Strategies for Cold-Starting Active Learning Loops Using Multiple Data Modalities

Anonymous authors
Paper under double-blind review

Abstract

In experimental materials science, every measurement counts. Discovering new materials in the context of compositionally complex materials is time-consuming and costly because of the high number of measurements required to screen the large composition-property space. To address this, there is a growing need for acceleration strategies that minimize data collection combined with acceptable surrogate model accuracy. Active learning can significantly reduce the number of labeled data points (measurements) required to train surrogate machine learning models while still achieving high predictive performance with low uncertainty. However, a major challenge in active learning is the cold-start problem: How to select informative initial points when no labeled data are yet available? We present and systematically evaluate multiple cold-start initialization strategies for active learning loops based on different existing "cheap" data modalities, as well as their multimodal combination. These strategies provide diverse and representative starting points and lead to rapid model convergence. Two acquisition functions, *Uncertainty Sampling* (US) and *Self-Adjusting Weighted Expected Improvement* (SAWEI), are compared for iterative point selection, automatically balancing exploration and exploitation. Active learning is stopped dynamically by monitoring the normalized mean predictive variance of the surrogate model. We apply our approach to eight experimental composition-spread materials libraries, a common setup for high-throughput screening, with different levels of compositional complexity. For those materials libraries we learn a surrogate model to predict electrical resistance as a function of composition. Our active learning framework significantly reduces the number of required measurements, achieving a reduction of 87% for some materials libraries using a single modality and a reduction of 85% on average for all materials libraries using a multimodal cold-start strategy. On average, we find that SAWEI outperforms uncertainty sampling. In summary, we demonstrate a practical, cold start active learning framework using a multimodal approach that accelerates autonomous experimental characterization on the path to autonomous materials discovery.

1 Introduction

The discovery of new materials is the search for a possible combination of chemical elements to find a specific composition with desired new properties or a comparable performance at a lower price compared to materials currently known, e.g., for use in electrocatalysis, electronics, or structural performance (McFarland & Weinberg, 1999). In experimental materials science, a common approach is to use composition-spread materials libraries, where thin films are deposited on a substrate exhibiting continuous spatial gradients of multiple chemical elements for a variety of different compositions on one sample (Ludwig, 2019). Each measurement area on such libraries corresponds to a distinct composition, typically defined by a predefined spatial sampling grid over the substrate. In our experiments, this is implemented using a 342-point grid on a materials library of approximately 10,cm. Various characterization techniques are applied (Ludwig, 2019), such as energy-dispersive X-ray spectroscopy (EDX) to analyze the chemical composition, X-ray diffraction (XRD) to determine the crystallographic structure, electrical resistance measurements, or UV-Vis spectroscopy to measure band gaps. Characterization of such libraries is time-consuming, and the speed

directly influences how many different compositions can be screened to find a new interesting property. A remedy is the use of active learning (AL), which guides the characterization toward measurements that maximally improve a surrogate model (Thelen et al., 2023; Stricker et al., 2025) and use it to predict the remaining ones with acceptable accuracy. However, typically the AL loop starts without labeled data ("cold start" (Maltz & Ehrlich, 1995; Schein et al., 2002)) and model convergence is not guaranteed. Here, we focus on one single property: the electrical resistance. Similar to (Thelen et al., 2023; Stricker et al., 2025), we build a surrogate model that maps the elemental composition to resistance using an AL approach. Our novelty lies in the systematic design, comparison, and experimental validation of multiple cold-start initialization strategies for active learning based on existing, cheap-to-obtain priors. AL guides the experiment to select the most informative next measurement area after initialization.

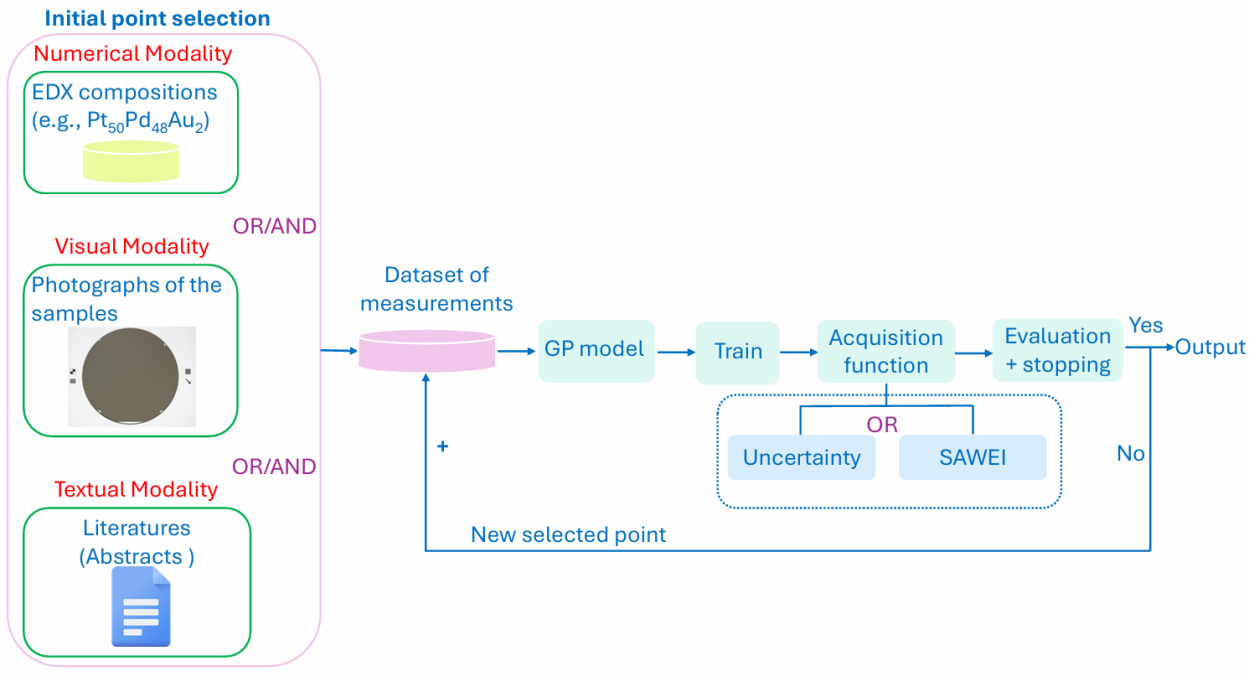


Figure 1: Overview of the active learning workflow combining single and multimodal initializations, a GP model, and acquisition functions (SAWEI or US with dynamic stopping.)

Figure 1 shows a schematic of our workflow: The algorithm begins with a small set of labeled data points ('Initial point selection'), typically on the order of 10, to train a surrogate model based on Gaussian Process Regression (GPR). We test two acquisition functions that decide on the next measurement area until a stopping criterion is reached.

