RESEARCH ARTICLE | JANUARY 12 2026

Predicting vacancy formation energies in refractory non-dilute random alloys using improved graph-based machine learning models trained on density functional theory data

Special Collection: Integrating Data Science and Computational Materials Science

Mahshad Fani ⓘ ; Oluwatimilehin Akinloye ⓘ ; Anvesh Nathani ⓘ ; Subah Mubassira ⓘ ; Iman Ghamarian ⓘ ; Shuozhi Xu ✉ ⓘ

Check for updates

CHORUS

View Online

Export Citation

Articles You May Be Interested In

A simple recipe for designing multicomponent ultra-high temperature ceramic classes by using structure maps coupled with machine learning

J. Appl. Phys. (July 2024)

Ab initio investigation of elastic properties of dilute Cu alloys for high-gradient accelerating structures

J. Appl. Phys. (November 2022)

Novel compounds of cerium binary alloys from high-throughput first-principles calculations

J. Appl. Phys. (June 2018)

Predicting vacancy formation energies in refractory non-dilute random alloys using improved graph-based machine learning models trained on density functional theory data

Mahshad Fani, [ID] Oluwatimilehin Akinloye, [ID] Anvesh Nathani, [ID] Subah Mubassira, [ID] Iman Ghamarian, [ID] and Shuozhi Xu[a) [ID]

AFFILIATIONS

School of Aerospace and Mechanical Engineering, University of Oklahoma, Norman, Oklahoma 73019-1052, USA

Note: This paper is part of the Special Topic on Integrating Data Science and Computational Materials Science.
[a)]Author to whom correspondence should be addressed: shuozhixu@ou.edu

ABSTRACT

This study employs four machine learning (ML) models—random forest (RF), XGBoost, graph convolutional networks (GCN), and graph attention networks (GAT)—to predict vacancy formation energies (VFEs) in refractory non-dilute random alloys including pure metals, binary, ternary, quaternary, and quinary systems based on Mo, Nb, Ta, V, and W. Training data are generated from density functional theory calculations. Among all elements, W has the highest VFE on average ($2.866 \rightarrow 3.517$ eV from pure metal to ternary), while Ta is the only element whose VFE rises monotonically ($2.841 \rightarrow 3.262$ eV from pure metal to quinary). In tree-based tabular models (RF, XGBoost), each chemical composition is encoded as a five-dimensional vector of elemental fractions on the Mo–Nb–Ta–V–W basis, augmented with a categorical indicator of the specific element in a certain alloy for which the vacancy is created, and features are standardized to balance input scales during training. In graph models (GCN, GAT), each composition is represented as a small graph whose nodes carry per-element descriptors (fraction, atomic number, and Magpie features), while edges summarize simple pairwise chemistry (such as fraction interactions and differences in atomic number, electronegativity, covalent radius, and Mendeleev number). All four models achieve high accuracy in predicting VFEs across alloy orders, with graph-based approaches capturing chemistry-aware trends beyond fixed tabular descriptors. Among the four, GCN attains the highest accuracy with $R^2 = 0.972$, root mean squared error $= 0.051$ eV, and mean absolute error $= 0.042$ eV, outperforming GAT and the tabular baselines. These results underscore the promise of graph-based ML for rapid, reliable prediction of VFEs in the refractory alloy design.

I. INTRODUCTION

Traditional alloy design has long centered on a single dominant metal forming the matrix, with small amounts of other elements added primarily to improve specific properties such as strength, corrosion resistance, and ductility.[1,2] Non-dilute random alloys (NDRAs) represent a breakthrough in alloy design where at least two principal elements are combined in near-equiatomic or substantial concentrations.[3,4] This fundamentally different approach creates a high degree of chemical complexity and atomic-level disorder, which profoundly affects the material's structural characteristics and physical properties.[5] We remark that the concept of NDRAs differs slightly from multi-principal element alloys (three or more principal elements) and high-entropy alloys (five or more principal elements).[6] The importance of NDRAs is underscored by their unique combination of desirable properties, often surpassing those of traditional alloys. They exhibit enhanced mechanical strength, ductility, thermal stability, corrosion resistance, and irradiation tolerance, making them

promising candidates for critical structural applications in aerospace, nuclear energy, and other extreme environments.[7–9] For example, refractory NDRAs (RNDRAs) containing elements like Mo, Nb, Ta, V, and W combine the high melting points and strength of refractory metals, potentially overcoming traditional limitations such as room-temperature brittleness and oxidation susceptibility.[10,11]

Like pure-metals and dilute alloys, the properties of NDRAs are greatly impacted by various defects within their structure. Among these defects, vacancies are the simplest and most frequently encountered.[12,13] The vacancy formation energy (VFE) measures the amount of energy needed to create such a vacancy by removing an atom from its regular lattice site and placing it in isolation.[14,15] VFE is critical because it directly influences the equilibrium concentration of vacancies, thereby affecting various physical and mechanical phenomena such as self-diffusion, creep, phase transformations, radiation tolerance, and thermal expansion.[16–19] Equilibrium vacancy concentrations at elevated temperatures significantly impact deformation mechanisms and diffusion processes, thus altering the material's performance and stability.[20–22] Therefore, calculating VFEs in RNDRAs is important for understanding thermophysical properties such as thermal conductivity, specific heat, and diffusion behavior.[13]

VFEs are typically determined through computational methods because of the technical challenges and inherent complexities in experimental measurement.[23] Classical atomistic simulation methods have commonly been utilized due to their lower computational cost and capacity for high-throughput modeling.[24,25] However, these methods rely on fitted interatomic potentials, which can lead to errors when predicting defect energies, especially in alloys with complex chemistry.[26,27] On the other hand, density functional theory (DFT), a quantum mechanical method, is widely regarded as a more accurate approach for calculating VFEs because it explicitly accounts for electronic structure and the distribution of electron density.[28] DFT inherently accounts for electronic charge redistribution and bonding interactions occurring during defect formation, making it particularly reliable for accurately predicting VFEs in RNDRAs.[17,18,29] Notably, Lin-Vines et al.[30] performed DFT calculations to explore defect energetics in MoNbTaVW, revealing significantly elevated VFEs compared to the constituent pure metals and highlighting the complex interplay of atomic-scale interactions in disordered environments. Their work provided foundational insights into how elemental mixing affects defect energetics but was limited in scope to a single quinary system.

However, DFT calculations are exceedingly time-consuming, making them unsuitable for high-throughput materials discovery work flows, particularly within closed-loop synthesis pipelines that demand efficient computational tools.[31] In recent years, machine learning (ML) surrogate models have become essential in materials science by significantly reducing computational costs.[32–34] To date, ML models have been applied to predict a variety of material properties, including bandgaps and elastic moduli across a wide range of systems.[35,36] A particularly important target is the prediction of defect formation energies. For instance, Sharma et al.[37] applied a random forest (RF) model using elemental and structural descriptors to estimate cation defect formation energies in perovskite oxides. Similarly, Arrigoni and Madsen[38] used a Gaussian process

metamodel with an evolutionary algorithm to search for low-energy point-defect configurations in semiconductors and oxides, incorporating uncertainty estimates to guide exploration. In another study, Baldassarri et al.[39] built a large DFT dataset of oxygen VFEs across diverse oxides and trained RF and kernel ridge regression models with site-specific features. For complex alloys, Manzoor et al.[40] proposed a regression framework using neighbor-based and bond-length descriptors to predict both VF and migration energies in NDRAs. Later, Mannodi-Kanakkithodi et al.[41] curated over 2500 DFT-calculated oxygen VFEs for nearly 1000 materials and evaluated several regressors, finding that an RF model with engineered features yielded the best performance. More recently, Tan et al.[42] developed an interpretable ML framework that combines high-throughput atomistic simulations with physics-informed local atomic environment descriptors to predict VFEs in CoNiCrFeMn, highlighting the dominant role of the first-nearest-neighbor shell and bond-strength competition in controlling VFEs with $R^2 = 0.98$.

ML has been widely applied to materials modeling tasks involving compositional complexity and property prediction. Gaussian process regression has been used to model transformation temperatures in NiTi-based shape memory alloys, capturing structure-sensitive trends relevant to alloy design.[43] Least-squares boosting has demonstrated effectiveness in predicting solubility behavior of super-critical CO_2 in ionic liquids, a task requiring accurate treatment of chemical interactions.[44] Support vector regression has been applied to estimate lattice parameters in spinel compounds, highlighting its utility in structure–property relationships.[45] Regression trees have been used to model the critical temperature of disordered MgB_2 superconductors, showing their suitability for disordered and defect-rich systems.[46] These methods provide context for our use of both tree-based tabular and graph-based ML approaches to predict VFEs in chemically complex random alloys.

