
Proceedings Paper:
Rowan-Robinson, R.M. orcid.org/0000-0002-3881-4064, Leong, Z., Carpio, S. orcid.org/0009-0003-5481-2653 et al. (2 more authors) (2024) Material informatics for functional magnetic material discovery. In: Li, C., Abert, C., Hadimani, R.L., Yan Law, J., Lupu, N., Phan, M-H., Piramanayagam, P., Visone, C., Amara, Y. and Bohn, F., (eds.) AIP Advances. 68th Annual Conference on Magnetism and Magnetic Materials, 30 Oct - 03 Nov 2023, Dallas, Texas. AIP Publishing .

https://doi.org/10.1063/9.0000657

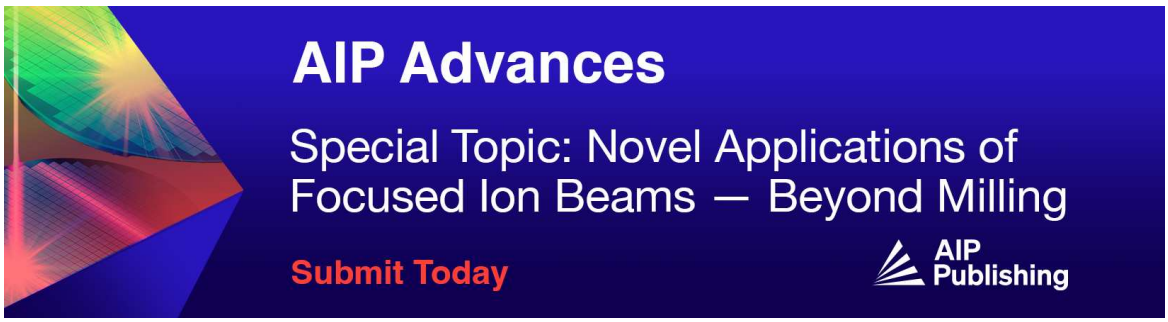
RESEARCH ARTICLE | JANUARY 17 2024

Material informatics for functional magnetic material discovery

Special Collection: 68th Annual Conference on Magnetism and Magnetic Materials

R. M. Rowan-Robinson ; Z. Leong; S. Carpio ; C. Oh ; N. A. Morley

AIP Advances 14, 015313 (2024)
https://doi.org/10.1063/9.0000657



Material informatics for functional magnetic material discovery

Cite as: AIP Advances 14, 015313 (2024); doi: 10.1063/9.0000657
Submitted: 29 September 2023 • Accepted: 30 October 2023 •
Published Online: 17 January 2024

R. M. Rowan-Robinson, Z. Leong, S. Carpio, C. Oh, and N. A. Morley[a)]

AFFILIATIONS

Department of Materials Science and Engineering, Mappin Street, University of Sheffield, Sheffield S1 3JD, United Kingdom

Note: This paper was presented at the 68th Annual Conference on Magnetism and Magnetic Materials.
[a)]**Author to whom correspondence should be addressed:** n.a.morley@sheffield.ac.uk

ABSTRACT

Functional magnetic materials are used in a wide range of "green" applications, from wind turbines to magnetic refrigeration. Often the magnetic materials used contain expensive and/or scarce elements, making them unsuitable for long term solutions. Further, traditional material discovery is a slow and costly process, which can take over 10 years. Material informatics is a growing field, which combines informatics, machine learning (ML) and high-throughput experiments to rapidly discover new materials. To prove this concept, we have devised a material informatics workflow and demonstrated the core components of natural language processing (NLP) to extract data from research papers to create a functional magnetic material database, machine learning with semi-heuristic models to predict compositions of soft magnetic materials, and high-throughput experimental evaluation using combinatorial sputtering and high-throughput magneto-optic Kerr effect (MOKE) magnetometry. This material informatics workflow provides a quicker, cheaper route to functional magnetic materials discovery.

I. INTRODUCTION

The drive to discover cheaper, more efficient materials for green technologies has moved away from traditional methodologies toward digital solutions and the emerging field of material informatics.[1] The rationale for this is that traditional methodologies are too slow to make a difference within a short time scale. Furthermore, they often involve tweaking existing compositions, which does not allow for the radical step changes needed to move material discovery forward at a faster pace. Magnetic materials are an ideal test bed for digital material discovery, with no new commercial hard or soft magnetic materials being discovered for over 30 years.[2,3]

Figure 1 gives the concept behind our material informatics workflow. Recent work has demonstrated that ML algorithms can be used to predict material compositions and properties that are of interest for specific applications, once trained on existing material databases.[4–6] High-throughput experiments can be carried out to validate the predictions and explore the surrounding composition space. These experimental results can then be added back to the materials database, creating a live document, where the more entries to the database the better the ML predictions are.

