Northeast Materials Database (NEMAD): Enabling Discovery of High Transition Temperature Magnetic Compounds

Suman Itani, Yibo Zhang, Jiadong Zang
Department of Physics and Astronomy, University of New Hampshire, Durham, NH 03824

Introduction

- Discovery of magnetic materials is crucial for technological advancement.
- Conventional ways to discover materials are generally time-consuming and costly.
- Current data-driven approaches are challenging and limited due to the lack of accurate, comprehensive, and feature-rich databases [2].
- This study used Large Language Models (LLMs) to automate magnetic materials data extraction from literature, creating an enhanced database.
- Using this database, we developed ML models to classify materials and predict properties.
- Using this models, we identified new materials.

Methods

Database Compilation:

- We collected more than 50,000 DOIs of scientific articles from Elsevier journal.
- Articles were downloaded initially in XML format.
- We developed an XML parser and table parser to convert them into a plain text.
- Next, we built a GPTArticleExtractor [3] workflow to extract required information from the articles in .json format using LLMs.
- Finally, we compiled a database of magnetic materials.

Model Development:

- We developed machine learning models, including Random Forest (RF), XGBoost, and Ensemble Neural Network (ENN) to classify materials and predict Curie and Néel temperatures.

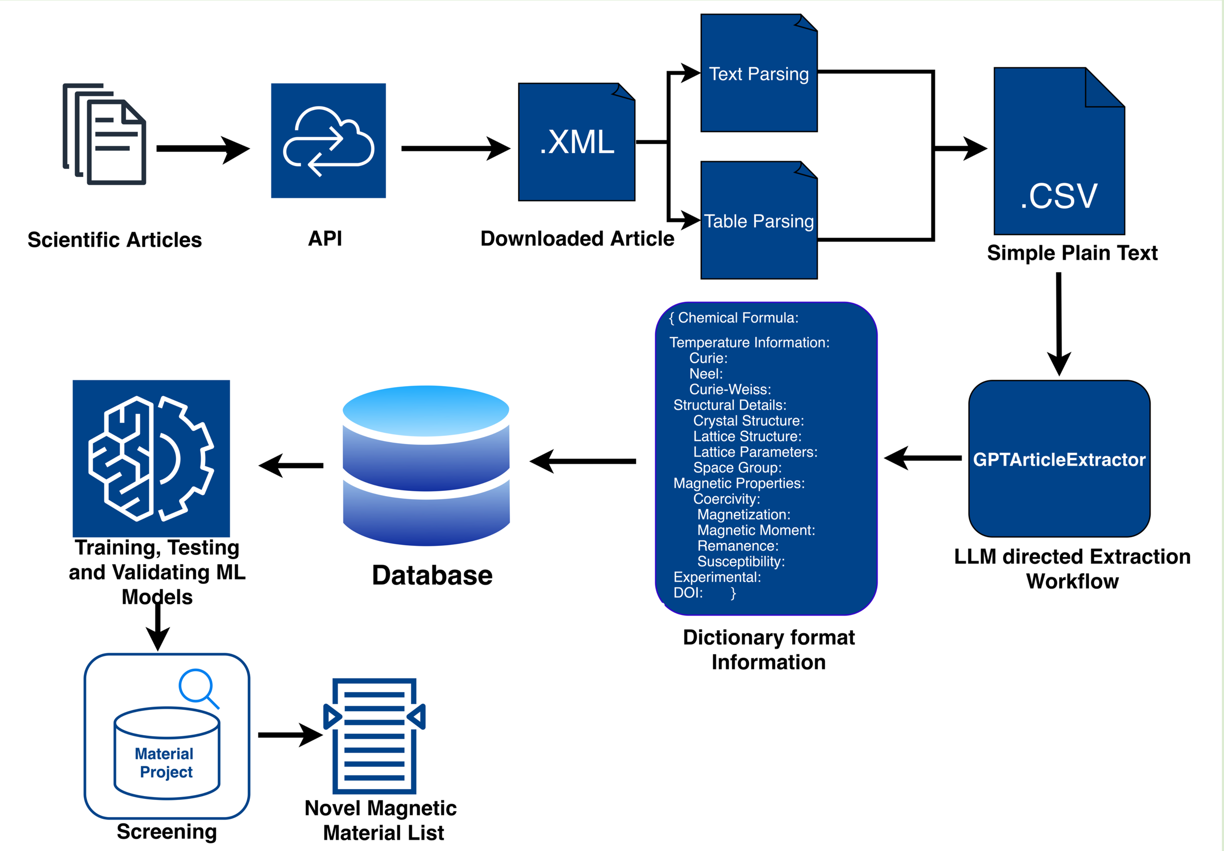
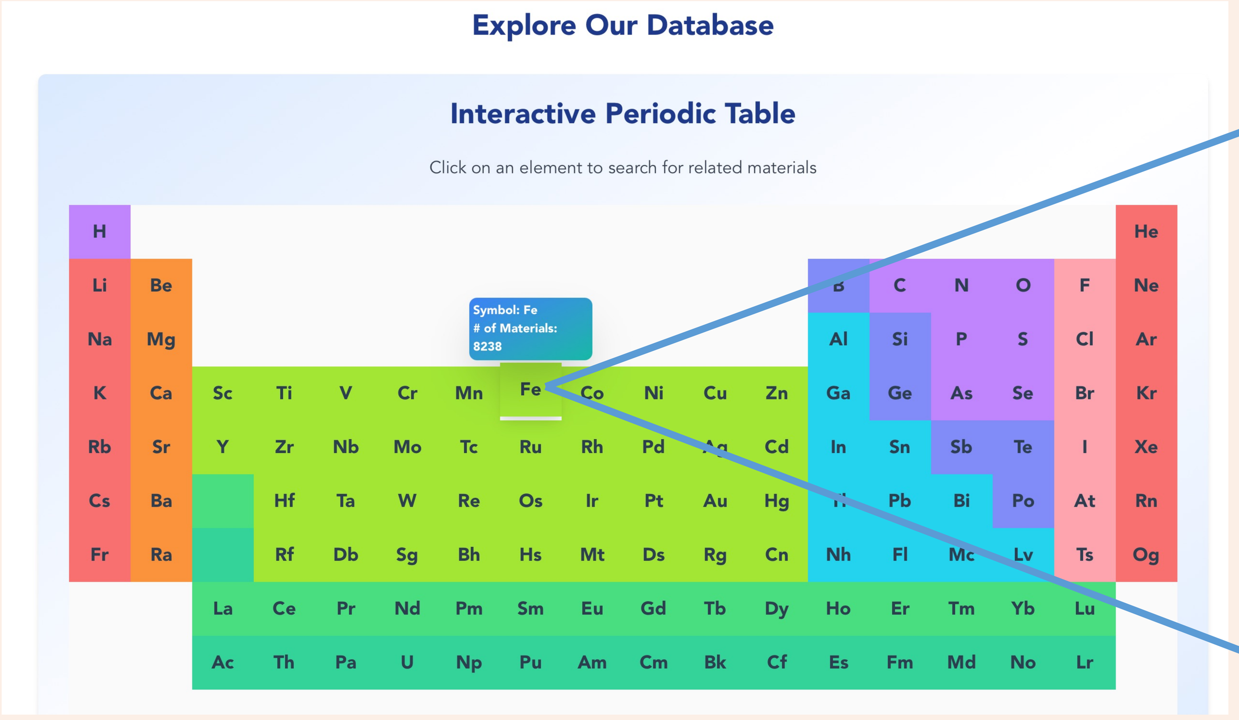


