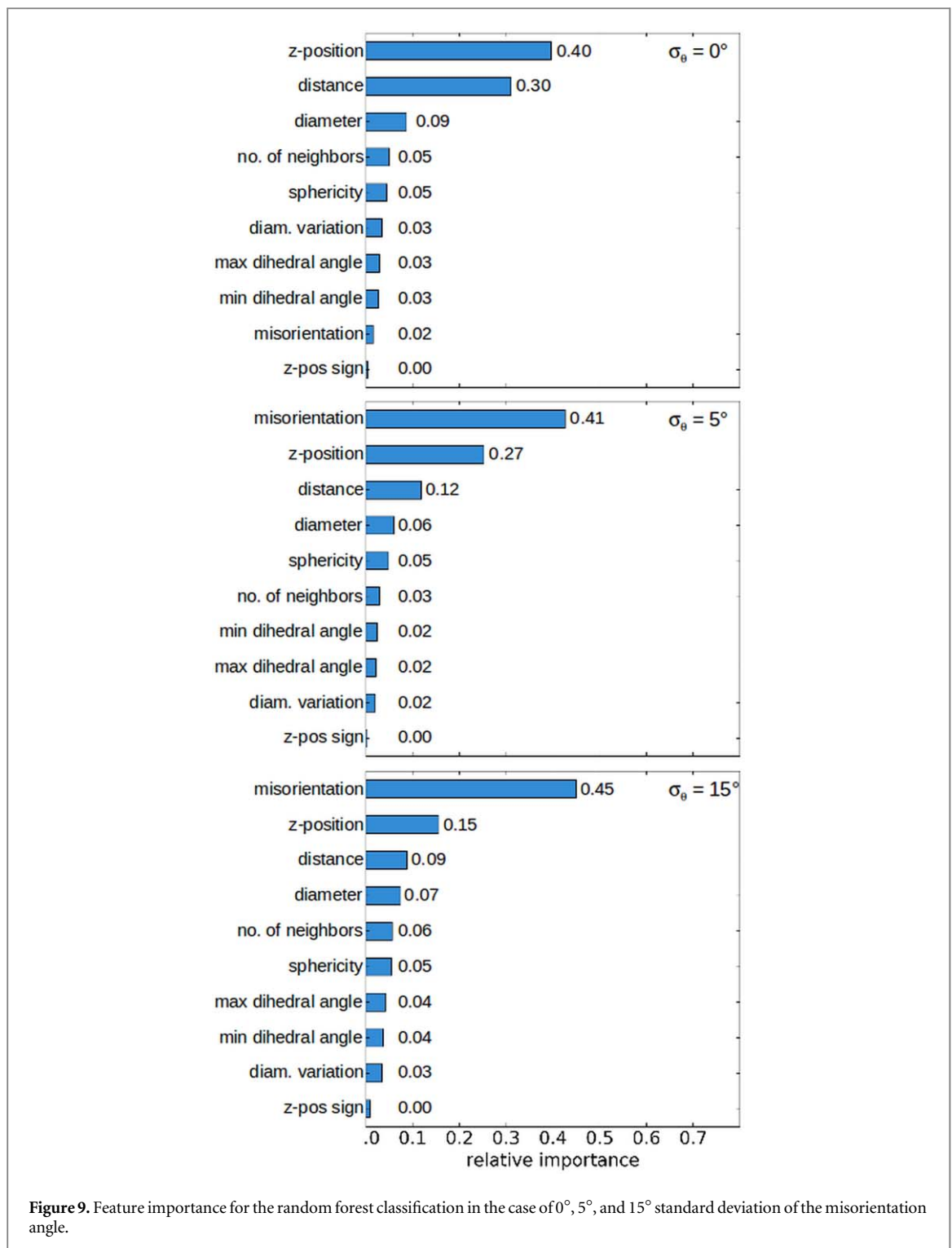


Figure 8. Receiver operating characteristic (ROC) curve for the random forest classification in the case of 0° , 5° , and 15° standard deviation of the misorientation angle.

the magnet, compare with sections 3.2 and 4, especially figures 13–15. We use the RF model for further analysis, and give some more validation next.

