npj | computational materials

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

Article

https://doi.org/10.1038/s41524-026-02013-0

Accurate screening of functional materials with machine-learning potential and transfer-learned regressions: Heusler alloy benchmark

Enda Xiao[1] ✉ & **Terumasa Tadano**[1,2] ✉

We present a machine learning-accelerated high-throughput (HTP) workflow for the discovery of functional materials. As a test case, quaternary and all-*d* Heusler compounds were screened for stable compounds with large magnetocrystalline anisotropy energy (E_{aniso}). Structure optimization and evaluation of formation energy and energy above the convex hull were performed using the eSEN-30M-OAM interatomic potential, while local magnetic moments, phonon stability, magnetic stability, and E_{aniso} were predicted by eSEN models trained on our DxMag Heusler database. A frozen transfer learning strategy was employed to improve accuracy. Candidate compounds identified by the ML-HTP workflow were validated with density functional theory, confirming high predictive precision. We also benchmark the performance of different uMLIPs, discuss the fidelity of local magnetic moment prediction, and demonstrate generalization to unseen elements via transfer learning from a universal interatomic potential.

The high-throughput (HTP) screening approach has emerged as a powerful strategy for accelerating the discovery of novel materials by systematically exploring large chemical spaces computationally or experimentally[1–3]. The density functional theory (DFT)-based HTP workflows have been widely employed to identify materials with target properties. However, as the search space increases, the associated computational cost becomes unsustainable, often restricting screening efforts to a manageable subspace[4,5]. To address this bottleneck, machine learning (ML) offers a promising route by drastically reducing computational costs. In this work, we demonstrate the robust integration of state-of-the-art ML techniques into the HTP workflow (ML-HTP) through a practical case study focused on screening quaternary and all-*d* Heusler compounds for stable candidates with strong magnetic anisotropy energy (E_{aniso}).

Initial realizations of the ML-HTP paradigm relied on ML models that utilize compositional descriptors as input features[6–9]. These models directly map chemical formulas to target properties, offering efficiency and simplicity. However, composition-based models are inherently unable to distinguish compounds with identical stoichiometry but different atomic arrangements. One workaround involves assigning layer indices to atomic sites, but this approach fixes the number of sites and can yield inconsistent predictions for symmetry-equivalent structures[10,11]. Crystal graph-based models do not have such drawbacks since they explicitly incorporate structure information as input, capturing structure-property relationships more accurately[12,13]. However, crystal graph-based models introduce an additional computational step, as geometry optimization must precede property prediction.

Although a single DFT optimization typically requires only a few minutes, the cumulative cost of screening a large number of candidate compounds becomes prohibitively expensive, particularly for magnetic systems where multiple magnetic configurations must be considered. A promising solution lies in leveraging universal machine learning interatomic potentials (uMLIPs), which can accelerate structure optimization by several orders of magnitude relative to DFT. The uMLIP field has witnessed rapid advancements in recent years, with many crystal graph-based models proposed. Despite the conceptual appeal, reliable and robust uMLIP-based structure optimization has only become practical recently, especially following the release of the large-scale and diverse Meta Open Materials 2024 Dataset (OMat24) training dataset in 2024[14]. This is demonstrated in the current work by benchmarking several uMLIPs, ranging from early-stage implementations to state-of-the-art developments.

With optimized structures, properties can be predicted using a machine learning regression model (MLRM). This approach substantially

---

[1]Research Center for Magnetic and Spintronic Materials, National Institute for Materials Science, Tsukuba, Ibaraki, Japan. [2]Digital Transformation Initiative Center for Magnetic Materials (DXMag), National Institute for Materials Science, Tsukuba, Ibaraki, Japan. ✉e-mail: Xiao.Enda@nims.go.jp; Tadano.Terumasa@nims.go.jp

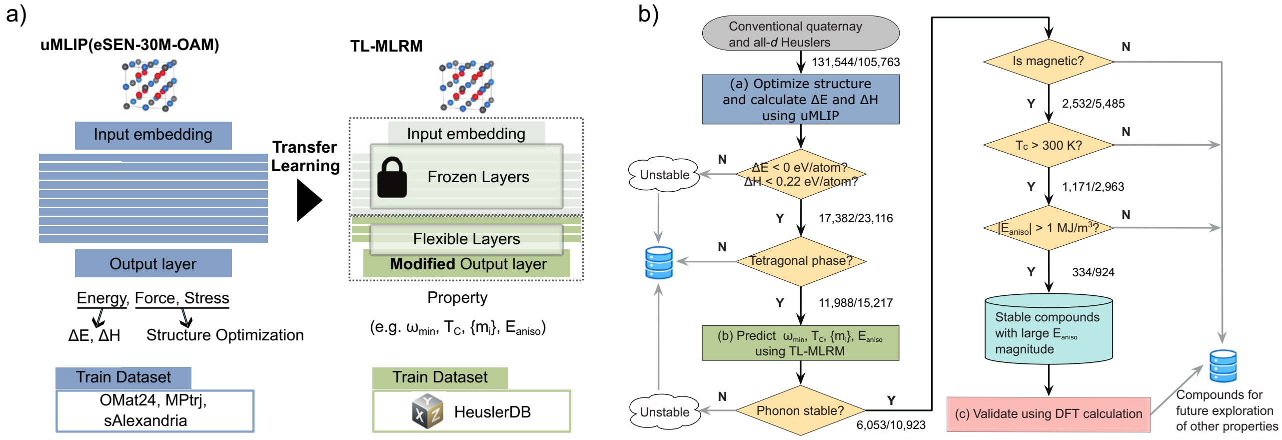


Fig. 1 | Frozen transfer learning overview and ML-HTP workflow. a Schematic of the development of the MLRM via frozen transfer learning, using eSEN-30M-OAM uMLIP as the base model. The uMLIP is used to perform structure optimization, formation energy calculation, and convex hull distance evaluation. The MLRM predicts properties from structures. **b** Workflow of the case study which identified stable conventional quaternary and all-*d* Heusler compounds exhibiting strong E_{aniso}. Counts of quaternary and all-*d* compounds at each stage are reported as quaternary/all-*d*.

reduces computational cost, particularly for properties that are expensive to compute via DFT, such as phonon spectra, conductivity, magnetic critical temperature (T_c), and E_{aniso}. However, training accurate MLRMs typically requires large, high-quality datasets. To overcome this challenge, transfer learning (TL) techniques can be employed to adapt pretrained ML models to new tasks[15–17]. TL leverages models that have already learned generalizable representations from extensive datasets and fine-tunes them using smaller, task-specific datasets. This strategy enhances predictive accuracy while substantially reducing data requirements.

As a case study, we conducted a ML-HTP screening on Heusler compounds, which have garnered significant attention due to their diverse functional properties, technological potential, and structure complexity[18]. Numerous DFT-HTP screenings have been carried out to identify candidates with different desirable properties[1,4,5,19–21]. In our previous work, we developed DXMag Computational Heusler Database (HeuslerDB), a comprehensive database encompassing nearly all conventional ternary Heusler compounds. The present study significantly extended the search space to include quaternary and all-*d* Heusler compounds, targeting stability and E_{aniso} as key screening criteria. In earlier DFT-HTP studies of Heusler compounds, 10 candidates with large E_{aniso} were identified out of 286 selected compositions, and 15 among 29,784 Co-based structures[4,5]. This low yield underscores the rarity of such materials and highlights the difficulty of this search problem, making it a stringent test case for ML-HTP approaches.

Previous studies attempted ML approaches to E_{aniso} in systems outside the Heusler family. They employed early ML models, such as crystal graph convolutional neural networks (CGCNN) and compositional-descriptor models, to predict E_{aniso} in Fe-Co-N alloys and physically motivated 2D materials selected based on domain knowledge, working on the order of hundreds of compounds[22–24]. Here, we employ state-of-the-art ML methods to extend the scope to hundreds of thousands of compounds with improved accuracy and practicality.

In this work, we demonstrate the use of uMLIP and TL-MLRMs as drop-in replacements for DFT structure optimization and property evaluation within the HTP framework, as illustrated in Fig. 1a. As a practical application, we employed this approach to identify conventional quaternary and all-*d* Heusler compounds with large E_{aniso}, while simultaneously satisfying thermodynamic, dynamic, and magnetic stability. Structure optimization and thermodynamic stability evaluation were performed using the eSEN-30M-OAM uMLIP[25]. The following predictions of local magnetic moment ($\{m_i\}$), minimum phonon frequency (ω_{min}), T_c, and E_{aniso} were performed using MLRMs. The MLRMs were trained via frozen transfer learning, using eSEN-30M-OAM uMLIP as the base model and fine-tuned using HeuslerDB data and newly computed data. ML-selected candidates were validated through DFT calculations to demonstrate the significant reliability of this ML-HTP approach. We further examine key factors that influence the performance of such ML-HTP workflows, including the accuracy of uMLIP-based structure optimization, the magnetic configuration prediction, the performance of the frozen transfer learning technique and generalization to unseen elements. The underlying code for the ML-HTP workflow is made available as open-source packages MLIP High-throughput Optimization and Thermodynamics (MLIP-HOT) and MLIP Frozen Transfer Learning (MLIP-FTL), which can be found on our group's website and Git repository.

