Data-Driven Property Prediction for Memristor Resistive Switching Layers

Ben D. Rowlinson [1] **Subramanian Ramamoorthy** [2] **Jacqueline M. Cole** [3] **Themis Prodromakis** [1]

[1]*Centre for Electronic Frontiers, School of Engineering, University of Edinburgh, Edinburgh, EH9 3BF, U.K.* [2]*Institute of Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, U.K.* [3]*Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, CB3 0HE, U.K.*. Correspondence to: Ben D. Rowlinson ben.rowlinson@ed.ac.uk.

Identifying optimal materials for the resistive switching layer of a memristor device is critical for advancing development of this technology. At present, active materials are explored by brute-force elemental substitution methods, time-consuming literature surveys, and costly trial-and-error experimentation. With the advent of data-driven and AI techniques for materials development, we apply these techniques to accelerate the scientific research and development of next-generation electronic devices. We benchmark data-driven ML models for materials in memristors against an experimental ground truth auto-extracted from the literature corpus.

1. Introduction

Memristors are a class of emerging two-terminal electronic devices that exhibit negative differential resistance and a characteristic "pinched-hysteresis" loop in the current-voltage plane [1]. Memristors are promising candidates for a number of key applications in ultra-low power electronics, including non-volatile memory [2], in-memory computing [3], analog computing [4], neuromorphic computing [5], and signal processing [6]. Each of these applications necessitate diverse operational parameters that are governed by microfabrication technique, defect engineering, and, critically, materials selection, particularly in the resistive switching layers (RSL) 1(a). Lightweight ML ($O(n)$) surrogate models are increasingly utilized to bypass computationally intensive *ab initio* techniques for materials selection. In this work, we present an ensemble of data-driven ML models to accelerate materials selection based on desirable properties for new RSL of memristor devices.

2. Methodology

An *in silico* inverted pyramid structure using lightweight ML property prediction is utilized for materials selection for the RSL, as per Fig. 1(b). Candidate materials are filtered from a large pool down to a handful of suitable candidate materials based on application-specific properties [7]. The gradient-boosted feature selection (GBFS) workflow, in Fig. 1(c), is based on a fully-interpretable gradient-boosted decision tree (GBDT) architecture and is employed to rank and prune chemical features based on their predictive power, as well as identifying non-linear feature interactions. Feature engineering and recursive feature elimination accommodate feature sparsity. These properties allow the GBFS workflow to avoid model overfitting and enhance generalizability.

For a memristor RSL, properties of interest include

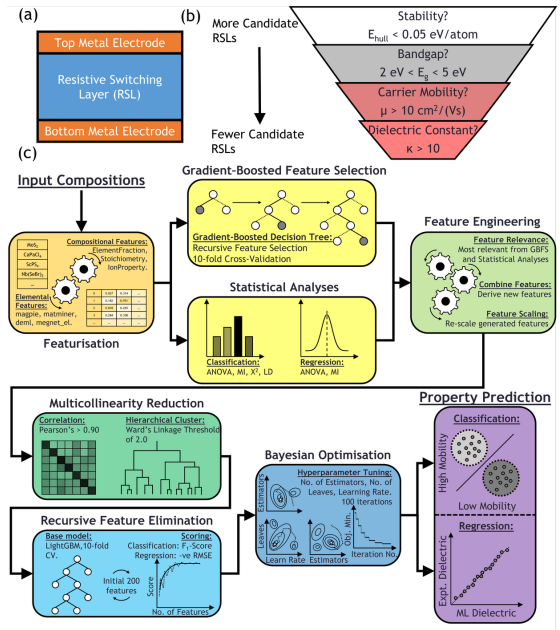

Fig. 1: (a) Memristor cross-section. (b) Inverted-pyramid property screening using surrogate models for E_g, κ, μ_e, and μ_h. (c) GBFS workflow for property prediction using chemical features.

electronic band gap (E_g), dielectric constant (κ), and charge carrier mobility (μ_e and μ_h, for electrons and holes, respectively). E_g and κ are framed as regression problems due to the wide availability of experimental data with which to assess these models. μ_e and μ_h predictions are framed as classification tasks and are shown in Appendix A.