Figure: Workflow for the database compilation to material screening

Results

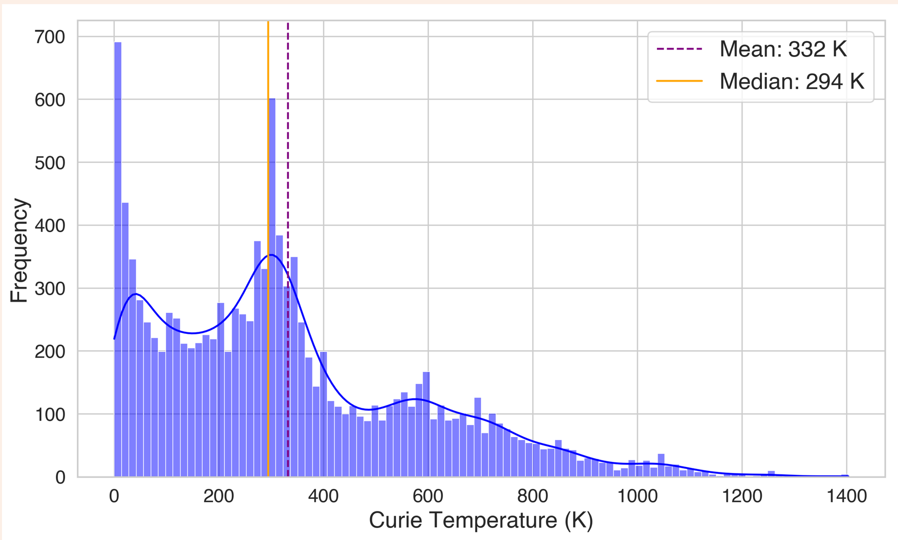
Database Statistics:

- We have developed a database of 26,706 magnetic materials.
- 65% of materials are ferromagnetic (FM) and 33% are antiferromagnetic (AFM).
- 22% of the FM compounds have a Curie temperature greater than 600K.



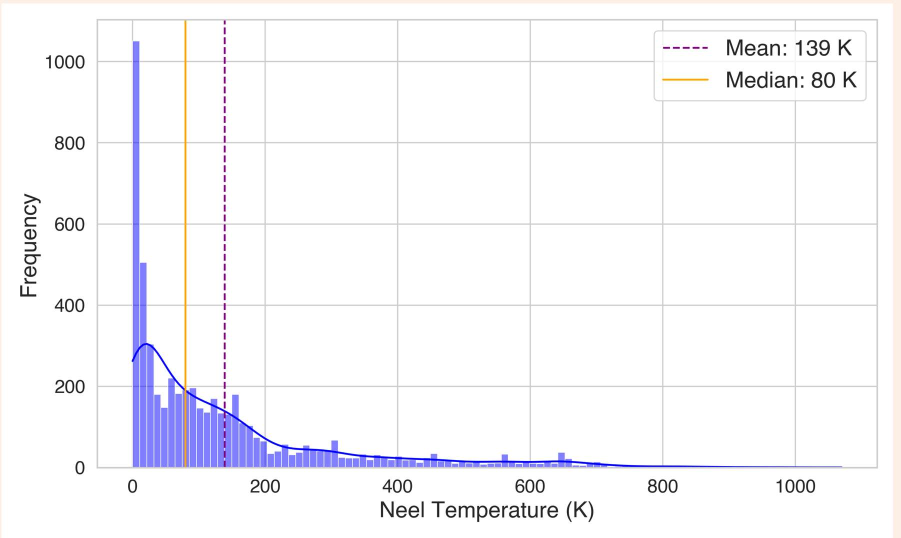
Material Detail


```
{
  "Material_Name": "SrFe12O19",
  "Curie": "700 K",
  "Neel": null,
  "Curie_Weiss": null,
  "Crystal_Structure": "M-type hexaferrite",
  "Lattice_Structure": "Hexagonal",
  "Lattice_Parameters": {
    "a": "5.88 Å",
    "b": "5.88 Å",
    "c": "23.05 Å"
  },
  "Space_Group": "P63/mmc",
  "Coercivity": "2.56 kOe",
  "Magnetic_Moment": null,
  "Magnetization": "49 emu/g",
  "Remanence": "35 emu/g",
  "Susceptibility": "0.00079 Oe·g/emu",
  "DOI": "10.1016/j.jmmm.2021.168195",
  "Experimental": true
}
```

(a) Curie temperature distribution plot

Scan me!
www.nemad.org



(b) Néel temperature distribution plot

Screening of High-Performance Compounds:

- Accessed stable magnetic compounds from the Material Project database
- Classified and predicted transition temperature using our models
- Identified 62 FM compounds with Curie temperatures greater than 500K
- Identified 19 AFM compounds with Néel temperature greater than 100K.