AL for predicting electrical resistance has been demonstrated to be effective for composition-spread materials libraries (Thelen et al., 2023). There, the initialization of the surrogate model involves a fixed set of uniformly spaced measurement areas. This leads to a generally good performance across different materials libraries, but this does not consider that each library exhibits different elemental gradients and, consequently, different property gradients. In other words, non-optimal points are typically chosen for a cold start. Using priors about the composition-property correlation has been shown to decrease the number of required measurements (Stricker et al., 2025), although only on one materials library and only using one existing modality.

Both examples show that there is a potential in using AL for accelerating experimental characterization and even more so by incorporating prior knowledge of "cheap" correlations to cold-start an AL loop. We apply AL to materials library characterization using multimodal experimental data that include photographs of the libraries, compositions obtained through EDX, and composition-property correlations derived from word embeddings. We address two key challenges: (1) how to effectively cold-start AL using one or more modalities

by combining different features; (2) how to guide subsequent measurements using acquisition functions uncertainty sampling (US) and Self-Adjusting Weighted Expected Improvement (SAWEI) (Benjamins et al., 2023). The two acquisition functions capture distinct sampling behaviors: US emphasizes exploration and is highly sensitive to the spatial distribution of the initial points, whereas SAWEI adaptively balances exploration and exploitation.

For stopping, we monitor both the model's prediction uncertainty and the convergence of its mean uncertainty by analyzing the gradient of the covariance history, inspired by (Thelen et al., 2023). We stop when the gradient flattens, which means that no significant improvements in the model accuracy are expected from more measurements. We focus on improving the initialization phase of active learning, which is critical in experimental settings where no labeled data exist and show that replacing expert-driven or heuristic initialization with multimodal, cheap priors leads to faster convergence without modifying the downstream active learning loop.

2 Related Work

In the past, intuition combined with trial and error was the prevalent approach to finding new materials with desired properties. With increasing chemical complexity (using four or more elements in concentration steps of 1 %), the combinatorial possibilities are too large for an Edisonian approach to be practical. This challenge has driven the increasing demand for machine learning (ML) to build adaptive models that can intelligently guide experiments and accelerate materials characterization and, thereby, discovery (Xue et al., 2016; Jain et al., 2013). ML has shown promising results in a wide range of research areas, but training models with predictive power requires large amounts of labeled data, which is expensive and time-consuming and, therefore, often not available in an experimental setting (Nandy et al., 2022). This is where active learning becomes essential, which uses less data without compromising model performance (Settles, 2009), for example, a closed-loop system integrating Bayesian AL with experimental execution, demonstrating rapid phase mapping and property optimization in materials discovery (Kusne et al., 2020). Similarly, the combined AL with generative models allows the efficient design of wide bandgap materials (Xin et al., 2021). Another example is AL applied in an autonomous measurement setup to accelerate high-throughput characterization of thin-film materials libraries for predicting the electrical resistance, starting from a predefined set of initial measurement areas (Thelen et al., 2023). Any AL loop needs an initial set of starting points to be labeled or measured to train the model, but selecting *informative* starting points is non-trivial, particularly when there are no prior labels.

This challenge, known as the cold-start problem, has a representative example in (Zhu et al., 2019), which addresses the item cold-start problem in recommendation systems by leveraging item attributes to guide active learning. Their attribute-driven strategy selects the most informative items to query, improving recommendation accuracy with minimal user feedback. In materials science, a Bayesian co-navigation framework was introduced to jointly explore experimental and theoretical spaces for material systems, where active learning and a nested optimization loop iteratively refine surrogate models and adjust theoretical model parameters to align simulations with experiments under limited initial data conditions (Slautin et al., 2024). A recent study examines how the choice of initial datasets affects Bayesian optimization performance in materials discovery, proposing domain-informed criteria to select diverse and representative seed points and demonstrating that expert-guided initialization significantly improves optimization efficiency compared to random strategies (Hastings et al., 2025).

3 Active Learning Framework

3.1 Gaussian Process Regression

GPR is typically used as a surrogate model for expensive optimization tasks (Ravi et al., 2024). It is a nonparametric and probabilistic model that not only provides accurate predictions but also quantifies the uncertainty of those predictions. This makes it particularly well-suited for guiding sampling decisions in high-cost experimental settings. GPR solves the problem as a Bayesian estimation task, where the predictive function is inferred from observed data assumed to be realizations of a Gaussian-distributed random

process (Pasolli & Melgani, 2011). The set of $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where x_i represents the composition and y_i is the corresponding resistance value, is used to infer predictions for unlabeled compositions x_*. The predicted output y_* is represented by the expectation of the output quantity conditioned on the known data $L = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ and the new input x_*:

$$\hat{y}_* | L, \mathbf{y}_* \sim \mathbb{E}\{y_* | L, \mathbf{x}_*\} = \int y_* p(y_* | L, \mathbf{x}_*) \, dy_*. \tag{1}$$

The probability density of the predictive distribution $p(y_* | L, \mathbf{x}_*)$ can be expressed by (Rasmussen & Williams, 2006):

$$p(y_* | L, \mathbf{x}_*) \sim \mathcal{N}\left(\mu_*, \sigma_*^2\right) \tag{2}$$

where

$$\mu_* = \mathbf{k}_*^t \left[K + \sigma_n^2 I\right]^{-1} \cdot \mathbf{y} + b \tag{3}$$

$$\sigma_*^2 = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^t \left[K + \sigma_n^2 I\right]^{-1} \mathbf{k}_* \tag{4}$$

where $k(\mathbf{x}, \mathbf{x}')$ is the covariance/kernel function, K is the covariance matrix of the training samples, and $\mathbf{k}_*$ represents the covariance vector. The bias factor, the variance of noise and the identity matrix are indicated by b, σ_n, and I, respectively. Equation 3 gives the best estimate for the prediction at $\mathbf{x}_*$, while Equation 4 quantifies the uncertainty of that prediction. The predictive variance σ_*^2 plays a central role in our framework, as it is used both to guide the selection of new measurement locations and to define a practical stopping criterion for terminating the AL loop. In all experiments, we employ Gaussian Process Regression (GPR) with an exponential (Ornstein–Uhlenbeck) kernel and automatic relevance determination (ARD). This kernel is suitable for modeling anisotropic relationships and less smooth response surfaces compared to squared-exponential kernels, making it appropriate for composition–resistance mappings that may exhibit sharp variations. GPR is deliberately chosen as a surrogate model because its predictive uncertainty directly enables both cold-start initialization strategies and uncertainty-aware acquisition and stopping decisions in an experimental setting.

