Project Proposal

Title: Multimodal ML Pipeline for Biocompatible Transparent Semiconductor Discovery

Group Members (Team 2):
Amala Vellappillil Biju, Brent Thorne, Dare Drouin, Jehad Saif, Michael Amaraut, Rachana Baskar

Abstract:

This project develops a multimodal machine learning pipeline to accelerate the discovery of biocompatible transparent semiconductors. By integrating structural, compositional, and electronic features from high-throughput DFT datasets, we aim to predict and rank candidate materials with tunable bandwidths, high optical transparency, and low biological toxicity. Our approach combines exploratory data analysis, graph-based representations, and multitask neural networks to model key properties such as band gap, dielectric constant, and formation energy. We incorporate biocompatibility screening and interpretability tools to ensure safe and actionable predictions. The pipeline supports applications in optoelectronics, biosensors, and sustainable materials, offering commercial and societal impact through reduced R&D costs and improved material safety.

Project Description:

We propose a multimodal ML pipeline that:

- Uses exploratory data analysis (EDA) to identify correlations between features and targets
- Trains graph neural networks (GNNs) on crystal and molecular structures
- Predicts three key targets: transparency, semiconductor behavior, and biocompatibility
- Scores materials using a composite metric: transparency + conductivity + biocompatibility

Problem Statement and Scope:

Transparent semiconductors are essential for displays, solar panels, and biosensors. However, discovering new candidates is computationally expensive and often overlooks biocompatibility.
We aim to:

- Identify materials with tunable electronic bandwidths (2.5-4.0 eV)
- Ensure stability and safety for biomedical and environmental use
- Reduce the search space for DFT validation using ML-based filters

Data Sources:

- NOMAD2018: space group, atoms in cell, relative composition, lattice vectors and angles
- JARVIS-DFT: band gaps, formation energies, dielectric constants
- Materials Project: crystal structures, electronic properties
- OQMD: compositional and thermodynamic data

Features and Representation:

- Structural: lattice parameters, symmetry, coordination
- Electronic: band gap, DOS, dielectric constant
- Compositional: atomic number, electronegativity, oxidation states
- Spectral: UV/Vis, Raman embeddings
- Biocompatibility: toxicity flags, degradation profile, surface energy
- OQMD: compositional and thermodynamic data