A common issue is that these existing databases are often limited, with too few entries, missing information and biased toward one class of material. For example, the NovaMag database,[7] which was utilised in this study for training our ML model to predict saturation magnetisation, is heavily biased toward Fe-based magnetic materials (over 1000 entries), with less than half the entries containing Co or Ni. This means that any ML algorithms trained on this database are likely to predict Fe-based materials. The arrival of much larger databases based on high-throughput DFT calculations such as AFLOW,[8] OQMD[9] and Material Project[10] have helped to address this. However, these are often limited in the range of magnetic properties calculated. New unbiased databases need to be produced, but this can be time consuming if done by hand. Furthermore, to realise the benefits of material informatics across magnetism, researchers need to have tools available to curate domain specific databases, since the specialist functional material properties (e.g., magnetostriction, Voigt constant etc.) are almost non-existent in current databases. One method of creating these databases is NLP, which uses algorithms to extract information such as composition, crystal structure and functional properties from papers.[11] This has been already used for simple database building, such as the critical temperature of superconductors.[12]

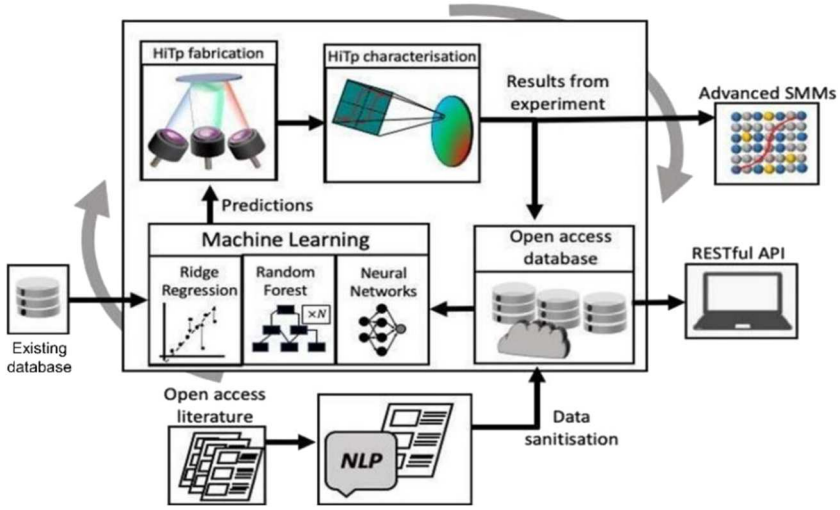


FIG. 1. Schematic of a proposed material informatics workflow for digital discovery of functional magnetic materials.

Our research has focused separately on the three different aspects of digital materials discovery: database building using NLP, materials composition and properties prediction using ML and material verification using high-throughput experiments for soft magnetic alloys.

II. METHODS

A. Natural language processing

Before developing our own NLP algorithm, we tested two available models on Github: Chemdataextractor[13] and MatSciBert.[14] Chemdataextractor is a named entity recognition NLP algorithm, which we adapted to extract composition and functional properties. MatSciBert is a semantic network NLP algorithm. We ran 8 papers[15] through both NLP softwares to extract the composition (Com) and magnetic properties (saturation magnetisation, M_s, coercive field, H_c and Curie temperature, T_c). Table I gives the results, where the efficiency, ε is defined as

$$\varepsilon = \frac{number\ \ of\ \ extracted\ \ parameters}{total\ \ number\ \ of\ \ parameters\ \ to\ \ extract} \times 100 \qquad (1)$$

Although this was a very small study, it is observed that neither NLP software was good at extracting the composition, while Chemdataextractor was good for the functional properties. Furthermore it was determined that if more than one composition was found within the paper, then the composition must be linked to the related functional properties. From these results, we developed our own NLP algorithm[16] that combined the best elements from both softwares, and added a correlation function to link the compositions

TABLE I. Results from the different NLP algorithms tested.