3.2 Cold Start Strategies

The selection of initial training data plays a critical role in the convergence speed of any AL pipeline (Yuan et al., 2011). For an effective cold-start, we use existing modalities for each materials library, enabling a library-specific strategy rather than relying on a fixed set of measurement areas (cf. (Thelen et al., 2023)). In short, we combine existing information about each materials library with data-driven methods to select a small but diverse set of measurement areas for optimal cold-start. All initialization strategies described below are applied under identical downstream AL conditions, allowing their impact on convergence speed and experimental efficiency to be directly compared.

3.2.1 Numerical Modality

The elemental composition, measured using EDX, constitutes the 'numerical modality' of each materials library. Characterization via EDX serves as a baseline measurement that is routinely performed to verify the quality and composition of the fabricated samples. As it is already part of the standard experimental workflow and can be acquired relatively quickly (e.g., on the order of a few hours per library), it can be considered *cheap* in the context of active learning. The composition also serves as the input to the surrogate model that maps composition to resistance. We explore it for use in cold-start using various techniques:

1. Random Sampling randomly picks N measurement areas from the entire set of measurement areas on the materials library U without replacement, where each measurement area has the same probability of being chosen. N is a user choice.

2. Latin Hypercube Sampling (LHS) (McKay et al., 1979) is a statistical method that selects N well-distributed samples across a multidimensional feature space by dividing each of the D dimensions into evenly spaced intervals. Each interval is represented exactly once in every dimension. N is a user choice.

3. K-Means Sampling (Bejarano et al., 2011) is an unsupervised clustering technique based on a user-defined number of K cluster centroids from the measurement areas in the materials library. K-Means iteratively minimizes the distance between each measurement area's composition and the centroid of its assigned cluster. Selected measurement areas explicitly capture the chemical diversity across a materials library, making this strategy representative-driven rather than coverage-driven. Here, $N = K$ and is a user choice.

4. Maximum Composition (Max Comp) (Stricker et al., 2025): initial measurement areas are selected where each individual element has its maximum in the composition. N is the same as the number of elements, hence dependant on the specific materials library (binary, ternary, quarternary, etc.).

5. Minimum Composition (Min Comp) (Stricker et al., 2025): initial measurement areas are selected where each individual element has its minimum in the composition. N is the same as the number of elements.

6. Farthest Point Sampling (FPS) (Eldar et al., 1997) is a greedy algorithm selecting a set of initial measurement areas by maximizing the distance between them. The first measurement area is a random selection, subsequent selections are *farthest* from all previously chosen points in the feature space. N is a user choice.

7. Greedy k-Center Sampling (Shi et al., 2022) selects a subset of k measurement areas as centers to represent K clusters and then iteratively picks the measurement area that has the maximum minimum distance to the selected center cluster. The objective is to minimize the maximum within-cluster distance, which is the distance from a sample to its closest center. $K = N$ is a user choice.

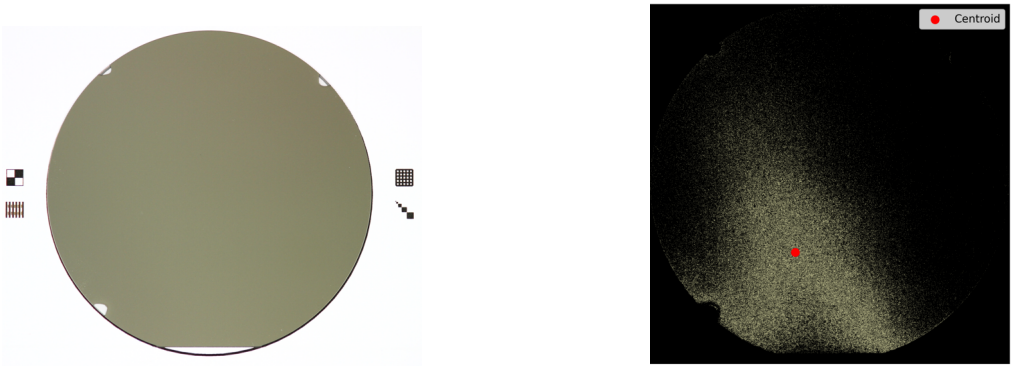
8. Outlier Discriminative Active Learning sampling (ODAL) is a warm-up strategy for improving the diversity of the initial training set in particular for imbalanced data scenarios (Barata et al., 2021). An outlier detection approach using Isolation Forest is applied. The method identifies measurement areas with *unusual* compositions and selects them as part of the starting points; the number of K selected points is user-defined (e.g., the top n_{init} outliers).

3.2.2 Visual Modality

The visual modality is called *visual* because the choice of initialization points is based on color photographs of the materials libraries. Photographs, like the EDX characterization, are also part of each library's baseline characterization and documentation procedure and are therefore readily available as a potential prior. K-Means clustering is applied to the color space of enhanced photographs of the materials libraries to identify visually distinct regions. This is based on the domain knowledge that differences in surface color can indicate differences in properties. Coincidentally, the color also correlates with film thickness: the film typically does not exhibit a constant thickness, so the color is also affected by the underlying substrate. In turn, the thickness of the film also affects the resistance, i.e., the color typically correlates not only with properties but particularly with the resistance. We use three levels of saturation enhancement (low, medium, and high) to bring out subtle color variations. Pixels are grouped on the basis of their RGB color values, and we use the centroids of these clusters. The color and spatial location within each group are mapped back to the materials library's measurement area grid and used as initialization candidates. The choice of $N = k$ is a user choice. Figure 2 shows examples of (a) the raw photograph of the materials library, subfigure (b) highlights one representative color cluster with its centroid marked by a red cross; all unrelated pixels and background regions are shown in black.

3.2.3 Textual Modality

The textual modality's name originates from the fact that the initialization candidates are derived from latent knowledge in scientific abstracts. A word embedding-based model (Zhang & Stricker, 2023) is used to compute similarities between the word "resistance" and individual elemental compositions. It is based on the observation that similarity in word embedding space between compositional representations and properties encodes physical/chemical correlations (Zhang et al., 2025). Once trained on a corpus of open



(a) (b)

Figure 2: (a) Raw materials library photograph used as input. (b) Processed materials library photograph with one cluster pixel group (grayish) and its centroid (red), one of K cluster center locations to be used as initialization points for active learning.

access abstracts, the model is used to represent elemental names and properties like 'resistance' as a vector in a high-dimensional space, here 200 dimensions. By comparing the cosine similarity between linearly weighted compositional vectors and property vectors, we estimate how closely each element's composition representation in vector space (e.g., $Pt_{50}Pd_{48}Au_2$) relates to the concept of 'resistance' in vector space through cosine similarity.