Figure 8 shows the ROC curves for the three different scenarios and random forest model. The model performance metrics as well as the AUC indicate very high performance of the trained random forest models, whereas a slight decline can be observed with increasing orientation angle. Figure 9 shows the feature importance for the three scenarios for the random forest model which was computed using the model agnostic approach [24] as implemented in *Skater* [31]. For perfectly aligned grains (0° misorientation) there are essentially two most important features, the absolute value of the vertical position of the grain in the magnet (z -position) and the distance of the grain from the center of the magnet. Clearly the sign of the z -position plays no important role, which indicates the symmetry of the problem. When misorientation is introduced, this becomes



the most important feature. One can clearly observe in figure 9 that the misorientation becomes more relevant with higher average misorientation angle. Whereas the dependence of the local switching field on the orientation is expected [19], the importance of the positions of the grain within the magnet is less obvious.

In a second step, we apply RF regression to predict the value of the local switching fields of the grains. Then we can get additional insight into feature dependence by *one-way partial dependence plots* for the random forest predictor. The partial dependence function represents the effect of a specific feature (for example the z -position) on the switching field after averaging out the influence of all other features [34]. Figures 10–12 show comparisons for different orientation scenarios by one-way dependency based on z -position, distance to center and misorientation angle, respectively.

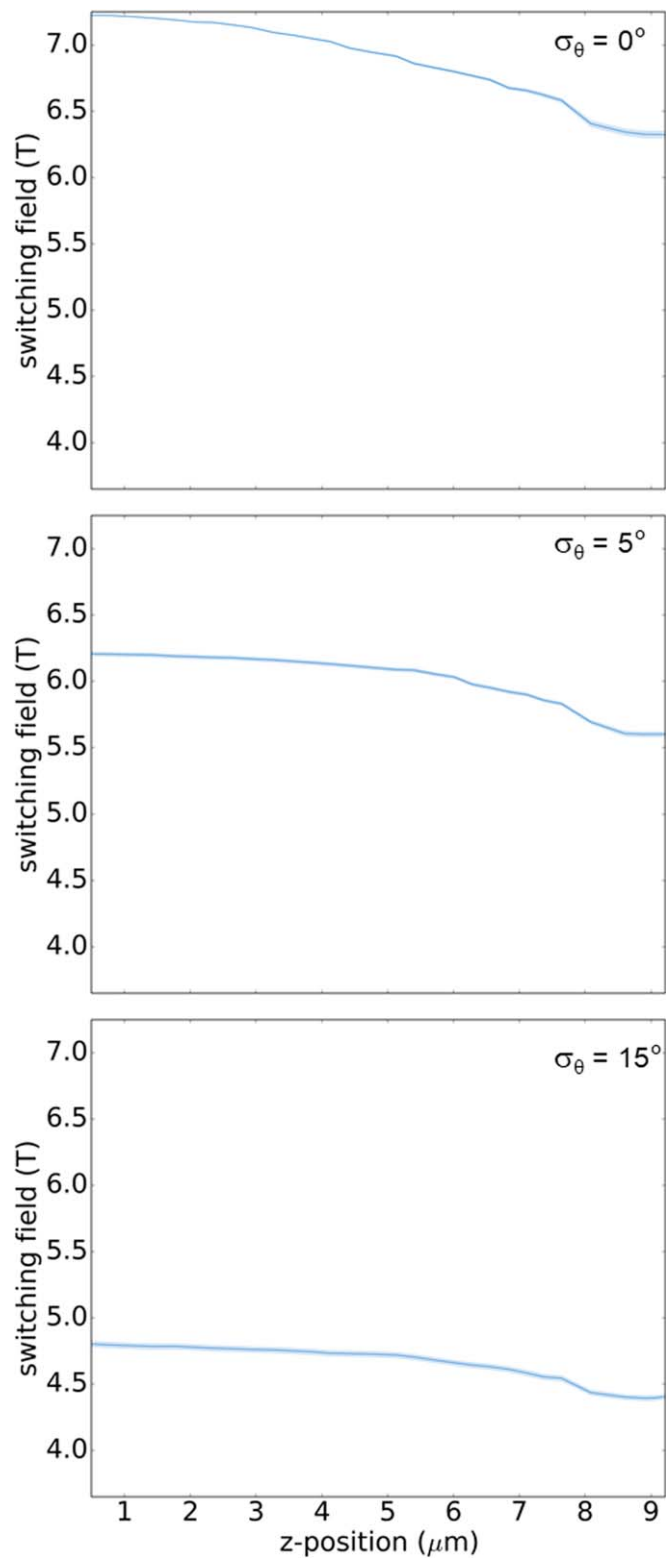


Figure 10. One-way partial dependence based on the vertical position of the grain within the magnet for 0° , 5° , and 15° standard deviation of the misorientation angle. A z-position close to $10\ \mu\text{m}$ indicates a grain near the bottom or top surface of the magnet, respectively.