Tree-based ensembles such as RFs and extreme gradient boosting (XGBoost) are standard baselines in materials informatics, particularly when the inputs are tabular, hand-engineered composition descriptors.[47] Standard tree-based models require fixed-size, precomputed descriptors and do not learn features directly from raw atomic structures.[48] They can capture local-environment effects only if such local descriptors are provided (and then pooled), unlike message-passing neural networks that learn them end-to-end.[49] As a result, they will underperform on properties governed by many-body or topology-dependent interactions if those interactions are not captured in the feature set; conversely, with informative local/graph-aware descriptors, they can perform strongly despite not learning representations end-to-end. For example, Ward et al.[50] introduced the Magpie descriptor library and showed that tree-based ensembles such as RFs can predict a range of materials properties using composition statistics alone.

Graph-based models such as the graph neural networks (GNNs) excel at predicting VFEs because they operate directly on a material's atomic graph, capturing the local bonding environments that govern defect energetics.[51] Unlike fixed descriptors, GNNs take atoms as nodes and bonds as edges, learning atom-level representations through message passing and convolutions.[52,53] Two forms of GNNs have been applied to materials science problems, including graph convolutional networks (GCNs) and graph

attention networks (GATs). Xie and Grossman[48] showed that a crystal GCN directly learns material properties from the connection of atoms in the crystal and can even extract the contributions from local chemical environments to global properties. Graph convolutions and attention mechanisms allow GCN/GAT to naturally capture complex short-range interactions. Each convolutional layer aggregates information from neighboring atoms, and attention weights let the model focus on the most chemically relevant neighbors. In this way, the models adaptively emphasize the local features that determine vacancy stability. It is shown by Lee et al.[54] that composition-only approaches often miss key variability in disordered alloys. Goodall and Lee[47] note that message-passing NNs can capture effects like inter-element interactions, which are obscured by traditional hand-engineered composition descriptors. Schmidt et al.[55] introduced crystal GATs and showed accurate, high-throughput prediction of thermodynamic stability, highlighting the benefit of attention over local neighborhoods in crystal graphs. Prior studies demonstrate that, after proper tuning, GATs rank among the top models on common materials datasets such as formation energies, adsorption energies, and band gaps.[56,57]

Despite these advances, two challenges remain. First, existing ML frameworks for predicting VFEs across chemically complex NDRAs remain scarce.[40] Most ML models either rely on hand-crafted descriptors that do not account for the many-body and topological interactions inherent in disordered systems,[48,50] or they are constrained to specific alloy types, thereby limiting generalizability. Second, ML-driven efforts to predict VFEs across varying alloy orders—from binaries to quinary—are lacking. A comprehensive, physically grounded surrogate model for VFE prediction that spans the full alloy complexity space is yet to be developed. In this study, we employ GCN and GAT models to predict the VFEs in RNDRAs and compare them with RF and XGBoost models. Our approach spans a broad range of alloy complexities from pure metals to quinary and leverages a DFT dataset consisting of 1530 calculations across 80 distinct host-element combinations. This comprehensive dataset enables quantitative mapping of how chemical complexity influences vacancy energetics. We also find that compared with XGBoost and RF, GCN and GAT perform better, suggesting that graph-based models hold strong potential as a general framework for accelerating the discovery and design of complex alloys where traditional descriptor-based ML methods may fall short. In particular, GCN best captures the complex compositional dependencies present in RNDRAs, indicating its advantage for VFE prediction in these complex alloys. Taken together, these contributions establish a new foundation for understanding and designing defect behavior in complex refractory alloys, highlighting the advantage of graph architectures for capturing composition-dependent, local bonding effects in defect energetics.

II. METHODS

A. DFT calculations

31 materials are considered in this work, including one quinary, five quaternaries, ten ternaries, ten binaries, and five pure elements (Mo, Nb, Ta, V, and W), as summarized in Table I. Special quasi-random structures (SQSs)[58] are employed to generate supercells of the alloys. Each cell contains $n = 54$ atoms for pure metals, binaries, ternaries, and quinary, while $n = 72$ atoms for quaternaries. The adequacy of this simulation size has been validated by Lin-Vines et al.[30]

To determine the VFEs, a vacancy is introduced into each relaxed SQS after removing a single atom. This process is repeated for all n unique atomic sites. The VFE (E_{vac}) is computed using the following relation:[19]

$$E_{vac} = E_{defective} + E_{atom} - E_{perfect}, \qquad (1)$$

where $E_{defective}$ is the total energy of the supercell containing a single vacancy, E_{atom} is the energy of an isolated atom of the corresponding element, and $E_{perfect}$ is the total energy of the relaxed, defect-free supercell. As a result, provided that E_{atom} is known, $n + 1$ VASP calculations are required for each material. For instance, in MoNbTaV, 73 VASP calculations lead to one $E_{perfect}$ for this quaternary and 72 VFE values for the four constituent elements (Mo, Nb, Ta, and V); for each element, there are 18 VFE values, and their average is taken as the final VFE for it in this particular alloy. In total, our study involves 1530 VASP simulations, with 80 VFEs derived across all cases.

First-principles calculations based on DFT are performed using VASP.[28,59] The simulations are conducted under constant pressure, permitting full relaxation of atomic positions, lattice shape, and volume. To treat partial occupancies, Methfessel–Paxton smearing[60] with a width of 0.05 eV is applied. A kinetic energy cutoff of 520 eV is used for the plane wave basis set, and Brillouin zone sampling is carried out using a $3 \times 3 \times 3$ Monkhorst–Pack k-point grid.[61] Exchange-correlation effects are described using the generalized gradient approximation in the Perdew–Burke–Ernzerhof formulation.[62] Projector-augmented wave pseudopotentials[63,64] provided with VASP version 6.4.0 are employed throughout. The convergence threshold for electronic self-consistency is set to 10^{-4} eV, and atomic relaxation was continued until the maximum force on atoms fell below 0.02 eV/Å. Upon relaxation, the defective configurations show localized vacancy defects without evidence of split vacancies or significant structural rearrangements.

TABLE I. All RNDRAs studied in this work, including pure metals, binaries, ternaries, quaternaries, and quinary. All alloy compositions are equiatomic.

Five pure metals	Ten binaries	Ten ternaries	Five quaternaries	One quinary
Mo, Nb, Ta, V, W	MoNb, MoTa,	MoNbTa, MoNbV,	MoNbTaV,	MoNbTaVW
	MoV, MoW,	MoNbW, MoTaV,	MoNbTaW,	
	NbTa, NbV,	MoTaW, MoVW,	MoNbVW,	
	NbW, TaV,	NbTaV, NbTaW,	MoTaVW,	
	TaW, VW	NbVW, TaVW	NbTaVW	

12 January 2026 13:43:23

TABLE II. Inputs and output used in the RF and XGBoost surrogates for VFE prediction.

Input	Output
Atomic fractions x_i, Magpie composition descriptors, and the element removed to form the vacancy	VFE

B. Development of the ML models

Based on the DFT calculation data presented earlier, we developed ML models to predict the VFEs of NRDRAs and pure metals. We convert each alloy's chemical formula into a fixed-length numerical representation in order to make it usable in ML pipelines. The elemental fractions are expressed on a predefined basis of five refractory elements [Mo, Nb, Ta, V, W], ensuring that every composition can be written as a vector of length five.[50,65] All input features are normalized using z-score standardization,

$$x_{\mathrm{norm}} = \frac{x - \mu}{\sigma}, \qquad (2)$$

where μ and σ denote the mean and standard deviation (SD), respectively, computed exclusively from the training split to prevent data leakage. We exclude pure-element VFEs from the inputs like our previous model,[34] avoiding target leakage and DFT priors, and enabling fully DFT-free screening.
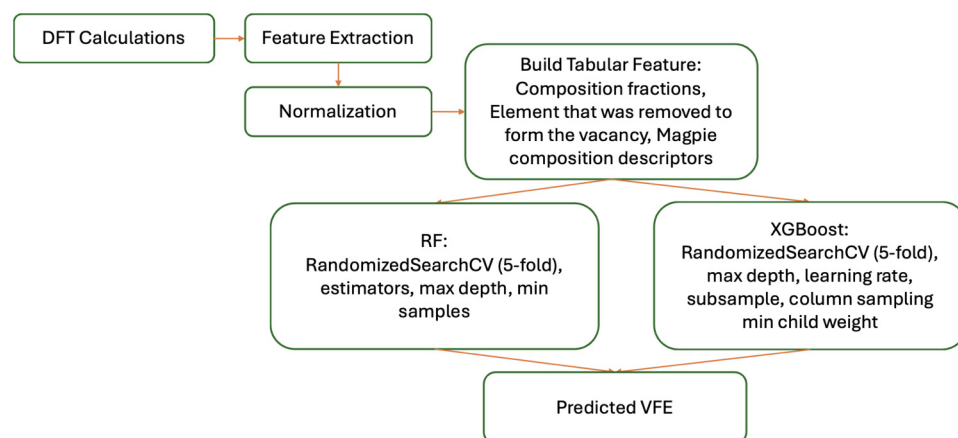
We first implemented two tree-based models: RF and XGBoost. Both models used the same fixed-length feature representation, which combined elemental fractions on the Mo–Nb–Ta–V–W basis, a categorical encoding of the element that was removed to form the vacancy, and 132 Magpie descriptors[50] generated with matminer's ElementProperty featurizer[66] to generate a comprehensive set of physically meaningful features for each alloy composition. These Magpie descriptors incorporate summaries of a wide range of elemental properties, such as the atomic size, electronegativity, and electronic structure for the constituent elements. This approach provides a robust and consistent input representation, allowing the ML models to effectively learn the relationship between alloy composition and VFE. Inputs and output of the RF and XGBoost models are summarized in Table II. Figure 1 summarizes the workflow of the RF and XGBoost surrogates, from DFT calculations and feature extraction through normalization and tabular feature construction to hyperparameter-tuned RF/XGBoost models that output the predicted VFE.