Figure 3 provides a schematic illustration of the relative positions of compositional and property-related terms in the word embedding space. For visualization purposes, only two dimensions are shown, while the actual word embedding space used has 200 dimensions. Importantly, the textual modality does not represent a causal or predictive model of electrical resistance. Instead, it provides a natural language-based prior that numerically encodes latent knowledge from the literature. Although these associations are inherently noisy, they are sufficient to positively bias early sample selection towards compositionally relevant regions. Similarities are used to guide the initial selection of measurement points in the AL setup by using the K top similarity measurement points.

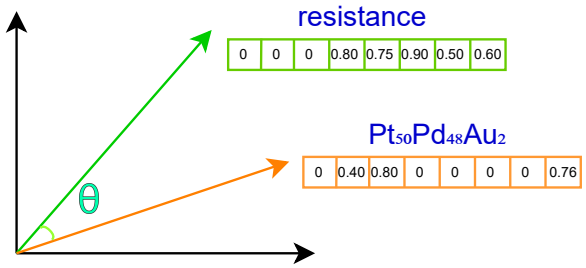


Figure 3: Illustration of cosine similarity between the word vector $\vec{\text{resistance}}$ and the composition vector $\vec{Pt_{50}Pd_{48}Au_2}$ in a shared embedding space. Smaller angles indicate stronger semantic relationships.

3.3 Acquisition Functions

The choice of acquisition function determines which data points are selected next by quantifying how promising each point is for model improvement. We test two acquisition strategies: *Uncertainty Sampling* (US) Lewis & Catlett (1994) and *Self-Adjusting Weighted Expected Improvement* (SAWEI) Benjamins et al. (2023) which are detailed below. We select US and SAWEI because they represent two fundamentally different ac-

tive learning behaviors. US performs pure uncertainty-driven exploration and is highly sensitive to the initial choice of training data points, making it well suited to evaluate the quality of cold-start strategies. In contrast, SAWEI adaptively balances exploration and exploitation by incorporating expected improvement and a dynamic weighting mechanism, reflecting a more realistic optimization-driven experimental setting. Evaluating cold-start strategies under both acquisition regimes ensures that observed improvements are not specific to a single sampling behavior, but remain robust across exploration-dominated and performance-driven active learning scenarios.

Uncertainty Sampling (US)

US selects the next point where the model is most uncertain, encouraging exploration of regions not well covered Lewis & Catlett (1994). In the context of Gaussian Process Regression, the acquisition function is:

$$\mathbf{x}_{\text{next}} = \arg\max_{\mathbf{x}} \sigma_*^2(\mathbf{x}) \tag{5}$$

The next point, $\mathbf{x}_{\text{next}}$, with the highest uncertainty is chosen to be measured.

Self-Adjusting Weighted Expected Improvement (SAWEI)

To adaptively balance exploration and exploitation, we use the SAWEI strategy (Benjamins et al., 2023), which builds on the expected improvement (EI) at a point x, defined as:

$$\text{EI}(\mathbf{x}) = \mathbb{E}[\max(f(\mathbf{x}) - f(\mathbf{x}^+), 0)] \tag{6}$$

where $f(x^+)$ is the best observed value so far and $f(x)$ is modeled as a GP with posterior mean $\mu(x)$ and standard deviation $\sigma(x)$. The closed-form expression for EI is:

$$\text{EI}(\mathbf{x}) = (\mu(\mathbf{x}) - f(\mathbf{x}^+))\Phi(z) + \sigma(\mathbf{x})\phi(z) \tag{7}$$

where:

$$z = \frac{\mu(\mathbf{x}) - f(\mathbf{x}^+) - \xi}{\sigma(\mathbf{x})} \tag{8}$$

and Φ and ϕ denote the cumulative distribution function (CDF) and the probability distribution function (PDF) of the standard normal distribution, respectively. ξ is an exploration hyperparameter in the acquisition function. Larger ξ values bias the selection towards the exploration of uncertain regions; smaller ξ values favor exploitation. SAWEI extends EI by setting a weighting factor α_t that controls the influence of the exploration-exploitation trade-off based on the *Upper Bound Regret* (UBR), which quantifies the potential loss due to not evaluating the globally optimal point. The acquisition function becomes

$$\text{SAWEI}_t(\mathbf{x}) = \alpha_t \cdot \sigma(\mathbf{x}) + (1 - \alpha_t) \cdot \text{EI}(\mathbf{x}) \tag{9}$$

The weighting parameter $\alpha_t \in [0, 1]$ is updated at each iteration based on the change in UBR. If the UBR remains constant for multiple iterations, α_t is changed towards the inverse of the currently dominant attitude. This self-adjusting mechanism allows SAWEI to respond dynamically to the optimization landscape without requiring manual parameter adjustment.

3.4 Stopping criteria

Defining an appropriate stopping criterion in AL avoids superfluous measurements while still achieving the intended goals of AL: a *good-enough* approximation of a function. *Good-enough* here refers to an error, e.g., the mean absolute error, below the uncertainty of the measurement technique itself. An *ideal* stopping criterion balances accurate predictions with resource efficiency. We use a dynamic stopping criterion based on the model's predictive uncertainty. For this, we monitor the normalized mean predictive variance (covariance) of the Gaussian process model across iterations (Thelen et al., 2023). In addition, we compute the Mean Absolute Error (MAE) using the full ground truth as part of the stopping criterion to ensure that the surrogate model achieves sufficient predictive accuracy as part of our method development. This verification

is essential now, especially because low uncertainty alone does not guaranty good prediction performance. After an initial warm-up phase with a fixed number of measurements (e.g., 5 points), we track the model predictive variance progression over subsequent iterations. If the normalized mean variance shows a consistent flattening trend, quantified by the gradient of the mean variance remaining below a predefined threshold (0.5% per iteration) for several consecutive iterations and the MAE less than a defined value (0.5Ω), the AL process is terminated. In our experience, this is a good heuristic criterion for model stability, and we observe little to no improvement from additional measurements. While the heuristic dynamic stopping criterion based on the predictive uncertainty provides a pragmatic stopping mechanism, we emphasize that further work is needed to establish a more general stopping criterion. The focus of the current study is to establish a theoretical framework and, therefore, uses uncertainty-based stopping in conjunction with MAE.

4 Results

In this section, we first present details of the data set. Subsequently, we evaluate the performance of single-modality cold-start strategies across multiple materials libraries. Then we analyze the effect of combining modalities for multimodal cold start and compare acquisition functions. Finally, we compare the proposed strategies against established baselines from previous work, assess robustness with respect to random seeds, and disentangle the effect of initialization quality from the number of initial points.