4. Discussion

We applied machine learning techniques in order to correlate the microstructure characteristics with the local magnetization reversal field of large-grained $\text{Nd}_2\text{Fe}_{14}\text{B}$ permanent magnets. In order to focus on general features of polycrystalline materials we assumed an ideal structure: (i) the grains are separated by a nonmagnetic grain boundary phase and (ii) there are no defects with reduced magnetocrystalline anisotropy. Though this setting is

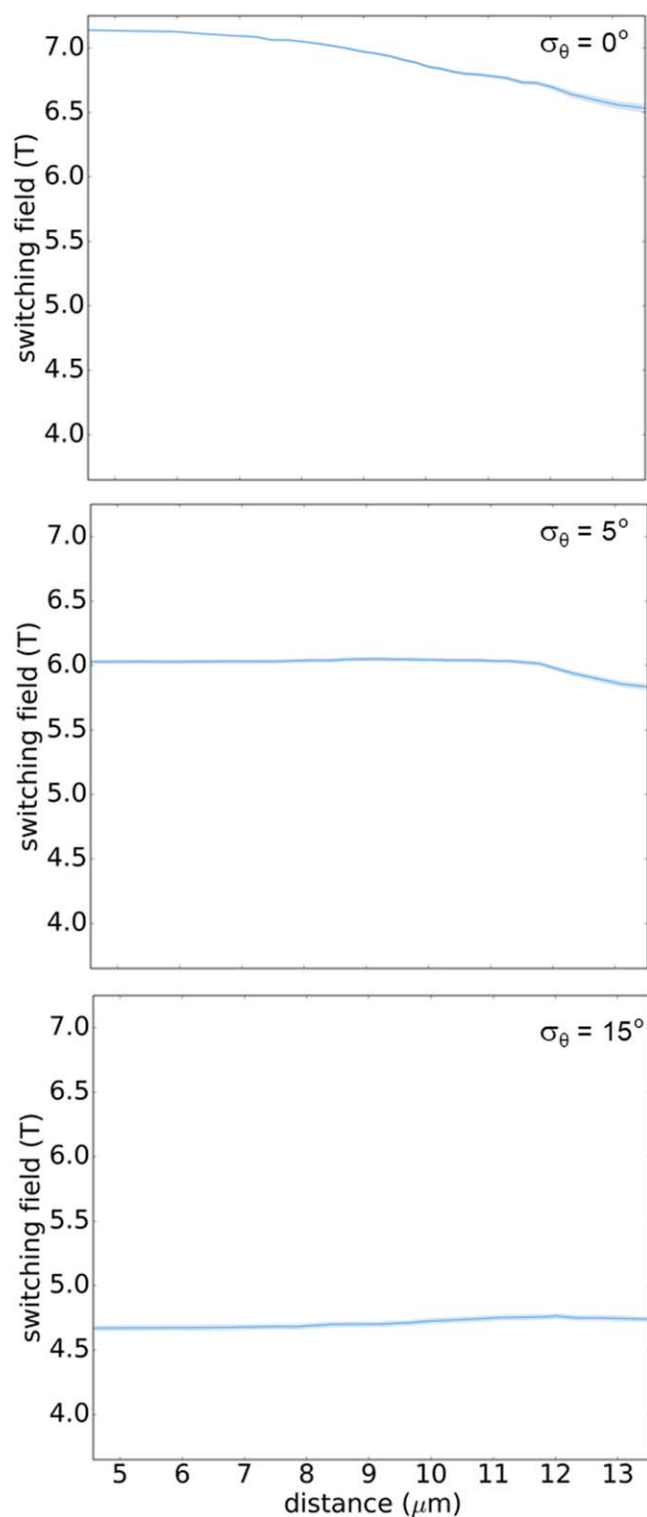


Figure 11. One-way partial dependence based on the distance of the grain to the center of the magnet for 0° , 5° , and 15° standard deviation of the misorientation angle.

unrealistic, it can provide clear insight what other features in addition to soft inclusions or ferromagnetic grain boundaries [35, 36] influence coercivity. The data used for machine learning was generated by a reduced order model that makes it possible to treat magnets which are much larger with respect to both grain size and number of grains than models suitable for conventional micromagnetic simulations. For small model size the prediction of the machine learning model can be compared with the results of full micromagnetic simulations. This comparison shows that a random forest classifier can predict the weakest grain in a magnet in 16 out of 20 test cases correctly.