Hyperparameters for both models were tuned via randomized fivefold cross-validation, optimizing the negated root mean squared error (RMSE) as the metric. Model performance was assessed on a held-out 20% test split using RMSE, mean absolute error (MAE), relative root mean squared error (RRMSE), and R^2. For the RF model, the best configuration consisted of 800 estimators, a maximum depth of 10, minimum samples split of 2, minimum samples per leaf of 1, and feature sampling with square-root. For the XGBoost model, the best configuration was a maximum depth of 6, learning rate of 0.2, subsample ratio of 0.8, column sampling ratio of 1, minimum child weight of 1, and gamma of 0. The final model was retrained with squared-error loss for up to 1000 boosting rounds, with early stopping (30-round patience) applied on a validation set to prevent overfitting.

Next, we focus on graph-based models, whereas VFE data from DFT are converted to features and normalized, after which each alloy is mapped to a graph (nodes, edges, and graph-level features), processed by GAT/GCN with edge attributes/weights, globally pooled (mean/max), fused with graph features, and passed to a final regressor to predict VFE.

Both graph models represent each alloy (and each pure metal) as a small composition graph. Figure 2 compares the input representations of graph-based model in this study vs our prior single-node one without internal edges.[34] In this work, nodes correspond to the distinct elements present and carry standardized features (atomic fraction and pure-element Magpie descriptors). For every unordered element pair $\{i, j\}$ with fractions x_i, x_j, atomic numbers Z_i, Z_j, electronegativities χ_i, χ_j,[67] covalent radii r_i^{cov}, r_j^{cov},[68] and Mendeleev numbers M_i, M_j,[69] we define a six-dimensional edge-attribute vector $\mathbf{e}_{ij}$ containing (i) a composition interaction term

FIG. 1. Schematic representation of the RF and XGBoost surrogate models used for VFE prediction.
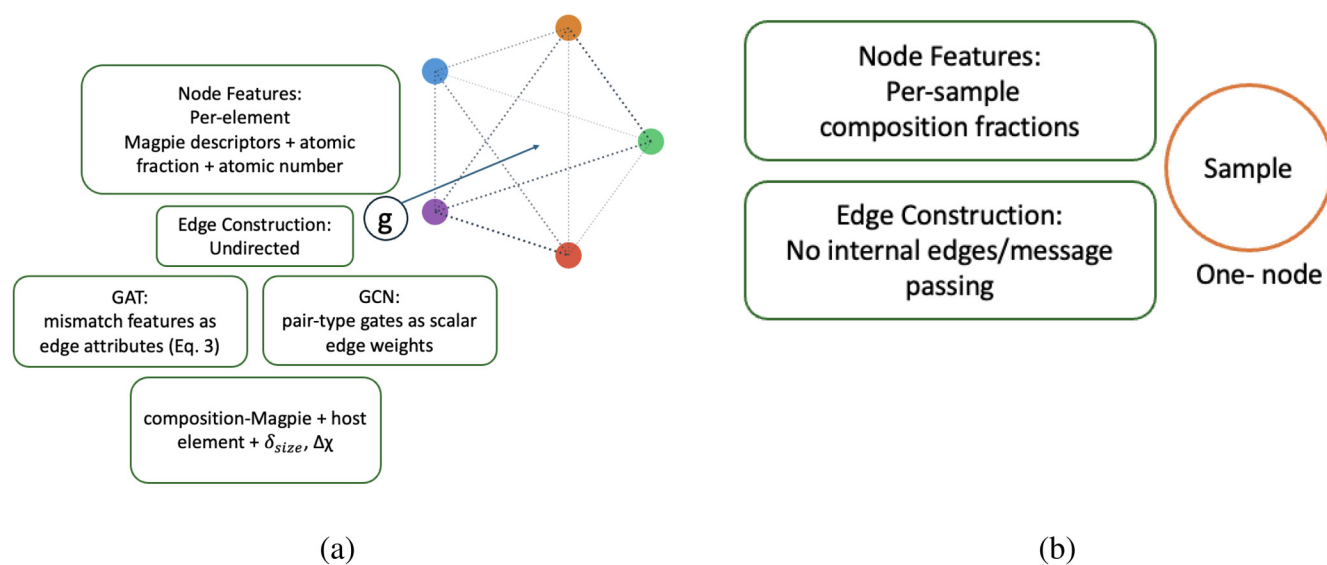
FIG. 2. Different input representations of the GAT and GCN models used for VFE prediction. (a) In this work, an undirected composition graph with per-element node features and a global graph feature labeled (g). (b) In our previous graph-based models,[34] each alloy sample is a single node (circle) whose features are the per-sample composition fractions; since the graph has only one node, there are no edges. Note that all parts of this diagram are new and are not taken from any previous work.

and (ii) five chemistry/size mismatch terms,

$$\mathbf{e}_{ij} = \left[x_i x_j, \ |x_i - x_j|, \ |Z_i - Z_j|, \ |\chi_i - \chi_j|, \ |r_i^{\text{cov}} - r_j^{\text{cov}}|, \ |M_i - M_j| \right],$$

(3)

where i and j are two different elements present in the same composition. The vector is six-dimensional because we include exactly six physically motivated pairwise terms; adding fewer would omit useful chemistry/size contrasts, and adding more would duplicate information in this setting. The pairwise mismatch features used as edge attributes in the graph network are designed to encode physically meaningful differences between neighboring elements, rather than arbitrary numerical distances.[70] For example, differences in atomic radius quantify the local size mismatch, which is directly related to lattice distortion and the elastic strain field around a defect.[71] In a graph-based ML model, encoding these pairwise mismatches lets the network sense how severe local strain or charge transfer effects will be around a site. In practice, features like atomic-size difference and mixing enthalpy strongly correlate with stability and defect behavior, especially in NDRAs.[72] Edges are constructed once for all unordered pairs and then mirrored to both directions for message passing; pure-metal graphs (single node) are handled by inserting a safe self-loop with zero edge features. At the graph level, we concatenate composition-level Magpie descriptors with a five-way host one-hot vector (the element removed to form the vacancy) and two simple physics terms δ_{size} and $\Delta\chi$ summarizing composition-weighted spreads in covalent radius and electronegativity. Each unordered chemical pair among {Mo, Nb, Ta, V, W} is also assigned a pair type with a learnable

scalar gate, allowing the model to modulate element–element interactions. All node, edge, and graph features, as well as the target, are standardized with z-scores; consistent with our code, scalers are fit on the full dataset prior to the 80/20 split, and predictions are inverse-transformed to physical units for reporting. In both GCN and GAT models, inputs and output are summarized in Table III, while Fig. 3 illustrates how the features are constructed from the DFT data and processed by the GAT/GCN workflow, including graph construction, message passing, global pooling, and fusion, to yield the final VFE prediction.

In the GCN model, two GCNConv layers (hidden width 256) are employed in lieu of two attention layers to isolate the effect of attention. Because GCNConv does not consume rich edge attributes, the 10 pair-type gates (one per unordered Mo/Nb/Ta/V/W pair) are mapped to scalar edge_weights that modulate degree-

TABLE III. Inputs and output used in the GAT and GCN surrogate models for VFE prediction.

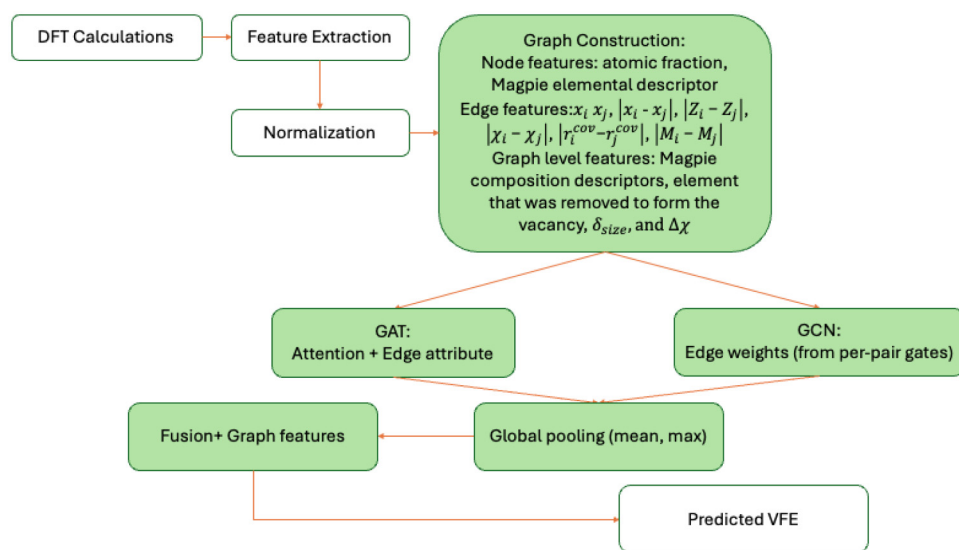| Input	Output
Node features: atomic fraction x_i, Magpie elemental descriptors;	VFE
Edge features: weights per Eq. (3): $x_i x_j$, $|x_i - x_j|$, $|Z_i - Z_j|$, $|\chi_i - \chi_j|$, $|r_i^{\text{cov}} - r_j^{\text{cov}}|$, $|M_i - M_j|$;	
Graph-level features: Magpie composition descriptors, element that was removed to form the vacancy, physics terms δ_{size} and $\Delta\chi$.	

