論文 / 著書情報
Article / Book Information

Title	Hydrogen permeability prediction in palladium alloys and virtual screening of B2 phase-stabilized Pd(100-x-y)CuxMy ternary alloys using machine learning
Authors	Eric Kolor, Edoardo Magnone, Muhammad Harussani Moklis, Md. Rubel, Sasipa Boonyubol, Koichi Mikami, Jeffrey S. Cross
Citation	Materials Today Communications, vol. 51, , Page 114875
Pub. date	2026, 2
DOI	https://dx.doi.org/10.1016/j.mtcomm.2026.114875
Creative Commons	Information is in the article.

Materials Today Communications 51 (2026) 114875

Contents lists available at ScienceDirect

Materials Today Communications

journal homepage: www.elsevier.com/locate/mtcomm

Hydrogen permeability prediction in palladium alloys and virtual screening of B2 phase-stabilized $Pd_{(100-x-y)}Cu_xM_y$ ternary alloys using machine learning

Eric Kolor [a,*], Edoardo Magnone [b], Muhammad Harussani Moklis [a], Md. Rubel [a], Sasipa Boonyubol [a], Koichi Mikami [a], Jeffrey S. Cross [a]

[a] Energy Science and Engineering, Department of Transdisciplinary Science and Engineering, Institute of Science Tokyo, 2−12−1, Ookayama, Meguro-ku, Tokyo 152−8550, Japan
[b] Department of Chemical and Biochemical Engineering, Dongguk University, Wonheung-gwan F619, 30, Pildong-ro 1 gil, Jung-gu, Seoul, 100-715, South Korea

ARTICLE INFO

Dataset link: Pd-membranes-permeability

Keywords:
Hydrogen-selective metallic membranes
Hydrogen permeability
Palladium–copper alloys (Pd–Cu)
B2 phase stabilization
Materials informatics
High-throughput virtual screening

ABSTRACT

Ordered B2 phase Pd–Cu alloys are promising candidates for dense metallic membranes for high-temperature hydrogen purification, but their practical deployment is hindered by thermally induced disordering to the *fcc* phase, which degrades hydrogen permeability. Here, we develop a machine-learning–assisted materials-screening framework to identify hypothesis-generating Pd–Cu–M ternary alloys with improved B2-phase stability and competitive hydrogen transport properties. CatBoost regressors were trained on literature-derived hydrogen permeability data restricted to bulk diffusion–controlled regimes, using composition-based descriptors combined with operating conditions. Multiple descriptor families were evaluated through systematic ablation, and a feature-selection strategy integrating Pearson filtering with fold-wise SHAP-driven recursive feature elimination was employed. Guided by the one-standard-error rule, a compact, domain-informed set of 13 features achieved a favorable accuracy–parsimony balance (R^2= 0.81), within 0.01 of the maximum performance obtained using 3 times more features. Model interpretability analysis indicates that high apparent permeability correlates with elevated temperature, lattice expansion relative to Pd, increased atomic size mismatch, and favorable alloy mixing tendencies. To enable responsible virtual screening, a *k*-nearest-neighbor applicability-domain analysis was combined with multi-objective Pareto optimization to distinguish interpolative predictions from extrapolative hypotheses. Within the model's applicability domain, several Pd–Cu–M systems at low dopant concentrations exhibit predicted permeabilities (1.06 – 1.09 $\times 10^{-8}$ mol·m^{-1}·s^{-1}·Pa$^{-0.5}$) comparable to the B2 $Pd_{47.25}Cu_{52.75}$ benchmark. Post hoc density functional theory calculations on selected compositions further support the stabilizing influence of Nb and Cr at dilute levels, and Y at higher concentrations, through reduced ground-state formation enthalpy. Overall, this work provides experimentally relevant targets and a scalable, data-informed pathway for the rational design of thermally stable Pd-based hydrogen separation membranes.

1. Introduction

Palladium (Pd) alloys possess uniquely active catalytic surface for dissociating and recombining H_2, and bulk lattice that accommodate facile thermally activated protons hops between interstices, making them promising metallic membrane candidates to unlock fuel cell-grade hydrogen generation. Yet, a central challenge is to reduce Pd content to lower membrane cost without compromising performance, while simultaneously tailoring material properties for specific applications [1]. Palladium−copper (Pd–Cu) membranes set a pragmatic compromise: Cu lowers cost, improves corrosion resistance, thermal expansion matching, tensile strength, sulfur-poisoning and embrittlement resistance [2]. However, near-equiatomic Pd–Cu alloys with an ordered body-centered cubic (*bcc*) structure referred to as B2 and associated with the highest theoretical permeability values and catalytic activity, undergoes a B2 → B2 + disordered face-centered cubic (*fcc*) or A1 transition upon H_2 uptake and heating, with substantial A1 formation by ~ 400–598 °C [3,4]. Maintaining stable B2-like surface chemistry and high

* Corresponding author.
E-mail address: kolor.k.aa@m.titech.ac.jp (E. Kolor).

https://doi.org/10.1016/j.mtcomm.2026.114875
Received 29 December 2025; Received in revised form 11 February 2026; Accepted 16 February 2026
Available online 17 February 2026

permeability coefficient upon temperature cycling is highly demanded [5,6].

The straightforward approach to stabilize the B2 phase is co-doping by judicious addition of a third metal, M, chosen to balance activity, stability, and mechanical properties. A typical eligible third metal M should stabilize B2 Pd–Cu–M by a mechanism of lowering the Gibbs free energy of the ordered state relative to disorder (B2 → A1), through a combination of stronger site-specific bonding, size-strain accommodation, reduced configurational entropy gain upon disordering, and non-trivial site-substitution effects [7]. Forming such B2 phase-stabilized Pd–Cu–M substitutional solid solution suggests probing a large design space: which elements improve B2 stability, prefer Pd or Cu sites substitution, which concentration should be added not to alter structure, and which trade-offs emerge (sulfur tolerance, hydrogen dissociation, ductility, CO resistance, etc.)? Prior works illustrated both opportunities and gaps. Earlier, patent by Benn et al. suggested by using Hume-Rothery rules and density functional theory (DFT)-based heat of formation criteria that the B2 phase in $Pd_{[35-55]}Cu_xM_{[1-15]}$ alloys could be stabilized using a *bcc* metal M ∈ {Fe, Cr, Nb, Ta, V, Mo, W} [8]. Likewise, U.S. NETL researchers proposed similar workflow for screening 37 metallic elements which could stabilize B2 $Pd_{8-x}Cu_8M_x$ and suggested M ∈ {Al, Ga, Hf, La, Mg, Sc, Ti, Y, Zn, Zr} as promising candidates [9–11]. The latter additionally reported experimentally that Mg might be the best B2 phase stabilizer at 400 ºC [11]. Experimental membrane-based Japanese patents have also reported dilute additions of M ∈ {Al, Ga, In} to form $Pd_{[41-50]}Cu_{[48-58.8]}M_{[0.2-2.0]}$ that maintain permeability $> 1.0 \times 10^{-8}$ mol·m^{-1}·s^{-1}·Pa$^{-0.5}$ above 600 °C [5,6]. Most recently, after showing that bespoke heat treatment can shift the permeability of B2 $Pd_{61}Cu_{39}$ (wt%) to 1.4 × that of B2 $Pd_{60}Cu_{40}$ at 300 °C, Horikawa & Ogawa et al. demonstrated that dilute Cu sites substitution in $Pd_{61}Cu_{39}$ by Mn and Al could yield B2 phase stabilized alloys at T ≥ 450 °C [3,12]. Yet, except few studies [5,6,12], the expected permeability ranges of B2-stabilized Pd–Cu–M alloys remain largely unmapped because this requires extensive syntheses, characterizations, testing, and *ab initio* simulations. Material informatics can offer a complementary route. Yang et al. recently evidenced this by showcasing a combined DFT and ML workflow to evaluate lattice–gas interactions [13]. Using 30 features spanning simple composition-based descriptors and DFT-derived quantities, they reported eXtreme Gradient Boosting (XGBoost) and Gradient Boosting Decision Trees (GBDT) as the strongest selectivity ($R^2 = 0.96$) and permeance ($R^2 = 0.99$) predictors, respectively, with the high scores arguably reflecting the homogeneity of the DFT-optimized structures.

In this work, we addressed the absence of reliable data-driven tools for predicting hydrogen permeability in experimental Pd alloys and for screening large pools of Pd alloy candidates. First, we constructed a dataset of experimentally observed Pd alloy membranes. Second, we trained an interpretable small-data-friendly algorithm *viz*. Categorical Boosting (CatBoost) to model the non-linear composition–testing condition–permeability relationship. Next, we used the model to map composition–permeability domains for 16 potential B2-stabilizing elements, revealing Pd-lean potential B2 phase-stabilized Pd–Cu–M candidates suitable for experimental validation. Lastly, we used DFT to analyze the ground-state stability of selected candidates.

2. Methodology

Throughout this work, we refer to our target variable as apparent hydrogen permeability rather than the intrinsic hydrogen permeability coefficient. The distinction between these membrane properties will be elucidated in the next sections.

2.1. Data curation

We manually assembled a tabular dataset enriched with sufficient metadata, covering experimentally observed dense crystalline Pd-alloy membranes and their operating conditions. To construct the dataset, we retained 71 sources that provided sufficiently complete experimental Material, ensuring minimal missing values aside from the lattice parameter. These sources included peer-reviewed journal articles, technical reports and patents, from which we compiled 333 distinct alloy compositions. We converted each alloy formula to atomic fractions and, whenever available, prioritized the experimentally determined compositions from solid-state characterization over the nominal targeted ratios. The analysis focused exclusively on pure H_2 permeation experiments (single-gas experiments) in non-composite dense alloy membranes, mostly planar and selected tubular configurations without any chemically bound porous substrate underlying the active diffusion area. Those non-composite Pd alloy dense membranes were reported to be mostly fabricated by arc melting/cold rolling, sputtering deposition on *e.g.* silicon wafers, electroless plating/peeling-off. We also have not accounted for heterogeneity and nuisance variables such as support-induced stress, membrane surface roughness, or different sealing methods which could influence permeability, because such information is unstructured and hardly quantifiable. We have also refrained to treat those as categorical variables to avoid biases. Instead, we mitigated the variability in experimental data using statistical strategies that will be presented next.

For each membrane sample, we extracted the experimental thickness, permeation temperature, pressure differentials $\left(\Delta P^n = P^n_{feed} - P^n_{permeate}\right)$ at which hydrogen fluxes (J_{H_2}) were measured, and the pressure exponent (n), which characterizes the rate-limiting step in hydrogen diffusion. Most flux data were obtained by digitizing $J_{H_2} - \Delta P^n$ plots using WebPlotDigitizer [14]. Lattice constants were either (*i*) taken directly when reported, or estimated (160 missing lattice parameters values, or 48 % of the unique alloys count) using (*ii*) non-linear least-squares refinement with UnitCell software [15,16], or (*iii*) Vegard's law when explicit or sufficient XRD data were unavailable. Specifically, for binary Pd–Cu alloys, cubic lattice constants were instead estimated using the B2 and *fcc* Pd–Cu specific Vegard's linear regression equations developed by Al-Mufachi et al. [17], applied to the relevant subdomain of the Pd–Cu phase diagram (Table S2) [4,18,19]. Experimental lattice parameters, when used, were obtained from independent structural characterization (*e.g.*, XRD), typically performed prior to permeability testing, and are therefore independent of the target permeability measurements. More details on lattice parameters calculations are provided in Supplementary Material.

Hydrogen permeability coefficient of dense metallic membranes is in theory an intrinsic material property defined as the product of the solubility and diffusion coefficients at a specific temperature. For its practical estimation, hydrogen fluxes (J_{H_2}) are first measured under varying pressure differentials (ΔP^n) at fixed temperature, followed by linear regression of J_{H_2} versus ΔP^n. Then, an approach is to take the slope of the regression line with the n value that maximizes the coefficient of determination R^2 which gives the permeance (permeability coefficient normalized by thickness) over a pressure range according to Richardson's equation [2,20]. In this work, we decided to maximize dataset size and to capture the full range of experimental flux values by taking permeability pointwise directly from reported fluxes and pressure differentials according to Eq. (1):

$$\text{Apparent Permeability}(\phi_{app}) = \frac{J_{H_2} \quad \times \text{Thickness}}{\Delta P^n} \tag{1}$$

The pressure exponent (n) was adopted from the values reported by their original authors, typically the exponent that maximized the coefficient of determination (R^2) in the $J_{H_2} - \Delta P^n$ regression. We chose to refer to the so-obtained pointwise permeability values as the apparent permeability, which should ideally be mathematically equivalent to the slope-derived values, but might implicitly inherit real-world experimental noise due to heterogeneous nature of the data. Although point-

wise (apparent) permeability (Eq. 1) is more sensitive to experimental noise than the slope-derived value, it allowed us to retain all available datapoints, thereby increasing the number of observations per membrane and enhancing the statistical robustness of the machine learning models. An aggregation-based robustness analysis (one record per alloy) is performed as sensitivity test (details in Supplementary Material).