Results

To accumulate data for training E_{aniso} MLRM, we first computed the E_{aniso} of conventional ternary Heusler compounds within HeuslerDB using DFT. The E_{aniso} of some Heusler compounds was reported in the previous work and the agreement between our DFT results and previous work is demonstrated in Fig. S1[4,5]. Among all conventional ternary Heusler compounds, 2190 (7.9 %) exhibit an E_{aniso} magnitude greater than 1 MJ/m^3. When further screened for thermodynamic, dynamical, and magnetic stability, only 135 compounds (0.5 %) meet both the high E_{aniso} and stability criteria, which are presentated in Table S2. These low percentages underscore the difficulty of identifying stable, high E_{aniso} compounds and highlight the need for more efficient screening methods, as demonstrated in current work as a case study. The ML-HTP workflow for this case study is summarized in Fig. 1(b); detailed computational procedures are provided in Sec. "Methods".

For conventional quaternary compounds, we enumerated all combinations where *X* and *Y* are transition metals from the *d*-block (excluding Tc and Hg), and *Z* is a main-group element from groups 13, 14, or 15 of the *p*-block. In addition, La and Lu were included for *X* and *Y* because their 4*f* orbitals are either empty or fully filled. This exhaustive enumeration, accounting for symmetry constraints, yielded 131,544 unique compositions. For the all-*d* Heuslers, we extended the screening space to include *d*-block transition metals together with La and Lu across all four sites (X_1, X_2, *Y*, and *Z*), resulting in a separate set of 105,763 unique compositions. A schematic of the screened chemical space is presented in Supplementary Information as Fig. S2.

Validation of ML-HTP selected candidates

Using eSEN-30M-OAM uMLIP in combination with MLRMs, we screened almost all conventional quaternary and all-*d* Heusler spaces for stable compounds with high E_{aniso}. As a result, 334 and 924 candidates were found,

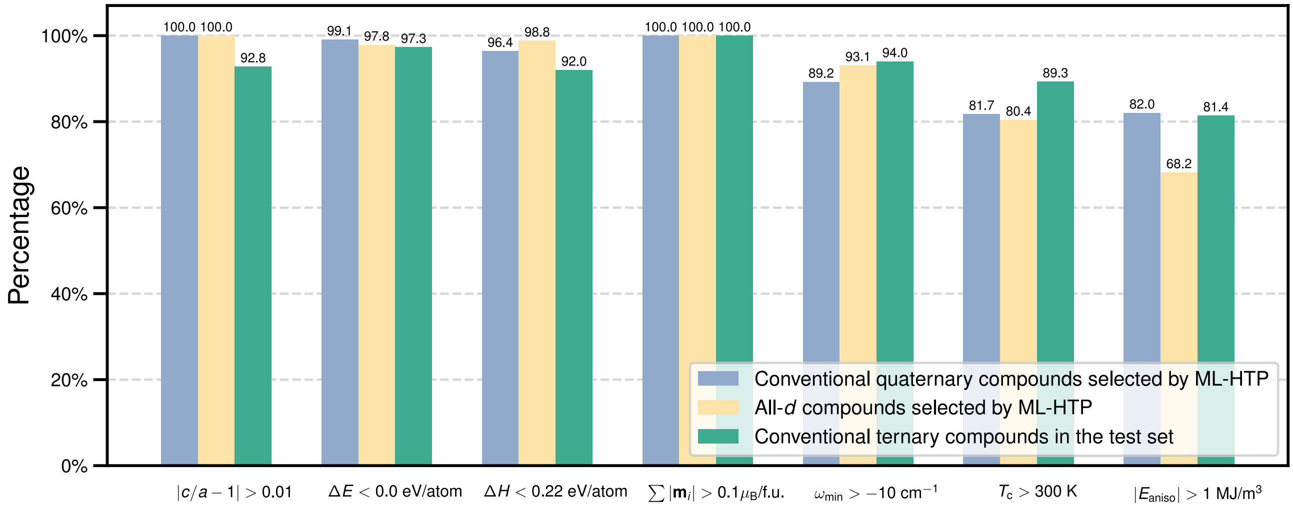


Fig. 2 | DFT validation summary of ML-HTP selected compounds. For ML-selected candidate lists of conventional quaternary (334) and all-d (924) Heusler compounds, the percentages that DFT results satisfy the screening criteria (i.e., the ML-HTP precision) are shown as blue and yellow bars. For comparison, the precision of the ML models measured on the test set of conventional ternary compounds is also shown as green bars. The test set size for c/a ratio, formation energy (ΔE), and energy above the convex hull (ΔH) is 10,000 and for $\{m_i\}$, $\omega_{\min}$, T_c, and E_{aniso} is 10% of the dataset size shown in Table 1.

respectively. To evaluate the reliability of this ML workflow, all candidates were validated using DFT calculations. The results are summarized in Fig. 2 and detailed data are provided in Tables S3 and S4. The percentages that DFT results satisfy the screening criteria (i.e., precision) are shown as blue and yellow bars for conventional quaternary and all-d Heusler compounds, respectively. The selection criteria include c/a ratio, ΔE, ΔH, $\{m_i\}$, $\omega_{\min}$, T_c, and E_{aniso}. For comparison, the precisions of the ML models, measured on the test set of conventional ternary compounds, are also provided as green bars.

In ML-HTP, the c/a ratio, ΔE, and ΔH are obtained from the optimized structure and the corresponding energy by eSEN-30M-OAM uMLIP. During the structure optimization process, relaxations were performed starting from multiple initial structures, and the relaxed structure with the lowest energy was selected. Since a non-zero E_{aniso} requires the Heusler compound to adopt a tetragonal phase, we applied a screening threshold of $|c/a-1|>0.01$ to identify tetragonality. Notably, all ML-selected candidates remained tetragonal in DFT validation, confirming that eSEN-30M-OAM reliably distinguishes between cubic and tetragonal phases. A more detailed discussion of structure optimization and performance for lattice parameters a, c, and the c/a ratio prediction, along with performance of other uMLIPs, is provided in Sec. "uMLIP optimization performance".

Using energies of candidate compounds, elements, and competing phases predicted by eSEN-30M-OAM, the ΔE and ΔH were calculated. The criteria of ΔE < 0 eV/atom and ΔH < 0.22 eV/atom were employed to identify thermodynamically stable candidates, following the thresholds established in our previous DFT-HTP study[21]. Among ML-selected candidates, 99.1% of conventional quaternary and 97.8% of all-d Heusler compounds were validated to have $\Delta E_{\text{DFT}}<0$ eV/atom. Similarly, 96.4% (quaternary) and 98.8% (all-d) of the compounds were found to have $\Delta H_{\text{DFT}}<0.22$ eV/atom. These high validation rates demonstrate that state-of-the-art uMLIP, such as eSEN-30M-OAM, can reliably assess thermodynamic stability.

It is important to note that the eSEN-30M-OAM uMLIP used is a general-purpose, pretrained model without any fine-tuning specific to the Heusler chemical space. Thus, these results highlight its strong generalization, making it an effective drop-in replacement for DFT-based optimization and thermodynamic stability assessment in the HTP workflow. Notably, the model achieves strong performance on the studied magnetic systems, despite not explicitly incorporating magnetic moments into its architecture or training. This strong performance and generalization are particularly valuable for the initial screening of novel material systems, where uMLIP can greatly reduce the search space by rapidly and reliably estimating optimized structures and thermodynamic stability.