FIG. 3. Schematic representation of the GAT and GCN model workflow. Each graph is built from nodes corresponding to the chemical species in an alloy and edges representing pairwise interactions between species, described by mismatch features. Graph-level features summarize the overall composition. Green-highlighted boxes indicate components that differ from our previous graph-based model.[34] Note that all parts of this diagram are new and are not taken from any previous work.

normalized aggregation; self-loops are added by the convolution. After the two GCN layers, we apply batch normalization, exponential linear unit (ELU), dropout, dual global pooling (mean and max), a linear projection of the graph-level features, and the same fusion head (pooled nodes, projected graph features, and their element-wise product) to regress VFE.

In the GAT model, we adopt a graph attention architecture[53] to capture chemistry-aware, anisotropic interactions beyond tabular descriptors. In this architecture, attention coefficients are edge-aware: the numeric edge attributes $\mathbf{e}_{ij}$ [Eq. (3)] directly influence the attention weights, and the pair-type gates further modulate these attributes before attention is computed. In many cases, high attention correlates with known descriptors such as atomic size or electronegativity mismatch, revealing how the model is effectively learning to emphasize the same factors that control VFEs in disordered alloys.[73] The network comprises a first multi-head attention layer (four heads, hidden width 256, concatenated) followed by a single-head attention layer (hidden 256), each with batch normalization, ELU, and dropout ($p = 0.05$). A linear skip projection aligns dimensions for a residual add into the second layer, and a squeeze–excitation block is applied afterward. Node embeddings are aggregated via global mean and max pooling; in parallel, graph-level features are linearly projected. The pooled node features, projected graph vector, and their element-wise interactions are fused and passed to a small multilayer perceptron (MLP) to predict VFE. Training uses an 80/20 split (seed 42), mini-batches of 16 graphs, AdamW (learning rate 10^{-3}, weight decay 10^{-4}), a ReduceLROnPlateau scheduler (factor 0.5, patience 30, floor 10^{-5}), and early stopping with a patience of 60 epochs; pure-metal cases are handled with the same safe self-loop.

For RF and XGBoost, the training time includes hyperparameter search plus the final fit. For GAT and GCN, it refers to the training loop only. We report throughput and average per-sample latency. For context, we adopt a DFT baseline of 2 h per VFE (7200 s) and report speed-up as $7200\,\mathrm{s}/t_{\mathrm{ML,per\text{-}sample}}$.

III. RESULTS AND DISCUSSION

A. DFT calculation results

Figure 4(a) shows how VFE changes as we add more element types to the material. The overall trend is that VFEs increase sharply from pure to binary and rise further into ternary; beyond three components, VFEs start to decrease, although changes are modest with only small element-dependent shifts. Mo, W, and V peak at ternary and soften slightly at quaternary and quinary, whereas Nb peaks at quaternary and then decreases at quinary, and Ta continues to creep upward into quaternary/quinary. The SD in VFEs, calculated among all host materials with the same number of elements, is the largest for binaries/ternaries and smaller for quaternaries; the quinary shows no SD because only one composition was sampled. With the same number of elements, the gap between W-based and V-based sets is as large as (and sometimes larger than) the jump from ternary to quaternary. Thus, beyond three elements, which elements are chosen matter more than adding another one. As shown in Fig. 4(b), the mean VFEs taken across all host materials (regardless of the number of elements) that contain a specific element are ranked as follows: W > Mo > Ta > Nb > V, indicating that W-based alloys exhibit the highest VFEs while V-based alloys are the lowest.[21,74] Because W, Mo, and often Ta create stronger, stiffer local bonding, removing an atom costs more energy, while Nb, and to a lesser extent V, yield comparatively softer local environments and a lower vacancy cost.[75] Taken together, our results demonstrate that both the removed element and the host material exert a stronger influence on VFE.

Figure 5 shows how the VFEs are distributed for all five elements in selected alloys. For each element, we compare the distribution of VFEs in a binary, ternary, quaternary, and quinary. Figure 5(a) concerns the VFE of Mo, for which MoNb shows a relatively narrow VFE distribution centered at higher energies, while adding Ta and V broadens the distribution and slightly shifts part of it to lower VFEs; the quinary's VFE distribution covers almost

FIG. 4. (a) VFE (in eV) vs number of elements in the material (1–5 for pure, binary, ternary, quaternary, and quinary). Each curve corresponds to an element (Mo, Nb, Ta, V, W). Markers are the mean VFEs; vertical bars show SD across materials of that order containing that element. (b) VFE for each element. Mean values and SDs are based on all materials that contain each element, regardless of the number of elements.

the full range of the other three, showing that many different local environments are present. In Fig. 5(b), which is for Nb, the binary NbTa has VFEs around 2.8–3 eV, whereas NbTaV shifts the distribution to noticeably lower energies; when W and then Mo are added (i.e., NbTaVW and MoNbTaVW), the distributions become wider and extend toward higher VFEs, again indicating more variation in local bonding in the quinary. In Fig. 5(c) which is for Ta-based alloys, TaV has the lowest VFEs among the four, concentrated below about 2.5 eV. As can be observed in Fig. 5(d), where all alloys contain the element V, VW and MoVW mostly lie at relatively high VFEs, whereas the quaternary MoNbTaV shows a much broader spread that reaches down to lower energies. In the last set, based on W, as shown in Fig. 5(e), MoW has a narrow distribution at the lowest VFEs, while adding Nb and later Ta (i.e., MoNbW and MoNbTaW) shifts the distributions upward and makes the distributions wider.

Table IV provides more quantitative data by presenting the VFEs of each element in each type of material, reporting both the mean (μ) and SD (σ). It highlights that μ increases most from pure to binary, rises modestly to ternary, and changes little thereafter, while the element ranking remains broadly consistent: W > Mo > Ta > Nb > V. The element with the highest μ is W for pure metals (2.866 eV), binaries (3.410 eV), and ternaries (3.517 eV). For quaternaries, the element with the top spot shifts narrowly to Mo (3.469 eV, only 0.007 eV above W), and at quinary, W again leads (3.301 eV). Ta is the only element whose mean VFEs increase monotonically from pure to quinary (2.841 $\rightarrow$ 3.262 eV). Among all elements, the largest increasing step in VFE is universally pure $\rightarrow$ binary, with gains of +0.603 (V), +0.544 (W), +0.415 (Mo), +0.276 (Ta), and +0.221 eV (Nb). The binary $\rightarrow$ ternary increment in VFE is smaller but still positive: +0.324 (V), +0.256 (Mo), +0.107 (W), +0.076 (Nb), and +0.032 eV (Ta). Changes in VFE beyond ternary

are modest and element-/host-material-dependent, ternary $\rightarrow$ quaternary: -0.019 (Mo), $+0.081$ (Nb), $+0.088$ (Ta), -0.114 (V), and -0.055 eV (W); quaternary $\rightarrow$ quinary: -0.211 (Mo), -0.064 (Nb), $+0.025$ (Ta), -0.051 (V), and -0.161 eV (W).

It is also found that the range in the mean VFE value across different elements that were removed to form the vacancy, $\max_i \mu_i - \min_i \mu_i$, decreases from $(2.866 - 2.303 = 0.563$ eV$)$ in the pure metals to $(3.301 - 3.065 = 0.236$ eV$)$ in the quinary. Thus, increasing the number of elements in the material homogenizes vacancy energetics even as the absolute VFEs remain above the pure-metal baseline.

As a result, the entire dataset spans a wide range of VFEs. While V typically lowers VFE, the V-based ternary TaVW reaches 4.07 eV, the highest VFE among all data points, because W and often Ta can dominate the local bonding and offset V's softening effect. In general, higher VFE is achieved by increasing W/Ta fractions[76] and reducing V, whereas lowering it is achieved by increasing the V content, with NbV emerging as a particularly effective route to low VFE.[77] Indeed, the binary NbV alloy shows the smallest VFE (2.28 eV).