2.2. Data processing

Only data points exhibiting a pressure exponent $n = 0.5$ were retained, corresponding to bulk-diffusion-controlled hydrogen transport. Measurements with $n \neq 0.5$, indicative of surface-limited or mixed-control regimes, were excluded. Therefore, the transport phenomena modelled in this paper obey the Sieverts' law. The latter law states that hydrogen solubility in dense bulk alloy membranes is proportional to the square root of its partial pressure. Consequently, n becomes a zero-variance variable, thus carried no information and was disregarded as feature.

Although the distribution of the point-wise permeability is left-skewed (Fig. S1, Supplementary Material) due to outlier data points that we deemed scientifically plausible, we confirmed that removing those data points is deleterious for further prediction. For example, we conserved copper-rich binary membranes with zero permeability coefficient [21], and iron-rich alloys with $\sim 10^{-14}$ $mol \cdot m^{-1} \cdot s^{-1} \cdot Pa^{-0.5}$ permeability [22]. We confirmed that general methods to handle skewed distribution (Yeo-Johnson and Box-Cox power transformations) have failed even after removing very low-magnitude permeability data points. In this work, we predicted $\log(1 + \textit{Apparent Permeability})$ where *Apparent Permeability* is the pointwise recomputed permeability (Eq. 1), but we always reported the results on the original scale. Features were systematically robust-scaled to temper outliers and mitigate scale variability, although gradient boosted decision trees models like CatBoost do not necessarily require any feature scaling nor target values transformations (see '3. Data scaling' section in the Supplementary Material for details).

2.3. Composition-aware data partitioning with column-similarity stratification

We grouped identical alloy formulas and stratified the obtained groups by column similarity in the periodic table. Grouping places all observations of the same canonical composition in one partition, enforcing mutual exclusivity and preventing leakage across training, cross-validation, and held-out test sets. Stratification exploits a column-similarity encoding using ascending IUPAC group numbers (e.g., $Pd_{91.12}Ti_{8.88} \rightarrow$ G4 −G10) to reflect shared valence-electron chemistry, preserve composition-family frequencies in train/test splits, and reduce covariate shift in a regression setting. Rare strata with fewer than 20 instances were merged into the "Others" bin (Fig. 1). Pure palladium (G10) served as a calibration reference during training and was excluded

Table 1
Nominal and actual split sizes for the training set and the held-out test set.

Nominal size	Actual size	Train (unique | samples)	Test (unique | samples)
80/20	80/20	258 | 1788	70 | 435
85/15	85/15	276 | 1883	52 | 340
90/10	~95/5	299 | 2092	29 | 131

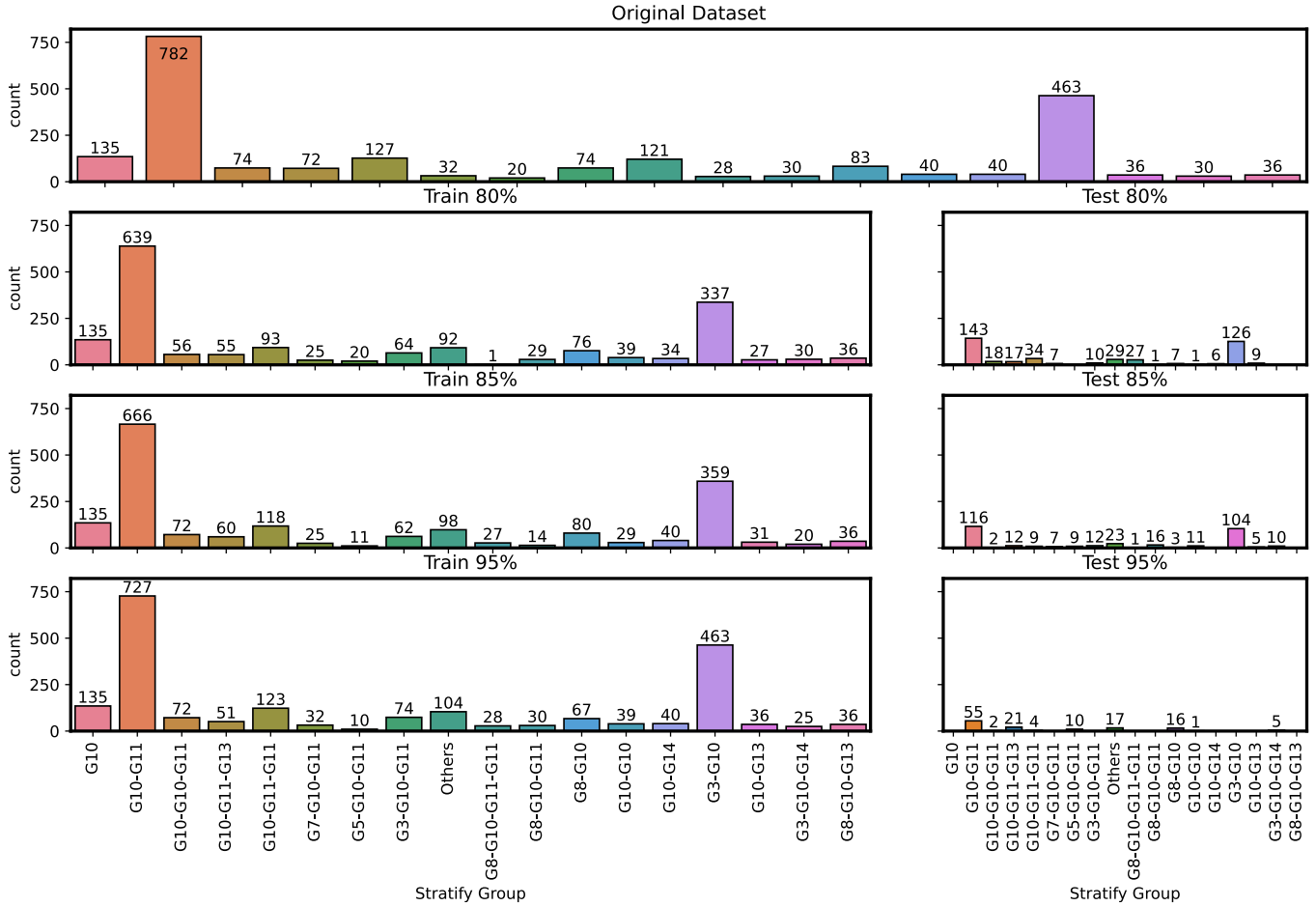

Fig. 1. Counts by stratum in the original dataset and across 80/20, 85/15, ~95/5 training and held-out test splits.

from test scoring. Splits of 80/20, 85/15, and ca. 95/5, in addition to distribution conservation across the various data sizes, are summarized in Table 1 and Fig. 1, respectively.

2.4. Feature engineering

We evaluated four descriptor families and their combinations. For simplicity, we referred to these by their abbreviated name. "Exp" captures experimental conditions (see Section 2.1) and also includes a physics-motivated lattice-misfit term given by:

$$\Delta a_{ss}/a_{Pd} = (a_{alloy} - a_{Pd}) / a_{Pd} \quad (2)$$

which quantifies Pd lattice dilation or contraction upon alloying. "Bond" targets alloy phase formation and bond characteristics using established descriptors [23–27]. Mean atomic packing efficiency (Mean APE) was computed with Miracle radii (Matminer-consistent); all other radius-based terms used metallic radii at coordination 12, consistent with crystalline solid solutions and Magnone et al. [24]. "CBFV" represents composition via rule-of-mixtures properties given by:

$$Property_{alloy} = \sum_{i=1}^{n} c_i P_i \quad (3)$$

where c_i are the elemental fractions and P_i the elemental property tables which we took from the 2024 Oliynyk dataset [28]. "Elemental" encodes elemental fractions, allowing the model to learn nonlinear interactions directly from stoichiometry. The full features list appears in Table S1 (Supplementary Material).

We formed pairwise, triplet, and quartet combinations to probe complementarity between physics-informed and composition-only Material: Exp_Bond, Exp_CBFV, Exp_Elemental, Exp_Bond_CBFV, Exp_Bond_Elemental, Exp_CBFV_Elemental, and Exp_Bond_CBFV_Elemental. Table 2 reports descriptor counts per set. This ablation design quantifies the incremental value of experimental context, bond and phase-formation descriptors, and composition-based vectors for permeability prediction.

2.5. Model development and validation

Given the heterogeneous, literature-derived nature of the dataset and the objective of robust, interpretable screening rather than explicit mechanistic law discovery, a gradient-boosted tree model was selected as a pragmatic and conservative choice. This motivated the use of CatBoost, which employs gradient boosting with symmetric ("oblivious") decision trees, to capture non-linear relationships between descriptors and apparent permeability, as evidenced by its strong predictive performance in recent composite materials studies [29]. CatBoost was selected because: (1) ordered boosting reduces prediction shift and target leakage; (2) symmetric depth-limited trees provide strong built-in regularization on small tabular datasets; (3) the library offers native handling of missing values and robust training controls (learning-rate shrinkage, L2 regularization, subsampling, early stopping) to limit overfitting [30]; (4) tree ensembles are comparatively insensitive to feature collinearity; (5) and CatBoost's regularization further mitigates variance when many descriptors are included.

Table 2
Type and cardinality of primary feature sets.

Feature set	Number of features
Exp	4
Bond	18
CBFV	75
Elemental	40

2.5.1. Modelling details

Our modeling pipeline follows a two-stage approach: (1) a diagnostic phase to evaluate the impact of all features, dataset sizes, and cross-validation folds without initial feature selection; and (2) a production phase employing feature selection strategies to identify robust models that balance predictive performance, and complexity. The models were trained without explicit phase labels; phase-related behavior is inferred indirectly through physically motivated descriptors such as lattice misfit. Throughout this work, model performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination R^2 (see the "4. Model Evaluation" section in the Supplementary Material for details).

In the first strategy, CatBoost regressors (CatBoost version 1.2.8) were trained directly on each raw feature set (see Section 2.4) to assess the model's robustness to multicollinearity and high-dimensional input spaces. A random seed of 42 is selected during each procedure described in this work. Model hyperparameters were optimized through randomized search using a stratified group K-fold cross-validation scheme with $k \in \{3, 5, 10\}$ to accommodate unequal group sizes. The folds were stratified based on the periodic group similarity encoding described earlier to preserve composition-family distributions, and mutual exclusivity of formula groups across folds was ensured. Early stopping was applied as a regularization technique to prevent overfitting, halting training when the validation error ceased to improve for 20 consecutive iterations. During hyperparameter optimization, the same held-out validation fold was used as internal evaluation set for both tracking the loss function and early stopping, without data leakage [31]. The initial number of estimators was set to 10,000 as an upper limit with shrinkage, while the effective model size was determined automatically by early stopping. Finally, after optimization, the model was retrained on the complete training set using the selected hyperparameters and the median optimal number of estimators identified during early stopping. The external test set was kept untouched until final evaluation.

In the second strategy, a two-stage feature selection scheme was applied to reduce multicollinearity and dimensionality, producing more compact and interpretable feature sets. In the first stage, feature redundancy was analyzed by calculating pairwise Pearson correlation coefficients. A threshold of $|r| > 0.90$ was used to identify strongly correlated pairs. When two features met this criterion, one was retained based on domain knowledge, prioritizing properties historically associated with hydrogen permeation in metals. Detailed information regarding feature filtering is provided in the "5. Pearson Correlation Filtering" section of the Supplementary Material. After reducing multicollinearity using the $|r| > 0.90$ filter, dimensionality was further reduced through a two-step procedure: A fold-optimized recursive feature elimination (RFE) guided by SHapley Additive Explanations (SHAP) values, followed by union aggregation of the top-ranked features across folds. The RFE was implemented using CatBoost's built-in *RecursiveBySHAP* algorithm and applied independently to each fold to determine the optimal number of important features identified by the SHAP framework. SHAP quantifies the contribution of each descriptor to model predictions, providing a consistent feature importance measure [32]. After RFE, a tree-specific SHAP analysis (*TreeSHAP*) was recomputed for each fold to rank descriptors from most to least influential.

For each target subset size, $m \epsilon \{10, 12, 15, 17, 20, 22, 25, 30\}$, the top m features were selected within each fold, and their union was taken across folds to form an aggregated feature set for that m. Because fold-specific lists overlapped only partially, the aggregated set size was typically close to m. Each aggregated feature set was then used to train CatBoost models with default hyperparameters for comparison. During training, early stopping with cross-validation (patience = 20) was applied to determine the optimal number of estimators per fold. The median value of these optimal estimators was subsequently used to retrain the model on the full training partition. The external test set remained untouched until the final prediction stage.

2.5.2. Selection of final parsimonious model

Each aggregated feature set produced a separate CatBoost regressor, resulting in multiple models that differed in feature set composition and number of estimators. Test predictions were evaluated using block bootstrap to estimate uncertainty in the performance metrics. Details on block bootstrap are available in section '10. Confidence interval' of the Supplementary Material. The one-standard-error rule was then applied to select the most parsimonious model whose mean absolute error (MAE) was within one standard error of the minimum MAE [33].