4.1 Data sets and Modalities

Our AL framework is evaluated using experimental data from composition-spread materials libraries. These libraries combine noble elements such as Ag (Silver), Au (Gold), Pd (Palladium), Pt (Platinum), Rh (Rhodium), and Ru (Ruthenium).

Table 1 provides an overview of the chemical systems of all used materials libraries. The chemical complexity ranges from ternary (Ag-Au-Pd) to quinary (Ir-Pd-Pt-Rh-Ru) alloy systems. These are typically referred to as 'compositionally complex solid solutions' or 'complex concentrated alloys'. Since the crystal structure is constant (face centered cubic), it is only the chemical composition that affects the property, here the resistance. This observation makes this materials class a good platform for tunable properties because the gradients of the properties typically follow gradients of compositions relatively smoothly (Batchelor et al., 2019; Löffler et al., 2021; Svane & Rossmeisl, 2022).

Elemental compositions are measured via EDX and serve as the numerical modality. Prior to resistance measurements, photographs of each library are available and provide the visual modality. The target property is the electrical resistance measured via a four-point probe setup. In addition, textual representations are derived from word embeddings trained on scientific abstracts and provide a global semantic prior relating compositions to material properties.

The individual data sets differ in compositional gradients, element combinations, and resistance distributions, enabling evaluation of robustness of our cold start strategies across heterogeneous chemical spaces.

Table 1: Materials systems used in the study

data set ID	**Materials System**
10272	Ag-Au-Pd
10275	Ag-Au-Pd-Pt-Rh
10304	Au-Pd-Pt-Rh
10311	Au-Pd-Pt-Rh-Ru
10402	Ag-Au-Pd-Pt
10403	Ag-Au-Cu-Pd-Pt
10374	Ir-Pd-Pt-Rh-Ru
10399	Au-Cu-Pd-Pt

4.2 Single-Modality Cold-Start Performance

To investigate the impact of the different initialization strategies on AL performance, we evaluate each modality present for cold start individually and in combination with other modalities. Different modalities offer different perspectives on the composition-property correlations and are used to select five *informative* starting points, approximately equal to 1.5% of the total measurement areas, similar to previous studies which serve as a baseline (Thelen et al., 2023; Stricker et al., 2025).

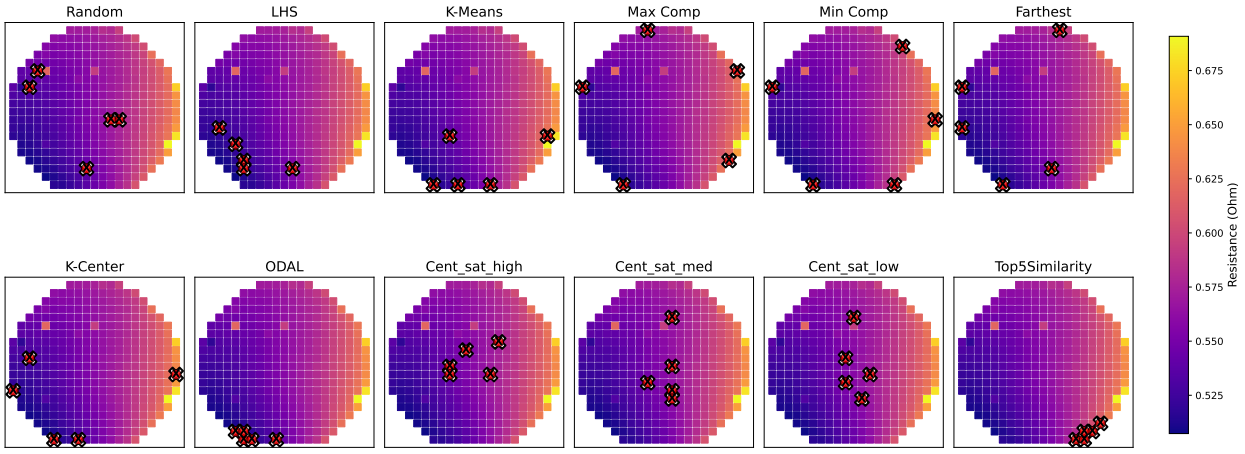


Figure 4: Comparison of different initialization strategies used in AL on materials library ID 10374. Each subplot shows the selected initial measurement areas marked with red crosses overlaid on the measurement area grid. Coloring is according to the target property 'resistance'.

Figure 4 illustrates how different initialization strategies select initial measurement areas across on the real space grid of a materials library. Each subfigure shows the selected points (red crosses) overlaid on the 342 measurement areas real space grid of one of the materials libraries and colored by resistance in Ohm as it was measured, i.e. the ground truth we want to approximate. These differences in the selected initial measurement areas demonstrate that the choice of initialization strategy fundamentally shapes the surrogate model's initial hypothesis space, as different strategies emphasize compositional extremes, spatial coverage, or semantic relevance.

For single-modality cold starts, we observe that several modalities converge and yield good average surrogate model performance. Figure 5 provides an overview of the performance of different initialization strategies with the acquisition functions US (a) and SAWEI (b). Values for each internal sample ID/initialization strategy denote the achieved reduction. Fields without values indicate no convergence within 100 iterations corresponding to about 30 % of the total number of grid points which is 342. For the US acquisition function, Greedy k-Center sampling achieves the highest overall reduction, for SAWEI, visual-based clustering is the best. The lowest reduction is observed for minimum composition using US, and FPS is the least effective under SAWEI. The cold-start strategies that perform best for the two acquisition functions in all materials libraries converge within the first 20 to 30 iterations; cf. Figure 6.

This demonstrates the strength of single-modality cold-start approaches. A prior-informed strategy provides an efficient initialization and the models reach accurate predictions within a few iterations.