Finally, we discuss the SD in VFE. We find that it is the largest for binaries/ternaries and smallest for quaternaries. Among all chemical elements, the smallest SD is associated with Mo in binaries ($\sigma = 0.334$ eV), and with W in ternaries ($\sigma = 0.295$ eV) and quaternaries ($\sigma = 0.189$ eV); the largest SD is associated with Ta in binaries ($\sigma = 0.518$ eV) and quaternaries ($\sigma = 0.288$ eV), and with V in ternaries ($\sigma = 0.534$ eV).

B. ML models

1. Performance

Table V illustrates the ML model performance results. First, we analyze the RF baseline. With enhanced tabular features (atomic

FIG. 5. Probability-density distributions of VFEs in selected refractory alloys. (a) Mo-based sets. (b) Nb-based sets. (c) Ta-based sets (d) V-based sets. (e) W-based sets.

TABLE IV. VFE (in eV) of different elements in different types of materials; values reported as mean ± SD ($\mu \pm \sigma$).

Host	Pure metal	Binary	Ternary	Quaternary	Quinary
Mo	2.817 ± 0.000	3.232 ± 0.334	3.488 ± 0.363	3.469 ± 0.226	3.258 ± 0.000
Nb	2.786 ± 0.000	3.007 ± 0.468	3.083 ± 0.348	3.164 ± 0.271	3.100 ± 0.000
Ta	2.841 ± 0.000	3.117 ± 0.518	3.149 ± 0.414	3.237 ± 0.288	3.262 ± 0.000
V	2.303 ± 0.000	2.906 ± 0.501	3.230 ± 0.534	3.116 ± 0.242	3.065 ± 0.000
W	2.866 ± 0.000	3.410 ± 0.436	3.517 ± 0.295	3.462 ± 0.189	3.301 ± 0.000

fractions, the element that was removed to form the vacancy, and 132 Magpie descriptors), the RF baseline achieves $R^2 = 0.905$, RMSE = 0.0952 eV, MAE = 0.076 eV, and RRMSE = 0.0294 on the 20% held-out test set. Figure 6(a) shows that errors are the largest near the lowest and highest VFEs because the model tends to pull extreme values back toward the average (a regression-to-the-mean effect), indicating that a tree-based tabular model captures global trends but struggles with composition-specific interactions.

Second, we evaluate XGBoost. Under the same features and split, XGBoost improves to $R^2 = 0.943$, RMSE = 0.0698 eV, MAE = 0.0594 eV, and RRMSE = 0.0212. This is a 26.7% reduction in RMSE and 21.8% reduction in MAE vs RF. This reflects boosting's ability to reduce bias by fitting residuals stage-by-stage. Persisting errors at very high/low VFEs indicate the limits of this model that only uses global/tabular descriptors. Results are shown in Fig. 6(b).

As seen in Fig. 7, switching from tabular inputs to a composition-graph representation yields the largest accuracy gains. In both graph models, nodes encode present elements with per-element Magpie descriptors and atomic fractions; edges carry simple pairwise mismatch features and the graph stream augments composition-level Magpie with two physics priors (δ_{size}, $\Delta\chi$). The two-layer GAT achieves $R^2 = 0.969$, RMSE = 0.0560 eV, MAE = 0.0453 eV, and RRMSE = 0.017, while the matched two-layer GCN achieves $R^2 = 0.972$, RMSE = 0.051 eV, MAE = 0.042 eV, and RRMSE = 0.016. The parity plots show tight clustering around the parity line with little systematic bias across the full VFE range; outliers at the extremes are notably reduced compared to the two tree baselines. The residual gap between the two graph models is modest: GCN attains a slightly lower RMSE than GAT (about ~9% reduction), indicating that attention is not strictly necessary to reach high accuracy, even

though it can help adaptively weight chemically distinct neighbors in other alloy systems.[57] Most of the overall improvement instead comes from representing local chemistry and explicit pairwise element mismatches, rather than from attention alone. This is especially useful for RNDRAs, where minor shifts in composition can cause large swings in properties.[34]

Compared with RF and XGBoost, the two graph-based models show outstanding performance. GAT (0.0453 eV) delivers a 40.4% MAE reduction vs RF and 23.7% vs XGBoost, while GCN (0.0420 eV) achieves 44.7% and 29.3% MAE reductions vs RF and XGBoost, respectively. GAT (0.056 eV) yields a 41.2% RMSE reduction vs RF and 19.8% vs XGBoost; GCN (0.0510 eV) yields a 46.4% RMSE reduction vs RF and 26.94% vs XGBoost. In absolute terms, the bes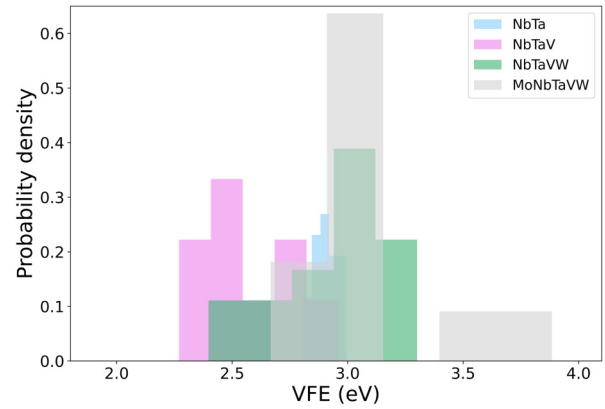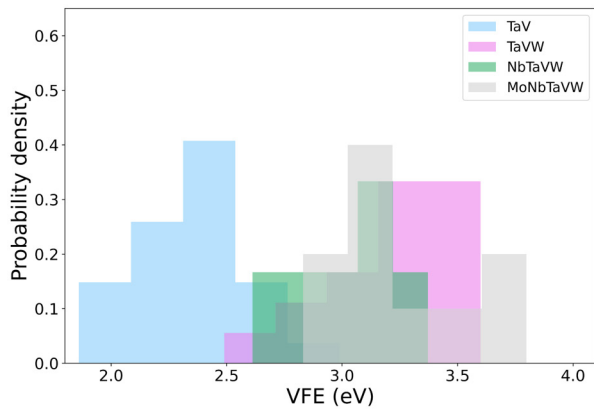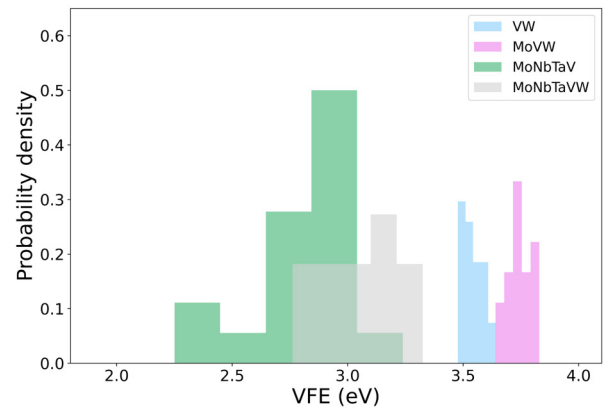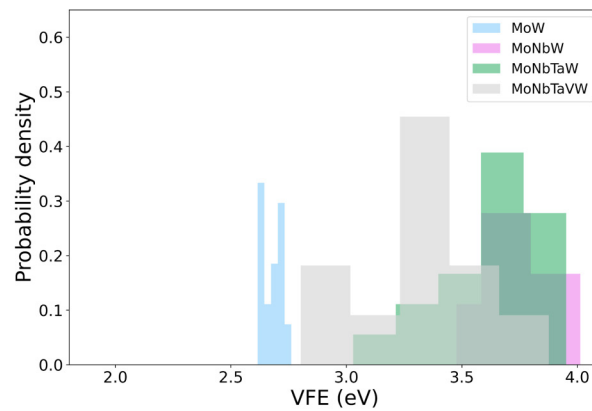t MAE (0.042 eV for the two-layer GCN) is only ~1%−2% of a typical VFE magnitude (3–3.5 eV), highlighting the accuracy gains from graph-based representations. It is also found that our improved GAT and GCN models outperform GAT-RB and GCN-RB, which were developed in our prior work.[34]

To place these errors in context, we compare our results with recent defect-prediction studies. Manzoor et al.[40] developed an ML framework for VFEs and vacancy migration energies in the Co–Cr–Cu–Fe–Ni system using a database generated from classical interatomic potentials. Their models achieve $R^2 > 0.81$ and RMSE < 0.1 eV for individual binary alloys, while predictions for the equiatomic CoCrCuFeNi alloy yield $R^2 = 0.6561$ and RMSE = 0.13 eV. From this perspective, our normalized errors are comparable to or smaller than those previously reported, despite the broader alloy diversity.

2. Model robustness

To assess robustness of the ML models with respect to the train/test partitioning, we repeated the 80/20 split 100 times with different random seeds and retrained all models based on the corresponding 25th-percentile values for each metric; the resulting distributions of R^2, RMSE, MAE, and RRMSE are summarized in Table VII in the Appendix. Specifically, instead of using the mean VFE out of each host-element combination, we sort the values of VFEs in each combination and take the value that is higher than 25% of all VFEs as the 25th-percentile value. Model performance based on these 25th-percentile values is a bit worse than those based on the mean values, yet it still shows good prediction accuracy.