3. Results and discussions

3.1. Dataset for modeling

After preprocessing, the final dataset contains 2223 records spanning 328 unique canonical compositions. By number of components, the set includes pure Pd, 186 binaries, 131 ternaries, 7 quaternaries, 0 quinary, 1 senary, and 2 septenary alloys. Thickness ranges from 1 μm to 1100 μm, temperature from ≈ 292 K to ≈ 1175 K, and apparent permeability from 0 to 9.54 × 10^{-8} mol·m^{-1}·s^{-1}·$Pa^{-0.5}$. Details of individual parameter distributions are provided in Fig. S1 (Supplementary Material). A dominant part of the dataset contains Pd-rich substitutionally random *fcc* solid solutions, while B2 and mixed *fcc*/B2 alloys were always Cu-containing alloys. However, although *fcc* records dominate the data set, group-aware validation and phase-sensitive descriptors mitigate trivial *fcc* bias in the screening results. The elemental space covers 40 elements including transition metals (3d–5d), a metalloid (B), selected post-transition metals, and lanthanides. Fig. 2 depicts a pictorial summary: panel A maps the elemental coverage, and panel B presents the distribution of Bravais lattice types based on column-similarity grouping. Due to the skewness and imbalance observed in the data distribution, chemically-oriented alleviating statistical transformations were carried out prior to modelling.

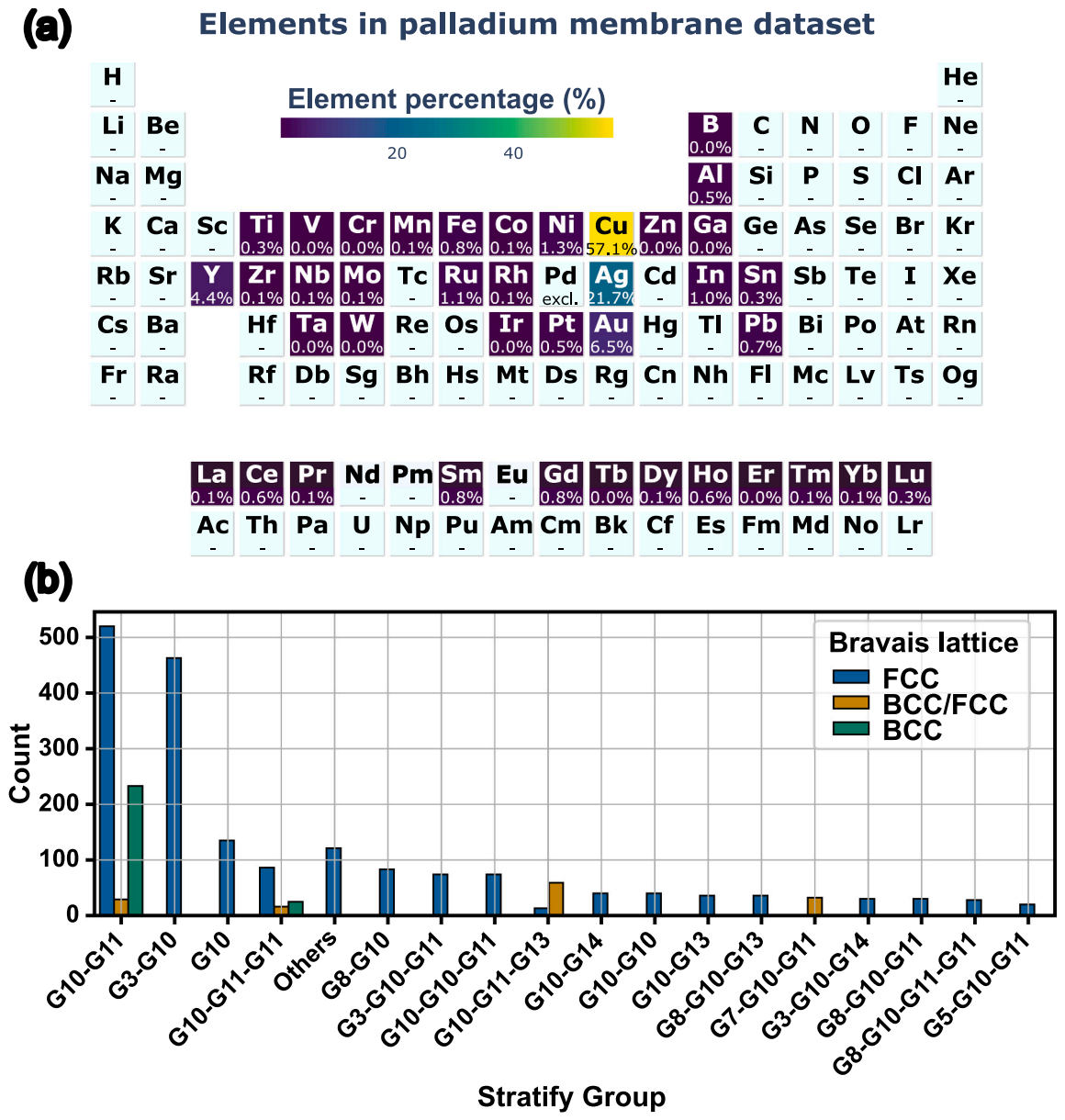

Fig. 2. (a) Elemental composition in the palladium alloy membrane dataset (plot generated using pymatviz [34]); percentages exclude Pd, the main element. Cells with 0.0 % indicate values < 0.1 %. (b) Bravais lattice type by stratum.

3.2. Column similarity encoding and stratified data splitting for regression

The mapping analysis (Fig. 2a and Fig. 2b) shows that the dataset is strongly imbalanced: four transition metals dominate (Cu >> Ag > Au > Y), with other elements sparsely represented. Lower-component alloys are overrepresented (binaries and ternaries >> higher-order systems). Copper alone accounts for 57.1 % of the entries, reflecting both the search for low-cost, sulfur-resistant membranes and the incremental adjustments traditionally applied to optimize Pd−Cu compositions and operating conditions. Such redundancy and imbalance are common in materials datasets, and a model trained directly on them risks misleading performance estimates, particularly for out-of-distribution tasks such as predicting underrepresented systems or alloys containing elements outside the current chemical space. Similar concerns about inflated accuracy from random splits in redundant inorganic datasets have been raised in prior works [35,36].

To mitigate these biases, we introduced a chemistry-aware partitioning scheme (Section 2.3). Elements were encoded by their column position in the periodic table, and the so obtained data were stratified to enforce approximate uniformity of encoded families across training, validation, and test sets. This design encourages the model to transfer learning across systems with similar valence-electron chemistry, improving robustness to underrepresented alloys. For example, as shown in Fig. 2b, Pd−(Cu, Ag, Au) alloys can be grouped as G10 −G11, and Pd−(Y, lanthanides) alloys as G3–G10, thereby partially rebalancing *fcc*-rich families and forcing the model to treat them as comparable chemical groups.

The data encoding and stratification approach introduced reduces covariate shift by maintaining similar encoded distributions across partitions, where covariate shift refers to cases in which the distributions of training and test inputs differ while the conditional output distribution remains unchanged [37]. Although residual shifts may persist in non-compositional variables such as thickness, the column-similarity encoding provides a more chemically meaningful and balanced split than random or purely stratified schemes. A systematic benchmarking against other material-specific splitting methods is beyond the scope of

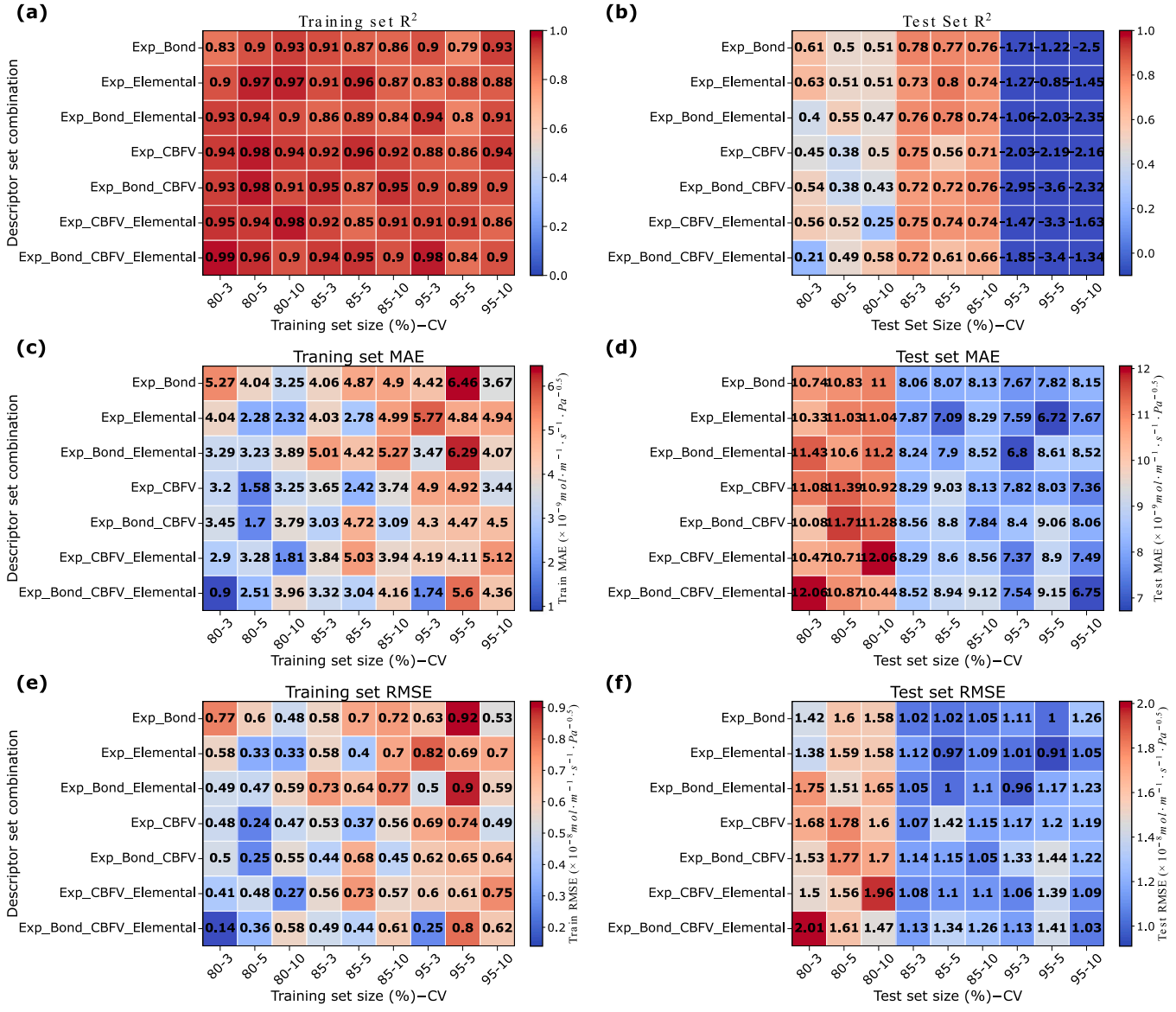

Fig. 3. (a) – (f) Fitting and prediction performance on the training and held-out test sets (without feature selection). The left panels show training results; the right panels show test results. For each descriptor set (*y*-axis), a cell corresponds to a specific train/test split ratio and cross-validation configuration (*x*-axis). In each grid, the descriptor sets are arranged from the lowest cardinality set to the highest (top-to-bottom). The negative R^2 values for the ~95/5 splits are kept for diagnostic purpose and are not be intended for comparison.

this work, but will be valuable in future studies.

3.3. Diagnostic of model performance without feature selection

In materials property prediction, two strategies are commonly adopted: a model-centric approach, which compares the performance of multiple algorithms with minimal preprocessing, and a feature-centric approach, which emphasizes advanced feature engineering with a single model to maximize predictive accuracy. We adopted the latter, selecting CatBoost regression under the heuristic that predictive power is primarily locked within the representation of the data rather than the model itself. This is consistent with the "no-free-lunch" theorem, which states that no algorithm universally outperforms others across all problems, and that accuracy can be improved through appropriate feature design [38]. Recent studies in materials informatics have also demonstrated the benefit of carefully crafted descriptors over brute-force model comparisons [39,40].

After hyperparameter tuning, the performance of the descriptor sets was compared across 80/20, 85/15, and ~ 95/5 train/test splits under identical stratified group K-fold cross-validation ($k \in \{3, 5, 10\}$). The performance metrics employed are complementary: MAE ($mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}$) average absolute distance between the actual and predicted values and is less sensitive to extreme values; RMSE ($mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}$) emphasizes large errors and is preferred when outliers are exponentially rare; R^2 (unitless) quantifies explained variance. For a physically sound permeability prediction, both MAE and RMSE should be comparable or lower than the experimental uncertainties, which cannot be readily obtained for our multi-source dataset.