An ensemble of data-driven models are trained and tested on computed properties (Alexandria [8] for E_g, Materials Project [9, 10] for κ, and MatHub-3d [11] for μ_e and μ_h). These resources are filtered for energy above convex hull of stability below 0.05 eV/atom. Elemental and compositional feature descriptors are generated using PyMatGen [12] as input. Structural features are omitted to accommodate experimental reports that lack accurate structural characterization. Experimental values auto-extracted from the literature corpus using ChemDataExtractor [13] are employed to benchmark the ML GBFS models for E_g [14] and κ [15]. Where multiple values of E_g or κ are extracted for the same material the mean is used to test the GBFS models. As these property values are auto-extracted there may be uncorrected measurements or parsing errors. SHAP analysis is conducted on all models and presented in Appendix B.

3. Results and Analysis

The surrogate model for E_g is presented in Fig. 2(a-d), achieving *MAE* of 0.301 eV, *RMSE* of 0.455 eV, and R^2 of 0.925 against the unseen test set derived from the distribution of the training data. The regressor demonstrates strong predictive performance, with a high R^2 indicating that most of the variance in E_g is captured by the surrogate model. The relatively low *MAE* and *RMSE* suggest good quantitative accuracy for in-distribution predictions.

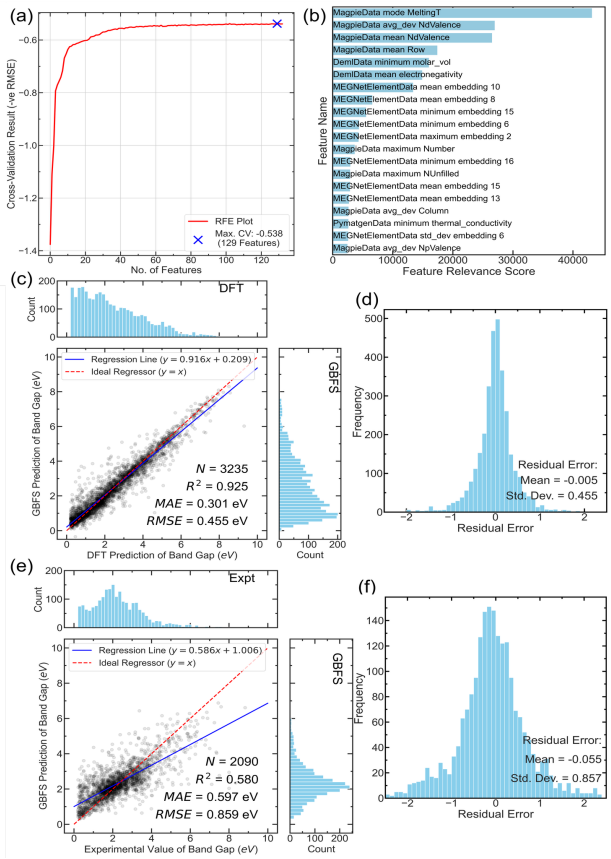

Fig. 2: GBFS E_g regressor trained on Alexandria. (a) Cross-validation, (b) top-20 relevant features, (c) parity plot, (d) histogram of residuals, (e) parity plot against experimental E_g, and (f) residuals.

The model trained on computed values of E_g is tested for predictive power against an experimental ground truth in Fig. 2(e-f). A *MAE* of 0.597 eV, *RMSE* of 0.859 eV and R^2 of 0.580 is achieved when tested against experimentally reported E_g. There is a marked, yet expected, degradation in predictive performance, with a substantially lower R^2. The *MAE* and *RMSE* indicate some transferability to predicting experimental E_g. While sufficient for materials filtering for RSLs, the errors remain non-negligible for other applications requiring fine E_g resolution.

The surrogate model for κ is presented in Fig. 3(a-d), demonstrating *MAE* of 2.97, *RMSE* of 4.54, and R^2 of 0.624 against the unseen, in-distribution test set. The GBFS model for κ exhibits moderate predictive capability on these test data, as reflected by an R^2 of 0.624, indicating that a substantial fraction of the variance is captured. However, the relatively large *MAE* (5.9%) and *RMSE* (9.1%) suggest limited quantitative accuracy despite reasonable trend reproduction in the parity plot.