3. Speed-up

Across all four surrogates, training and inference are fast with large speed-ups relative to a 2 h/DFT–VFE baseline. For the tree-

TABLE V. Comparison of test-set performance of the RF, XGBoost, GAT, and GCN models for VFE prediction. Results from our previous GAT and GCN models trained on the same dataset, named after the first author, Richard Brinlee (RB), are also shown.

Model	R^2	RMSE (eV)	MAE (eV)	RRMSE
RF	0.905	0.0952	0.0760	0.0294
XGBoost	0.943	0.0698	0.0594	0.0212
GAT	0.969	0.0560	0.0453	0.0170
GCN	0.972	0.0510	0.0420	0.0160
GAT-RB[34]	0.898	0.098	0.08	0.030
GCN-RB[34]	0.888	0.103	0.082	0.031

FIG. 6. A comparison between actual and predicted VFEs based on the (a) RF and (b) XGBoost models.

FIG. 7. A comparison between actual and predicted VFEs based on the (a) GCN and (b) GAT models.

12 January 2026 13:43:23

based tabular models, RF required 69.071 s total training time (67.612 s tuning + 1.459 s final fit) and achieved an inference time of 0.091 299 s over 16 compositions (∼175.2 comps/s; 5.706 ms per composition), yielding a DFT-to-ML speed-up of 1 261 788 per VFE. XGBoost trained in 15.305 s (15.222 s tuning + 0.083 s final fit) and inferred in 0.000 501 s over 16 compositions (∼31 936 comps/s; 0.031 ms per composition), corresponding to a speed-up of 2.298×10^8 per VFE.

For the graph models, the two-layer GCN trained in 3.77 s (mean 0.05 s/epoch over 78 epochs) and inferred in 0.004 s over 16 graphs, which is exactly 0.25 ms per graph (∼4000 graphs/s). The two-layer GAT trained in 3.76 s (mean 0.04 s/epoch over 86 epochs) and likewise inferred in 0.004 s over 16 graphs (∼4000 graphs/s; 0.25 ms per graph). Under the same 2 h baseline, these inference latencies imply speed-ups of ∼2.88×10^7 per VFE for both GNNs. These results show that all ML surrogates deliver high-throughput screening capability, with XGBoost giving the highest raw inference throughput among the tested models and both GNNs sustaining sub-millisecond per-graph latency.

4. Uncertainty quantification

It is essential to include uncertainty or confidence estimates for ML predictions, especially in our case, where VFEs depend sensitively on small energy differences and local chemical environment. In what follows, we employ model-appropriate approximate procedures for uncertainty quantification, and comparisons across models should rely on the calibration statistics (fractions of points within $\pm 1\sigma$ and $\pm 2\sigma$) rather than on the absolute magnitudes of σ. Results of uncertainty quantification for all models are shown in Table VI.

First, we quantified the predictive uncertainty of the RF model. We used the spread of predictions across individual trees to obtain an empirical 1σ uncertainty for each test sample. The RF yields a mean predicted 1σ uncertainty of 0.1550 eV (median 0.1468 eV), with 75% of test points lying within $\pm 1\sigma$ and 93.8% within $\pm 2\sigma$. For comparison, if the errors were perfectly Gaussian, about 68% and 95% of the points would be expected within $\pm 1\sigma$ and $\pm 2\sigma$, respectively, so the RF uncertainties are slightly conservative but reasonably well calibrated. This indicates that the RF not only achieves good average accuracy but also provides informative uncertainty estimates for VFEs.

Second, for XGBoost, we formed an ensemble of models with identical hyperparameters but different random seeds and took the SD of their predictions as the uncertainty. Although the resulting

mean and median 1σ uncertainties are small (0.0175 and 0.0172 eV, respectively), only 12.5% and 18.8% of test points fall within the corresponding $\pm 1\sigma$ and $\pm 2\sigma$ intervals, indicating that this ensemble is under-dispersed and strongly underestimates the true error.

Finally, we quantified the predictive uncertainty of the two graph-based models using Monte Carlo (MC) dropout. For each test composition, we performed $T = 50$ (T is the number of MC-dropout samples) stochastic forward passes with dropout active and took the μ and the SD across passes as an empirical 1σ uncertainty. For the GAT model, the mean and median predicted 1σ uncertainties are 0.0178 and 0.0161 eV, respectively. However, only 18.8% of the test points fall within the corresponding $\pm 1\sigma$ interval and 50.0% within $\pm 2\sigma$, indicating that the MC-dropout ensemble is under-dispersed and substantially underestimates the true error. The GCN model shows a similar behavior: the mean and median predicted 1σ uncertainties are 0.0147 and 0.0136 eV, with 25.0% and 43.8% of test points lying within $\pm 1\sigma$ and $\pm 2\sigma$, respectively. Thus, while both GAT and GCN deliver accurate VFE predictions, their MC-dropout uncertainties are small compared to the more conservative uncertainty estimates obtained for the RF model.

IV. CONCLUSIONS

In this paper, we first employed DFT to compute VFEs across a targeted set of 31 refractory materials, including five pure BCC metals, ten binaries, ten ternaries, five quaternaries, and one quinary. It is found that VFEs increase with the number of elements in the material and with the W/Ta content, whereas V depresses VFE. These trends rationalize the interplay between strong-bonding species (W/Ta) and the softening role of V in disordered BCC chemistries.

We then trained four ML models based on DFT data. Among these models, tree-based tabular baselines capture global composition trends, with XGBoost outperforming RF. Switching to composition graphs with pairwise mismatch features and two physics priors brings the largest gains. Based on RMSE/MAE on the 20% test set, the performance ordering is GCN ⪞ GAT > XGBoost > RF. Thus, graph-based models substantially outperform the tree baselines, with the two-layer GCN achieving slightly smaller errors than the two-layer GAT. This suggests that most of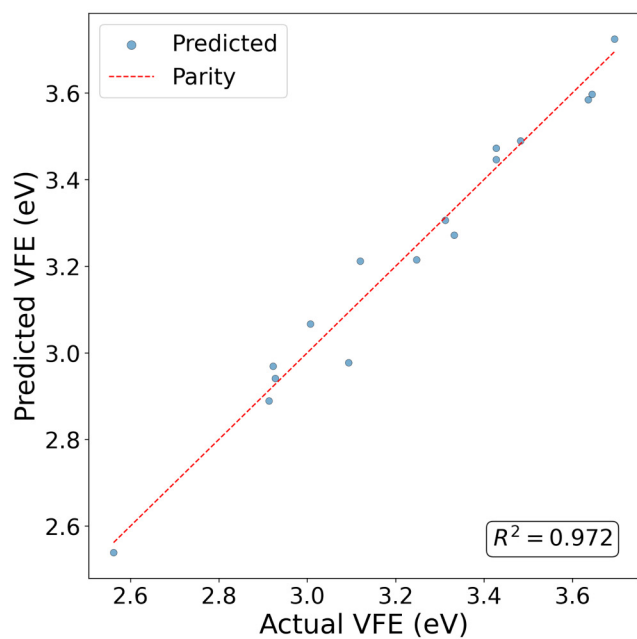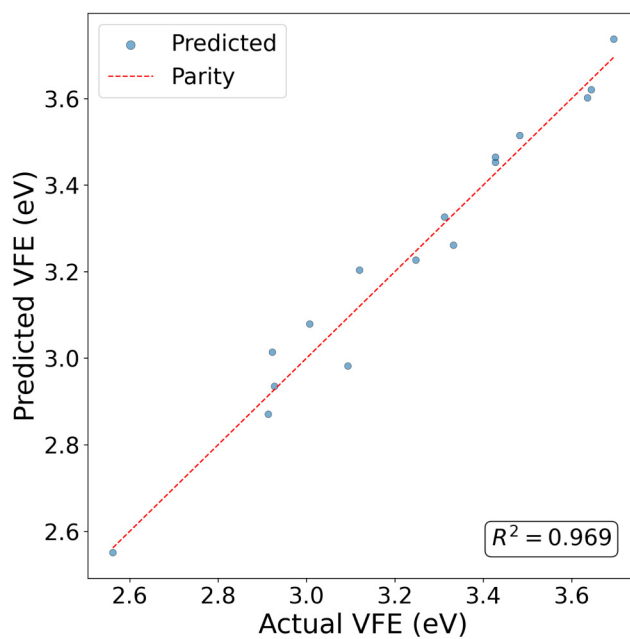 the gain comes from explicitly encoding local chemistry and pairwise element mismatches in a graph representation, while attention provides only a modest additional benefit for these RNDRAs rather than being strictly required to capture composition-dependent interactions.

To conclude, (i) the curated DFT dataset establishes consistent, interpretable VFE trends across pure metals and alloys; (ii) graph-based models, most effectively capture composition-dependent local bonding effects; and (iii) the combined physics-aware featurization and graph architectures provide accurate, orders-of-magnitude-faster surrogates for VFE prediction. These advances should facilitate targeted exploration of RNDRAs, where modest shifts in composition can produce outsized changes in defect energetics.