We arranged the results obtained at this stage in Fig. 3 ((a)–(f)), where prediction results on the unseen test set are vertically stacked at right and those on the training set on the left. Training R^2 values ranged from 0.79 to 0.99, indicating good fitting capacity. However, the 95/5 split produced deceptively low MAE and RMSE but consistently negative R^2 on the held-out test set. This is caused by the very small test partition, limited target variance and the fact that the compositions in the 5 % test set are rare and do not cover a representative part of the compositional distribution of the 95 % training set. Thus, 95/5 is interpreted as a diagnostic split and is not considered for comparison (more details in Supplementary Material, section 9). The 80/20 split yielded the highest overall errors, reflecting an unfavorable bias–variance trade-off. The 85/15 split provided the most reliable balance: sufficient training size, a large enough test block for stable estimates, and good distributional concordance across column-similarity strata. We therefore adopted 85/15 for subsequent analyses.

Within the 85/15 split, no consistent dependence was observed between scoring metrics and the number of *k*-folds. Importantly, combined MAE and RMSE values fall within the interquartile range of the full apparent permeability distribution ($[5.90 \times 10^{-9}$, 2.57×10^{-8} $mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}]$), confirming that the models capture the correct magnitude of the property space. However, prediction quality remains unsatisfactory for very low-permeability alloys such as Pd–Fe and Cu-rich *fcc* Pd–Cu [21,22], which lie in the skewed distribution tail and are likely treated as outliers. This is reflected in RMSE values being an order of magnitude larger than MAE, consistent with the higher sensitivity of RMSE to outliers. For context, most apparent permeability values in the dataset range from 10^{-9}–10^{-8} $mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}$, while peak values for Pd-based membranes (composite or self-standing) typically lie near 10^{-8} [34].

Among the descriptor families, the simplest encoding (Exp_Elemental) achieved the strongest held-out performance ($R^2 = 0.80$, MAE = 7.09×10^{-9} $mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}$, RMSE = 0.97×10^{-8} $mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}$, based on 5-fold CV). This is unexpected, given fractional stoichiometry representations are the simplest and are often regarded as more efficient in large datasets, as shown in the ElemNet deep-learning system [41]. In contrast, the domain-grounded Exp_Bond set (22 descriptors) delivered stable performance across each CV folds [$k = 3$ (R^2_{train}=0.91, R^2_{test}=0.78), $k = 5$ (0.87, 0.77), $k = 10$ (0.86, 0.76)], underscoring the robustness of alloy phase-formation descriptors. These findings align with Magnone et al. (2023), who emphasized the predictive value of valence electron concentration (VEC), electronegativity difference, mixing entropy, and atomic radius difference, and with Yan et al. (2025), who applied similar principles to design Nb-based membranes [24,26]. Overall, Fig. 3 highlighted a trade-off: Exp_Elemental provides the highest raw accuracy but limited generalization power, while Exp_Bond offers consistent, interpretable predictions grounded in alloy chemistry. The Exp_Bond_Elemental combination strikes a practical middle ground, capturing the benefits of both. These results justify our focus on domain-knowledge-containing descriptor sets for subsequent feature selection and screening analyses.

From the ablation tests conducted, it is concluded that a data split size of 85/15 with a feature set formed by experimental and bond properties are promising for modelling apparent hydrogen permeability.

3.4. Dimensionality–complexity reduction

After identifying the 85/15 split as the most stable partition (Section 3.3), we next sought to balance accuracy and model simplicity. Each feature set was first filtered for redundancy ($|r| \geq 0.90$) and then subjected to fold-wise SHAP-based recursive feature elimination with union aggregation across folds. This procedure yielded 135 models differing in cross-validation scheme, feature pool, and subset size, enabling a systematic assessment of the performance−complexity trade-off.

Fig. 4 shows the evolution of R^2 on the held-out test set as a function of feature count. It can be seen that parsimonious sets consistently achieved higher R^2 than larger ones, underscoring the value of compact, physics-informed descriptors. The best models reached $R^2_{test} = 0.82$, including: (*i*) 38 features from the Exp_Bond_CBFV_Elemental pool (3-fold CV), (*ii*) 15 features from Exp_Bond (3-fold CV, $R^2_{train,mean} = 0.94$), and (*iii*) 19 features from Exp_Bond_CBFV (3-fold CV). Relative to predictions without feature selection, these models showed reduced overfitting and slightly improved generalization, while requiring fewer features. Notably, the 15-feature Exp_Bond set delivered equivalent test performance to larger sets, highlighting the sufficiency of alloy phase-formation descriptors. Our selected subset includes 15 of the 20 alloy chemistry parameters identified by Wen et al. (2019) [25], extending beyond those emphasized by Magnone et al. [24,25]. It should be emphasized that although models using elemental fractions as direct inputs performed strongly, they can struggle to generalize to alloys containing elements outside the training space. Such models are therefore better suited for interpolation (prediction for an element in the same elemental space) within known systems rather than out-of-distribution prediction.

Overall, these results demonstrate that rigorous feature selection enables reduction of model complexity without loss of performance, and even modest gains, re-emphasizing the use of compact domain-informed descriptors for virtual screening.

3.5. Final model selection

To select a final model for virtual screening, we prioritized parsimony and generalization. Model selection was guided by Occam's razor and the one-standard-error (one-SE) rule, which recommend choosing the simplest model whose accuracy is statistically indistinguishable from the best [42–44]. Performance uncertainty was quantified by block bootstrapping of test set predictions ($n = 50{,}000$ resamples, grouped by formula), yielding mean scores, standard errors, and 95 % confidence intervals.

As shown in Fig. 5, the one-SE criterion identified the Exp_Bond model with 13 features as the most parsimonious and efficient. This

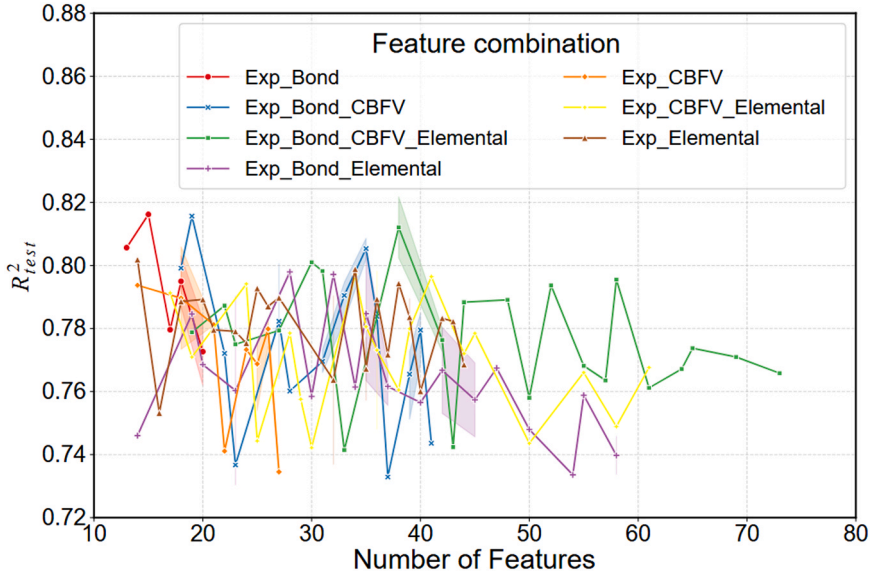

Fig. 4. Coefficient of determination (R^2) on the held-out test set as a function of the number of features.

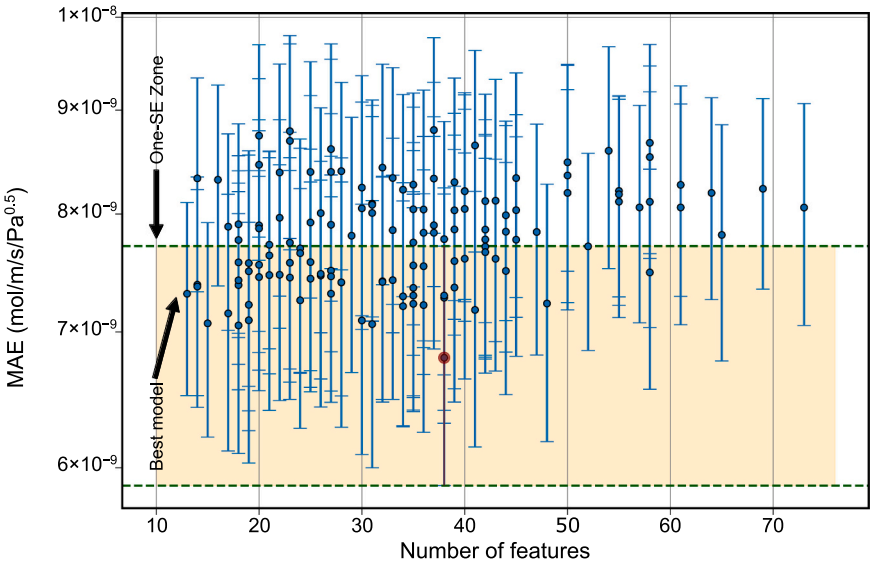

Fig. 5. Mean MAE (o-shaped markers with black edge color) with standard errors; the shaded band from the mean denotes the one-SE region. Error bars represent standard errors obtained using bootstrap. The model achieving the lowest MAE, highest R^2, and lowest RMSE is highlighted in red, and the selected model is indicated by a pointing black arrow.

domain-informed model achieved MAE $= 7.4 \times 10^{-9}$ mol·m^{-1}·s^{-1}·Pa$^{-0.5}$and RMSE $= 9.5 \times 10^{-9}$ mol·m^{-1}·s^{-1}·Pa$^{-0.5}$, values comparable to the lowest errors from larger feature sets but with far fewer descriptors. Notably, RMSE is now in the same order of magnitude as MAE, indicating improved robustness to outliers.

The parity plot in Fig. 6 (point predictions) shows that the model explains 81 % of the variance in H_2 apparent permeability ($R^2 = 0.81$) prediction across diverse alloy groups. Element-wise error statistics, provided in the Supplementary Material (Figure S6 ((a) & (b)), show that predictive uncertainty increases for underrepresented solute elements, consistent with data imbalance in literature-derived datasets.

To estimate the uncertainty on the point prediction (Fig. 6), block bootstrap resampling was used and confirmed the previous performance with a cross-validated mean $R^2 = 0.78$ and a 95 % confidence interval [0.61, 0.88] (Table 3). Note that the difference in the results of block bootstrap (Table 3) and point predictions (Fig. 6) is expected since block bootstrap is a stochastic method, corresponding to a resampling with replacement of groups of formulas and associated prediction from the test set, followed by scoring multiple times. The asymmetric interval, with greater uncertainty on the lower side, reflects the model's reduced confidence for very low-apparent permeability alloys, which act as statistical outliers. This highlights an opportunity: expanding the dataset with additional low-apparent permeability alloys would likely improve calibration and reduce skewness in future iterations.

Overall, the block bootstrap method combined with the One-SE parsimony approaches oriented the selection of the model trained on the Exp_Bond feature set which balances accuracy, interpretability, and parsimony, making it a robust tool for virtual screening of Pd-based alloys.

3.6. SHAP for feature importance analysis

We applied TreeSHAP to the final 13-feature Exp_Bond model to interpret the outcome of the apparent hydrogen permeability prediction

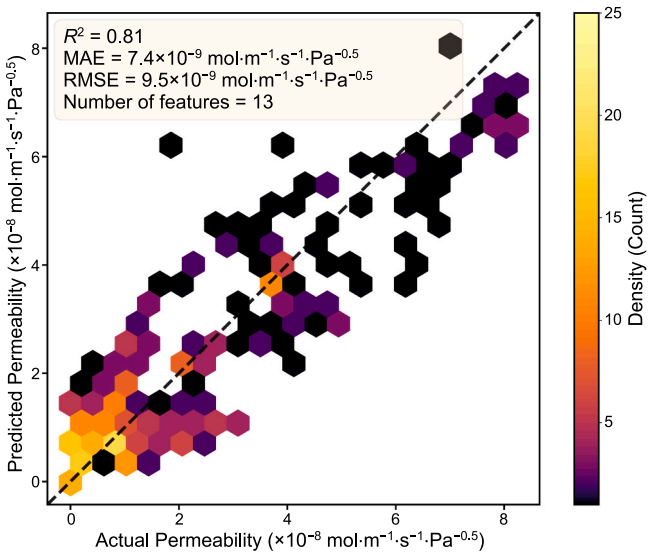

Fig. 6. Parity plot of point predictions on the held-out test set for the 13-feature Exp_Bond model selected by the one-SE rule. Uncertainty at the individual-composition level is not shown.

Table 3
Results of block bootstrap on the test set predictions by using the most parsimonious model trained on Exp_Bond (13 features). The mean, standard error, the [2.5, 97.5] percentiles to compute the 95 % CI ([lower, upper]) bounds are reported. The results are obtained after $n = 50{,}000$ resampling with replacement. The model is trained once and remained fixed.

