Deep learning for thermoelectric property prediction: A data-informed approach to skutterudites

Victor Posligua[a], **Karina Landivar**[b], **Gerda Rogl**[c], **Peter F. Rogl**[c], **Jose J. Plata**[d], **Antonio M. Marquez**[d]

[a] *Institute for Functional Intelligent Materials, National University of Singapore (NUS), Block S9, 4 Science Drive 2, 117544 Singapore* victorhp@nus.edu.sg

[b] *Department of Physics, Università degli Studi de Milano, Via Festa del Perdono, 7, 20122 Milan, Italy*

[c] *Institute of Materials Chemistry, University of Vienna, Waehringerstraße 42, A-1090 Wien, Austria*

[d] *Department of Physical Chemistry, University of Seville, 41012, Spain*

1. Introduction

The accurate prediction of thermoelectric properties, particularly the figure of merit (zT), is crucial for the discovery and optimisation of new thermoelectric materials. Skutterudites, a class of promising thermoelectric compounds, exhibit complex composition-property relationships that challenge traditional predictive methods. In this study, we leverage a large dataset of experimentally curated [1] and computationally derived skutterudites to develop a deep learning framework for precise zT prediction.

2. Methodology

Our approach combines domain-specific feature engineering, statistical outlier detection using LightGBM [2] and a deep neural network implemented in TensorFlow [3] to capture complex composition-property relationships within a standardised dataset. The model was optimised using Scikit-Learn [4], systematically exploring the hyperparameter space to optimise predictive accuracy and generalisation.

3. Results and Discussion

The model achieved a training RMSE of 0.069 and validation RMSE of 0.077, with a R^2 score of 0.947 and 0.956, respectively, demonstrating strong predictive performance (Fig. 1). Unlike brute-force approaches that rely on predefined statistical assumptions, our model captures physicochemical meaningful correlations between input features and zT, enhancing interpretability and guiding material design. These findings highlight the potential of data-driven approaches to accelerate the discovery of high-performance thermoelectric materials.

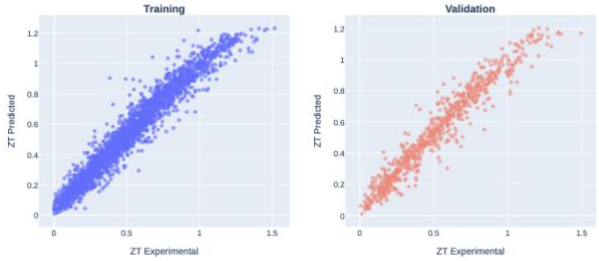

Fig. 1: Parity plots comparing predicted and experimental zT values for the training (blue) and validation (red) datasets. The strong alignment along the diagonal indicates high predictive accuracy and generalisation capability of the model.

4. Conclusions

This study presents a scalable and interpretable deep learning framework for predicting thermoelectric properties in skutterudites. By integrating advanced machine learning techniques with domain-informed feature engineering, our approach enables accurate property predictions while uncovering physicochemical trends relevant to material design. Future work will expand the dataset and incorporate first-principles simulations to further enhance model reliability and transferability.

References

[1] G. Rogl and P. F. Rogl. Filled Sb-Based Skutterudites from 1996-2022. Crystals 1843 (12), 1-30, 2022.

[2] G. Ke, *et al*. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems, 30, 2017.

[3] Martin Abadi, *et al*. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv preprint. arXiv:1603.04467.

[4] F. Pedregosa, *et al.* Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830, 2011.