While our models achieved high accuracy in predicting VFEs, they are trained on a relatively small dataset of 1530 DFT

TABLE VI. Uncertainty metrics for ML models, based on ensemble (RF, XGBoost) or MC-dropout (GAT, GCN) estimates of the predictive SD (σ), and $\varepsilon = y_{\text{true}} - y_{\text{pred}}$ is the prediction error.

Model	$\bar{\sigma}$ (eV)	Median σ (eV)	% ($|\varepsilon| \leq 1\sigma$)	% ($|\varepsilon| \leq 2\sigma$)
RF	0.1550	0.1468	75.0	93.8
XGBoost	0.0175	0.0172	12.5	18.8
GAT	0.0178	0.0161	18.8	50.0
GCN	0.0147	0.0136	25.0	43.8

calculations. This may limit their generalizability to compositions or structures not included in the training set. Additionally, the models do not account for temperature effects or dynamic behavior, which could influence vacancy formation in real materials. In future work, we aim to expand the dataset to include more complex chemistries and explore transfer learning or active learning strategies to improve model robustness for practical relevance.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Mahshad Fani: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal). **Oluwatimilehin Akinloye:** Investigation (equal). **Anvesh Nathani:** Methodology (equal); Software (equal); Validation (equal). **Subah Mubassira:** Formal analysis (equal); Investigation (equal); Validation (equal). **Iman Ghamarian:** Conceptualization (equal); Funding acquisition (equal); Supervision (equal). **Shuozhi Xu:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in GitHub at https://github.com/shuozhixu/JAP_2026, Ref. 78.

APPENDIX: MORE ON ML MODEL ROBUSTNESS

Table VII summarizes the 25th-percentile test performance of all four models (RF, XGBoost, GAT, and GCN) over 100 random 80/20 train–test splits. For GAT, while the R^2 value based on the mean VFE values is 0.969 (Table V), the model based on 25th-percentile VFE values achieves $R^2 = 0.939$, RMSE = 0.075 eV, and MAE = 0.0653 eV. Similar behavior is observed for the RF, XGBoost, and GCN models, whose 25th-percentile metrics remain worse than, yet close to, their respective means and preserve the same ranking of models. These results show that the reported test accuracies are not driven by a single favorable split and that the

TABLE VII. Comparison of test-set performance of the RF, XGBoost, GAT, and GCN models based on the 25th-percentile VFE values.

Model	R^2	RMSE (eV)	MAE (eV)	RRMSE
RF	0.83	0.123	0.108	0.0393
XGBoost	0.885	0.103	0.0905	0.0330
GAT	0.939	0.075	0.0653	0.0240
GCN	0.942	0.073	0.060	0.0230

learned surrogates are statistically robust with respect to reasonable variations in the train/test partition.

REFERENCES

[1] A. Brezini, R. Bouamrane, F. Hamdache, and C. Depolier, "Theoretical model of the density of states for random dilute binary alloys," Phys. Status Solidi B **188**(2), 697–710 (1995).

[2] Y. F. Ye, Q. Wang, J. Lu, C. T. Liu, and Y. J. M. T. Yang, "High-entropy alloy: Challenges and prospects," Mater. Today **19**(6), 349–362 (2016).

[3] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, and S.-Y. Chang, "Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes," Adv. Eng. Mater. **6**(5), 299–303 (2004).

[4] B. Cantor, I. T. H. Chang, P. Knight, and A. J. B. Vincent, "Microstructural development in equiatomic multicomponent alloys," Mater. Sci. Eng. A **375–377**, 213–218 (2004).

[5] O. N. Senkov, G. B. Wilks, D. B. Miracle, C. P. Chuang, and P. K. Liaw, "Refractory high-entropy alloys," Intermetallics **18**(9), 1758–1765 (2010).

[6] D. B. Miracle and O. N. Senkov, "A critical review of high entropy alloys and related concepts," Acta Mater. **122**, 448–511 (2017).

[7] S. Wei, S. J. Kim, J. Kang, Y. Zhang, Y. Zhang, T. Furuhara, E. S. Park, and C. C. Tasan, "Natural-mixing guided design of refractory high-entropy alloys with as-cast tensile ductility," Nat. Mater. **19**(11), 1175–1181 (2020).

[8] S. Mubassira, M. Fani, A. Raj, C. Hirt, R. S. Brinlee, A. Poozesh, W.-R. Jian, S. Z. Chavoshi, C. Lee, and S. Xu, "Chemical short-range order and its influence on selected properties of non-dilute random alloys," Comput. Mater. Sci. **248**, 113587 (2025).

[9] M. Y. Araghi, A. Dashti, M. Fani, I. Ghamarian, C. Ruiz, and S. Xu, "Melt-based additive manufacturing of refractory metals and alloys: Experiments and modeling," J. Mater. Res. Technol. **37**, 870–892 (2025).

[10] C. Hampel, "Refractory metals. Tantalum, niobium, molybdenum, rhenium, and tungsten," Ind. Eng. Chem. **53**(2), 90–96 (1961).

[11] O. N. Senkov, S. Gorsse, and D. B. Miracle, "High temperature strength of refractory complex concentrated alloys," Acta Mater. **175**, 394–405 (2019).

[12] C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, "First-principles calculations for point defects in solids," Rev. Mod. Phys. **86**(1), 253–305 (2014).

[13] Y. Kraftmakher, "Equilibrium vacancies and thermophysical properties of metals," Phys. Rep. **299**(2-3), 79–188 (1998).

[14] S. C. Middleburgh, R. E. Voskoboinikov, M. C. Guenette, and D. P. Riley, "Hydrogen induced vacancy formation in tungsten," J. Nucl. Mater. **448**(1–3), 270–275 (2014).

[15] G. S. Smirnov and V. V. Stegailov, "Formation free energies of point defects and thermal expansion of bcc U and Mo," J. Phys.: Condens. Matter **31**(23), 235704 (2019).

[16] A.-Y. Gao, Y.-L. Liu, Z.-H. Dai, and C. Duan, "Elucidating hydrogen assisting vacancy formation in metals: Mo and Nb as examples," Europhys. Phys. J. B **86**, 355 (2013).

[17] D. J. M. King, S. C. Middleburgh, A. G. McGregor, and M. B. Cortie, "Predicting the formation and stability of single phase high-entropy alloys," Acta Mater. **104**, 172–179 (2016).

12 January 2026 13:43:23

[18] P. A. Burr and M. W. D. Cooper, "Importance of elastic finite-size effects: Neutral defects in ionic compounds," Phys. Rev. B **96**(9), 094107 (2017).

[19] Y. Pan and Y. Lin, "Influence of vacancy on the mechanical and thermodynamic properties of IrAl₃ compound: A first-principles calculations," J. Alloys Compds **684**, 171–176 (2016).

[20] M. L. Fullarton, R. E. Voskoboinikov, and S. C. Middleburgh, "Hydrogen accommodation in α-iron and Nickel," J. Alloys Compds **587**, 794–799 (2014).

[21] T. Korhonen, M. J. Puska, and R. M. Nieminen, "Vacancy-formation energies for fcc and bcc transition metals," Phys. Rev. B **51**(15), 9526 (1995).

[22] R. A. Johnson, "Point-defect calculations for tungsten," Phys. Rev. B **27**(4), 2014 (1983).

[23] S. K. Nayak, S. Danayak, and K. N. Kulkarni, "Experimental determination of vacancy formation energies in multicomponent systems," Metall. Mater. Trans. A **54**(8), 3341–3348 (2023).

[24] M. S. Daw and M. I. Baskes, "Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals," Phys. Rev. B **29**(12), 6443 (1984).

[25] R. A. Johnson and D. J. Oh, "Analytic embedded atom method model for bcc metals," J. Mater. Res. **4**(5), 1195–1201 (1989).

[26] M. W. Finnis and J. E. Sinclair, "A simple empirical n-body potential for transition metals," Philos. Mag. A **50**(1), 45–55 (1984).

[27] R. R. Zope and Y. Mishin, "Interatomic potentials for atomistic simulations of the Ti-Al system," Phys. Rev. B **68**(2), 024102 (2003).

[28] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," Comput. Mater. Sci. **6**(1), 15–50 (1996).

[29] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, "Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD)," JOM **65**, 1501–1509 (2013).

[30] A. X. Lin-Vines, J. A. Wilson, A. Fraile, L. J. Evitts, M. J. D. Rushton, J. O. Astbury, W. E. Lee, and S. C. Middleburgh, "Defect behaviour in the MoNbTaVW high entropy alloy (HEA)," Results Mater. **15**, 100320 (2022).

[31] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. L. Billinge *et al.*, "Recent advances and applications of deep learning methods in materials science," npj Comput. Mater. **8**(1), 59 (2022).

[32] J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost," Chem. Sci. **8**(4), 3192–3203 (2017).

[33] M. Fani, W. Chadwell, N. Tasnim, X. Wang, M. Y. Araghi, K. Lu, Z. Zhou, T. Gu, and S. Xu, "Constructing surrogates for atomistic simulations via deep learning and generative large language models," J. Mater. Res. **2025**, 1–17.