Test metrics	Mean	Std. error	Lower bound	Upper bound
R^2	0.78	0.070	0.614	0.876
MAE	7.31×10^{-9}	7.97×10^{-10}	5.68×10^{-9}	8.77×10^{-9}
RMSE	9.37×10^{-9}	8.75×10^{-10}	7.48×10^{-9}	1.09×10^{-8}

(Fig. 7). The normalized lattice size difference relative to pure Pd ($\Delta a_{ss}/a_{Pd}$) emerged as the dominant factor (Fig. 7a). Larger lattice expansion generally increased apparent permeability (Fig. 7b), consistent with established behavior in *fcc* Pd-rich solid solutions [45–48]. Lattice expansion makes H_2 dissolution more exothermic, improving the solubility−diffusivity balance, whereas contraction enhances diffusivity but reduces solubility, leading to modest decreases in apparent permeability. This explains why Pd−rare earth solid solutions, with significantly larger solute radii, often outperform Pd−noble metal alloys. Given the fact that the lattice-misfit descriptor ($\Delta a_{ss}/a_{Pd}$) is semi-empirical, as lattice parameters are not uniformly reported and so, are partially estimated using Vegard's law; a sensitivity analysis excluding $\Delta a_{ss}/a_{Pd}$ confirms that the model's predictive trends remain robust with $R^2_{test} = 0.79$ (see Supplementary Material, Table S6).

Environmental variables were highlighted by the TreeSHAP analysis. Pressure difference $\Delta P^{0.5}$ and thickness also emerged as influential for the apparent permeability, although the permeability coefficient itself is theoretically an intrinsic material property and independent of these terms in a perfect Sieverts regime [2]. By accessing the Richardson equation ($J = \frac{\phi}{l}\Delta P^{0.5}$), this translates into the permeability coefficient being constant over a specific pressure range. In that case the derived permeance instead would have an inverse dependence on the thickness when the diffusion is bulk-controlled [49]. Their importance in the apparent permeability prediction likely reflects not only experiment vs theory gap, but also likely noise in data distribution, which TreeSHAP correctly detected. In other words, point-wise (apparent) permeability, recomputed from digitized $J-\Delta P^n$ curves, inherits residual dependence on ΔP^n and thickness. The model correctly captured this artefact, highlighting its sensitivity to experimental noise and data processing rather than a violation of Sieverts' law. While permeability is formally an intrinsic material property under ideal diffusion-limited conditions, thickness and pressure difference capture residual regime-dependent and experimental effects in heterogeneous literature datasets and should not be interpreted as intrinsic material descriptors. To assess whether the predictive trends persist when operational variables are excluded, we trained an intrinsic-only model using composition, lattice, bonding descriptors, while removing pressure difference and thickness. As expected, predictive performance decreases (cross-validated mean $R^2 \approx 0.73$; $R^2_{test} \approx 0.67$), reflecting the removal of operational variance embedded in pointwise reconstructed permeability data. Importantly, the model retains substantial predictive power, indicating that intrinsic material descriptors still capture meaningful trends, albeit with reduced variance explained. Moreover, to assess whether repeated measurements of the same alloy inflate predictive performance, we evaluated model predictions at the composition level (one observation per unique composition) by aggregating predicted and experimental permeability values for identical compositions within each fold. The performance on the independent test set retains robust composition-level accuracy ($R^2 \approx 0.82$) (Supplementary Material, Table S8).

In contrast, temperature displayed a physically meaningful effect: permeability jointly depends on the diffusivity (D) and solubility (S) which increase with the temperature, in line with Arrhenius behavior

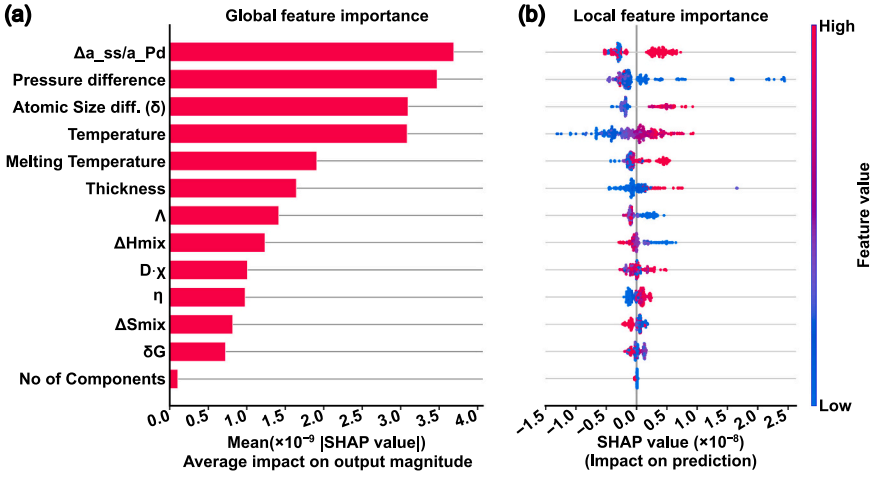

Fig. 7. (a) Global feature importance (bar plot), and (b) Local feature importance (bee swarm plot) of the simplest−efficient 13-feature Exp_Bond model used to predict apparent permeability.

($\phi = D{\cdot}S = \phi_0 \exp(\frac{-E_a}{RT})$). We have also demonstrated that predictive performance remains steady when using a physics-motivated 1000/T, which can be justified by the fact that tree-based models are not very sensitive to monotonic transformations of single features (Supplementary Material, Table S5).

Among alloy descriptors, the atomic size difference (δ) showed a two-regime effect: medium-to-large δ values enhanced apparent permeability, while small δ reduced it. The melting temperature (T_m) also correlated positively with apparent permeability in many cases, though not universally. For Group 11 alloys (Cu, Ag, Au), higher T_m does not correspond to higher apparent permeability ($\phi_{Pd-Ag} > \phi_{Pd-Au} \approx \phi_{Pd-Cu}$). However, when comparing Pd−Ag and Pd−Y, the correlation reemerges. We suggest that higher T_m raises Tammann ($\sim 0.5T_m$) and Hüttig ($\sim 0.3T_m$) temperatures, delaying bulk atomic mobility and surface segregation that degrade H_2 transport. Thus, while T_m is not a universal predictor, it captures bond-strengthening effects in certain systems.

Thermodynamic descriptors also showed systematic influence: higher mixing enthalpy and entropy were associated with reduced apparent permeability, reflecting the destabilization of solid solutions. This highlights solid solution stability as a key prerequisite for high hydrogen transport. Interestingly, the VEC, previously emphasized by Magnone et al. [24], was not ranked decisive in this study, likely because tree-based models prioritize nonlinear combinations of descriptors, which can diminish the marginal role of VEC. Instead, TreeSHAP identified secondary features such as the mixing entropy-to-squared size ratio (Λ), modulus mismatch in the strengthening model (η), mismatch of local electronegativity (D·χ), and shear modulus difference (δG) as contributing factors, suggesting that apparent permeability depends on a broader interplay of structural, thermodynamic, and electronic descriptors rather than a single dominant parameter.

Taken together, the Shap explanability analysis show that the ML model captures both expected physical trends (lattice expansion, Arrhenius temperature dependence, stability penalties) and subtler nonlinear effects introduced by environmental parameters. The analysis provides mechanistic insight and highlights which alloy properties most consistently favor enhanced apparent permeability.

3.7. Mapping apparent $\mathbf{H_2}$ permeability of potential B2 phase stabilized $\mathbf{Pd_{(100-x-y)}Cu_xM_y}$

We used the simplest-and-most-efficient model (*viz.* the final 13-feature Exp_Bond; see Section 2.5.2. and 3.5) to construct composition–predicted apparent permeability contour maps for 13 ternary systems of the form $Pd_{(100-x-y)}Cu_xM_y$ ($35 \leq 100 - x - y \leq 50$, $0.1 \leq y \leq 20$) at 673 K and 131.0 $Pa^{0.5}$. The scan, performed at 0.1 at% resolution, covered 392,600 virtual alloys, and M ∈ {Al, Cr, Mn, Fe, Ga, La, Nb, Ta, Ti, V, Y, Zn, Zr}. The chosen co-dopants M were drawn from first-principles studies as elements favoring negative formation enthalpies in B2 Pd–Cu–M ternaries [9–11]. The virtual screening was restricted to elements present in the training dataset to mitigate unsupported out-of-domain extrapolation. The search space respects the known B2 stability window of PdCu (~36 −47 at% Pd) [4].

3.7.1. Applicability domain and reliability of virtual screening

Because gradient-boosted decision tree models are known to interpolate reliably but extrapolate poorly, an explicit applicability-domain (AD) analysis was conducted to delineate the compositional regions where model predictions can be considered trustworthy.

We employed a *k*-nearest-neighbor (*k*NN) distance analysis in standardized descriptor space, using robustly scaled features identical to those used for model training. The AD boundary was defined using the 97.5th percentile of nearest-neighbor distances within the training set, beyond which predictions are labeled extrapolative.

Using the full training dataset as reference, the median nearest-neighbor distance between training alloys is 0.156, indicating that chemically similar alloys in the literature are typically separated by ~0.16 in descriptor space. Approximately 23.5 % of the held-out test samples lie near or beyond the AD boundary; nevertheless, the model retains strong predictive performance on the test set ($R^2 = 0.81$), indicating robustness to moderate extrapolation. In contrast, ~99 % of the unconstrained virtual alloys fall outside the densely sampled training manifold, reflecting the combinatorial size of the Pd–Cu–M composition space relative to the sparsity of available experimental data.

To reduce unsupported extrapolation and align screening with the B2-relevant chemistry, we performed a second AD analysis using a composition-bracketed subset of the training data corresponding to the virtual screening window ($35 \leq$ Pd ≤ 50 at% and $0 <$ M ≤ 20 at%). The resulting fraction represents ~6.27 % of the training dataset. When distances were recomputed relative to this subset, the fraction of extrapolative virtual compositions decreased to ~91 %, demonstrating that most dilute ternary alloys remain weakly represented in the experimental literature despite chemistry-informed constraints.

Importantly, multi-objective screening was done separately for in-domain alloys (~8.72 % of the virtual alloys set) identified under this second, more conservative AD definition. Predictions outside the AD are treated as hypothesis-generating only and are reported separately in the Supplementary Material. Fig. 8, mapping both interpolative and extrapolative compositions, revealed several plausible hypothesis-generating alloy families with promising predicted apparent permeabilities. Some out-of-distribution rare-earth and group IV solutes (Y, La, Zr) systems seem to achieve maxima above 3×10^{-8} $mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}$, nearly double the reference ($\phi_{Pd_{47.25}Cu_{52.75}} = 1.6 \times 10^{-8}$ $mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}$ [3]).

The *k*-NN Euclidian distance analysis performed demonstrated that virtual screening using relatively weak extrapolators such as CatBoost may yield contentious results if a clear applicability domain of the model is not identified.

3.7.2. Multi-objective Pareto screening within the applicability domain

Multi-objective optimization was separately performed on in-domain and out-of-domain (shown in Supplementary Material, section 14. Complementary analysis for virtual screening) virtual alloys, using Pareto dominance to identify non-dominated trade-offs among three competing objectives: (*i*) maximum predicted apparent permeability, (*ii*) minimum Pd content, and (*iii*) minimum Matminer-calculated intermetallic Miedema formation enthalpy (trends comparison with DFT-calculated formation enthalpy values for $Pd_{43.75}Cu_{50}M_{6.25}$ families are shown in Supplementary Material, Fig. S5) [50]. All objectives were treated equally without scalar weighting, as implemented in the Paretoset library, to avoid subjective prioritization.

For each Pd–Cu–M system, the resulting Pareto set was further summarized by selecting three representative compositions: (*i*) the alloy with the highest permeability, (*ii*) the alloy with the lowest formation enthalpy, and (*iii*) the alloy with the lowest Pd content. These representatives capture complementary extremes of performance and stability within each dopant system.

The final list of candidate alloys (in the defined AD) is reported in Table 4, while full Pareto sets and results for extrapolative compositions are provided in the Supplementary Material (Table S9, Fig. S7–S9).

For virtual compositions classified as in-domain, the expected prediction error is estimated using the 95 % confidence interval of the test-set MAE obtained via block bootstrap. This interval reflects model performance uncertainty rather than pointwise predictive uncertainty.

To better represent the results of Table 4, we plotted the predicted apparent permeability for each system as shown in Fig. 9. It can be seen that for most of the Pareto selected hypothesis-generating virtual alloy candidates labelled as interpolative samples, the co-dopants M concentrations were low (< 0.5 at% for the most permeable alloys) while their permeability remained $\leq 1.09 \times 10^{-8}$ $mol{\cdot}m^{-1}{\cdot}s^{-1}{\cdot}Pa^{-0.5}$. This very dilute additions are in line with reported literature strategies to prevent

Table 4
Pareto set of extracted in-domain compositions (in atom% and weight%) with the highest predicted apparent permeability at 673 K (max. ϕ), minimum Pd content (min. Pd) and minimum enthalpy (min ΔH_{mied}) for each system. The predictions are interpolative (expected absolute error on in-domain predictions is in the order (95 % CI: $|5.68 \times 10^{-9}$–8.77×10^{-9} mol·m^{-1}·s^{-1}·Pa$^{-0.5}|$).