4.3 Multimodal Cold-Start Performance

After evaluating each modality individually, we investigate the effect of combining multiple modalities for cold-start on model convergence. The idea behind combining multiple modalities is to leverage complementary information from different 'dimensions' that might affect composition-property relationships. We explore whether modality diversity accelerates model convergence and further minimizes the number of necessary measurements compared to single-modality cold-start. The results are shown in Figures 7 and

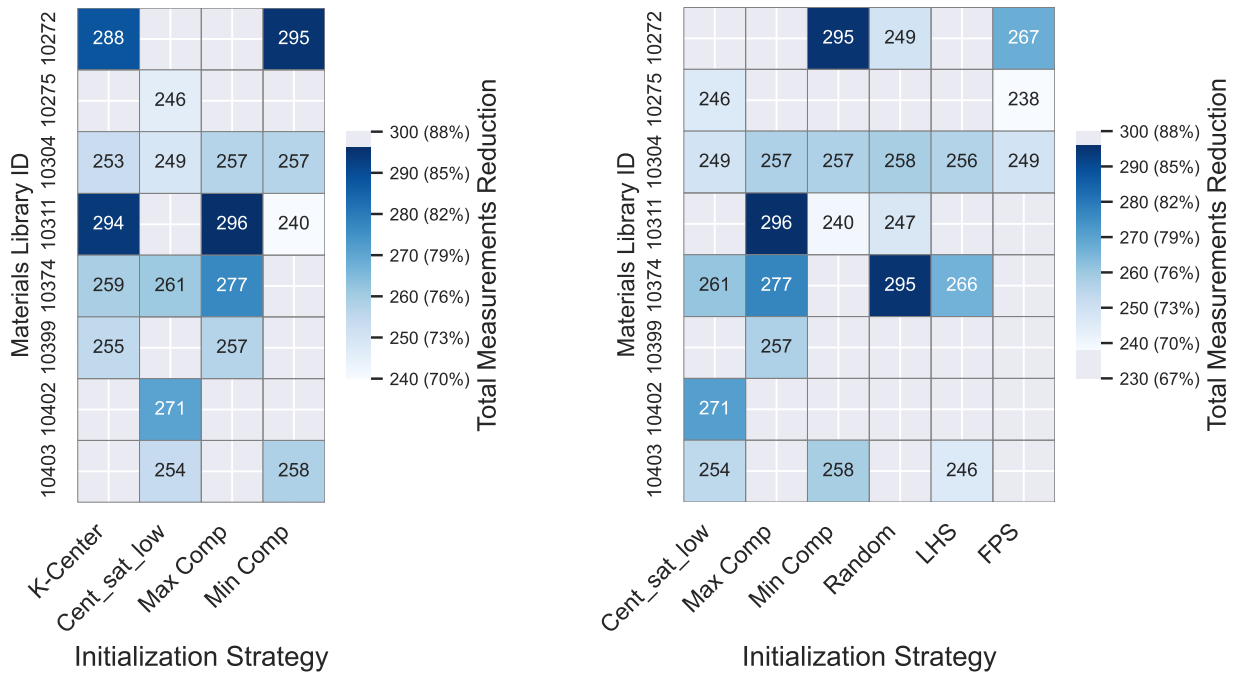
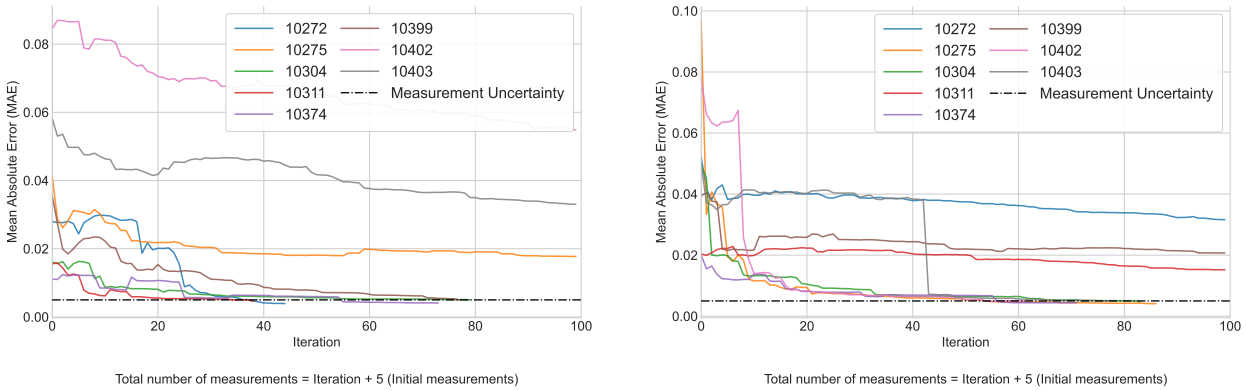


Figure 5: Performance of single-modality strategies across eight materials libraries: (a) US and (b) SAWEI acquisition functions. Both strategies reduce the number of measurements by over 50% (maximum 342 per materials library). Abbreviations: K-Center – greedy k-center sampling; Cent-Sat-Low – image-based centroid clustering with low saturation; Random – random sampling.



(a) (b)

Figure 6: Mean Absolute Error (MAE) over 100 iterations for predicting resistance (materials library ID 10311). (a) Top-5 Similarity (textual) with US acquisition across all initialization strategies. (b) Minimum Composition (numerical) with SAWEI acquisition across all initialization strategies. The Four-point probe baseline [4] is shown for comparison.

Figure 8. These show the best-performing combinations for the US and SAWEI acquisition functions as a heatmap, where darker red means a higher reduction of required measurements, additionally shown as values. Both heatmaps only include combinations of initialization strategies that decrease the number of measurements on average by approximately 50% to 80% across all tested materials libraries. The level of reduction is consistently achieved across all materials libraries, each with distinct compositions supporting

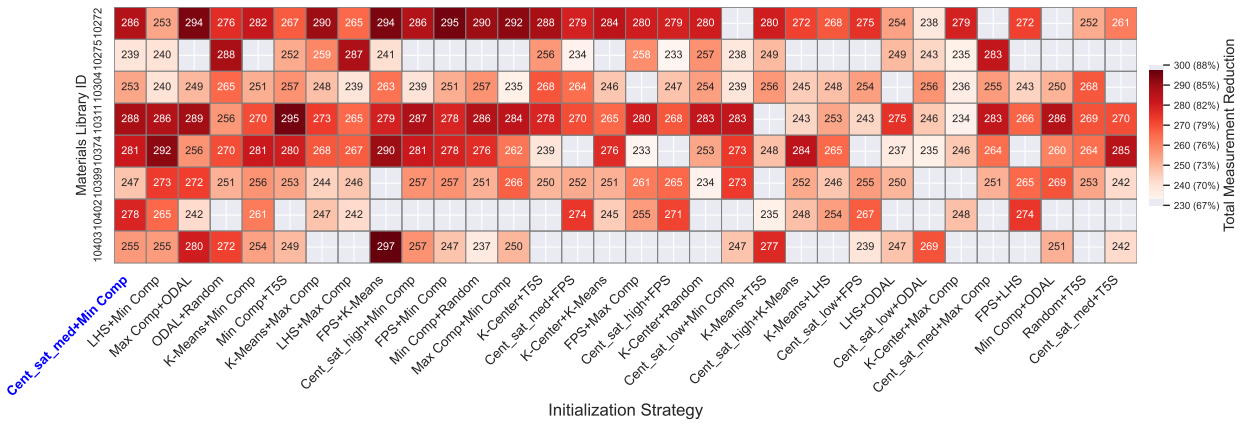


Figure 7: Heatmap of measurement reduction for initialization strategies that reduce the measurements needed by 85% down to 50% across materials libraries with US (maximum 342 per materials library). Abbreviations: Cent_sat_high – high saturation centroids; Cent_sat_med – medium saturation centroids; Cent_sat_low – low saturation centroids; T5S – Top-5 similarity; Random – Random sampling. Boxes without color and numbers indicate non-converged tests.

the robustness and generalizability of our multimodal cold-start strategies. Among them, the combination of visual-based clustering and minimum composition consistently performs best regardless of the acquisition strategy, as highlighted in the heatmaps. We attribute this behavior to the complementary roles of the two modalities: minimum composition defines (lower) boundaries in composition space, whereas surface color-based clustering captures correlations with resistance-related variations.