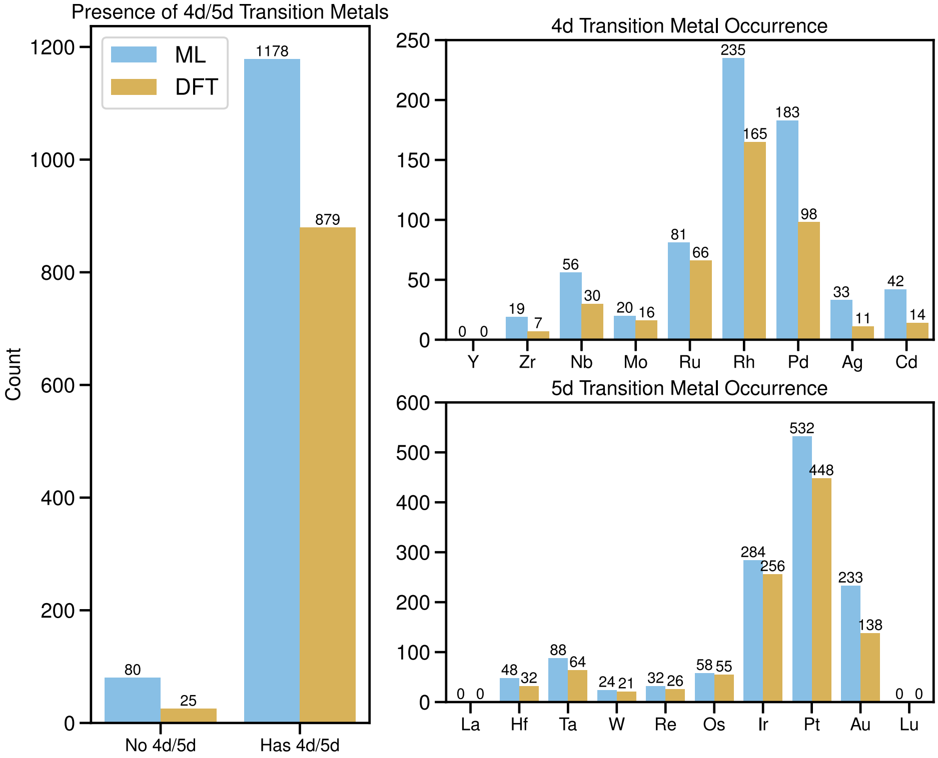
The properties $\{m_i\}$, $\omega_{\min}$, T_c, and E_{aniso} were predicted using MLRMs applied to uMLIP-optimized structures. Because E_{aniso} is a magnetic property, $\{m_i\}$ was predicted, and a screening threshold of $\sum|\mathbf{m}_i|>0.1\ \mu_B$/f.u. was applied to identify magnetic compounds. DFT validations confirmed all ML-selected candidates to be magnetic. Moreover, the $\{m_i\}$ MLRM accurately predicts both the magnitude and sign of local moments, as discussed in detail in Sec. "Prediction of local magnetic moment". Magnetic system identification is a critical yet computationally demanding step in DFT-HTP, as multiple initial $\{m_i\}$ values must be tested, and the low fraction of magnetic systems in some material families can lead to substantial wasted computation. By incorporating the $\{m_i\}$ MLRM method as a pre-screening step, the search space can be substantially reduced.

To identify compounds with dynamic stability, magnetic stability, and large E_{aniso}, we applied the criteria $\omega_{\min}>-10$ cm^{-1}, $T_c>300$ K, and $|E_{\text{aniso}}|>1$ MJ/m^3. Among the ML-selected candidates, 89.2% of conventional quaternary and 93.1% of all-d Heusler compounds were validated to have $\omega_{\min}$ above -10 cm^{-1}. For magnetic stability, 81.7% of conventional quaternary and 80.4% of all-d Heusler compounds were validated to have T_c above 300 K. For the target property E_{aniso}, the validation rates were 82.0% and 68.2%, respectively. To assess the sensitivity to the criteria values, we also evaluated the precision using a range of more stringent thresholds. The results, summarized in Fig. S3, show that the selection precision is not sensitive to the threshold values in the investigated range.

The MLRMs were trained exclusively using train set of conventional ternary Heusler compounds, yet were applied to evaluate quaternary and all-d compositions. By comparing validation rates to the precisions calculated using the test set of conventional ternary compounds, the MLRMs for $\{m_i\}$ and T_c generalize well to these expanded chemical spaces. In contrast, the E_{aniso} model exhibits lower performance for all-d compounds. This discrepancy can be attributed to the difference in chemical environments; while conventional quaternary compounds retain Z-site elements from the p-block—consistent with the training set—all-d compounds introduce Z elements from the d-block, which were absent during training. E_{aniso} is a sensitive property, influenced by subtle details of the electronic structure, and thus more susceptible to domain shifts than $\{m_i\}$ or T_c.

Relaxing the screening thresholds increases the pool of candidate compounds and might capture promising cases missed initially at the expense of more false positives. Additionally, the curated list of compounds that meet stability criteria and further magnetic system criterion serves as an

Fig. 3 | Distribution of ML-selected compounds on elements contained. Distribution of ML-selected candidate compounds based on whether 4d or 5d elements are present and distribution over 4d and 5d elements contained. The distribution of DFT validated strong E_{aniso} candidates is also shown.



efficient starting point for the investigation of other functional properties. For readers interested in exploring an expanded candidate list, the full set of ML-predicted data for 131,544 conventional quaternary and 105,763 all-*d* Heusler compounds will be accessible through HeuslerDB.

Distribution of strong E_{aniso} candidates

In addition to validating predictive precision, it is essential to determine whether the ML models capture known physical trends. A well-established insight is that compounds containing 4*d* and 5*d* elements typically exhibit larger E_{aniso} than those composed of 3*d* elements, owing to the stronger spin–orbit coupling associated with the heavier atomic nuclei of 4*d* and 5*d* elements. This behavior is clearly reproduced in the ML-HTP results, as shown in Fig. 3, which presents the distribution of candidate compounds according to the presence of 4*d*/5*d* elements. The figure further highlights the specific 4*d*/5*d* elements that appear in the identified compounds. For comparison, Fig. 3 also includes the distribution of compounds with DFT-calculated E_{aniso} magnitudes exceeding 1 MJ/m^3.

uMLIP optimization performance

In recent years, uMLIP has advanced rapidly, with numerous new models proposed and trained. To identify the most suitable model for our screening workflow, we benchmarked representative uMLIP models developed since 2023. These include ALIGNN-FF, CHGNet, SevenNet-l3i5, SevenNet-MF-ompa, HIENet, MatterSim-v1, eqV2-S-OAM, eqV2-M-OAM, eqV2-L-OAM, and eSEN-30M-OAM[14,25–30]. The evaluation focused on structure optimization for 10,000 conventional ternary compounds randomly selected from the ground states in HeuslerDB. To identify the global minimum, 14 initial structures were generated by applying strain to the conventional cell (two formula units) and converting it to the primitive cell (one formula unit). Specifically, the *a*, *b*, and *c* axes were uniformly scaled by ± 10% and ± 30%, or the *c*-axis alone was varied by ±10%, ±20%, ±30%, ±40%, and ±50%. The lowest-energy structure from these relaxations was selected as the predicted ground state. Convergence tests for all evaluated models are shown in Fig. S4, demonstrating that the selected ground states are well-converged.

The performance of structure optimization was assessed by comparing the predicted lattice constants *a* and *c*, and the resulting *c*/*a* ratio, with corresponding DFT values. The results are summarized in Fig. 4. The relative error (RE) is defined as the maximum of $\left|\frac{x_{\mathrm{ML}}}{x_{\mathrm{DFT}}}-1\right|$ and $\left|\frac{x_{\mathrm{DFT}}}{x_{\mathrm{ML}}}-1\right|$, where x_{ML} and x_{DFT} denote the values predicted by the ML and DFT, respectively. We report the fractions of compounds within 5% and 1% RE tolerances. Among the evaluated models, the eSEN-30M-OAM and eqV2 models achieved the highest accuracy at 5% RE, with eSEN-30M-OAM showing slightly better performance at the stricter 1% RE threshold. A key distinction between the two models is the number of local minima encountered: eqV2-L-OAM identified 91,585 local minima, whereas eSEN-30M-OAM identified 32,606. Both counts are much greater than 10,000, confirming the presence of multiple local minima, but are still significantly less than 10,000 × 14. This suggests that many different initial distortions ultimately converge to the same local minimum. Importantly, eSEN-30M-OAM found substantially fewer local minima than eqV2-L-OAM. This difference is attributed to the smoother energy landscape of eSEN-30M-OAM. Convergence tests also demonstrate that eSEN-30M-OAM achieves convergence with fewer initial structures, resulting in lower computational cost for HTP purpose.

The predictive performance of total energy (E), formation energy (ΔE), and energy above the convex hull (ΔH) using uMLIP was assessed by comparing the uMLIP predictions with DFT values, using the absolute error (AE), $|x_{\mathrm{ML}} - x_{\mathrm{DFT}}|$. The fractions of compounds with AE below 0.01 and 0.05 eV/atom are shown in Fig. 4. Among the benchmarked models, eSEN-30M-OAM and the eqV2 variants showed the highest accuracy for ΔE and ΔH at the 0.05 eV/atom threshold, with eSEN-30M-OAM displaying a slight drop in accuracy at the more stringent 0.01 eV/atom level. Predicted total energies from uMLIP were found to be systematically lower than DFT values, reducing direct agreement; however, this offset also applies to elemental references and competing phases, so the relative quantities ΔE and ΔH remain in strong agreement with DFT. Given its robust performance in both structure optimization and thermodynamic stability, eSEN-30M-OAM was selected for integration into the ML-HTP workflow.