Virtual system		Virtual candidates (at%)	Virtual candidates (wt%)	Permeability (mol·m^{-1}·s^{-1}·Pa$^{-0.5}$)
Reference$^{exp., 673\ K}$	[3]	$Pd_{47.25}Cu_{52.75}$	$Pd_{60}Cu_{40}$	$\sim 1.60 \times 10^{-8}$
Reference$^{exp., 673\ K}$	[12]	$Pd_{48.3}Cu_{51.7}$	$Pd_{61.01}Cu_{38.99}$	$\sim 2 \times 10^{-8}$
		$Pd_{48.3}Cu_{51.2}Ag_{0.5}$	$Pd_{60.85}Cu_{38.51}Ag_{0.64}$	$\sim 2 \times 10^{-8}$
		$Pd_{48.3}Cu_{51.2}Al_{0.5}$	$Pd_{61.14}Cu_{38.70}Al_{0.16}$	$\sim 1.9 \times 10^{-8}$
		$Pd_{48.3}Cu_{51.2}Mn_{0.5}$	$Pd_{61.03}Cu_{38.64}Mn_{0.33}$	$\sim 1.75 \times 10^{-8}$
Al	max. ϕ	$Pd_{47.0}Cu_{52.9}Al_{0.1}$	$Pd_{59.79}Cu_{40.18}Al_{0.03}$	1.08×10^{-8}
	min Pd	$Pd_{35.0}Cu_{61.7}Al_{3.3}$	$Pd_{48.16}Cu_{50.69}Al_{1.15}$	8.67×10^{-9}
	min ΔH_{mied}	$Pd_{45.1}Cu_{49.3}Al_{5.6}$	$Pd_{59.37}Cu_{38.76}Al_{1.87}$	8.60×10^{-9}
Ga	max. ϕ	$Pd_{46.9}Cu_{53.0}Ga_{0.1}$	$Pd_{59.66}Cu_{40.26}Ga_{0.08}$	1.08×10^{-8}
	min Pd	$Pd_{35.2}Cu_{64.6}Ga_{0.2}$	$Pd_{47.63}Cu_{52.19}Ga_{0.18}$	5.81×10^{-9}
	min ΔH_{mied}	$Pd_{44.4}Cu_{52.3}Ga_{3.3}$	$Pd_{57.07}Cu_{40.15}Ga_{2.78}$	7.92×10^{-9}
Cr	max. ϕ	$Pd_{47.6}Cu_{52.3}Cr_{0.1}$	$Pd_{60.35}Cu_{39.59}Cr_{0.06}$	1.05×10^{-8}
	min Pd	$Pd_{35.2}Cu_{64.7}Cr_{0.1}$	$Pd_{47.64}Cu_{52.29}Cr_{0.07}$	5.92×10^{-9}
	min ΔH_{mied}	$Pd_{45.3}Cu_{54.6}Cr_{0.1}$	$Pd_{58.12}Cu_{41.82}Cr_{0.06}$	9.87×10^{-9}
Mn	max. ϕ	$Pd_{46.4}Cu_{53.5}Mn_{0.1}$	$Pd_{59.18}Cu_{40.75}Mn_{0.07}$	1.08×10^{-8}
	min Pd	$Pd_{35.0}Cu_{64.8}Mn_{0.2}$	$Pd_{47.43}Cu_{52.43}Mn_{0.14}$	5.75×10^{-9}
	min ΔH_{mied}	$Pd_{43.4}Cu_{55.0}Mn_{1.6}$	$Pd_{56.32}Cu_{42.61}Mn_{1.07}$	9.14×10^{-9}
Fe	max. ϕ	$Pd_{48.2}Cu_{51.7}Fe_{0.1}$	$Pd_{60.91}Cu_{39.02}Fe_{0.07}$	1.05×10^{-8}
	min Pd	$Pd_{35.2}Cu_{64.6}Fe_{0.2}$	$Pd_{47.65}Cu_{52.21}Fe_{0.14}$	5.76×10^{-9}
	min ΔH_{mied}	$Pd_{45.3}Cu_{54.6}Fe_{0.1}$	$Pd_{58.11}Cu_{41.82}Fe_{0.07}$	9.90×10^{-9}
Ti	max. ϕ	$Pd_{44.0}Cu_{55.7}Ti_{0.3}$	$Pd_{56.86}Cu_{42.97}Ti_{0.17}$	1.09×10^{-8}
	min Pd	$Pd_{35.0}Cu_{63.4}Ti_{1.6}$	$Pd_{47.57}Cu_{51.45}Ti_{0.98}$	8.84×10^{-9}
	min ΔH_{mied}	$Pd_{41.3}Cu_{57.0}Ti_{1.7}$	$Pd_{54.27}Cu_{44.73}Ti_{1.00}$	9.40×10^{-9}
Zr	max. ϕ	$Pd_{46.2}Cu_{53.7}Zr_{0.1}$	$Pd_{58.96}Cu_{40.93}Zr_{0.11}$	1.09×10^{-8}
	min Pd	$Pd_{35.7}Cu_{64.2}Zr_{0.1}$	$Pd_{48.16}Cu_{51.72}Zr_{0.12}$	6.16×10^{-9}
	min ΔH_{mied}	$Pd_{45.1}Cu_{54.4}Zr_{0.5}$	$Pd_{57.81}Cu_{41.64}Zr_{0.55}$	1.00×10^{-8}
V	max. ϕ	$Pd_{46.0}Cu_{53.9}V_{0.1}$	$Pd_{58.80}Cu_{41.14}V_{0.06}$	1.09×10^{-8}
	min Pd	$Pd_{35.0}Cu_{64.8}V_{0.2}$	$Pd_{47.43}Cu_{52.44}V_{0.13}$	5.96×10^{-9}
	min ΔH_{mied}	$Pd_{41.4}Cu_{56.8}V_{1.8}$	$Pd_{54.35}Cu_{44.52}V_{1.13}$	9.76×10^{-9}
Nb	max. ϕ	$Pd_{44.2}Cu_{55.5}Nb_{0.3}$	$Pd_{56.95}Cu_{42.71}Nb_{0.34}$	1.08×10^{-8}
	min Pd	$Pd_{35.0}Cu_{61.8}Nb_{3.2}$	$Pd_{46.86}Cu_{49.40}Nb_{3.74}$	1.02×10^{-8}
	min ΔH_{mied}	$Pd_{35.3}Cu_{59.4}Nb_{5.3}$	$Pd_{46.82}Cu_{47.04}Nb_{6.14}$	8.44×10^{-9}
Ta	max. ϕ	$Pd_{46.2}Cu_{53.7}Ta_{0.1}$	$Pd_{58.90}Cu_{40.88}Ta_{0.22}$	1.08×10^{-8}
	min Pd	$Pd_{35.0}Cu_{62.5}Ta_{2.5}$	$Pd_{45.71}Cu_{48.74}Ta_{5.55}$	1.02×10^{-8}
	min ΔH_{mied}	$Pd_{35.2}Cu_{60.7}Ta_{4.1}$	$Pd_{44.89}Cu_{46.22}Ta_{8.89}$	8.44×10^{-9}
Y	max. ϕ	$Pd_{44.4}Cu_{55.5}Y_{0.1}$	$Pd_{57.20}Cu_{42.69}Y_{0.11}$	1.09×10^{-8}
	min Pd	$Pd_{39.1}Cu_{60.8}Y_{0.1}$	$Pd_{51.80}Cu_{48.09}Y_{0.11}$	7.55×10^{-9}
	min ΔH_{mied}	$Pd_{45.1}Cu_{54.5}Y_{0.4}$	$Pd_{57.84}Cu_{41.73}Y_{0.43}$	1.01×10^{-8}
La	max. ϕ	$Pd_{44.4}Cu_{55.5}La_{0.1}$	$Pd_{57.16}Cu_{42.67}La_{0.17}$	1.07×10^{-8}
	min Pd	$Pd_{40.3}Cu_{59.6}La_{0.1}$	$Pd_{53.01}Cu_{46.82}La_{0.17}$	8.45×10^{-9}
	min ΔH_{mied}	$Pd_{45.2}Cu_{54.5}La_{0.3}$	$Pd_{57.85}Cu_{41.65}La_{0.50}$	1.02×10^{-8}
Zn	max. ϕ	$Pd_{47.1}Cu_{52.8}Zn_{0.1}$	$Pd_{59.85}Cu_{40.07}Zn_{0.08}$	1.08×10^{-8}
	min Pd	$Pd_{35.0}Cu_{64.4}Zn_{0.6}$	$Pd_{47.41}Cu_{52.09}Zn_{0.50}$	6.01×10^{-9}
	min ΔH_{mied}	$Pd_{43.9}Cu_{52.8}Zn_{3.3}$	$Pd_{56.68}Cu_{40.70}Zn_{2.62}$	8.21×10^{-9}

segregation and transport properties damage that will arise if high co-dopants concentrations were added to the membrane. Conversely, for out-of-distribution alloys, the co-dopants concentrations are high (0.1 – 20 at%), with permeability values reaching in some cases twice that of the reference for La, Y, and Zr systems (Fig S7). Such, higher values might likely reflect a mimicry of *fcc* systems, wherein an increase of the concentration of the dopant is accompanied by permeability increase up to a certain critical concentration, where it starts dropping (inverted U-shape behavior).

Overall, *k*-NN distance analyses combined to virtual screening and multi-objective Pareto optimization is an effective strategy to generate plausible hypothetical B2 phase-stabilized alloys which ground state stability can be confirmed using DFT.

3.8. Post hoc DFT evaluation of stability

To provide atomistic support for selected in-domain candidates, the ground-state stability of two representative Pd–Cu–M compositions ($Pd_{44.2}Cu_{55.5}Nb_{0.3}$ and $Pd_{47.6}Cu_{52.3}Cr_{0.1}$) was evaluated using plane-wave density functional theory (DFT) as implemented in VASP [51,52].

Because the machine-learning–suggested compositions involve sub-percent dopant concentrations, they were approximated using the smallest practically representable concentrations in a 128-atom $4 \times 4 \times 4$ B2 supercell (one dopant atom per supercell, corresponding to 0.78 at%). The resulting proxy compositions were $Pd_{57}Cu_{70}Nb$ ($Pd_{44.53}Cu_{54.69}Nb_{0.78}$ at%) and $Pd_{61}Cu_{66}Cr$ ($Pd_{47.67}Cu_{51.56}Cr_{0.78}$ at%).

Dopant site preference was accounted for following Ref. [7], with Nb and Cr substituting Cu at the $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ sublattice position of the B2 PdCu parent lattice. To maintain the overall stoichiometry of the supercell, compensating Cu antisite defects were introduced. This configuration represents a symmetry-consistent, low-energy substitutional model intended to evaluate relative formation enthalpies at 0 K rather than finite-temperature configurational equilibria. Calculations employed the PAW formalism with the PBE exchange–correlation functional. A plane-wave cutoff of 520 eV was used throughout. Structural relaxations and static calculations employed Monkhorst–Pack meshes corresponding to reciprocal-space resolutions of 0.03 Å^{-1} and 0.02 Å^{-1}, respectively, as generated using VASPKIT [53]. Atomic positions were relaxed until residual forces were below 0.01 eV·Å^{-1}. Final total energies were obtained using the tetrahedron method. Electronic convergence thresholds were set to 10^{-4} eV (relaxation) and 10^{-5} eV (static). Spin polarization was included; final magnetic moments converged to negligible values. Formation enthalpies were computed as:

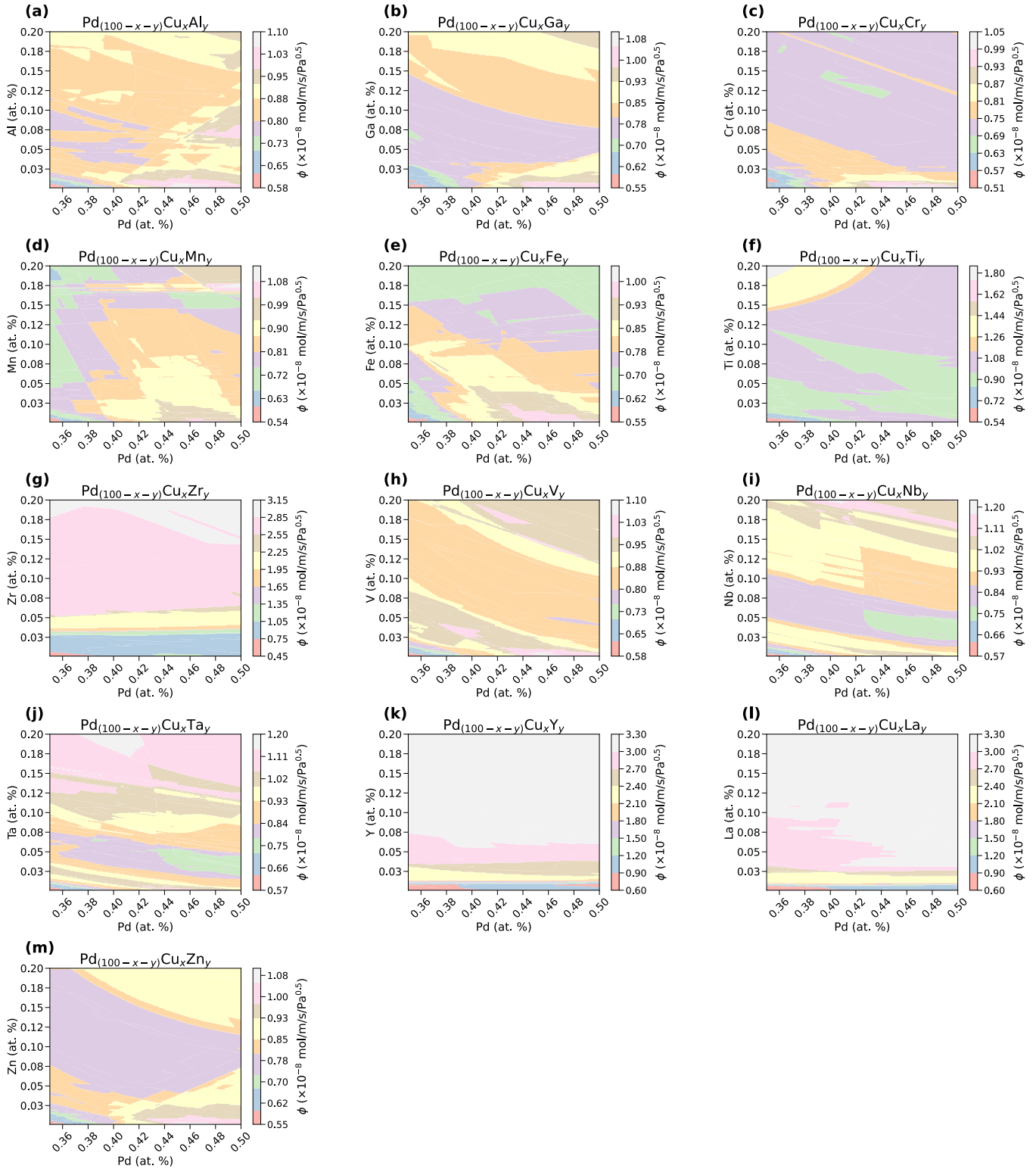

Fig. 8. (a) – (m) Predictive apparent H_2 permeability at 673.15 K, 131.0 $Pa^{0.5}$ for 15 μm-thick 13 virtual ternary $Pd_{(100-x-y)}Cu_xM_y$ alloy systems based on the 13-feature Exp_Bond model. (a) – (p) Ternary system of $Pd_{(100-x-y)}Cu_xM_y$ (M∈{Al, Ga, Cr, Mn, Fe, Ti, Zr, V, Nb, Ta, Y, La, Zn}). Predicted apparent permeability values correspond to point estimates from the trained model; uncertainty at the individual-composition level is not shown.

$$\Delta H_{f,i}(\mathrm{Pd}_m\mathrm{Cu}_n\mathrm{M}_t) = \frac{(E_i - mE_{\mathrm{Pd}} - nE_{\mathrm{Cu}} - tE_{\mathrm{M}})}{(m+n+t)} \quad (4)$$

where E_i is the total energy of configuration i, and E_{Pd}, E_{Cu}, and E_M are the elemental reference energies.

As validation, the calculated formation enthalpy of stoichiometric B2 PdCu (−11.53 kJ mol^{-1}) agrees well with prior reports (−12 kJ mol^{-1} [9]) and Materials Project data (−11.87 kJ mol^{-1}).

Both ternary proxies exhibit negative formation enthalpies (Table 5), indicating thermodynamic stability at 0 K.

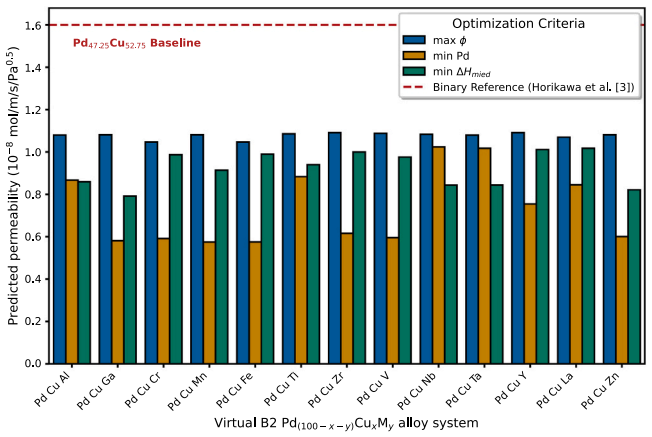

Fig. 9. Predictive apparent H_2 permeability for top in-domain (interpolative) $Pd_{(100-x-y)}Cu_xM_y$ hypothesis-generating alloys given three-objective optimization (maximum permeability (max ϕ), minimum Pd content (min Pd) and minimum intermetallic Miedema enthalpy of formation (min ΔH_{mied}).

Table 5
Ground state (0 K) formation enthalpy of B2 PdCu and a configuration of $Pd_{57}Cu_{70}Nb$.

Structure	ΔH_f (KJ/mol)	Lattice parameter (Å)
B2 PdCu (1:1)	−11.53	3.0173
B2 $Pd_{57}Cu_{70}Nb$	−12.43	3.0039
B2 $Pd_{61}Cu_{66}Cr$	−7.94	3.0087

The slightly more negative value for $Pd_{57}Cu_{70}Nb$ (−12.43 kJ mol^{-1}) relative to stoichiometric PdCu suggests a modest stabilization of the B2 phase at dilute Nb concentrations. In contrast, $Pd_{61}Cu_{66}Cr$ (–7.94 kJ mol^{-1}) exhibits reduced stabilization relative to binary B2 PdCu, consistent with prior reports of composition-dependent phase tendencies [7].

Overall, these results support Nb and Cr as viable dilute co-dopants for B2 PdCu, while underscoring that quantitative phase equilibria at operating temperatures require future finite-temperature treatments.

4. Limitations and future outlook

In this study, we trained CatBoost on experimentally observed Pd alloy data. Inevitably, this constrains model generalization, as literature-sourced experimental data are inherently sparse and noisy compared to the highly homogeneous datasets generated using DFT. However, the more the experimental data the higher the accuracy. Given its restriction to regimes with $n = 0.5$, the so-obtained model's applicability is constrained to bulk diffusion-limited transport regimes and is not intended to describe surface-limited or ultra-thin membrane operation, where additional kinetic and interfacial effects dominate. As a consequence, the model does not distinguish between surface-controlled and bulk-diffusion-controlled enhancements. In addition, the hypothesis-generating Pd–Cu–M virtual alloys proposed for validation may operate in either bulk-diffusion limited or surface limited regime depending on their fabrication techniques, the presence of defects, the surface morphology and their thickness. The determination of the onset of either condition cannot be controlled by the model per se. The lattice-misfit descriptors are treated as semi-empirical, and their limitations are explicitly acknowledged. Especially, the lattice misfit descriptor for mixed *fcc*–B2 phases will gain in robustness with the development of a mixed *fcc*–B2 phase-specific Vegard's law. Furthermore, the model does not explicitly resolve phase fractions in mixed *fcc*–B2 alloys. Ideally, pointwise permeability values are mathematically equivalent to slope-derived values when the same pressure exponent is used; however, heterogeneous reporting practices across the literature motivate the use of pointwise extraction combined with aggregation-based robustness analyses. The Miedema formation enthalpy is employed as a semi-empirical, mean-field stability descriptor and does not explicitly capture atomic ordering effects in B2 structures at DFT accuracy; its use is therefore restricted to high-throughput screening-level trend analysis. The present study considers hydrogen permeability under pure H_2 conditions only and does not address sulfur or CO tolerance, which would require explicit surface–adsorbate modeling or mixed-gas experiments. While bond-related descriptors inform chemical cohesion, they do not provide a direct assessment of ductility, embrittlement resistance, or ductile-to-brittle transition behavior. The Pareto-based screening identifies trade-off–optimal compositions without imposing application-specific cost thresholds; techno-economic feasibility must be assessed separately for each target application. Finally, given the high-temperature application of the proposed virtual alloys, cluster expansion and Monte Carlo simulations accelerated using fine-tuned Machine Learning Interatomic Potentials (MLIPs), represent accessible methods to assess finite-temperature thermodynamic properties of the hypothesis-generating candidates.

5. Conclusion

This study demonstrates how carefully constrained machine learning can be leveraged to extract physically meaningful insights and generate experimentally relevant hypotheses for Pd-based hydrogen separation membranes. Using a literature-derived dataset of 328 unique Pd-alloy membranes (2223 permeability records) restricted to bulk diffusion–controlled transport, CatBoost regressors were trained to capture composition–condition–property relationships governing apparent hydrogen permeability.

To mitigate biases arising from data imbalance and heterogeneity, chemically informed data splitting and stratified group cross-validation were employed, improving robustness and generalization. A systematic descriptor selection strategy—combining Pearson filtering, fold-wise SHAP-based recursive feature elimination, and cross-fold aggregation—yielded compact, interpretable models. Guided by the one-standard-error rule, a 13-variables, domain-informed feature set preserved predictive accuracy ($R^2 = 0.81$) while substantially reducing model complexity.

Model interpretation reveals physically consistent trends: higher apparent permeability correlates with elevated temperature (Arrhenius behavior), lattice expansion relative to Pd, increased atomic-size mismatch, and favorable alloy mixing tendencies. Building on this foundation, composition–permeability landscapes were constructed for $Pd_{(100-x-y)}Cu_xM_y$ systems (M = Al, Ga, Cr, Mn, Fe, Ti, Zr, V, Nb, Ta, Y, La, Zn), explicitly restricting screening to dopants represented in the training data to control extrapolation. Applicability-domain analysis coupled with multi-objective Pareto screening enabled the identification of interpolative candidate compositions balancing permeability, Pd content, and thermodynamic stability proxies.

Beyond screening, the trained model provides a practical surrogate for estimating permeability trends and inferring activation energies from predicted pressure- and temperature-dependent behavior. More broadly, this framework establishes a scalable pathway for integrating curated experimental data, interpretable machine learning, and targeted first-principles validation to guide the rational design of thermally stable, Pd-lean metallic membranes for high-temperature hydrogen separation and membrane-reactor applications.

Appendix A. Supplementary materials

The Supplementary Material includes: details on the evaluated feature sets; details on lattice parameters calculation, data scaling procedures; details on the Pearson correlation filtering; details on the selected model hyperparameters; discussion of negative coefficients of

determination; block bootstrap confidence-interval estimation; Miedema formation-enthalpy calculations; details on the sensitivity analyses; details on element-wise error statistics; and supplementary informations on virtual screening.

CRediT authorship contribution statement

Kwami Aldo Eric Kolor: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Edoardo Magnone:** Writing – review & editing, Validation, Methodology. **Muhammad Harussani Moklis:** Writing – review & editing. **Md Rubel:** Writing – review & editing. **Sasipa Boonyubol:** Writing – review & editing, Validation, Methodology. **Koichi Mikami:** Writing – review & editing, Validation, Supervision, Methodology. **Jeffrey S. Cross:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the first author used OpenAI's ChatGPT Plus to assist with the language refinement, readability improvement, and scientific structure enhancement. After using these services, the author carefully reviews, edits, and verifies all the AI-generated content for factual accuracy, completeness, and adherence to scientific standards. The author takes full responsibility for the content of the published article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors express their gratitude to the Ministry of Education, Culture, Sports, Science, and Technology (Monbukagakusho): MEXT scholarship and the Institute of Science Tokyo (formerly Tokyo Institute of Technology) for their financial assistance and support in conducting this research. The authors also thank Professor Atsushi Ishikawa for providing access to the VASP software.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.mtcomm.2026.114875.

Data availability

I shared the link to the code and data in the manuscript. The github link contains all the codes, datasets, VASP Inputs and CONTCAR, and methods to reproduce the visualizations

Pd-membranes-permeability (Github)

References

[1] O.V. Akimova, I.A. Ovchenkova, E.C. Kulikova, A.V. Ovcharov, N.R. Roshan, Hydrogen-tuning of magnetic properties of Pd based membrane alloys, Mater. Today Commun. 35 (2023) 106156, https://doi.org/10.1016/j.mtcomm.2023.106156.

[2] E. Kolor, M. Usman, S. Boonyubol, K. Mikami, J.S. Cross, Advances in palladium-based membrane research: high-throughput techniques and machine learning perspectives, Processes 12 (2024) 2855, https://doi.org/10.3390/pr12122855.

[3] D. Horikawa, T. Matsumura, M. Ebisugi, S. Kubota, Hydrogen permeable membrane made of PdCu alloy and hydrogen purification method due to hydrogen permeable membrane, JP2023039770A, 2023. ⟨https://patents.google.com/patent/JP2023039770A/en⟩ (accessed September 15, 2025).