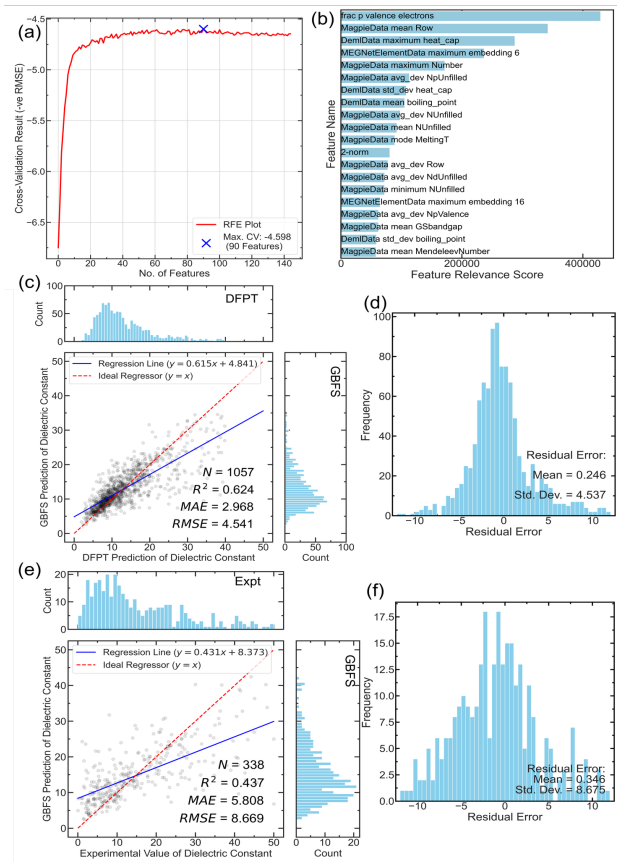

Fig. 3: GBFS κ regressor trained on Materials Project. (a) Cross-validation, (b) top-20 relevant features, (c) parity plot, (d) histogram of residuals, (e) parity plot against experimental κ, and (f) residuals.

The model trained on computed values of κ is tested for predictive power against 338 experimental values in Fig. 3(e-f). A *MAE* of 5.81, *RMSE* of 8.67, and R^2 of 0.437 is achieved by the model when tested against experimentally reported κ. These metrics indicate weak predictive performance and limited generalization across data sources. The comparatively poor accuracy is likely due to material inhomogeneities, measurement inconsistencies, and noise in the reported κ values. However, for course-grain materials filtering using lightweight and scalable prediction models the errors remain tolerable for RSLs.

4. Conclusions

We demonstrate fully-interpretable and lightweight data-driven ML property predictors for accelerating selection and development of memristor resistive switching layers. This workflow can be extended to other properties of interest possessing sufficient datasets, such as direct/indirect band gap, density of states at Fermi level, and ion/impurity transport. The inverted-pyramid will be used to select materials for experimental implementation.

Acknowledgments

All authors gratefully acknowledge UK Research and Innovation (UKRI) and Engineering and Physical Sciences Research Council (EPSRC) funding for the AI Hub for Productive Research and Innovation in Electronics (APRIL) AI Hub [grant number EP/Y029763/1].

References

[1] Leon Chua. If it's pinched it's a memristor. *Semiconductor Science and Technology*, 29(10), 2014.

[2] L. Zhong, L. Jiang, R. Huang, and C. H. de Groot. Nonpolar resistive switching in cu/sic/au nonvolatile resistive memory devices. *Applied Physics Letters*, 104(9), 2014.

[3] Daniele Ielmini and H. S. Philip Wong. In-memory computing with resistive switching devices. *Nature Electronics*, 1(6):333–343, 2018.

[4] Yanan Zhong, Jianshi Tang, Xinyi Li, Xiangpeng Liang, Zhengwu Liu, Yijun Li, Yue Xi, Peng Yao, Zhenqi Hao, Bin Gao, He Qian, and Huaqiang Wu. A memristor-based analogue reservoir computing system for real-time and power-efficient signal processing. *Nature Electronics*, 5(10):672–681, 2022.

[5] Z. Y. He, T. Y. Wang, J. L. Meng, H. Zhu, L. Ji, Q. Q. Sun, L. Chen, and D. W. Zhang. Cmos backend compatible memristors for in situ digital and neuromorphic computing applications. *Mater Horiz*, 8(12):3345–3355, 2021.

[6] Caterina Sbandati, Spyros Stathopoulos, Patrick Foster, Noam D. Peer, Cristian Sestito, Alex Serb, Stefano Vassanelli, Dana Cohen, and Themis Prodromakis. Single-trial detection of auditory cues from the rat brain using memristors. *Science Advances*, 10(eadp7613), 2024.

[7] Edward O. Pyzer-Knapp, Jed W. Pitera, Peter W. J. Staar, Seiji Takeda, Teodoro Laino, Daniel P. Sanders, James Sexton, John R. Smith, and Alessandro Curioni. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Computational Materials*, 8(1), 2022.

[8] J. Schmidt, H. C. Wang, T. F. T. Cerqueira, S. Botti, and M. A. L. Marques. A dataset of 175k stable and metastable materials calculated with the pbesol and scan functionals. *Sci Data*, 9(1):64, 2022.