	Chemdataextractor					MatScibert					Our NLP code				
Paper	Com	M_s	H_c	T_c	ε (%)	Com	M_s	H_c	T_c	ε (%)	Com	M_s	H_c	T_c	ε (%)
A	✓	✓	✓	···	100	x	x	x	···	0	✓	✓	✓	···	100
B	✓	✓	✓	✓	80[a]	x	x	x	x	0	✓	✓	✓	✓	100
C	x	✓	···	···	50	x	x	···	···	0	✓	✓	···	···	100
D	In part	✓	···	···	75	x	x	···	···	0	✓	✓	···	···	100
E	x	✓	✓	···	66	x	x	x	···	0	✓	✓	✓	···	75[b]
F	In part	✓	···	···	75	x	x	···	···	0	✓	✓	···	···	75[b]
G	✓	✓	···	···	100	x	x	···	···	0	✓	✓	···	···	75[b]
H	✓	✓	···	···	100	x	x	···	···	0	✓	✓	···	···	100
Average			80.8					0					90.6		

[a] An additional magnetisation was extracted.
[b] Composition and saturation magnetisation were both extracted, but were not linked together.

TABLE II. Composition of the four CoFe–Ni–Mn–Ti films studied, with the predicted values from the model and the measured values from Energy Dispersive Spectrum (EDS) data.

	CoFeNi		Mn		Ti	
Sample	Predicted (%)	Measured (%)	Predicted (%)	Measured (%)	Predicted (%)	Measured (%)
r46c28	58	60	32	33	10	7
r55c32	62	65	28	29	10	6
r61c48	65	69	25	24	10	7
r74c52	67	73	23	20	10	7

and functional properties together. The results were then exported to a structured database.

B. Machine learning

There are several different ML algorithms that can be adapted to be trained on databases for material properties and compositions predictions. For this work, we utilised the Random Forest algorithm, as it is one of the better techniques for ML predictions when trained on databases with limited entries. Our model was trained on the Novamag database, as it contains calculated values for a larger range of magnetic properties, including magnetocrystalline anisotropy, exchange stiffness and Curie temperature. Saturation magnetisation was the most populated field with only 0.4% of the dataset (7 compounds) missing and consequently was selected as our target variable.

C. High-throughput experiments

One way to test the predictions of ML is to use high-throughput experiments.[17] Here, we use combinatorial sputtering to achieve over 50 different compositions on a three-inch silicon wafer. Before sputtering, the wafer was cleaned using acetone followed by IPA. It was then placed within a 3-gun sputtering chamber. Three different targets were used, for example when exploring soft magnetic materials, a single CoFeNi target with two different elemental targets (i.e. Mn, Al, Ti). We adapted the model of Frisk *et al.*[18] for the geometry of our sputtering system, to predict the composition across the substrate, from the power and position of the three sputter guns. The model was also able to predict the film thickness across the substrate [Fig. 3(a)]. For demonstration of the technique, a CoNiFe–Mn–Ti film was fabricated. Once fabricated, the substrate was cut into three 2 cm wide strips and measured in a high-throughput MOKE magnetometer. This MOKE magnetometer was

a.

	doi	sentence	composition	parameter	unit	value
0	10.1016/j.jallcom.2020.158450	12	FeCrCoNi-	T_c	K	420
1	10.1016/j.jallcom.2020.158450	12	Al_0.5CoCrFeNi	T_c	K	215
2	10.1016/j.jallcom.2020.158450	23	AlCoCrFeMnNi	M_s	Am^2/kg	60
3	10.1016/j.jallcom.2020.158450	26	AlCoCrFeMnNi	T_c	°C	550
4	10.1016/j.jallcom.2020.158450	95	AlCoCrFeNi)	M_s	kA/m	348