Both modalities are ideally suited for AL cold-start because they exist before any costly measurements are made, as they are part of the standard characterization procedure for characterizing experimental samples and are comparatively cheap. Other effective multimodal initialization strategies, such as maximum composition with ODAL and textual modality with minimum composition, also show fast model convergence. Overall, our tests demonstrate that the combination of multiple modalities or different strategies from the same modality is more efficient than using a single modality or strategy.

Figure 9 shows the MAE (w.r.t. the ground truth) evolution as a function of measurements for two combined data modalities: Top 5 Similarity (T5S), a textual modality, and minimum composition strategy, a numerical modality with US and SAWEI acquisition function, respectively, for materials library ID 10311. When we combine each of the two data modalities with the different initialization strategies, we find that some combinations work particularly well: Pairing semantic relevance (Top-5 Similarity) with numerical diversity (Min Comp) or outlier-based sampling (ODAL) with compositional extremes (Max Comp) leads to faster convergence and better model performance early in the AL process. Altogether, these results confirm the robustness of multimodal initialization.

A summary of all acquisition functions combined with the different initialization strategies is shown in Figure 10. The figure provides a compact overview of the performance of all cold-start strategies evaluated with the two acquisition functions. A total of 12 base strategies (x-axis) are combined with each other, and we measure the impact of bi-modal cold-start. Note, the x-axis shows two bars, one for each acquisition function. For each acquisition function, we measure if the pairing of two modalities for cold start affects the base strategy positively (decrease) or negatively (increase) or not at all.

The chart compares both individual and combined strategies under the US and SAWEI acquisition functions. Some strategies stand out: Min Comp, ODAL, K-Means, FPS, and T5S frequently reduce the number of necessary measurements across different data sets independent of the acquisition function. Random sampling (based on a single run with a fixed seed) or visual-based centroid methods show less consistent results. While small improvements are observed in some cases, they occasionally lead to an increased number of necessary measurements for convergence except if the number of initialization points is doubled. We attribute this

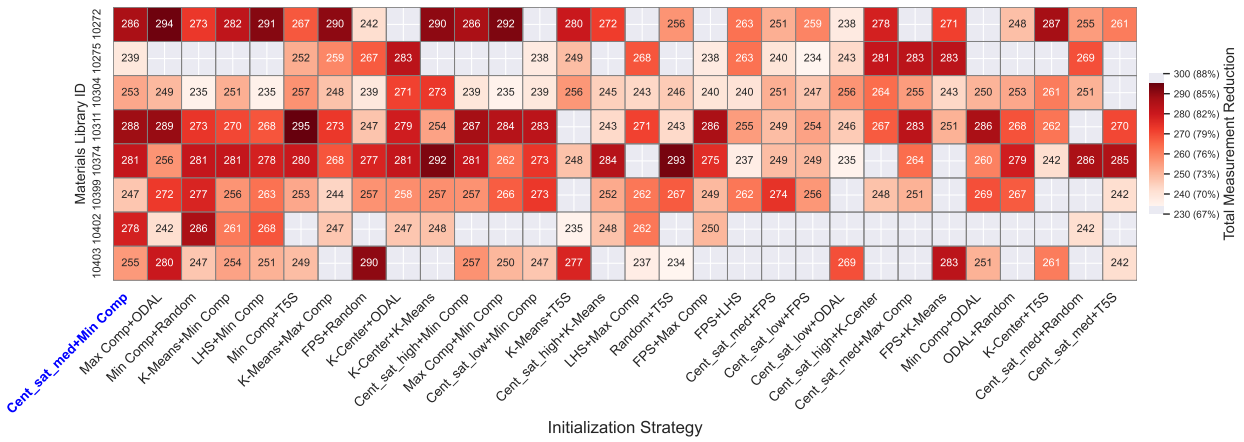
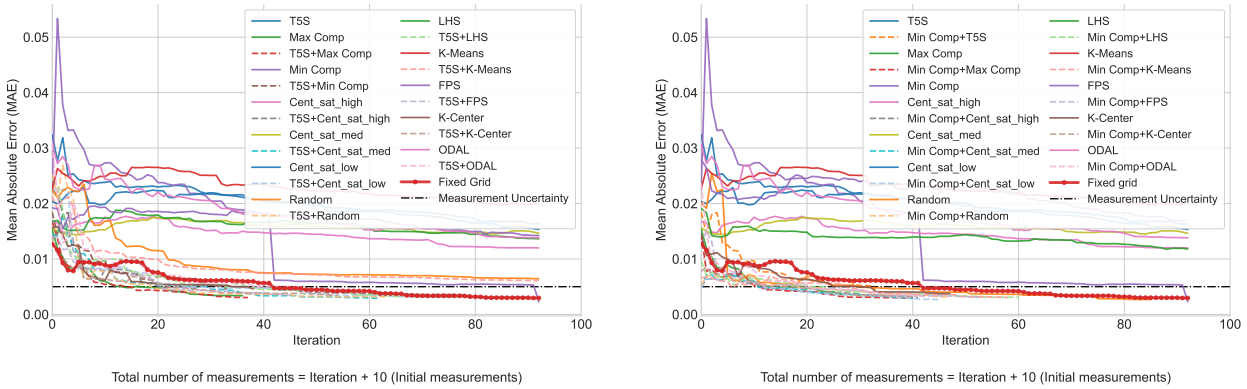


Figure 8: Same as Fig. 7 but for the SAWEI acquisition function.