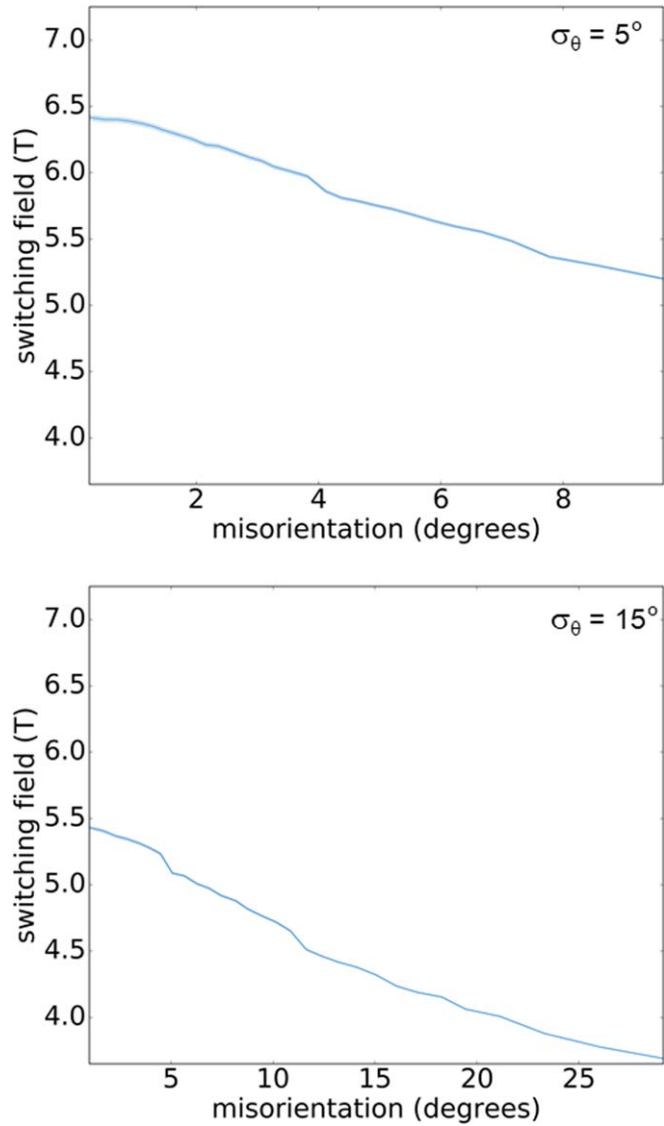


Figure 12. One-way partial dependence based on the misorientation of the anisotropy axes for 5 (top) and 15 (bottom) degrees standard deviation of the misorientation angle.

In order to find out what microstructure features are most significant, we computed the feature importance of a random forest classifier trained with the switching field distribution of 7 polycrystalline samples consisting of 1000 grains each. The feature importance was found to depend on the degree of alignment. For a scenario with a standard deviation of the misorientation angle of 15° the most important feature is the crystallographic orientation. As expected [19] the switching field decreases with increasing misorientation angle. The second and third most important features are the vertical position of the grain, and the distance of the magnet from the magnet's center. For perfect alignment (zero degree misorientation) these two are the most important features followed by the grain diameter. Local interpretable model agnostic explanation [37] shows that the switching field of a grain is smaller the closer the grain is located to the top or bottom surface of the magnet. This dependence is more pronounced for the perfectly aligned grains where the switching field of a grain near the top or bottom is more than 11% smaller than that of a grain near the center. For the scenario with 15° mean misalignment the decrease of the switching field based on the vertical position is 7%. Similarly, the switching field of a grain decreases with increasing distance from the center of the magnet. A two-way partial dependence plot of the switching field as function of *z-position* and *distance from center* shows that the lowest switching fields occur near the top and bottom edges of the magnet (see figure 13). These are the locations where the local demagnetizing field of the magnet reach the highest values [38]. Furthermore, near these edges the demagnetizing field is tilted with respect to the magnetization direction which reduces the local Stoner–Wohlfarth switching field according to (3). While the dependence of the local switching field on the grain orientation is known from the basic micromagnetic theory [19], the influence of the position of the grain on the local coercive field strength is less obvious. One may argue that strong local demagnetizing fields may also occur