[9] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1), 2013.

[10] I. Petousis, D. Mrdjenovich, E. Ballouz, M. Liu, D. Winston, W. Chen, T. Graf, T. D. Schladt, K. A. Persson, and F. B. Prinz. High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials. *Sci Data*, 4:160134, 2017.

[11] M. Yao, Y. Wang, X. Li, Y. Sheng, H. Huo, L. Xi, J. Yang, and W. Zhang. Materials informatics platform with three dimensional structures, workflow and thermoelectric applications. *Sci Data*, 8(1):236, 2021.

[12] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

[13] J. Mavracic, C. J. Court, T. Isazawa, S. R. Elliott, and J. M. Cole. Chemdataextractor 2.0: Autopopulated ontologies for materials science. *J Chem Inf Model*, 61(9):4280–4289, 2021.

[14] Q. Dong and J. M. Cole. Auto-generated database of semiconductor band gaps using chemdataextractor. *Sci Data*, 9(1):193, 2022.

[15] J. Zhao and J. M. Cole. A database of refractive indices and dielectric constants auto-generated using chemdataextractor. *Sci Data*, 9(1):192, 2022.

Appendix A. Charge Carrier Mobility Models

The classifier for μ_e uses 58 input features and performs with an F1-score of 0.758. The classifier for μ_h uses 46 input features and performs with an F1-score of 0.800. RFE curves, feature relevance plots, receiver-operator characteristics (ROC), and confusion matrices for μ_e and μ_h prediction are shown in Fig. A1 and A2, respectively. The μ_e and μ_h classifiers perform well on the test set of given the relatively small training/cross-validation set (N = 5554).

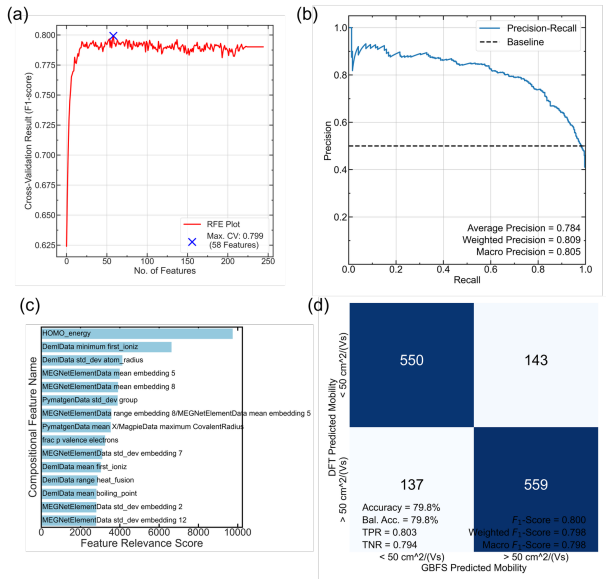

Fig. A1: GBFS μ_e classifier with MatHub-3D ground truth. (a) Cross-validation, (b) AUC-ROC plot, (c) top-20 relevant features, (d) confusion matrix.

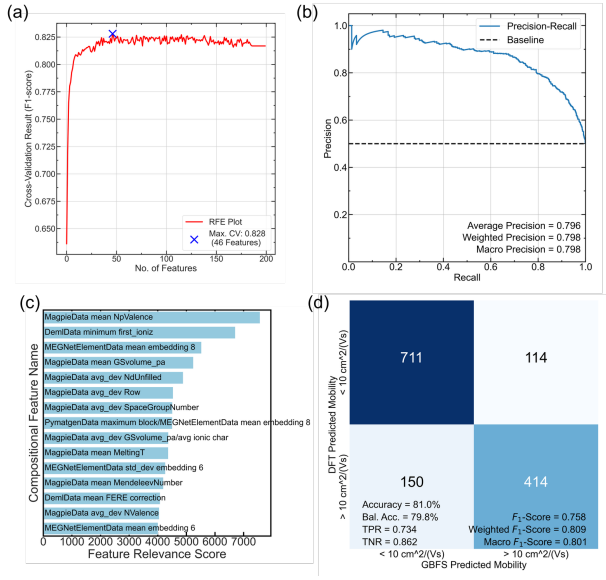

Fig. A2: GBFS μ_h classifier with MatHub-3D ground truth. (a) Cross-validation, (b) AUC-ROC plot, (c) top-20 relevant features, (d) confusion matrix.

