

AI-Guided Discovery of Large-Band-Gap Dielectric Materials for Next-Generation Power Grids

Shivansh Raj

BSc Computer Science | University of Liverpool

sraj@liverpool.ac.uk

Supervisors: Dr. Xue Yong (Dept. of EEE) | Dr. Bengt Tegner (Research IT, University of Liverpool)

N8 CIR Summer Internship 2026 | 8 Weeks

The Problem



Why does this matter?

- Power grids carry electricity at extremely high voltages
- Current insulator SF₆ gas is 23,000 times more harmful than CO₂
- A solid-state replacement is urgently needed

Our Research Question

Can AI predict which materials have the largest band gaps and screen new candidates faster than any lab?

- Lab testing takes years and millions of pounds
- AI screens thousands of materials in seconds

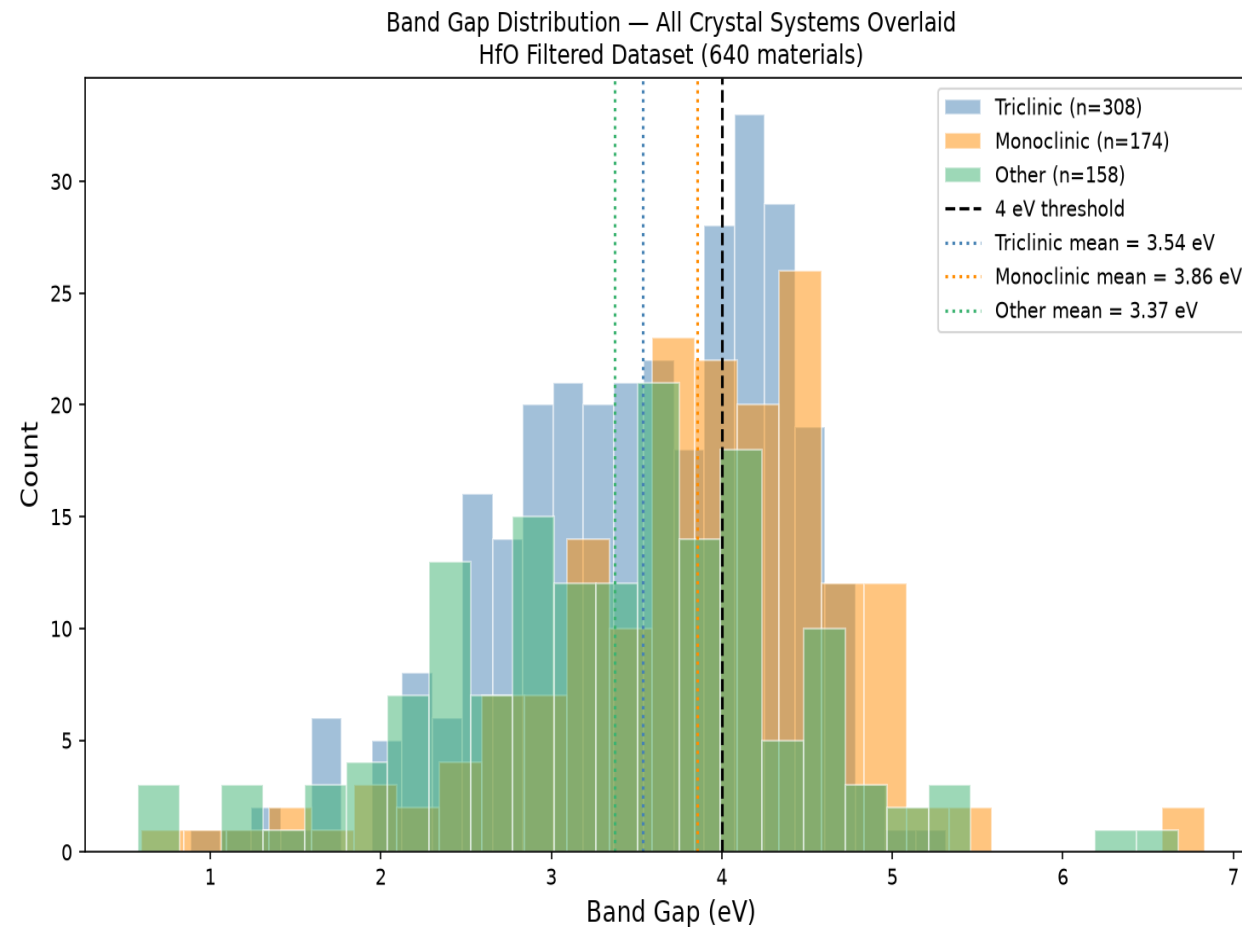
The Data

Where did the data come from?

- 4,766 materials downloaded from the Materials Project database
- Two chemical systems: hafnium oxide and zinc oxide
- 165 numerical features per material
- Band gaps calculated by quantum mechanics (DFT)

How did we clean the data?