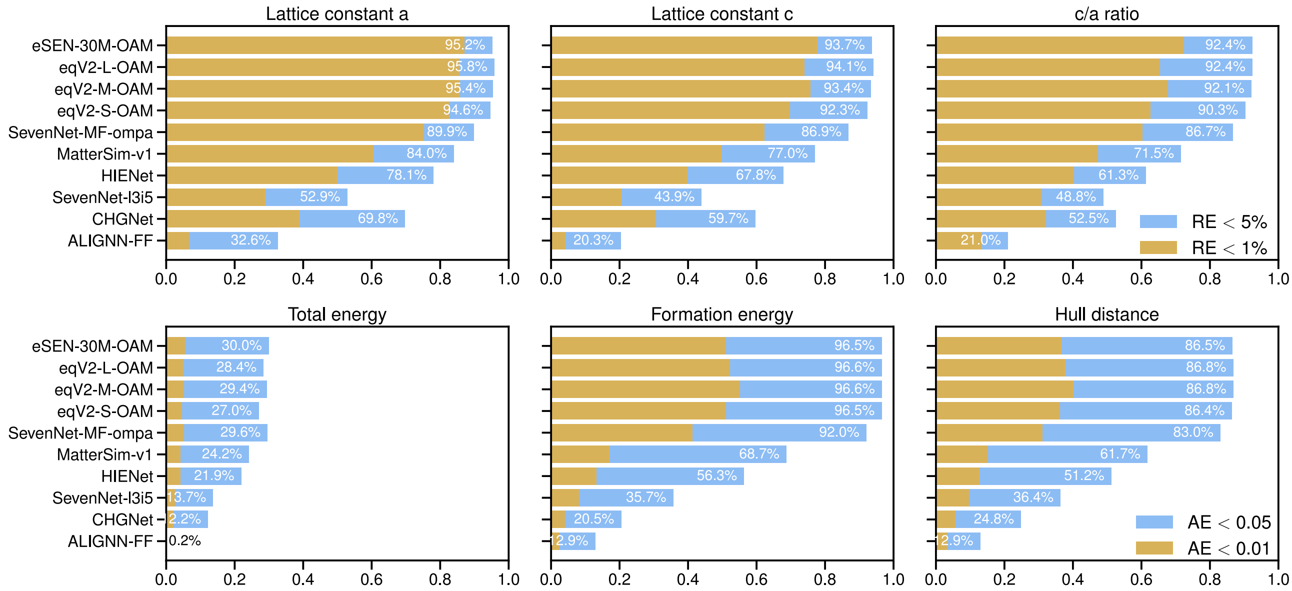


Fig. 4 | Benchmark of uMLIP performance. Lattice constants *a* and *c*, *c*/*a* ratio, total energy (*E*), formation energy (ΔE), and convex hull distance (ΔH) predicted by various uMLIP are benchmarked against DFT references. For each property, the fraction of compounds with predictions falling within specified relative error (RE) or absolute error (AE) thresholds is reported. Energetic quantities (E, ΔE, and ΔH) are expressed in eV/atom. The test set consists of 10,000 ground-state compounds randomly sampled from HeuslerDB.

Table 1 | Performance comparison of the eSEN uMLIP and MLRM with ALIGNN models and previously reported results

Property	Metric	eSEN-30M-OAM	ALIGNN-FF	Previous reports
a	R^2	0.994	0.128	0.80–0.94[a], 0.94[b], 0.987[c], 0.979[d]
	RMSE	0.023	0.330	0.11–0.12[e]
ΔE	R^2	0.995	0.453	0.80–0.88[a], 0.93[b], 0.982[c]
	RMSE	0.029	0.310	0.117[e]
ΔH	R^2	0.980	0.330	0.91[f], 0.969[c]

Property	Metric	eSEN MLRM	ALIGNN MLRM	Previous reports	Dataset size
$\{m_i\}$	R^2	0.989	—	—	27,864
m_{total}	R^2	0.986	0.904	0.75–0.89[a], 0.82[g], 0.927[c]	27,864
$\sum\lvert\mathbf{m}_i\rvert$	R^2	0.989	0.891	—	27,864
ω_{min}	R^2	0.750	0.734	—	8198
T_c	R^2	0.910	0.844	0.76[h], 0.73[i]	2106
	Accu.	0.910	—	0.73[h]	
E_{aniso}	R^2	0.680	0.592	—	6123

The size of dataset used for MLRM is also listed.

[a][8] Dataset size is about 1000,
[b][31] Dataset size is about 65,000,
[c][11] Dataset size is about 500,000 for *a* and m_{total}, and about 450,000 for ΔE and ΔH,
[d][32] Dataset size is 143,
[e][10] Dataset size is 16,272,
[f][33] Dataset size is 426,148,
[g][34] Dataset size is 1153,
[h][6] Dataset size is 408,
[i][35] Dataset size is 6500.

Improvements over existing approaches

Previous studies have reported the performance of composition-based models in predicting lattice constants, ΔE, and ΔH for cubic Heusler compounds. For comparison, we evaluated metrics of eSEN-30M-OAM on the cubic Heusler subset, with results summarized in Table 1. The R^2 score for the lattice constant *a* is 0.994, surpassing the previously reported ranges of 0.80–0.94 across different Heusler types and the values of 0.94, 0.979, and 0.987 in other works[8,11,31,32]. Similarly, the R^2 for ΔE reaches 0.995, outperforming earlier results of 0.80–0.88 and 0.93, 0.982[8,11,31]. The R^2 for ΔH is 0.98, exceeding prior values of 0.91 and 0.969[11,33]. The root mean squared errors (RMSE) for *a* and ΔE are 0.023 Å and 0.029 eV/atom, respectively, which are significantly lower than the 0.11–0.12 Å and 0.117 eV/atom reported in earlier work[10].

The MLRMs used in screening were trained on the data from HeuslerDB, supplemented with newly computed T_c and E_{aniso} values using optimized structures in HeuslerDB. Test set metrics are summarized in Table 1 and benchmarked against previously reported results. The MLRM for $\{m_i\}$ achieved an R^2 score of 0.989. For comparison with prior studies that used total magnetization (m_{total}) as the target property, our model yielded $R^2 = 0.986$ for m_{total}, exceeding earlier values of 0.75–0.89, 0.82, and 0.927[8,11,34]. For T_c, the model attained $R^2 = 0.91$ and classification accuracy of 0.91, both substantially higher than previously reported values of $R^2 = 0.76$ and 0.73 and accuracy of 0.73[6,35].

To the best of our knowledge, no previous study has predicted phonon stability or E_{aniso} of Heusler compounds using ML. The effectiveness of the ω_{min} and E_{aniso} models developed here is supported by the validation results in Sec. “Results”. While phonon stability could also be assessed using uMLIP combined with phonon calculation methods, our regressor-based approach is motivated by both efficiency and accuracy. Only the minimum frequency is needed for phonon stability assessment, thus, a regressor is sufficient and much faster than calculating the full spectrum. We also evaluated uMLIP +phonon using CHGNet, MatterSim, and eSEN-30M-OAM on a test set of 1000 conventional ternary compounds, and found stability/instability classification accuracies of 62.5%, 74.9%, and 80.2%, respectively, which are substantially below the regressor’s 93.6%. Notably, CHGNet misclassifies 67.2% of stable compounds as unstable, showing a strong tendency to underestimate stability, while eSEN weakly overestimates it and MatterSim exhibits a more balanced performance. Additional details, including example phonon spectra by uMLIPs and DFT, are in the Supplementary Information. The E_{aniso} model achieved an R^2 of 0.68, lower than those of $\{m_i\}$, ω_{min}, and T_c, highlighting the higher sensitivity and complexity of

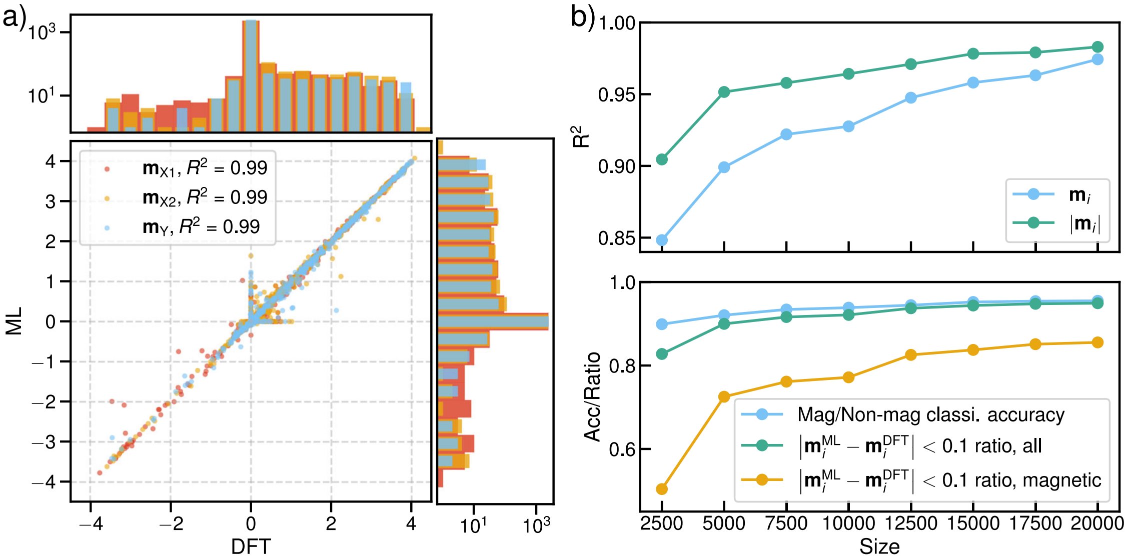


Fig. 5 | Local magnetic moment prediction performance. **a** Scatter plot comparing ML-predicted $\{m_i\}$ with DFT values for the test set. **b** Learning curves for $\{m_i\}$ prediction. The top panel shows R^2 scores for both local moments and their magnitudes. The bottom panel reports the magnetic/nonmagnetic classification accuracy, and the fraction of compounds with absolute prediction error below 0.1 μ_B for all compounds and for the magnetic subset.