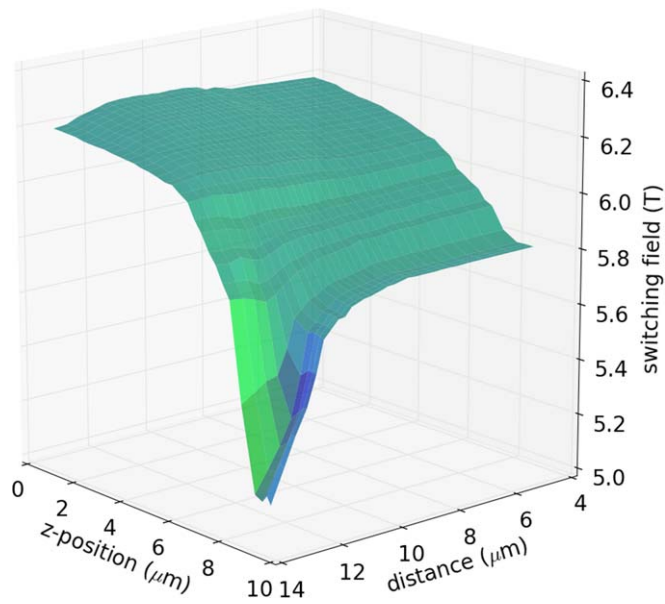


Figure 13. Two-way partial dependence based on *z-position* and *distance from center* for 5° standard deviation of the misorientation angle. A *z-position* close to 10 indicates a grain near the bottom or top surface of the magnet.

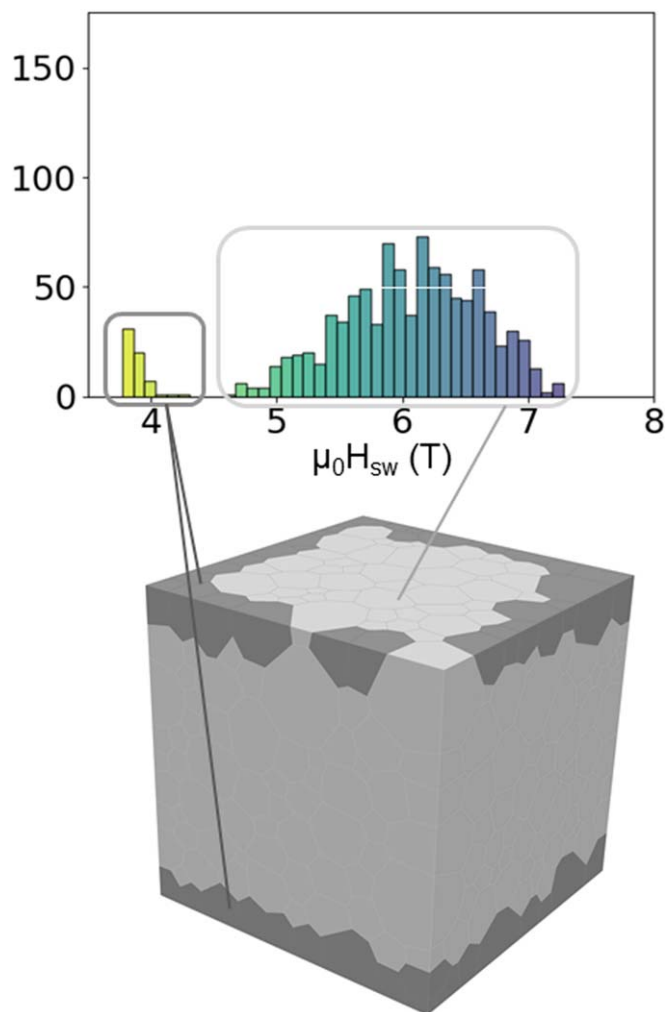
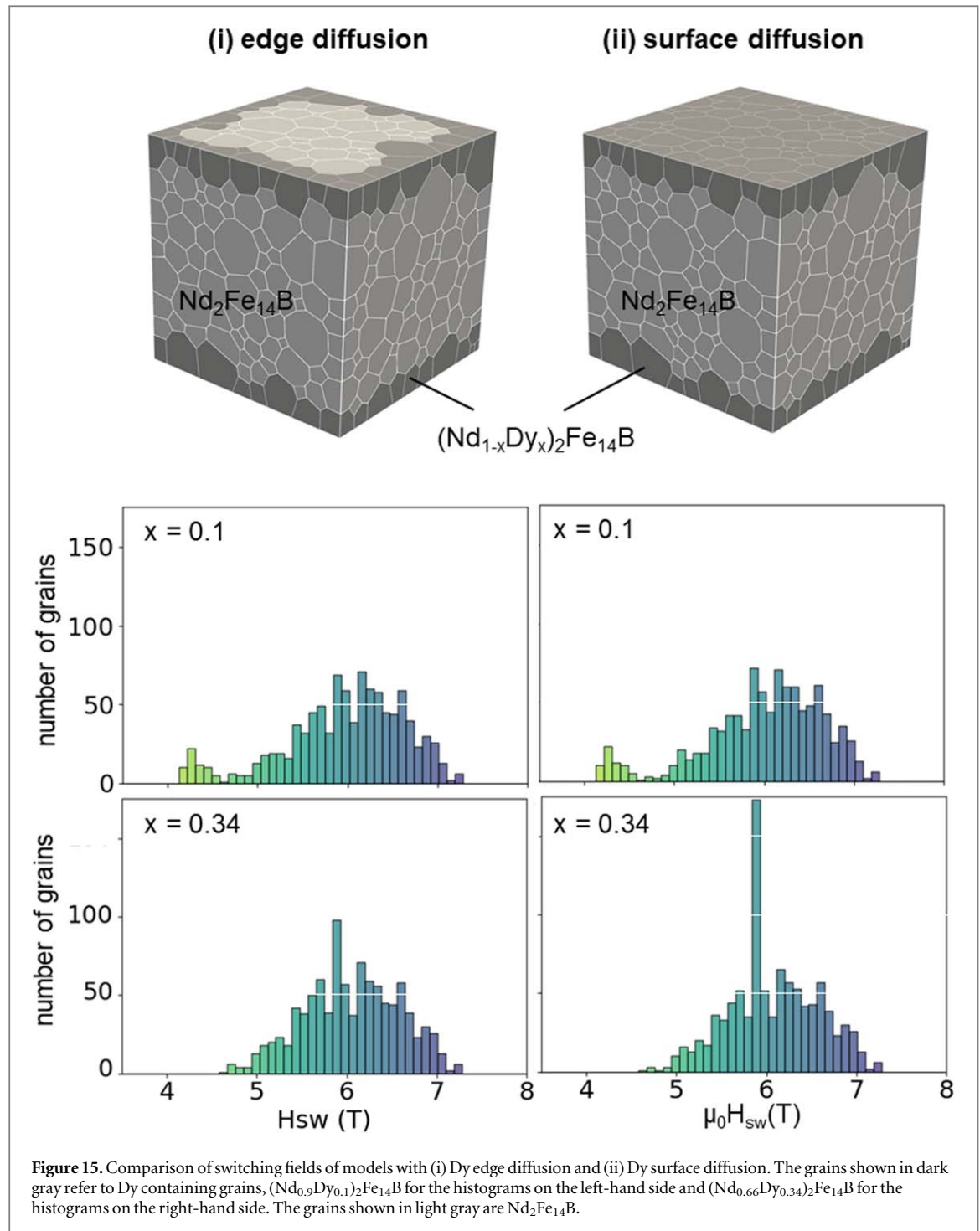


