The accelerating integration of artificial intelligence (AI) into materials science promises to shorten the discovery cycle for energy-relevant compounds.

Paolo DE ANGELIS - *Polytechnic of Turin*

Here, we introduce Energy-GNoME, an open and "living" repository containing over 33,000 candidate materials extracted from the larger DeepMind GNoME dataset of more than 380,000 predicted-stable crystals. The protocol leverages and addresses a common issue in AI applications for materials science: most specialized materials databases are built on human expertise, which, when used for AI training, results in cross-domain bias and unreliable out-of-distribution predictions. We introduce a novel protocol based on supervised gradient-boosted decision trees (GBDT) that utilizes these biases to identify the domains of applicability in the high-dimensional feature space. In this way, the ML classifiers learn from human experience and work as "AI-experts" to identify the material with potential similar functionality, and thus, property regressors are expected to yield robust and reliable results.

The proposed workflow rapidly screens relevant to thermoelectric, perovskite, and battery cathode materials, and estimates target properties such as the thermoelectric figure of merit (zT), electronic band gap (Eg), and average reduction voltage (ΔV) using Graph Neural Networks (GNNs) trained ad hoc.

The resulting Energy-GNoME database currently includes 13,069 screened thermoelectric candidates, 4,259 perovskites suitable for photovoltaics applications, and 20,454 potential cathodes for lithium and *post*-lithium technologies. Furthermore, with shared Python libraries and an interactive dashboard, we aim to open the database to continue integration and improvement for experimental and computational community feedback.