Table 1: Lists of High Curie Temperature FM Compounds Screened Out by Our Models

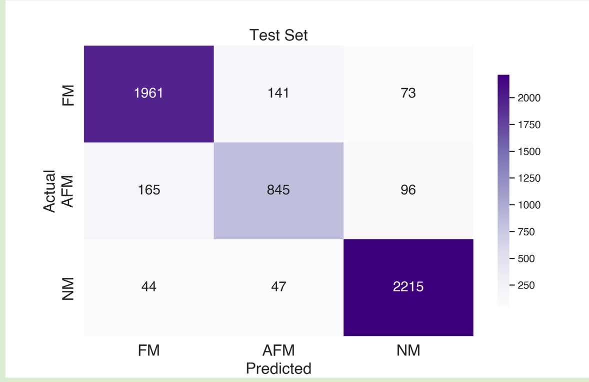
Material Composition	Predicted Type	Ensemble NN (K)	XGBoost (K)	RF (K)
VFeCoGe	FM	609	710	729
FeCo2Ge	FM	1051	1068	999
LiFeRh2	FM	636	690	577
AlFeCo2	FM	912	928	907
Ba(FeO2)2	FM	660	659	658
GaFe2Co4Si	FM	1042	1013	1007

Conclusion

- A comprehensive database of magnetic materials was built.
- This database can be used to train future ML models to predict material properties.
- Our ML Models can be employed by researchers to develop next-generation magnetic materials through large-scale materials screening.
- Our approach is versatile and can be applied to other areas of material science, such as superconducting, thermoelectric, photovoltaic, and ferroelectric materials.
- Overall, this study presents an effective method for identifying high-performance magnetic materials using large language models and ML techniques.

Classification Model:

- Model: RF Classifier
- Classify: FM, AFM, and NM
- Overall accuracy: 90%
- Dataset Size: 28,128
- Features: Generate from Chemical Composition



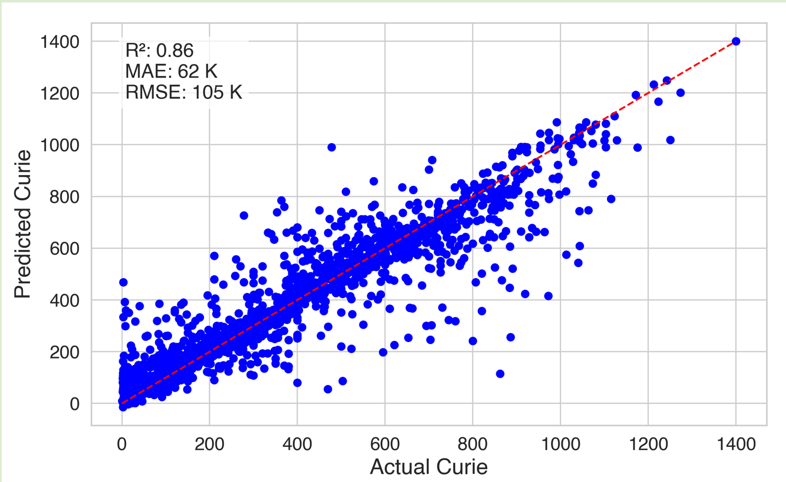
(a) Confusion Matrix

Table 2: Performance Statistics of Model

Dataset	Class	Precision	Recall	F1-Score	Accuracy
Training (60%)	FM	1.00	1.00	1.00	0.99
	AFM	0.99	0.98	0.99	
	NM	1.00	1.00	1.00	
Validation (20%)	FM	0.92	0.90	0.91	0.90
	AFM	0.81	0.75	0.78	
	NM	0.92	0.97	0.94	
Testing (20%)	FM	0.90	0.90	0.90	0.90
	AFM	0.82	0.76	0.79	
	NM	0.93	0.96	0.94	