Figure 14. Grain structure showing the weakest grains (dark gray) together with the switching field distribution of a homogeneous $\text{Nd}_2\text{Fe}_{14}\text{B}$ model.



near the nonmagnetic grain boundary phase inside the magnet that may initiate magnetization reversal. The machine learning model shows that this is not the case and the lowest reversal fields always occur near the edges of the magnet. These results indicate that local variation of the magnetic properties, which enhances the switching field near the surfaces or edges of the magnet, is sufficient to improve the magnet's performance. Possible routes to achieve higher coercive grains locally are grain boundary diffusion [39, 40] and additive manufacturing [41, 42]. Thompson *et al* [43] used electron probe microanalysis to analyze the Dy-concentration in diffusion treated sintered magnets and showed that the highest heavy rare-earth concentration occurred near the corners of the magnet. A similar local variation of the magnetic properties may be achieved by additive manufacturing.

As shown above, machine learning revealed a strong effect of the position of the grain within the magnet on the switching field. Indeed, figure 13 shows that the lowest switching fields occur for grains located at the edges (near the top and bottom of the magnet and at a large distance of the center). We now take a grain structure from the test set with 5° misorientation and analyze its switching field distribution. Figure 14 shows the switching field

distribution of the grains and the location of the weakest grains. The distribution shows a small peak for $\mu_0 H_{sw} < 4$ T whereas the mean switching field is at 5.9 T and the maximum switching field is at 7.2 T. We can identify the grains with low switching field, which are shown in figure 14. As predicted by the machine learning algorithm these are the grains at the top and bottom edges of the magnet.