- Removed metals as they conduct electricity
- Kept only materials with band gap above 1.0 eV
- Kept only thermodynamically stable materials
- Result: 2,076 high quality training materials



Band gap distribution by crystal system: HfO dataset shown

ZnO follows a similar pattern with a lower mean band gap of 2.88 eV

The Approach



Why these choices?

Why HfO and ZnO?

Both are known dielectric oxide families with promising band gaps, making them a good starting point

Why filter for stability?

Unstable materials cannot be synthesised, so only physically realisable candidates are useful

Why group by space group?

Materials with similar crystal symmetry share similar bonding patterns and are more predictable as a group

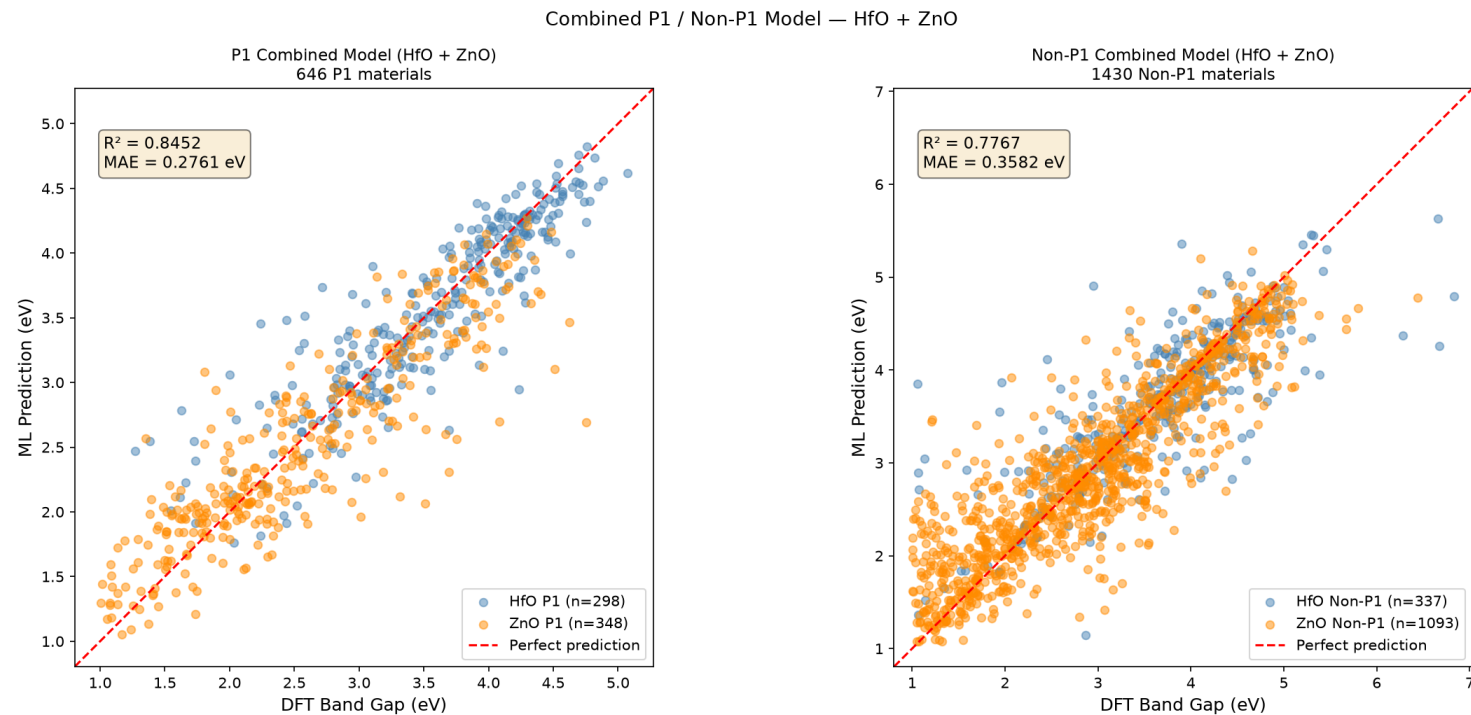
Why cross validation?

The model is always tested on data it has never seen, giving an honest and realistic performance estimate

Why Barkla HPC?

Training 2,000-tree models on 2,076 materials requires significant computing power

Results: How Well Does the Model Predict?



Predicted vs true band gap: P1 model on the left (R^2 of 0.845) and Non-P1 model on the right (R^2 of 0.776)

Key Metrics

Best Model

XGBoost (P1)