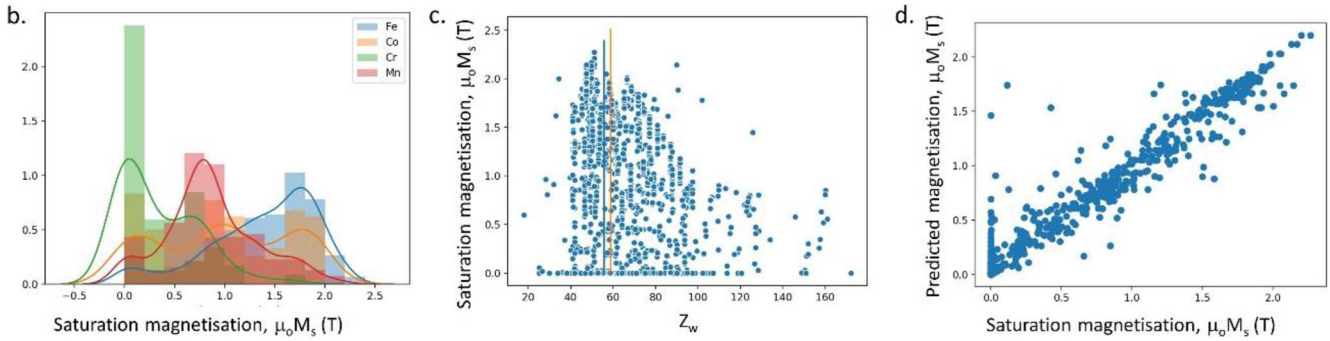


FIG. 2. (a) Results taken using our developed NLP code. (b) Frequency of transition metal element as a function of saturation magnetisation. (c) Saturation magnetisation as a function of the composition weighted atomic mass, with solid blue and orange lines the values for Fe and Co respectively. (d). Predicted saturation magnetisation versus actual saturation magnetisation.

developed to measure a MOKE loop every 4 mm on the film surface. The software then automatically determined the coercive field (H_c) and Kerr voltage (V_K) from each loop and plotted them against the composition predicted. This provides a ternary map of the magnetic properties [Figs. 3(b) and 3(c)], allowing for compositions of interest to be identified and studied further. For instance compositions with high V_K, and a low H_c can be determined and are good candidates for soft magnetic material. For this study, four films with the same Ti concentration but different Mn to CoFeNi ratio were studied (Table II). The compositions are in good agreement with the predicted values.

III. RESULTS

A. Natural language processing

We ran our newly developed NLP code on the 8 papers previously used in Ref. 15 and attained a >90% efficiency (Table I). To further test the code, it was run on 1000 magnetic papers[16] and achieved >76% efficiency. The code was able to distinguish between the different compositions and assign the correct properties to each composition. Figure 2(a) shows an image of the code output that was exported to a spreadsheet. The output parameters are Digital Object Identifier (doi) of the paper; sentence number; composition