In order to show how $\text{Nd}_2\text{Fe}_{14}\text{B}$ magnets can be improved by Dy-diffusion, we compare the switching field distribution for different scenarios: (i) A sample where the grains near top and bottom edges have higher anisotropy field, and (ii) a sample where the grains near top and bottom surfaces have higher anisotropy field. The grains with higher anisotropy field have a composition $(\text{Nd}_{1-x}\text{Dy}_x)_2\text{Fe}_{14}\text{B}$. Following Oikawa *et al* [44] we decrease the spontaneous magnetization M_s linearly with increasing Dy-content. For the grains with higher anisotropy field we used $(\text{Nd}_{0.9}\text{Dy}_{0.1})_2\text{Fe}_{14}\text{B}$ or $(\text{Nd}_{0.66}\text{Dy}_{0.34})_2\text{Fe}_{14}\text{B}$ with a magnetization $\mu_0 M_s = 1.52$ T and 1.3 T, respectively. When the grains at the top and bottom edges are hardened by Dy diffusion the peak at low fields disappears gradually. The minimum switching field increases from $\mu_0 H_{sw,min} = 3.8$ T without Dy-diffusion (see figure 14) to $\mu_0 H_{sw,min} = 4.14$ T and 4.62 T for a Dy-content of $x = 0.1$ and $x = 0.34$ in the grains near the top and bottom edges, respectively (see figure 15(i)).

In Dy-free magnets the grains near the top and bottom surface have reduced switching field which in turn reduce the coercive field of the magnet. Hardening of the grains near the top and bottom edges by Dy-diffusion avoids these low coercive grains. A similar result was achieved by hardening the grains near the top and bottom surface, see figure 15(ii). This effect may be used in magnet production and may further reduce the heavy rare-earth content while keeping a high coercive field.

5. Conclusion

In summary, we showed that machine learning techniques can be applied to characterize the role of microstructure features in permanent magnets. The results derived from the machine learning model show that the position of the grain within the magnet is important. The grains near the top and bottom edges of the magnets have lower switching fields than grains located elsewhere. Other properties like number of neighbors, dihedral angle, or sphericity play a minor role. For future applications of machine learning in permanent magnet design we can envision several scenarios ranging from the structure optimization, guided rapid prototyping by additive manufacturing to the use of machine learning models as building blocks for the multiscale simulation of hysteresis.

In the example given in this work we identified the location of the weakest grains in ideally structured $\text{Nd}_2\text{Fe}_{14}\text{B}$ magnets without any defects. The grains with the lowest switching fields are located at the top and bottom edges of the magnet. This suggests that localizing grain boundary diffusion of heavy rare-earth elements to these specific regions only may be sufficient to increase coercivity. Thus, the magnet's performance and temperature stability may be improved with a minimum amount of heavy rare-earth elements.

Acknowledgments

Financial support by the Austrian Science Fund (FWF) via the project 'ROAM' under grant No. P31140, the SFB 'ViCoM' under grant No. F41 and the SFB 'Complexity in PDEs' under grant No. F65 is acknowledged, as well as support by Toyota Motor Corporation, and the future pioneering program Development of Magnetic Material Technology for High-efficiency Motors commissioned by the New Energy and Industrial Technology Development Organization (NEDO). The computations were partly achieved by using the Vienna Scientific Cluster (VSC) via the funded project No. 71140.