R^2

0.845

MAE

0.269 eV

Materials

646 P1

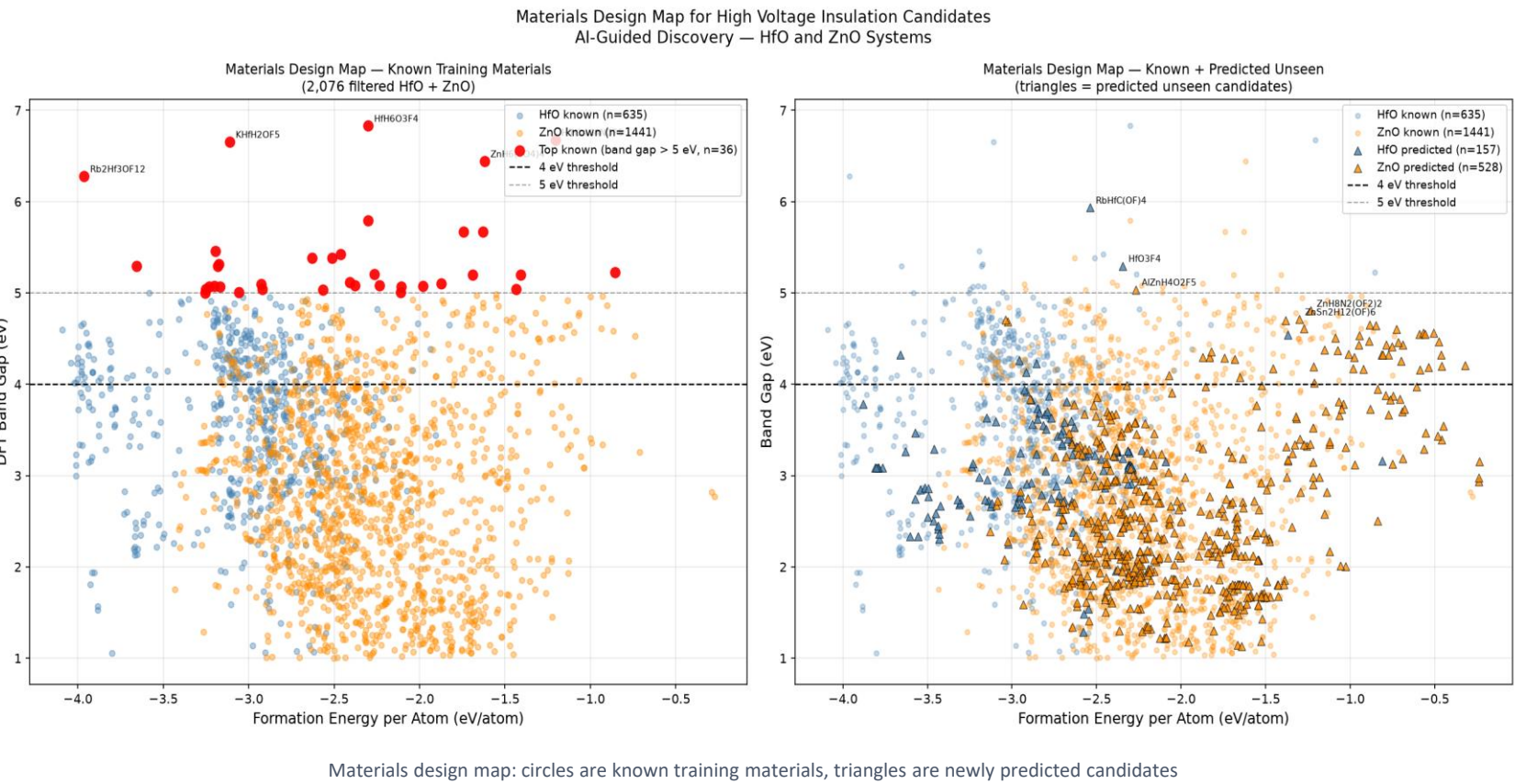
Systems

HfO and ZnO

What does this mean?

- Points close to the red diagonal line indicate accurate predictions
- An R^2 of 0.845 means the model explains 84.5% of all band gap variation
- An MAE of 0.269 eV means predictions are within a quarter of an electron volt on average

Results: Discovering New Candidates



Top Candidates

685 unseen materials screened

55 predicted above 4 eV

RbHfC(OF)4

5.94 eV | HfO

HfO3F4

5.30 eV | HfO

AlZnH4O2F5

5.04 eV | ZnO

Fluorine compounds dominate the top candidates

How to read this map:

- Higher up means larger band gap which means a better insulator. Further left means more stable and easier to synthesise.
- The most promising candidates appear in the top region, above the 4 eV dashed threshold line.

Bringing It All Together

The Question

- Can AI predict which oxide materials make the best electrical insulators?
- The answer is yes, with an R2 of 0.845 and good generalisation to new chemical systems

How We Approached It

- Downloaded and cleaned 4,766 materials down to 2,076 quality training examples
- Grouping by space group and isolating P1 materials was the key breakthrough

What We Found

- Space group number is the most important predictor, accounting for 14% of feature importance
- 55 new candidates were identified above 4 eV, with the top candidate RbHfC(OF)_4 at 5.94 eV

Next Steps

- DFT verification of the top predicted candidates and adding AIO data to the training set
- Experimental synthesis of the most stable high band gap candidates

Acknowledgements

Dr. Xue Yong and Dr. Bengt Tegner

Department of EEE and Research IT, University of Liverpool

Project supervision, scientific direction and computing guidance

N8 CIR

N8 Centre of Excellence in Computationally Intensive Research

Funding and organising this internship programme

Barkla HPC Team

University of Liverpool Research Computing

High performance computing resources that made this project possible