E_{aniso} as a target property. Nevertheless, despite the reduced R^2, its classification accuracy remains satisfactory and sufficient for integration into the ML-HTP workflow.

To benchmark advances in ML techniques since 2023, we applied the ALIGNN-FF uMLIP and ALIGNN MLRM to identify conventional quaternary candidate compounds and evaluated the validation rates of strong E_{aniso} compounds[13,26]. In this test, the scalar quantity $\sum|\mathbf{m}_i|$ was used directly as the target property rather than being calculated from the $\{m_i\}$ prediction. The metrics of ALIGNN-FF and ALIGNN MLRM are summarized in Table 1. Using all screening thresholds, only 17 compounds qualified as candidates. To improve statistical robustness, we removed the phonon stability criterion, expanding the candidate list to 107 compounds, of which 26 (24.3%) exhibit $|E_{\mathrm{aniso}}| > 1$ MJ/m^3. While this yield is notably higher than the 7.9% obtained from direct DFT-HTP screening, it remains far below the 82.0% success rate achieved with the eSEN-based ML-HTP workflow. These results highlight the substantial improvements in screening precision enabled by the state-of-the-art eSEN model.

We further tested a hybrid workflow in which structure optimization was performed with eSEN-30M-OAM uMLIP, while property prediction was carried out using ALIGNN MLRMs. This approach yielded 276 candidate compounds, of which 149 (54.0%) were confirmed by DFT to exhibit strong E_{aniso}. The improved yield relative to ALIGNN-FF-based optimization underscores the critical importance of accurate structure optimization with eSEN-30M-OAM for enhancing ML-HTP screening. However, the yield still falls short of the 82.0% achieved by the fully eSEN-based workflow, indicating that progress in both the uMLIP and MLRM components is essential for maximal efficiency. We also tested the inverse hybrid configuration, using ALIGNN-FF uMLIP for structure optimization combined with eSEN MLRMs for property prediction. This workflow identified 243 candidates, of which only 76 (31.3%) were validated as strong E_{aniso} compounds. This marked reduction in performance highlights the pivotal role of selecting an accurate uMLIP for structure optimization.

Prediction of local magnetic moment

Since the goal of this study is to identify compounds with large E_{aniso}, it is first necessary to determine whether a compound is magnetic. Relying solely on the total magnetization is inadequate, as it cannot capture antiferromagnetic (AFM) or low-moment ferrimagnetic (FiM) compounds. To address this, we employed the total absolute magnetic moment, defined as $\sum_i|\mathbf{m}_i|$, where m_i denotes the local magnetic moment at atomic site i.

Local magnetic moments were predicted using an MLRM based on the eSEN architecture, trained to output the moment at each atomic site. Restricting to collinear configurations in which all moments are aligned along the z-axis, each moment is represented by a scalar whose sign encodes the direction, with $\ell = 0$ per site in the output head. To account for the z-direction ambiguity-where a configuration and its sign-inverted counterpart (i.e., all local moments flipped) are physically equivalent—we modified the loss function to compute losses for both the predicted $\{m_i\}$ and its sign-inverted counterpart, and take the smaller value as the loss. This ensures invariance under global spin inversion.

Figure 5a shows a scatter plot comparing $\{m_i\}$ from the MLRM and DFT for compounds in the test set, with histograms along the axes illustrating their distributions. Local moments at the atom site Z are omitted as they are nonmagnetic. Because 73.4% of the test compounds are nonmagnetic, both histograms exhibit a pronounced peak at zero. For magnetic systems, the global sign is adjusted so that the total magnetic moment is positive; since most magnetic compounds are ferromagnetic, positive moments dominate in the distribution. Nearly all points fall along the diagonal, and only 1.4% of points lie along either axis, demonstrating that the model accurately predicts both the magnitude and sign of local moments, and reliably distinguishes ferromagnetic (FM) and ferrimagnetic (FiM) systems. For the magnetic/nonmagnetic classification, we evaluated performance using receiver operating characteristic (ROC) and precision-recall (PR) curves, as shown in Fig. S6. The area under the curve (AUC) values are 0.98 and 0.97, respectively, indicating highly accurate classification.

Predicting $\{m_i\}$ is a common yet computationally demanding step in HTP studies of magnetic materials. The approach developed in this work achieves high accuracy in $\{m_i\}$ prediction and is readily transferable to other systems. A central question, however, is how many training compounds are required to reach satisfactory accuracy. To address this, we performed a learning-curve analysis by training the MLRM on progressively larger subsets of the dataset and evaluating performance on a fixed test set of 5000 compounds, of which 1486 are magnetic. For evaluation, local moments at the X_1, X_2, and Y sites were concatenated across samples into a single array, while the Z site was excluded since it is nonmagnetic in conventional Heusler compounds.

The learning curve is shown in Fig. 5b, illustrating how model performance improves with increasing training set size. In the top panel, the R^2 scores for both local moments and their magnitudes are presented. The gap

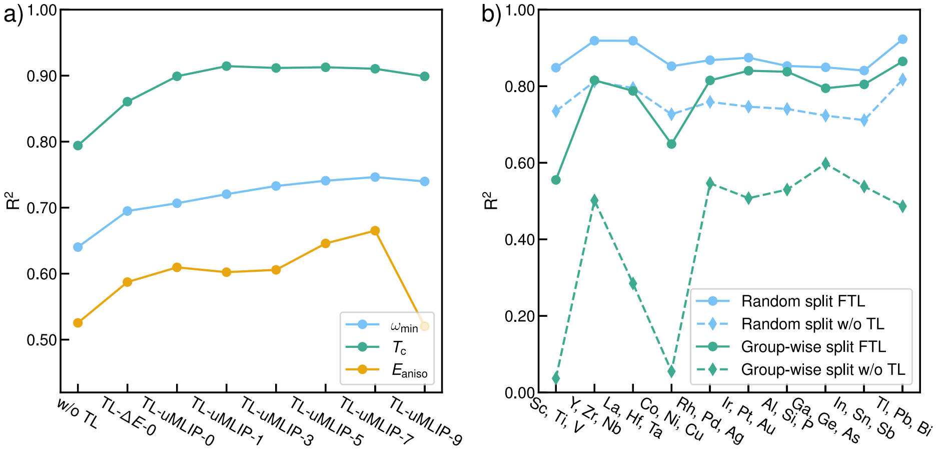


Fig. 6 | Frozen transfer learning performance and generalization. a R^2 of models initialized from the eSEN-30M-OAM uMLIP and trained with different numbers of frozen layers (denoted TL-uMLIP-*n*). In the $n = 0$ case, the embedding layer is also trainable. A model trained from scratch (denoted w/o TL) and a transfer learning variant initialized from a base model pre-trained on ΔE data from the HeuslerDB database (denoted TL-ΔE-0) are included for comparison. **b** Comparison of model performance on group-wise splits versus random splits. Models were evaluated on predicting T_c. Results are shown for frozen transfer learning (FTL) and models trained from scratch (w/o TL). For each test, three neighboring elements within the same period were used as the holdout elements. Six sets of d-block elements and four sets of p-block elements were tested.

between the two curves indicates that, while the model generally captures the magnitude accurately, it sometimes assigns the incorrect sign. For example, Mn_2ScGe with DFT-computed local moments {2.62, 3.03, −0.29, −0.10} μ_B is predicted as {−2.71, 3.04, 0.02, −0.02} μ_B by MLRM. The bottom panel reports two key metrics: (i) classification accuracy for identifying magnetic systems and (ii) the fraction of local moments with absolute error below 0.1 μ_B, evaluated across all compounds and within the magnetic subset. With 5000 training samples, the model achieves a classification accuracy of 0.92, and 90% of all compounds fall within the 0.1 μ_B error threshold. However, this fraction decreases to 72% when restricted to the magnetic compounds subset, indicating that the model identifies whether a site is magnetic with high reliability but remains less accurate in predicting exact $\{m_i\}$ values. Increasing the training set to 125,000 samples improves this fraction to 82%, while relaxing the threshold to 0.2 μ_B further raises it to 92%. Although performance improves with larger datasets, the gains become progressively smaller. These results highlight the critical role of dataset size in improving $\{m_i\}$ accuracy and inform the selection of training size in future work focusing on other magnetic systems.