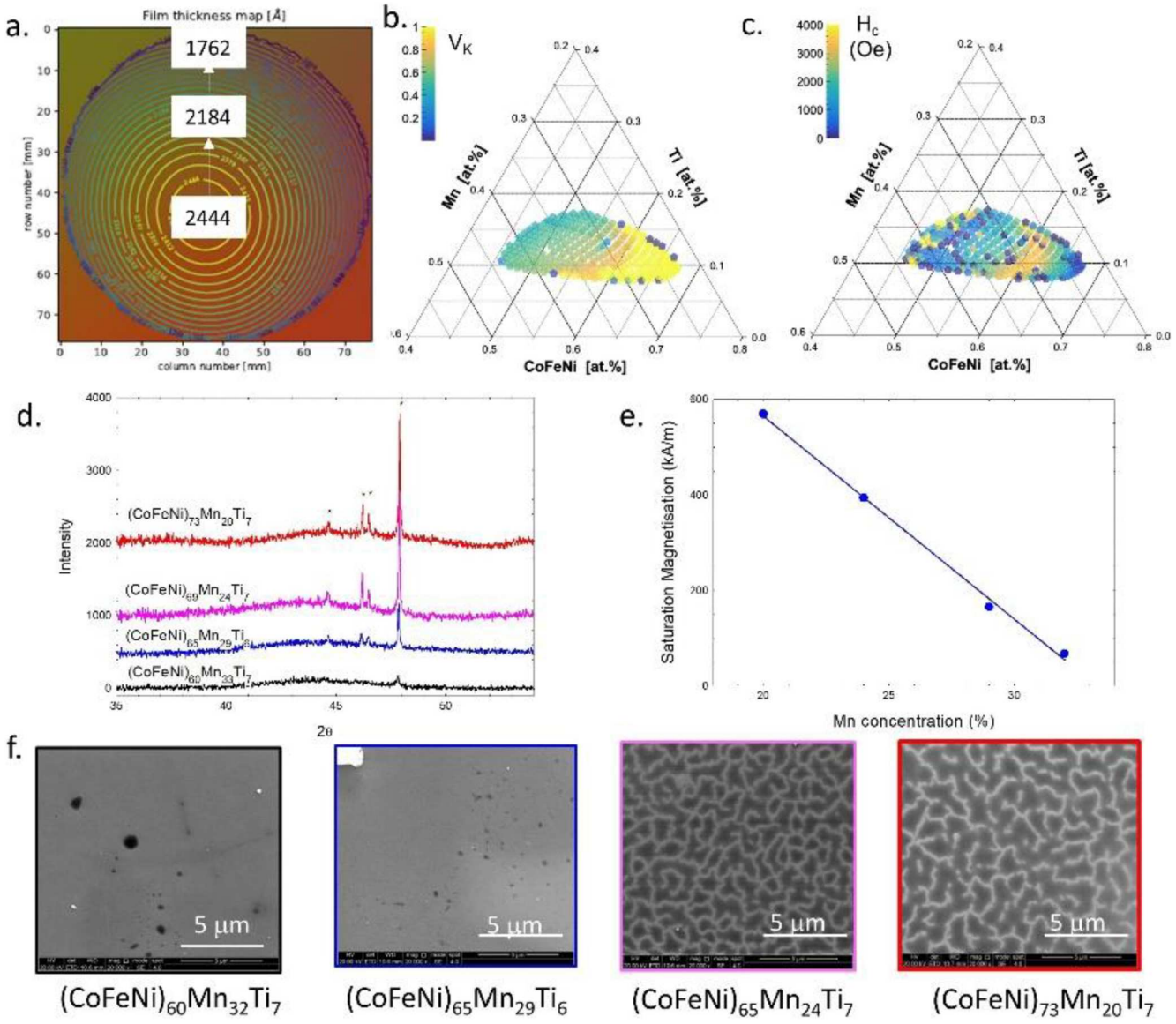


FIG. 3. (a) Predicted film thickness across the substrate; (b) and (c) ternary maps for the normalised Kerr voltage and coercive field for the CoFeNi–Mn–Ti film. (d) X-ray diffraction (XRD). (e). Magnetisation as a function of Mn concentration and (f). Scanning electron microscopy (SEM) data for four selected CoFeNi–Mn–Ti films.

and property parameter, unit, and value. It is observed that for the one paper shown, the NLP code was able to extract two compositions and two T_c from the same sentence. Further, it extracted a range of compositions and properties from the same paper. This newly designed NLP code is likely to speed up the process of materials database building.

B. Machine learning

A number of different exercises were performed before the algorithm was trained on the Novamag database.[7] The first task demonstrated how the saturation magnetisation of each alloy was linked to the transition metal elements (i.e., Fe, Co, Cr and Mn) within its composition [Fig. 2(b)]. It is observed that the alloys with the highest M_s mainly contain Fe, while most alloys containing Cr have $M_s \sim 0$. Although this is well established, it demonstrates that the underlying dataset is largely consistent with expected results.

The second task performed was to design our input features for the Random Forest model. All input features were derived from the chemical formula of the alloy using readily available data. We calculated several composition weighted values, including the atomic number, atomic mass, melting temperature, valence electron number and electronegativity. For example, the composition weighted atomic mass: $Z_w = x_i Z_i$; e.g., for $Fe_8Co_2N_8 \Rightarrow (8/18)Z_{Fe} + (2/18)Z_{Co} + (8/18)Z_N$. From Fig. 2(c), it is observed that the highest M_s occur when combining the transition metal elements with lighter elements, with the maximum value for $Z_w \sim 50$, i.e., combining Fe ($Z_{Fe} \sim 55$) with elements such as Al ($Z_{Al} \sim 27$) or B ($Z_B \sim 11$). There is a second peak for $Z_w \sim 65$, then M_s decreases as the Z_w increases, until a plateau region is reached for $Z_w = 150$.

The third task was to train the algorithm to link the M_s to the input features derived from the alloy composition, and then predict M_s for known compositions. Figure 2(d) shows the ML predicted M_s against the actual measured M_s. It is observed that there is a strong correlation between the two, demonstrating that the ML code used was able to perform the task. There are some obvious deviations as expected. Since the code can predict saturation magnetisation for a given composition, it can be adapted to predict compositions for a given M_s.

C. High-throughput experiments

Figure 3 provides an overview of the magnetic and structural data for the four films. It is observed that as the Mn concentration increases within the composition, the films go from two phases to one phase. From the X-ray diffraction (XRD) data [Fig. 3(d)], one phase is represented by a broad peak at ~44.5°, which is likely to be a disordered BCC or FCC phase often found in multi-component alloys. The second phase at the lower Mn concentration produces sharper peaks due to a Co_2Ti-like phase [indicated by * in Fig. 3(d)]. These two phases are observed in backscattered electron (BSE) images from the scanning electron microscope (SEM), as light and dark phases [Fig. 3(f)]. The increase in Mn removes this second phase (Co_2Ti), as only a single phase is observed in the SEM micrographs. The single phase corresponds to the decrease in magnitude of the XRD peaks for these films. The formation of secondary phases is often due to the enthalpy of mixing of the different elements within the alloy. Ti has a high enthalpy of mixing with Co and Ni, hence the second phase appears when Mn concentrations are low. By increasing Mn, the effect of the high enthalpy of mixing of Ti is reduced, and only a single phase is observed. The M_s decreases linearly as the Mn concentration increases [Fig. 3(e)], which is expected as the magnetic elements concentration was reduced. The addition of the second phase does not appear to affect the M_s.