Appendix B. SHAP Analysis of GBFS Models

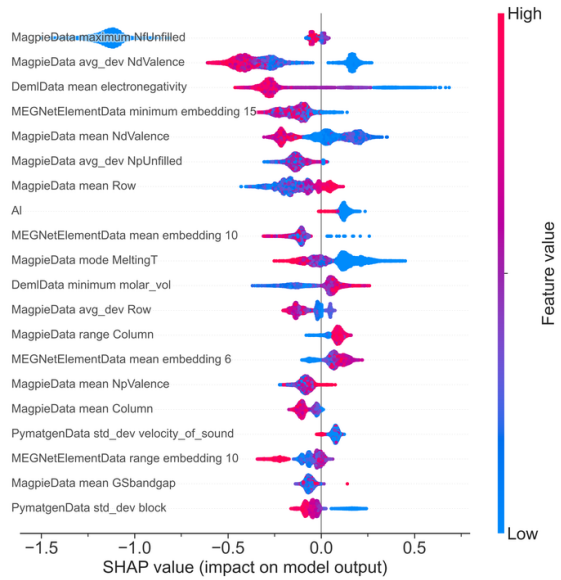

Fig. A3: SHAP analysis for GBFS model for band gap.

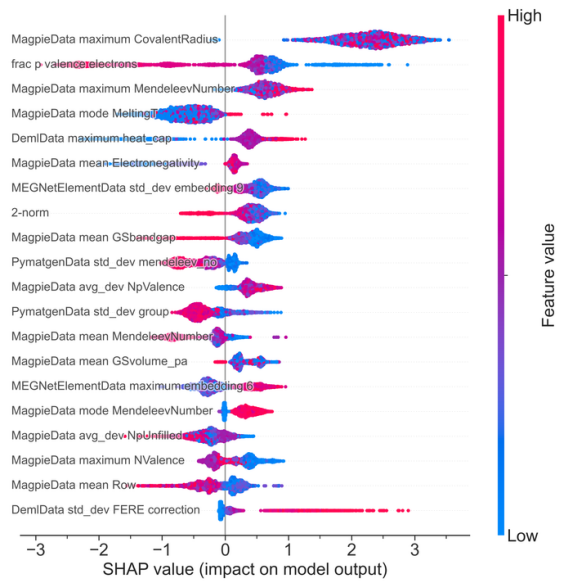

Fig. A4: SHAP analysis for GBFS model for dielectric constant.

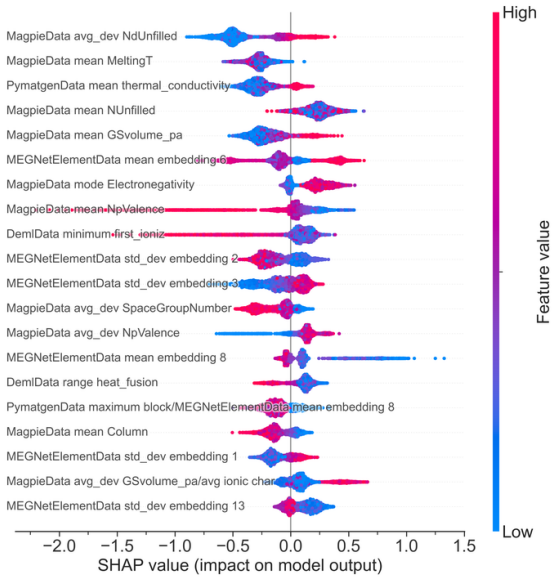

Fig. A5: SHAP analysis for GBFS model for electron mobility classification.

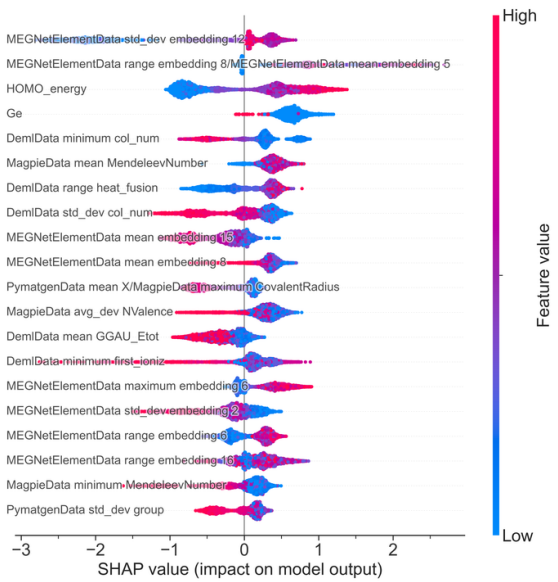

Fig. A6: SHAP analysis for GBFS model for hole mobility classification.