[4] A.A. Popov, Yu.V. Shubin, P.E. Plyusnin, M.R. Sharafutdinov, S.V. Korenev, Experimental redetermination of the Cu–Pd phase diagram, J. Alloy. Compd. 777 (2019) 204–212, https://doi.org/10.1016/j.jallcom.2018.10.332.

[5] K. Sato, H. Imamura, N. Yuki, Hydrogen permeable alloy and hydrogen permeation membrane utilizing the same, JP2011202258A (2011). ⟨https://patents.google.com/patent/JP2011202258A/en?oq=JP+2012-201974⟩ (accessed September 15, 2025).

[6] H. Imamura, Hydrogen permeable copper alloy, hydrogen permeable film, and steam reformer, JP2012201974A (2012). ⟨https://patents.google.com/patent/JP2012201974A/en?oq=JP+2012-201974⟩ (accessed September 15, 2025).

[7] Y.T. Hu, L. Chen, H. Gong, Structural stability, phase transition, and hydrogen diffusion of PdCu phases with additions of Nb and Cr, Int. J. Hydrog. Energy 42 (2017) 3051–3057, https://doi.org/10.1016/j.ijhydene.2016.09.186.

[8] R.C. Benn, S.M. Opalka, T.H. Vanderspurt, Durable Pd-based alloy and hydrogen generation membrane thereof, US7655183B2, 2010. ⟨https://patents.google.com/patent/US7655183B2/en?inventor=Raymond+Benn&oq=Raymond+Benn⟩ (accessed September 15, 2025).

[9] M.C. Gao, L. Ouyang, Ö.N. Doğan, First principles screening of B2 stabilizers in CuPd-based hydrogen separation membranes: (1) Substitution for Pd, J. Alloy. Compd. 574 (2013) 368–376, https://doi.org/10.1016/j.jallcom.2013.05.126.

[10] Ö.N. Doğan, M.C. Gao, R.H. Young, D.N. Tafen, Cu–Pd–M hydrogen separation membranes, US8608829B1, 2013. ⟨https://patents.google.com/patent/US8608829B1/en?oq=US8608829⟩ (accessed September 15, 2025).

[11] R. Hu, M.C. Gao, Ö.N. Doğan, Accelerated Materials Design for Hydrogen Separation Membranes, in: I. Escobar, B. Van Der Bruggen (Eds.), ACS Symp. Ser., American Chemical Society, Washington, DC, 2011, pp. 27–38, https://doi.org/10.1021/bk-2011-1078.ch003.

[12] K. Ogawa, D. Horikawa, K. Hakamata, T. Nakazawa, S. Kubota, Study for Hydrogen-Permeable Pd Membrane Using First-Principles Calculation, Defect Diffus. Forum 439 (2025) 63–72, https://doi.org/10.4028/p-q7TCoW.

[13] D. Yang, P. Xu, Q. Xiang, W. Xue, N. Liao, Machine learning enabled efficient prediction and accelerated discovery of palladium alloys membranes for hydrogen separation, J. Membr. Sci. 720 (2025) 123777, https://doi.org/10.1016/j.memsci.2025.123777.

[14] A. Rohatgi, WebPlotDigitizer, (2011). ⟨https://automeris.io/⟩.

[15] T.J.B. Holland, S.A.T. Redfern, UnitCell: Least squares cell refinement program with regression diagnostics, (2006). ⟨https://filedn.com/lU1GlyFhv3UuXg5E9dbnWFF/TJBHpages/unitcell.html⟩.

[16] T.J.B. Holland, S.A.T. Redfern, Unit cell refinement from powder diffraction data: the use of regression diagnostics, Mineral. Mag. 61 (1997) 65–77, https://doi.org/10.1180/minmag.1997.061.404.07.

[17] N.A. Al-Mufachi, S. Nayebossadri, J.D. Speight, W. Bujalski, R. Steinberger-Wilckens, D. Book, Effects of thin film Pd deposition on the hydrogen permeability of $Pd_{60}Cu_{40}$ wt% alloy membranes, J. Membr. Sci. 493 (2015) 580–588, https://doi.org/10.1016/j.memsci.2015.07.015.

[18] P.R. Subramanian, D.E. Laughlin, Cu-Pd (Copper-Palladium), J. Phase Equilibria 12 (1991) 231–243, https://doi.org/10.1007/BF02645723.

[19] P. Huang, S. Menon, D. De Fontaine, On the Cu–Pd phase diagram, J. Phase Equilibria 12 (1991) 3–5, https://doi.org/10.1007/BF02663663.

[20] F. Gallucci, Richardson Law (Berlin Heidelberg), in: E. Drioli, L. Giorno (Eds.), Encycl. Membr, Springer, Berlin, Heidelberg, 2012, pp. 1–2, https://doi.org/10.1007/978-3-642-40872-4_523-1 (Berlin Heidelberg).

[21] D.L. Mckinley, Method for hydrogen separation and purification, US3439474A, 1969. ⟨https://patents.google.com/patent/US3439474A/en⟩ (accessed September 18, 2025).

[22] K. Bryden, Nanostructured palladium–iron membranes for hydrogen separation and membrane hydrogenation reactions, J. Membr. Sci. 203 (2002) 29–42, https://doi.org/10.1016/S0376-7388(01)00736-0.

[23] S. Guo, Phase selection rules for cast high entropy alloys: An overview, Mater. Sci. Technol. 31 (2015) 1223–1230, https://doi.org/10.1179/1743284715Y.0000000018.

[24] E. Magnone, M.C. Shin, J.I. Lee, J.H. Park, Relationship between hydrogen permeability and the physical-chemical characteristics of metal alloy membranes, J. Membr. Sci. 674 (2023) 121513, https://doi.org/10.1016/j.memsci.2023.121513.

[25] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, Y. Su, Machine learning assisted design of high entropy alloys with desired property, Acta Mater. 170 (2019) 109–117, https://doi.org/10.1016/j.actamat.2019.03.010.

[26] E. Yan, L. Tao, K. Zhang, H. Chen, G. Huang, J. Bai, Y. Li, K. Liao, Y. Zou, H. Zhang, L. Sun, A new strategy for designing Nb-based H_2-selective alloys based on physical-chemical parameter matching rules, J. Mater. Res. Technol. 38 (2025) 2782–2793, https://doi.org/10.1016/j.jmrt.2025.08.125.

[27] X. Yang, Y. Zhang, Prediction of high-entropy stabilized solid-solution in multi-component alloys, Mater. Chem. Phys. 132 (2012) 233–238, https://doi.org/10.1016/j.matchemphys.2011.11.021.

[28] S. Lee, C. Chen, G. Garcia, A. Oliynyk, Machine learning descriptors in materials chemistry used in multiple experimentally validated studies: Oliynyk elemental property dataset, Data Brief. 53 (2024) 110178, https://doi.org/10.1016/j.dib.2024.110178.

[29] M. Katlav, M.E. Tabar, K. Turk, AI-guided design framework for bond behavior of steel-concrete in steel reinforced concrete composites: from dataset cleaning to feature engineering, Mater. Today Commun. 42 (2025) 111286, https://doi.org/10.1016/j.mtcomm.2024.111286.

[30] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, (2017). https://doi.org/10.48550/ARXIV.1706.09516.
[31] J. Macaluso, Hyperparameter Tuning XGBoost with early stopping, Jeff Macaluso (2024). ⟨https://macalusojeff.github.io//post/HyperparameterTuningXGB/⟩ (accessed September 16, 2025).
[32] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (2020) 56–67, https://doi.org/10.1038/s42256-019-0138-9.
[33] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning (New York), Springer, New York, NY, 2009, https://doi.org/10.1007/978-0-387-84858-7.
[34] J.J. Conde, M. Maroño, J.M. Sánchez-Hervás, Pd-based membranes for hydrogen separation: review of alloying elements and their influence on membrane properties, Sep. Purif. Rev. 46 (2017) 152–177, https://doi.org/10.1080/15422119.2016.1212379.
[35] J. Hu, D. Liu, N. Fu, R. Dong, Realistic material property prediction using domain adaptation-based machine learning, Digit. Discov. 3 (2024) 300–312, https://doi.org/10.1039/D3DD00162H.
[36] Q. Li, N. Fu, S.S. Omee, J. Hu, MD-HIT: machine learning for material property prediction with dataset redundancy control, npj Comput. Mater. 10 (2024) 245, https://doi.org/10.1038/s41524-024-01426-z.
[37] M. Sugiyama, M. Krauledat, K.-R. Müller, Covariate Shift Adaptation by Importance Weighted Cross Validation, J. Mach. Learn. Res. 8 (2007) 985–1005.
[38] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Trans. Evol. Comput. 1 (1997) 67–82, https://doi.org/10.1109/4235.585893.
[39] J. Gong, S. Chu, R.K. Mehta, A.J.H. McGaughey, XGBoost model for electrocaloric temperature change prediction in ceramics, npj Comput. Mater. 8 (2022) 140, https://doi.org/10.1038/s41524-022-00826-3.
[40] A. Roy, M.F.N. Taufique, H. Khakurel, R. Devanathan, D.D. Johnson, G. Balasubramanian, Machine-learning-guided descriptor selection for predicting corrosion resistance in multi-principal element alloys, npj Mater. Degrad. 6 (2022) 9, https://doi.org/10.1038/s41529-021-00208-y, n.
[41] J. Damewood, J. Karaguesian, J.R. Lunger, A.R. Tan, M. Xie, J. Peng, R. Gómez-Bombarelli, Representations of materials for machine learning, Annu. Rev. Mater. Res. 53 (2023) 399–426, https://doi.org/10.1146/annurev-matsci-080921-085947.
[42] M. Dubova, S. Chandramouli, G. Gigerenzer, P. Grünwald, W. Holmes, T. Lombrozo, M. Marelli, S. Musslick, B. Nicenboim, L.N. Ross, R. Shiffrin, M. White, E.-J. Wagenmakers, P.-C. Bürkner, S.J. Sloman, Is Ockham's razor losing its edge? New perspectives on the principle of model parsimony, Proc. Natl. Acad. Sci. 122 (2025) e2401230121, https://doi.org/10.1073/pnas.2401230121.
[43] D. Krstajic, L.J. Buturovic, D.E. Leahy, S. Thomas, Cross-validation pitfalls when selecting and assessing regression and classification models, J. Chemin.. 6 (2014) 10, https://doi.org/10.1186/1758-2946-6-10.
[44] P. Sarkar, STAT 383C: Statistical Modeling, (2015). ⟨https://www.cs.cmu.edu/~psarkar/sds383c_16/lecture9_scribe.pdf⟩.
[45] Z. Chen, Z. Yang, Y. Tong, Z. Yin, S. Li, High hydrogen permeability of Pd-Ru-In membranes prepared by electroless co-deposition, Sep. Purif. Technol. 343 (2024) 127073, https://doi.org/10.1016/j.seppur.2024.127073.
[46] S. Ramaprabhu, R. Leiberich, A. Weiss, Solubility of hydrogen in solid solution palladium alloys, Z. F. üR. Phys. Chem. 161 (1989) 83–105, https://doi.org/10.1524/zpch.1989.161.Part_1_2.083.
[47] Y. Shirasaki, T. Tsuneki, T. Seki, I. Yasuda, T. Sato, N. Itoh, Improvement in HYdrogen Permeability of Palladium Membrane by Alloying with Transition Metals, J. Chem. Eng. Jpn. 51 (2018) 123–125, https://doi.org/10.1252/jcej.17we128.
[48] A. Suzuki, H. Yukawa, Analysis for reverse temperature dependence of hydrogen permeability through Pd–X (X = Y, Ho, Ni) alloy membranes based on hydrogen chemical potential, Membranes 10 (2020) 123, https://doi.org/10.3390/membranes10060123.
[49] Y.H. Ma, I.P. Mardilovich, E.E. Engwall, Thin composite palladium and palladium/alloy membranes for hydrogen separation, Ann. N. Y. Acad. Sci. 984 (2003) 346–360, https://doi.org/10.1111/j.1749-6632.2003.tb06011.x.
[50] L. Ward, A. Dunn, A. Faghaninia, N.E.R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G. J. Snyder, I. Foster, A. Jain, Matminer: an open-source toolkit for materials data mining, Comput. Mater. Sci. 152 (2018) 60–69, https://doi.org/10.1016/j.commatsci.2018.05.018.
[51] G. Kresse, J. Hafner, Ab initio molecular dynamics for open-shell transition metals, Phys. Rev. B 48 (1993) 13115–13118, https://doi.org/10.1103/PhysRevB.48.13115.
[52] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. 6 (1996) 15–50, https://doi.org/10.1016/0927-0256(96)00008-0.
[53] V. Wang, N. Xu, J.-C. Liu, G. Tang, W.-T. Geng, VASPKIT: a user-friendly interface facilitating high-throughput computing and analysis using VASP code, Comput. Phys. Commun. 267 (2021) 108033, https://doi.org/10.1016/j.cpc.2021.108033.