Thus, using high-throughput experiments, has allowed for over 50 different compositions of CoFeNi–Mn–Ti to be studied. As part of the material informatics workflow, this data would then be fed back into the database before re-training the ML model.

IV. CONCLUSION

Material informatics is likely to revolutionise how new materials are discovered. The proposed material informatics workflow, demonstrates the three core components (NLP, ML and high-throughput experiments) in isolation with the potential for these to be joined up in the cyclic workflow outlined by Fig. 1. Magnetic materials are a perfect test bed for such technology, with new soft and hard magnetic materials being required for the advancement of green technology. ML is proven as an efficient route to predicting new magnetic materials, but it relies on large unbiased databases. We have demonstrated how NLP algorithms can be developed to extract compositions and properties from a range of different sources (papers, abstracts etc.) creating such databases. NLP will have an important role in generating bespoke databases for less commonly modelled functional properties important for subfields of magnetocalorics, magneto-optics etc. allowing ML models to obtain domain specific knowledge. High-throughput experiments can then be used to investigate the compositions predicted and validate the ML predictions in a short period of time, while further building the training database. This allows compositions of interest to be identified quickly for further study.

ACKNOWLEDGMENTS

This work was supported by the Leverhulme Trust (RPG-2018-324) and the Henry Royce Institute summer internship scheme. We wish to acknowledge the Henry Royce Institute for Advanced Materials, funded through EPSRC grants EP/R00661X/1, EP/S019367/1, EP/P02470X/1 and EP/P025285/1, for access to the SQUID at The University of Sheffield.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

R. M. Rowan-Robinson: Conceptualization (lead); Investigation (equal); Methodology (equal); Software (lead); Writing – review & editing (equal). **Z. Leong**: Investigation (equal); Methodology (equal); Software (equal). **S. Carpio**: Data curation (equal); Investigation (equal); Software (equal); Writing – review & editing (equal). **C. Oh**: Data curation (equal); Investigation (equal); Software (equal); Writing – review & editing (equal). **N. A. Morley**:

Funding acquisition (lead); Project administration (lead); Supervision (lead); Writing – original draft (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] R. Jose and S. Ramakrishna, Appl. Mater. Today **10**, 127–132 (2018).
[2] J. M. Silveyra *et al.*, Science **362**, eaao0195 (2018).
[3] J. M. Coey, Scr. Mater. **67**, 524–529 (2012).
[4] A. O. Oliynyk *et al.*, Chem. Mater. **28**, 7324 (2016).
[5] W. Huang *et al.*, Acta Mater. **169**, 225 (2019).
[6] J. M. Rickman *et al.*, JAP **128**, 221101 (2020).
[7] P. Nieves *et al.*, Comp Mats Sci **168**, 188 (2019).
[8] http://www.aflowlib.org/.
[9] https://oqmd.org/.
[10] https://next-gen.materialsproject.org/about.
[11] J. H. Lee *et al.*, Int. J. of Precis. Eng. and Manuf.-Green Tech. **10**, 1337–1349 (2023).
[12] C. J. Court and J. M. Cole, npj Comput. Mater. **6**, 18 (2020).
[13] M. C. Swain and J. M. Cole, J. Chem. Inf. Model. **56**, 1894 (2016); http://chemdataextractor.org/.
[14] MatSciBert, arXiv:2109.15290.
[15] N. A. Morley, Feasibility study to assess natural language processing for functional magnetic materials, https://doi.org/10.15131/shef.data.c.6373155, 2023.
[16] N. A. Morley, Royce summer internship project on NLP for materials discovery, https://doi.org/10.15131/shef.data.c.6800097.v1, 2023.
[17] F. Ren *et al.*, Sci. Adv. **4**, eaaq1566 (2018).
[18] https://uu.diva-portal.org/smash/get/diva2:1039332/FULLTEXT01.pdf.