Frozen transfer learning for MLRM construction

To improve the performance of the MLRM, we employed a frozen transfer learning strategy using the eSEN-30M-OAM uMLIP as the base model. The eSEN-30M-OAM uMLIP was trained on the OMat, MPtrj, and sAlexandria databases, providing a comprehensive training set spanning the periodic table[14,27,36–38]. Through this training, the embedding and the first several layers learn general chemical and structure representations. To leverage this, we transferred the parameters of the embedding and the first several layers into our MLRM and kept them fixed (frozen layers), while only updating the remaining layers and the output layer (flexible layers). This approach is analogous to a recent work in which the initial layers of ORB, EqV2, or MACE uMLIP were used to generate feature vectors that were subsequently passed to property prediction models such as MODNet, XGBoost, and MLP[39].

The eSEN-30M-OAM model consists of 10 layers. Figure 6a shows the R^2 scores of models trained with different numbers of frozen layers, denoted TL-uMLIP-*n*. In the $n = 0$ case, the embedding layer is also flexible. Results for $\omega_{\min}$, T_c, and E_{aniso} are presented. Model performance improves as the number of frozen layers increases up to $n = 7$, after which it declines when more layers are frozen. This trend reflects the balance between transferring knowledge from the base model and maintaining sufficient flexibility to adapt to the new task. Based on this analysis, TL-uMLIP-7 was used in the ML-HTP case study for $\omega_{\min}$, T_c, E_{aniso}, and $\{m_i\}$ MLRM. For comparison, a model trained from scratch (w/o TL) is included, which yields lower R^2 scores and highlights the benefit of transfer learning. We also tested a variant initialized from a base model pre-trained on formation energy (ΔE) data from the HeuslerDB database (denoted TL-ΔE-0). TL-uMLIP-0 outperforms TL-ΔE-0, indicating that initialization from the eSEN-30M-OAM model, trained on a comprehensive dataset, is more advantageous.

Besides improving overall performance, transfer learning from uMLIP can significantly enhance model generalization to unseen elements. To demonstrate this, we evaluated model performance using group-wise split analysis. For each group-wise split test, three neighboring elements within the same period were used as the holdout elements. Any compound containing at least one of the holdout elements was reserved as the holdout test data; the remaining compounds were split into train/val with a 9:1 ratio. This setup simulates realistic scenarios where models are applied to material systems containing elements not present in the training data, a typical domain shift in materials science. For direct comparison, we also measured performance using train/val/test sets created using random splits while keeping the counts of train/val/test the same as in the group-wise split. We tested six sets of d-block elements and four sets of p-block elements as holdout sets. We used T_c as the target property for efficiency. Results obtained using frozen transfer learning and models trained from scratch are shown in Fig. 6b.

Using FTL, performance on group-wise splits is generally lower than on random splits, indicating that the model performance on unseen elements is reduced. The drop in performance for p-block holdouts is smaller, which is consistent with the fact that p-block elements are typically nonmagnetic in these compounds and thus less directly related to the target property T_c. For group-wise split tests with models trained from scratch, a much larger performance decrease is observed. This occurs because FTL transfers the general chemical representation learned by uMLIP to property regressors, whereas models trained from scratch do not benefit from this prior knowledge. This highlights the advantage of transfer learning from uMLIP for improving generalization to unseen elements. Variation in random-split performance is due to different train set sizes since the number of holdout compounds varies with the chosen holdout elements.

It should be noted that TL is not always effective. It can be limited when the representations learned by the base model do not generalize to the target domain. For example, source and target may differ substantially in material class (e.g., bulk versus surface or molecular systems), or downstream task concerns different physics (e.g., static energetics versus dynamic properties)[15,40,41]. Furthermore, a recent work, which transfers from a model trained using computational data to learn experimental data, revealed that the error of the transfer-learned model decreases according to a power-law as the size of the computational data for the base model increases[42]. This suggests that transfer learning may also be less effective when the base model is trained on a small dataset, even if the base model performance within its own domain is good.

Discussion

We demonstrated the feasibility of combining uMLIPs and MLRMs for HTP screening. As a case study, we identified 334 conventional quaternary Heusler compounds and 924 all-*d* Heusler compounds that exhibit thermodynamic, dynamical, and magnetic stability, together with large E_{aniso}. The precision of this workflow was confirmed through DFT validation of the candidate list.

For other material systems, if a database for a subset is available, the same workflow can be applied to explore the remaining chemical space. For novel materials, uMLIP can be used directly to reduce the search space by filtering for thermodynamic stability, while target properties need to be computed for a representative subset of compounds to train MLRMs for ML-HTP screening of the broader space.

The MLRMs in this study involve a domain shift, as the train dataset does not contain quaternary or all-*d* Heusler compounds, yet the models are applied to these systems. While performance is satisfactory, further gains are possible using an active learning approach. Iteratively refining the models by selecting informative compounds can improve performance and help identify candidates missed by the current one-shot approach. In such an iterative framework, current DFT validation results can be added to the training set, forming the first iteration of model refinement. This iterative strategy naturally extends the present work, enabling more thorough exploration of chemical space while progressively improving model accuracy. Such an approach is especially valuable when the target material space differs substantially from the training data.

In the framework of this study, uMLIP perform the critical task of structure optimization, which is traditionally handled by DFT-derived methods, while MLRMs predict the properties of the optimized structures, tasks that are typically carried out using DFT or DFT-derived methods. Together, uMLIP and MLRMs enable a drop-in replacement for DFT in conventional HTP workflows. This replacement is general and can be readily extended to other properties, material classes, and DFT-based HTP pipelines, enabling accelerated HTP screening and discovery of new materials.

Methods

ML-HTP workflow

The ML-HTP workflow is schematically illustrated in Fig. 1(b) and described in detail below.

In step (a) of the ML-HTP workflow, the structures were optimized and the ΔE and ΔH were calculated using uMLIP. The initial lattice constant was estimated as the average value of known X_2YZ Heusler compounds in the HeuslerDB database that share two elemental species with the target composition. A conventional cell in cubic phase with this estimated lattice constant was then constructed. To generate initial structures, the lattice parameters of this cell were systematically varied: the *a*, *b*, and *c* were uniformly scaled by ± 10% or ± 30%, and, alternatively, the *c* alone was scaled by ±10% or ±30%. All generated structures were subsequently converted to the primitive unit cell and relaxed using the eSEN-30M-OAM uMLIP. The structure with the lowest energy after relaxation was identified as the ground state. The choice of eSEN-30M-OAM was motivated by its superior performance relative to other uMLIP, as discussed in Sec. "uMLIP optimization performance". The selection of initial structures was validated by a convergence test, which is provided in Fig. S4.

For a compound to be thermodynamically stable against decomposition into its constituent elements or competing phases, the formation energy must be negative ($\Delta E < 0$), and the distance to the convex hull must be zero ($\Delta H = 0$). In practice, however, metastable phases ($\Delta H > 0$) at 0 K may become stabilized at finite temperature[43]. Following our previous work, we adopt a practical stability criterion of $\Delta E < 0.0$ eV/atom and $\Delta H < 0.22$ eV/atom, which has been shown to effectively capture experimentally accessible compounds[21]. Using the energies of ground state candidates, constituent elements and competing phases predicted by eSEN-30M-OAM, we computed ΔE and ΔH to assess thermodynamic stability.

In step (b) of the workflow, the $\omega_{\min}$, $\{m_i\}$, T_c, and E_{aniso} were predicted with MLRMs trained on the HeuslerDB and additional computed data. The uMLIP-optimized structures were used as inputs. The construction of these MLRMs is described in Sec. "Prediction of local magnetic moment" and Sec. "Frozen transfer learning for MLRM construction". Compounds were classified as dynamically stable for $\omega_{\min} > -10\ \text{cm}^{-1}$, magnetic for $\sum_i |\boldsymbol{m}_i| > 0.1\mu_B$/f.u., magnetically stable for $T_c > 300$ K, and strong E_{aniso} candidates for $|E_{\text{aniso}}| > 1$ MJ/m^3. Tetragonal compounds satisfying all of these conditions were designated as promising stable materials with strong E_{aniso}.

In step (c) of workflow, the candidate list was validated with DFT calculations to assess the reliability of the ML-HTP workflow. Structure optimizations were performed starting from various initial spin configurations, consistent with our previous DFT-HTP work. For conventional Heusler compounds, the magnetic moments at the X_1, X_2, and *Y* sites were initialized in configurations where they were either parallel or antiparallel to each other. For all-*d* Heuslers, spin arrangements on all four sites were considered. To capture possible high-spin and low-spin states, two initial magnitudes of the local moments ($|\boldsymbol{m}_i| = 1$ and $4\ \mu_B$) were tested, along with a nonmagnetic configuration ($|\boldsymbol{m}_i| = 0$). The uMLIP-optimized structures served as the starting geometries. After structure relaxation, the ground state was identified by comparing total energies. For the resulting ground states, we computed ΔE, ΔH, E_{aniso}, phonon, and T_c using VASP, OQMD, ALAMODE, and SPRKKR[44–49].