[34] R. Brinlee, A. Poozesh, A. Nathani, A. Raj, X.-G. Li, I. Ghamarian, and S. Xu, "Integrating atomistic simulations and machine learning to predict unstable stacking fault energies of refractory non-dilute random alloys," JOM **77**(11), 8127–8136 (2025).

[35] C. Chen, D. T. Nguyen, S. J. Lee, N. A. Baker, A. S. Karakoti, L. Lauw, C. Owen, K. T. Mueller, B. A. Bilodeau, V. Murugesan *et al.*, "Accelerating computational materials discovery with machine learning and cloud high-performance computing: From large-scale screening to experimental validation," J. Am. Chem. Soc. **146**(29), 20009–20018 (2024).

[36] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery," Nature **624**(7990), 80–85 (2023).

[37] V. Sharma, P. Kumar, P. Dev, and G. Pilania, "Machine learning substitutional defect formation energies in ABO₃ perovskites," J. Appl. Phys. **128**(3), 034902 (2020).

[38] M. Arrigoni and G. K. H. Madsen, "Evolutionary computing and machine learning for discovering of low-energy defect configurations," npj Comput. Mater. **7**(1), 71 (2021).

[39] B. Baldassarri, J. He, A. Gopakumar, S. Griesemer, A. J. A. Salgado-Casanova, T.-C. Liu, S. B. Torrisi, and C. Wolverton, "Oxygen vacancy formation energy in metal oxides: High-throughput computational studies and machine-learning predictions," Chem. Mater. **35**(24), 10619–10634 (2023).

[40] A. Manzoor, G. Arora, B. Jerome, N. Linton, B. Norman, and D. S. Aidhy, "Machine learning based methodology to predict point defect energies in multi-principal element alloys," Front. Mater. **8**, 673574 (2021).

[41] A. Mannodi-Kanakkithodi, X. Xiang, L. Jacoby, R. Biegaj, S. T. Dunham, D. R. Gamelin, and M. K. Y. Chan, "Universal machine learning framework for defect predictions in zinc blende semiconductors," Patterns **3**(3), 100450 (2022).

[42] F. Tan, X. Liu, X. Liang, and Y. Cui, "Data-inspired atomic environment-dependence of vacancy formation energy in high-entropy alloys," Int. J. Plast. **195**, 104545 (2025).

[43] Y. Zhang and X. Xu, "Transformation temperature predictions through computational intelligence for NiTi-based shape memory alloys," Shape Mem. Superelasticity **6**(4), 374–386 (2020).

[44] Y. Zhang and X. Xu, "Solubility predictions through LSBoost for supercritical carbon dioxide in ionic liquids," New J. Chem. **44**(47), 20544–20567 (2020).

[45] I. Olanrewaju Alade, Y. Zhang, and X. Xu, "Modeling and prediction of lattice parameters of binary spinel compounds (AM 2 × 4) using support vector regression with Bayesian optimization," New J. Chem. **45**(34), 15255–15266 (2021).

[46] Y. Zhang and X. Xu, "Disordered MgB₂ superconductor critical temperature modeling through regression trees," Physica C **597**, 1354062 (2022).

[47] R. E. A. Goodall and A. A. Lee, "Predicting materials properties without crystal structure: Deep representation learning from stoichiometry," Nat. Commun. **11**(1), 6280 (2020).

[48] T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," Phys. Rev. Lett. **120**(14), 145301 (2018).

[49] G. Cong and V. Fung, "Improving materials property predictions for graph neural networks with minimal feature engineering," Mach. Learn. Sci. Technol. **4**(3), 035030 (2023).

[50] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials," npj Comput. Mater. **2**(1), 1–7 (2016).

[51] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, "Graph networks as a universal machine learning framework for molecules and crystals," Chem. Mater. **31**(9), 3564–3572 (2019).

[52] X. Fang, L. Liu, J. Lei, D. He, S. Zhang, J. Zhou, F. Wang, H. Wu, and H. Wang, "Geometry-enhanced molecular representation learning for property prediction," Nat. Mach. Intell. **4**(2), 127–134 (2022).

[53] P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer *et al.*, "Graph neural networks for materials science and chemistry," Commun. Mater. **3**(1), 93 (2022).

[54] S. R. Lee, A. F. Wright, N. A. Modine, C. C. Battaile, S. M. Foiles, J. C. Thomas, and A. Van der Ven, "First-principles survey of the structure, formation energies, and transition levels of as-interstitial defects in InGaAs," Phys. Rev. B **92**(4), 045205 (2015).

[55] J. Schmidt, L. Pettersson, C. Verdozzi, S. Botti, and M. A. L. Marques, "Crystal graph attention networks for the prediction of stable materials," Sci. Adv. **7**(49), eabi7948 (2021).

[56] V. Fung, J. Zhang, E. Juarez, and B. G. Sumpter, "Benchmarking graph neural networks for materials chemistry," npj Comput. Mater. **7**(1), 84 (2021).

[57] S.-Y. Louis, Y. Zhao, A. Nasiri, X. Wang, Y. Song, F. Liu, and J. Hu, "Graph convolutional neural networks with global attention for improved materials property prediction," Phys. Chem. Chem. Phys. **22**(32), 18141–18148 (2020).

[58] A. Zunger, S.-H. Wei, L. G. Ferreira, and J. E. Bernard, "Special quasirandom structures," Phys. Rev. Lett. **65**(3), 353–356 (1990).

[59] G. Kresse and J. Hafner, "Ab initio molecular dynamics for liquid metals," Phys. Rev. B **47**(1), 558–561 (1993).

[60] M. Methfessel and A. T. Paxton, "High-precision sampling for Brillouin-zone integration in metals," Phys. Rev. B **40**(6), 3616–3621 (1989).

[61] H. J. Monkhorst and J. D. Pack, "Special points for Brillouin-zone integrations," Phys. Rev. B **13**(12), 5188–5192 (1976).

[62] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**(18), 3865–3868 (1996).

[63] P. E. Blöchl, "Projector augmented-wave method," Phys. Rev. B **50**(24), 17953–17979 (1994).

[64] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," Phys. Rev. B **59**(3), 1758–1775 (1999).

[65] D. Jha, L. Ward, A. Paul, W.-K. Liao, A. Choudhary, C. Wolverton, and A. Agrawal, "Elemnet: Deep learning the chemistry of materials from only elemental composition," Sci. Rep. **8**(1), 17593 (2018).

[66] L. Ward, A. Dunn, A. Faghaninia, N. E. R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla et al., "Matminer: An open source toolkit for materials data mining," Comput. Mater. Sci. **152**, 60–69 (2018).

[67] L. C. Allen, "Electronegativity is the average one-electron energy of the valence-shell electrons in ground-state free atoms," J. Am. Chem. Soc. **111**(25), 9003–9014 (1989).

[68] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, and S. Alvarez, "Covalent radii revisited," Dalton Trans. **2008**(21), 2832–2838 (2008).

[69] D. G. Pettifor, "A chemical scale for crystal-structure maps," Solid State Commun. **51**(1), 31–34 (1984).

[70] S. Liu and C. Yang, "Machine learning design for high-entropy alloys: Models and algorithms," Metals **14**(2), 235 (2024).

[71] C. Tandoc, Y.-J. Hu, L. Qi, and P. K. Liaw, "Mining of lattice distortion, strength, and intrinsic ductility of refractory high entropy alloys," npj Comput. Mater. **9**(1), 53 (2023).

[72] X. Zhao, Z. Wei, J. Zhao, Y. Jia, S. Cao, D. Wang, and Y. Lei, "Machine-learning-assisted design of novel TiZrNbVAl refractory high-entropy alloys with enhanced ductility," Metals **14**(8), 894 (2024).

[73] B. Medasani, A. Gamst, H. Ding, W. Chen, K. A. Persson, M. Asta, A. Canning, and M. Haranczyk, "Predicting defect behavior in B2 intermetallics by merging ab initio modeling and machine learning," npj Comput. Mater. **2**(1), 1 (2016).

[74] P. A. Korzhavyi, I. A. Abrikosov, B. Johansson, A. V. Ruban, and H. L. Skriver, "First-principles calculations of the vacancy formation energy in transition and noble metals," Phys. Rev. B **59**(18), 11693 (1999).

[75] S. Zheng and S. Wang, "First-principles design of refractory high entropy alloy VMoNbTaW," Entropy **20**(12), 965 (2018).

[76] S. Giusepponi and M. Celino, "The effects of vacancies in the mechanical properties of tungsten: A first-principles study," Nucl. Instrum. Methods Phys. Res., Sect. B **342**, 70–75 (2015).

[77] X. Zhou, S. He, and J. Marian, "Vacancy energetics and diffusivities in the equiatomic multielement Nb-Mo-Ta-W alloy," Materials **15**(15), 5468 (2022).

[78] GitHub, https://github.com/shuozhixu/JAP_2026

12 January 2026 13:43:23