ORCID iDs

Lukas Exl  <https://orcid.org/0000-0002-5343-693>

Johann Fischbacher  <https://orcid.org/0000-0002-0815-5379>

Alexander Kovacs  <https://orcid.org/0000-0002-8413-1950>

Harald Oezelt  <https://orcid.org/0000-0002-3754-3565>

Markus Gusenbauer  <https://orcid.org/0000-0002-3540-3964>

Gino Hrkac  <https://orcid.org/0000-0001-7284-135X>

Thomas Schrefl  <https://orcid.org/0000-0002-0871-0520>

References

- [1] Constantinides S 2016 *Magnetics Magazine* Spring 2016, p 6
- [2] Binnemans K, Jones P T, Müller T and Yurramendi L 2018 *J. Sustain. Metall.* **4** 126–46
- [3] Yang Y et al 2016 *J. Sustain. Metall.* **3** 122–49
- [4] Skokov K and Gutfleisch O 2018 *Scr. Mater.* **154** 289–94
- [5] Mangal A and Holm E A 2018 *Int. J. Plast.* **111** 122–34
- [6] Mangal A and Holm E A 2018 *Integr. Mater. Manuf. Innov.* **7** 87–95
- [7] Fidler J and Knoch K 1989 *J. Magn. Magn. Mater.* **80** 48–56
- [8] Sasaki T, Ohkubo T, Takada Y, Sato T, Kato A, Kaneko Y and Hono K 2016 *Scr. Mater.* **113** 218–21
- [9] Kronmüller H, Durst K D and Sagawa M 1988 *J. Magn. Magn. Mater.* **74** 291–302
- [10] Quey R, Dawson P and Barbe F 2011 *Comput. Methods Appl. Mech. Eng.* **200** 1729–45
- [11] Quey R and Renversade L 2018 *Comput. Methods Appl. Mech. Eng.* **330** 308–33
- [12] Coey J M D 2009 *Magnetism and Magnetic Materials* (Cambridge: Cambridge University Press)
- [13] Fischbacher J, Kovacs A, Exl L, Kühnel J, Mehofer E, Sepehri-Amin H, Ohkubo T, Hono K and Schrefl T 2018 *Scr. Mater.* **154** 253–8
- [14] Schrefl T and Fidler J 1992 *J. Magn. Magn. Mater.* **111** 105–14
- [15] Guptasarma D and Singh B 1999 *Geophysics* **64** 70–4
- [16] Rave W, Ramstöck K and Hubert A 1998 *J. Magn. Magn. Mater.* **183** 329–33
- [17] Donahue M and McMichael R 1997 *Physica B* **233** 272–8
- [18] Stoner E C and Wohlfarth E P 1948 *Phil. Trans. R. Soc. A* **240** 599–642
- [19] Kronmüller H, Durst K D and Martinek G 1987 *J. Magn. Magn. Mater.* **69** 149–57
- [20] Kronmüller H 2007 *Handbook of Magnetism and Advanced Magnetic Materials* (Vol 2) (New York: Wiley)
- [21] Bance S et al 2014 *J. Appl. Phys.* **116** 233903
- [22] Rodgers J L and Nicewander W A 1988 *Am. Stat.* **42** 59
- [23] Géron A 2017 *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (Sebastapol: O'Reilly Media Inc.)
- [24] Sarkar D, Bali R and Sharma T 2018 *Practical Machine Learning with Python* (New York: Apress)
- [25] Breiman L 2001 *Mach. Learn.* **45** 5–32
- [26] Mangal A and Holm E A 2018 (arXiv:1804.05924)
- [27] Kuhn M and Johnson K 2013 *Applied Predictive Modeling* vol 26 (Berlin: Springer)
- [28] Kohavi R 1996 Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid *KDD* vol 96 (Citeseer) pp 202–7
- [29] Molnar C 2018 *Interpretable Machine Learning* <https://christophm.github.io/interpretable-ml-book/> (Victoria: Leanpub)
- [30] Fisher A, Rudin C and Dominici F 2018 arXiv:1801.01489
- [31] Choudhary P, Kramer A and Datasciencecom Team 2018 Skater: Model Interpretation Library <https://github.com/datascienceinc/Skater/>
- [32] Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F and Mueller A 2015 *GetMobile: Mobile Comput. Commun.* **19** 29–33
- [33] Fischbacher J, Kovacs A, Gusenbauer M, Oezelt H, Exl L, Bance S and Schrefl T 2018 *J. Phys. D: Appl. Phys.* **51** 193002
- [34] Hastie T, Tibshirani R and Friedman J 2008 *The Elements of Statistical Learning: Data Mining, Inference and Prediction* 2nd edn (Berlin: Springer)
- [35] Murakami Y, Tanigaki T, Sasaki T, Takeno Y, Park H, Matsuda T, Ohkubo T, Hono K and Shindo D 2014 *Acta Mater.* **71** 370–9
- [36] Zickler G A, Fidler J, Bernardi J, Schrefl T and Asali A 2017 *Adv. Mater. Sci. Eng.* **2017** 6412042
- [37] Ribeiro M T, Singh S and Guestrin C 2016 Why should I trust you?: explaining the predictions of any classifier *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (New York: ACM) pp 1135–44
- [38] Grönfeld M and Kronmüller H 1989 *J. Magn. Magn. Mater.* **80** 223–8
- [39] Sepehri-Amin H, Ohkubo T and Hono K 2013 *Acta Mater.* **61** 1982–90
- [40] Soderžnik M, Korent M, Soderžnik K Ž, Katter M, Üstüner K and Kobe S 2016 *Acta Mater.* **115** 278–84
- [41] Huber C et al 2016 *Appl. Phys. Lett.* **109** 162401
- [42] Li L, Post B, Kunc V, Elliott A M and Paranthaman M P 2017 *Scr. Mater.* **135** 100–4
- [43] Thompson M P, Chang E, Foto A, Citron-Rivera J G, Haddad D, Waldo R and Pinkerton F E 2017 *IEEE Electrific. Mag.* **5** 19–27
- [44] Oikawa T, Yokota H, Ohkubo T and Hono K 2016 *AIP Adv.* **6** 056006