Computational methods

The uMLIP-based structural optimizations were performed using the Atomic Simulation Environment (ASE) package[50]. The Fast Inertial Relaxation Engine (FIRE) optimizer was employed, with symmetry constraints enforced throughout the relaxation process[51]. To ensure consistency in reference energies for computing ΔE and ΔH, the elemental phases and competing phases were also optimized using the same uMLIP. Their initial geometries were taken from DFT-optimized structures in the Open Quantum Materials Database (OQMD) database[44,45].

The MLRMs were developed to predict $\omega_{\min}$, $\{m_i\}$, T_c, and E_{aniso}. To leverage prior knowledge, we employed a frozen transfer learning strategy, as illustrated in Fig. 1a. Each MLRM was initialized from the pre-trained eSEN-30M-OAM uMLIP, with the embedding layer and the first seven message-passing layers kept frozen, while the final three layers and the output layer were fine-tuned. This framework was implemented using a modified version of the FAIRChem package (v1)[52]. For the $\{m_i\}$ MLRM training, the loss function was adapted to address the global sign ambiguity of magnetic moments, as described in Sec. "Prediction of local magnetic moment". The training dataset consisted of HeuslerDB together with newly computed E_{aniso} and T_c values based on the optimized structures in HeuslerDB. The data were randomly partitioned into training, validation, and test sets in an 8:1:1 ratio. For $\{m_i\}$, we used all ground-state entries, resulting in 27,864 data points. The $\omega_{\min}$ data were available for all thermodynamically stable ground states, yielding 8198 entries. For T_c, 2106 data points were used, including 750 newly computed values. Since E_{aniso} data were not included in HeuslerDB, we calculated E_{aniso} for all magnetic tetragonal ground-state systems, obtaining 6123 entries.

DFT calculations were performed primarily with Vienna ab initio Simulation Package (VASP)[53,54], using the projector augmented wave (PAW) method and the generalized gradient approximation (GGA) with the Perdew-Burke-Ernzerhof (PBE) functional[55,56]. ΔE, ΔH, phonon, and T_c were computed using OQMD, ALAMODE, and SPRKKR following the methodology of our previous DFT-HTP study of ternary Heusler compounds[21,44–49]. The agreement of magnetic moments between VASP and SPRKKR is demonstrated in Fig. S7. Phonon calculations for the all-*d* compounds MnOsMnRe and MnReMnRu failed due to convergence issues in DFT, and these compounds were treated as unstable in the validation rate analysis. E_{aniso} was calculated as $E_{\text{aniso}} = E_{\perp} - E_{\parallel}$ using the force theorem[57–59]. Calculations were performed in the primitive cell with $\boldsymbol{k}$-meshes generated using Python Materials Genomics (pymatgen) at a density of 6000 Å^{-3}, and

Table 2 | Computational cost of each task in the HTP workflow, reported in node-minutes

	DFT (mean/ median/IQR)	ML training (mean)	ML prediction (mean)
Structure relaxation	5.6/3.8/3.0	—	0.12
$\{m_i\}$	—	876	<1e-03
Phonon stability	4542/4008/1442	258	<1e-03
T_c	70/66/14	78	<1e-03
E_{aniso}	142/133/34	192	<1e-03

DFT entries show per-job statistics (mean/median/IQR). ML entries show mean values. For structure relaxation, one job corresponds to a single relaxation from an initial distortion. For structure optimization, the cost per compound should be multiplied by the number of initial structures, which varies across material systems; here, we report the cost for a single structure relaxation. The uMLIP is pre-trained and used without fine-tuning, so its training time is excluded. DFT structure optimizations directly yield local magnetic moments, so no separate timing is reported.

the tetrahedron method with Blöchl corrections was applied[60,61]. Input generation, structure manipulation, and symmetry analysis were carried out using pymatgen, ASE, ASE2SPRKKR, and spglib[45,50,60,62,63].

Computational cost

The computational cost for each task in the HTP workflow is summarized in Table 2. DFT timings are based on validation runs for ML-selected candidates, while ML timings are taken from the ML-HTP screening. For DFT, we report per-job statistics such as mean, median, and interquartile range (IQR). uMLIP structure optimization and MLRM predictions were performed in batches, and mean values are reported as individual per-job timings are not available. MLRM training times correspond to the wall time required to train a single model on one GPU. DFT calculations except phonon were performed on dual-socket Intel Xeon Platinum 8268 (Cascade Lake, 24 cores per CPU, 2.9 GHz, 48 cores per node), phonon calculations were performed on Fujitsu A64FX processors (Armv8.2-A SVE 512 bit, 48 compute cores per CPU, 2.2 GHz), uMLIP structure optimizations were performed on dual-socket Intel Xeon Gold 6230 (Cascade Lake, 20 cores per CPU, 2.1 GHz, 40 cores per node), and MLRMs were trained and applied on NVIDIA RTX 6000 Ada GPUs. All DFT calculations, except for phonon calculations, were performed on 2 nodes; phonon calculations were performed on 6 nodes. The uMLIP structure optimizations were performed on 1 node, and MLRM training and prediction were performed on 1 GPU.

In the ML-HTP case study, uMLIP structure optimization, MLRM training, and MLRM prediction used 1,835 node-hours, 23 GPU-hours, and 4 GPU-hours, respectively. DFT calculations of ML-selected candidates consumed 1,645 node-hours for optimizations and 99,680 node-hours for property evaluations. Ignoring the node difference, the total ML-HTP cost is 103,160 node-hours plus 27 GPU-hours. A DFT-HTP workflow screening the same chemical space would need about 256,160 node-hours for structure optimizations and 18,802,625 node-hours for property evaluations, estimated using statistics reported in Table 2. This comparison shows an estimated speed-up of 185 times for ML-HTP with DFT validation, or 10^4 times if used without DFT validation at the end. Please note that these estimates are approximate, and waiting time for job execution, human time spent on debugging and workflow management are not included. The actual speed-ups also vary based on material systems, target properties, and computational resources.

Data availability

The ML-HTP candidate list and DFT validation results are included in the Supplementary Information as Tables S3 and S4. The complete set of all screened compounds, along with ML-predicted properties, will be made available through the HeuslerDB database at https://www.nims.go.jp/group/spintheory/.

Code availability

The developed packages MLIP-HOT and MLIP-FTL will be made available through the Spin Theory Group GitHub repository at https://github.com/nims-spin-theory and our group website at https://www.nims.go.jp/group/spintheory/.

Received: 2 September 2025; Accepted: 8 February 2026;
Published online: 19 February 2026

References

1. Sanvito, S. et al. Accelerated discovery of new magnets in the Heusler alloy family. *Sci. Adv.* **3**, e1602241 (2017).
2. Zhang, H. High-throughput design of magnetic materials. *Electron. Struct.* **3**, 033001 (2021).
3. Barwal, V. et al. Large magnetoresistance and high spin-transfer torque efficiency of $Co_2Mn_xFe_{1-x}Ge$ ($0 \leq x \leq 1$) Heusler alloy thin films obtained by high-throughput compositional optimization using combinatorially sputtered composition-gradient film. *APL Mater.* **12**, 111114 (2024).
4. Faleev, S. V. et al. Heusler compounds with perpendicular magnetic anisotropy and large tunneling magnetoresistance. *Phys. Rev. Mater.* **1**, 024402 (2017).
5. Hu, K. et al. High-throughput design of Co-based magnetic Heusler compounds. *Acta Mater.* **259**, 119255 (2023).
6. Hilgers, R., Wortmann, D. & Blügel, S. Machine Learning-based estimation and explainable artificial intelligence-supported interpretation of the critical temperature from magnetic ab initio Heusler alloys data. *Phys. Rev. Mater.* **9**, 044412 (2025).
7. Baigutlin, D. R., Sokolovskiy, V. V., Buchelnikov, V. D. & Taskaev, S. V. Machine learning algorithms for optimization of magnetocaloric effect in all-d-metal Heusler alloys. *J. Appl. Phys.* **136**, 183903 (2024).
8. Mitra, S., Ahmad, A., Biswas, S. & Kumar Das, A. A machine learning approach to predict the structural and magnetic properties of Heusler alloy families. *Comput. Mater. Sci.* **216**, 111836 (2023).
9. Liu, C. et al. Machine learning to predict quasicrystals from chemical compositions. *Adv. Mater.* **33**, 2102507 (2021).
10. Xie, R., Crivello, J.-C. & Barreteau, C. Screening new quaternary semiconductor Heusler compounds by machine-learning methods. *Chem. Mater.* **35**, 7615–7627 (2023).
11. Lu, Y., Sun, Y., Hou, C., Li, Z. & Ni, J. Explainable attention CNN for predicting properties of Heusler alloys. *J. Phys. Chem. C* **129**, 14958–14967 (2025).
12. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
13. Choudhary, K. & DeCost, B. Atomistic line graph neural network for improved materials property predictions. *npj Comput. Mater.* **7**, 1–8 (2021).
14. Barroso-Luque, L. et al. Open materials 2024 (OMat24) inorganic materials dataset and models.Preprint at arXiv https://doi.org/10.48550/arXiv.2410.12771 (2024).
15. Yamada, H. et al. Predicting materials properties with little data using shotgun transfer learning. *ACS Cent. Sci.* **5**, 1717–1730 (2019).
16. Lee, J. & Asahi, R. Transfer learning for materials informatics using crystal graph convolutional neural network. *Comput. Mater. Sci.* **190**, 110314 (2021).
17. Hoffmann, N., Schmidt, J., Botti, S. & Marques, M. A. L. Transfer learning on large datasets for the accurate prediction of material properties. *Digit. Discov.* **2**, 1368–1379 (2023).
18. He, J., Rabe, K. M. & Wolverton, C. Computationally accelerated discovery of functional and structural Heusler materials. *MRS Bull.* **47**, 559–572 (2022).
19. Noky, J., Zhang, Y., Gooth, J., Felser, C. & Sun, Y. Giant anomalous Hall and Nernst effect in magnetic cubic Heusler compounds. *npj Comput. Mater.* **6**, 1–8 (2020).