Curie Temperature Prediction Models:

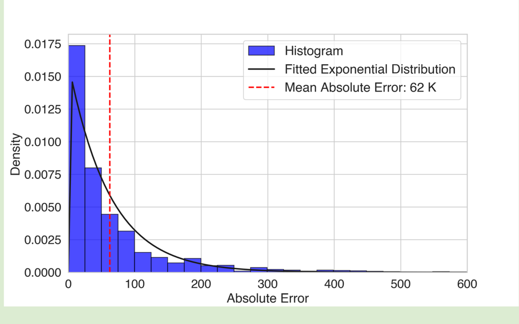
- Models: RF, ENN, & XGBoost
- Dataset Size:
  Original: 11,923
  Balanced: 8,249
- Features: Generate from Chemical Composition



(a) Predicted vs Actual Curie Temperature

Table3: Metrics to evaluate regression models predicting Curie temperature

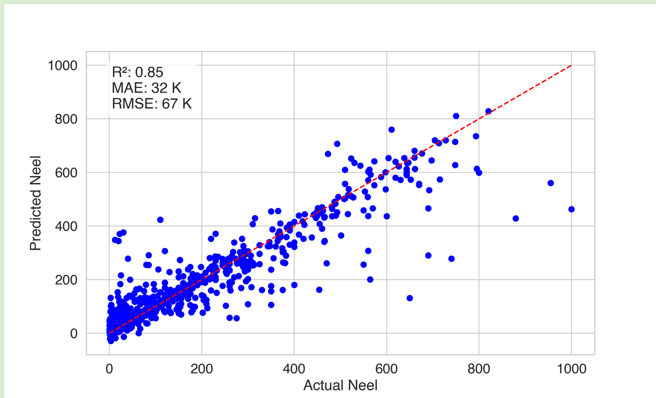
Model	Dataset	R-squared (Test)	MAE (Test)	RMSE (Test)
Random Forest	Original	0.80	70K	114K
	Balanced	0.83	72K	113K
Ensembled Neural Network	Original	0.79	63K	118K
	Balanced	0.84	64K	110K
XGBoost	Original	0.80	66K	114K
	Balanced	0.86	62K	105K



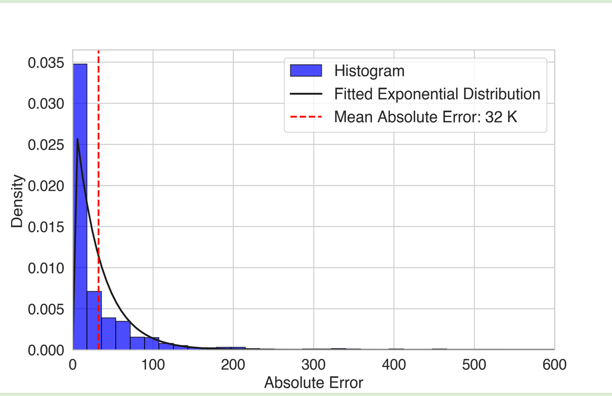
(b) Absolute Error Distribution Plot

Néel Temperature Prediction Models:

- Model: RF, ENN, XGBoost
- Dataset Size: 5,389
- Feature: Generate from Chemical Composition
- Best (XGBoost): R2 = 0.85, MAE = 32K, RMSE = 67K



(a) Predicted vs Actual Néel Temp. Plot



(b) Absolute Error Distribution Plot

References

[1] Coey, J. M. Magnetism and magnetic materials (Cambridge university press, 2010).
[2] Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: status, challenges, and perspectives. Adv. Sci. 6, 1900808 (2019)
[3] Zhang, Y. et al. GPTArticleExtractor: An automated workflow for magnetic material database construction J. Magn. Magn. Mater. 597, 172001 (2024)
[4] Itani, S., Zhang, Y., & Zang, J. Northeast Materials Database (NEMAD): Enabling Discovery of High Transition Temperature Magnetic Compounds. *arXiv:2409.15675* (2024)

This work was supported by the Department of Energy, Basic Energy Science under Grant No. DE-SC0020221.