20. Xing, G., Masuda, K., Tadano, T. & Miura, Y. Chemical-substitution-driven giant anomalous Hall and Nernst effects in magnetic cubic Heusler compounds. *Acta Mater.* **270**, 119856 (2024).
21. Xiao, E. & Tadano, T. High-throughput computational screening of Heusler compounds with phonon considerations for enhanced material discovery. *Acta Mater.* **297**, 121312 (2025).
22. Xie, Y., Tritsaris, G. A., Grånäs, O. & Rhone, T. D. Data-driven studies of the magnetic anisotropy of two-dimensional magnetic materials. *J. Phys. Chem. Lett.* **12**, 12048–12054 (2021).
23. Liao, T. et al. Predicting magnetic anisotropy energies using site-specific spin-orbit coupling energies and machine learning: application to iron-cobalt nitrides. *Phys. Rev. Mater.* **6**, 024402 (2022).
24. Dutta, A. & Sen, P. Machine learning assisted hierarchical filtering: a strategy for designing magnets with large moment and anisotropy energy **10**, 3404–3417 (2022).
25. Fu, X. et al. Learning smooth and expressive interatomic potentials for physical property prediction. *Proc. Mach. Learn. Res.* 267, 17875–17893 (2025)..
26. Choudhary, K. et al. Unified graph neural network force-field for the periodic table: Solid state applications. *Digit. Discov.* **2**, 346–355 (2023).
27. Deng, B. et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
28. Kim, J. et al. Data-efficient multifidelity training for high-fidelity machine learning interatomic potentials. *J. Am. Chem. Soc.* **147**, 1042–1054 (2025).
29. Yan, K. et al. A materials foundation model via hybrid invariant-equivariant architectures. Preprint at arXiv https://doi.org/10.48550/arXiv.2503.05771 (2025).
30. Yang, H. et al. MatterSim: a deep learning atomistic model across elements, temperatures and pressures. Preprint at arXiv https://doi.org/10.48550/arXiv.2405.04967 (2024).
31. Hu, X. et al. Searching high spin polarization ferromagnet in Heusler alloy via machine learning. *J. Phys. Condens. Matter* **32**, 205901 (2020).
32. Miyazaki, H. et al. Machine learning based prediction of lattice thermal conductivity for half-Heusler compounds using atomic information. *Sci. Rep.* **11**, 13410 (2021).
33. Kim, K. et al. Machine-learning-accelerated high-throughput materials screening: Discovery of novel quaternary Heusler compounds. *Phys. Rev. Mater.* **2**, 123801 (2018).
34. Liu, K. et al. Machine learning assisted development of Heusler alloys for high magnetic moment. *Comput. Mater. Sci.* **250**, 113692 (2025).
35. Hirohata, A. et al. Machine learning for the development of new materials for a magnetic tunnel junction. *npj Spintron.* **3**, 1–9 (2025).
36. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
37. Schmidt, J. et al. Machine-learning-assisted determination of the global zero-temperature phase diagram of materials. *Adv. Mater.* **35**, 2210788 (2023).
38. Schmidt, J. et al. Improving machine-learning models in materials science through large datasets. *Mater. Today Phys.* **48**, 101560 (2024).
39. Kim, S. Y., Park, Y. J. & Li, J. Leveraging neural network interatomic potentials for a foundation model of chemistry (2025). https://doi.org/10.48550/arXiv.2506.18497 (2025).
40. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
41. Chang, R., Wang, Y.-X. & Ertekin, E. Towards overcoming data scarcity in materials science: unifying models and datasets with a mixture of experts framework. *npj Comput. Mater.* **8**, 242 (2022).
42. Minami, S. et al. Scaling law of Sim2Real transfer learning in expanding computational materials databases for real-world predictions. *npj Comput. Mater.* **11**, 146 (2025).
43. Sun, W. et al. The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225 (2016).
44. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).
45. Bahn, S. & Jacobsen, K. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **4**, 56–66 (2002).
46. Tadano, T. & Tsuneyuki, S. Self-consistent phonon calculations of lattice dynamical properties in cubic $SrTiO_3$ with first-principles anharmonic force constants. *Phys. Rev. B* **92**, 054301 (2015).
47. Tadano, T., Gohda, Y. & Tsuneyuki, S. Anharmonic force constants extracted from first-principles molecular dynamics: applications to heat transfer simulations. *J. Phys.: Condens. Matter* **26**, 225402 (2014).
48. Ebert, H., Ködderitzsch, D. & Minár, J. Calculating condensed matter properties using the KKR-Green's function method—recent developments and applications. *Rep. Prog. Phys.* **74**, 096501 (2011).
49. Liechtenstein, A. I., Katsnelson, M. I., Antropov, V. P. & Gubanov, V. A. Local spin density functional approach to the theory of exchange interactions in ferromagnetic metals and alloys. *J. Magn. Magn. Mater.* **67**, 65–74 (1987).
50. Hjorth Larsen, A. et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
51. Bitzek, E., Koskinen, P., Gähler, F., Moseler, M. & Gumbsch, P. Structural relaxation made simple. *Phys. Rev. Lett.* **97**, 170201 (2006).
52. Chanussot, L. et al. Open Catalyst 2020 (OC20) dataset and community challenges. *ACS Catal.* **11**, 6059–6072, https://doi.org/10.1021/acscatal.0c04525 (2021).
53. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
54. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
55. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
56. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
57. Daalderop, G. H. O., Kelly, P. J. & Schuurmans, M. F. H. First-principles calculation of the magnetocrystalline anisotropy energy of iron, cobalt, and nickel. *Phys. Rev. B* **41**, 11919–11937 (1990).
58. Xing, G., Miura, Y. & Tadano, T. Lattice dynamics and its effects on magnetocrystalline anisotropy energy of pristine and hole-doped YCo_5 from first principles. *Phys. Rev. B* **105**, 104427 (2022).
59. Xing, G., Miura, Y. & Tadano, T. First-principles prediction of phase transition of YCo_5 from self-consistent phonon calculations. *Phys. Rev. B* **108**, 014304 (2023).
60. Ong, S. P. et al. Python materials genomics (pymatgen): a robust, open-source Python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
61. Blöchl, P. E., Jepsen, O. & Andersen, O. K. Improved tetrahedron method for Brillouin-zone integrations. *Phys. Rev. B* **49**, 16223–16233 (1994).
62. ASE2SPRKKR software package — ASE2SPRKKR documentation. https://ase2sprkkr.github.io/ase2sprkkr/.
63. Togo, A., Shinohara, K. & Tanaka, I. Spglib: a software library for crystal symmetry search. *Sci. Technol. Adv. Mater. Methods* **4**, 2384822 (2024).

Acknowledgements

This study used computational resources of the supercomputer Fugaku provided by the RIKEN Center for Computational Science (Project ID:

hp250229), the computer resources provided by ISSP, U-Tokyo under the program of SCCMS, and the computer resources at NIMS Numerical Materials Simulator. This study was supported by MEXT Program: Data Creation and Utilization-Type Material Research and Development Project (Digital Transformation Initiative Center for Magnetic Materials) Grant Number JPMXP1122715503 and as "Program for Promoting Researches on the Supercomputer Fugaku" (Data-Driven Research Methods Development and Materials Innovation Led by Computational Materials Science, JPMXP1020230327).

Author contributions

T.T. conceptualized, designed, and supervised the project; reviewed and edited the manuscript. T.T. and E.X. developed the methodology and code implementation; performed the calculations and analysis; E.X. drafted the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-026-02013-0.

Correspondence and requests for materials should be addressed to Enda Xiao or Terumasa Tadano.