(a) (b)

Figure 9: Mean Absolute Error (MAE) with respect to the ground truth over 100 iterations for resistance prediction on materials library ID 10311. (a) Results using the Top-5 Similarity (T5S, textual modality) initialization combined with the US acquisition function across all initialization strategies. (b) Results using the Minimum Composition (MinComp, numerical modality) initialization combined with the SAWEI acquisition function across all initialization strategies. In both panels, we include the fixed-grid baseline from (Thelen et al., 2023) and a single-run random sampling baseline (performed with a fixed seed) for direct comparison. Strategies that outperform the random and fixed-grid baselines demonstrate more stable and data-efficient convergence behavior.

increase to the randomness of the point selection or the limitations of certain methods, which bear no correlation to properties, i.e., 'hit-and-miss'. Using the 'low saturation' in the processing of photographs of metallic thin-film libraries typically fails to capture visual differences, making the selected points less suitable for initialization. This comparison shows that thoughtful, informed initialization using diverse modalities can make a large difference in how fast AL-guided characterization converges in real materials characterization. More specifically, SAWEI generally outperforms US. However, US works better with strategies like random sampling or ODAL when it is combined with another modality, as it works by selecting the uncertain or under-represented areas in the feature space to explore. Ultimately, the best choice of acquisition function depends on the initialization strategy used.

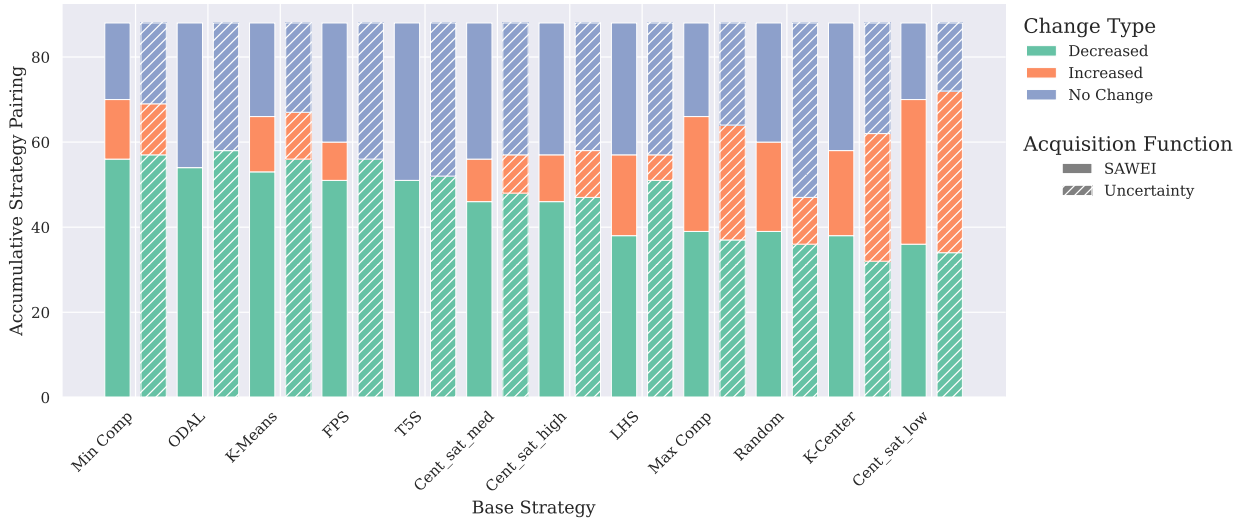


Figure 10: Summary of performance across all data sets for both SAWEI (full bars) and US (shaded bars) acquisition functions for multimodal initialization. The x-axis denotes the base strategy, and y-axis values indicate the number of total strategies combined with the base strategy. Coloring indicates a decrease (green), no change (blue) or increase (orange) in the number of measurements needed to converge.

4.4 Comparison with Previous Work

Active learning (AL) has previously been applied to predict electrical resistance in high-throughput materials libraries (Thelen et al., 2023) and later integrated into a workflow environment (Stricker et al., 2025).

In (Thelen et al., 2023), the AL loop was initialized using a fixed grid of uniformly spaced compositions across the 342 available measurement areas. While this strategy effectively reduced the number of required measurements and serves as an important proof of concept, it does not account for variations in elemental gradients and, consequently, property gradients across different materials libraries.

In our study, we improve upon the previous benchmark by employing multimodal cold-start. We find that between 30% and 80% of the mixed initialization strategies outperform the baseline; the numbers are reported in Table 2. This shows the advantage of using multimodal priors that bring together different views on the available data space of the materials library. By including information from composition along with visual features from materials library photographs taken of materials libraries after synthesis and textual insights from literature, the initial selection of measurement areas becomes much more diverse and specific to the individual materials library.

Our approach leads to faster model training and ultimately fewer experiments needed to reach acceptable prediction performance.

The benefits of this approach are further demonstrated in Table 2, which compares the performance of mixed initialization strategies across eight materials libraries. For each library, we report two metrics: (1) the number of mixed strategies that result in a lower number of measurements compared to the baseline using a fixed grid as initialization (Thelen et al., 2023), and (2) the corresponding percentage out of all proposed mixed strategies. For example, the materials library of ID 10311, 53 strategies, amounting to 80.3%, outperforms the baseline. Similarly, libraries 10304 and 10399 show improvement rates of 72.7% and 69.7%, respectively.

Figure 11 visually compares how different initialization strategies select their starting points. The baseline (Thelen et al., 2023) fixed grid method (left) uses fixed, evenly spaced positions on the materials library. The minimum composition and visual strategies (middle) adaptively choose points that capture structural and spatial variations. The multimodal approach (right) brings these perspectives together, resulting in more

diverse, library-specific initial measurements. Consequently, sampling is more efficient and convergence is faster compared to single-modality or fixed initializations.

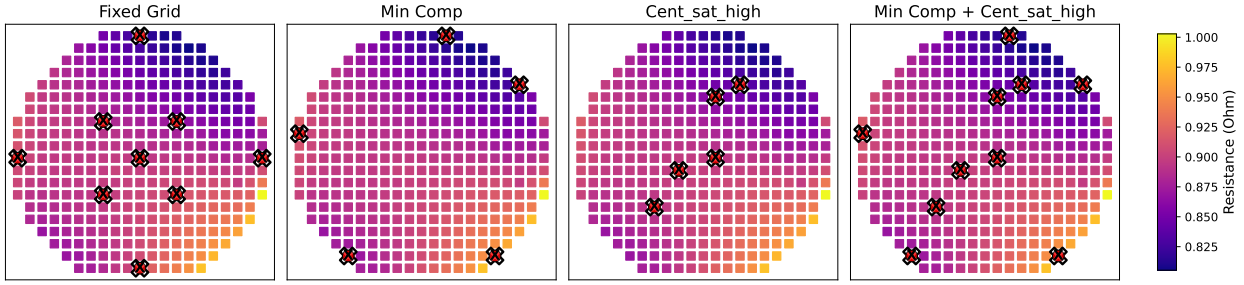


Figure 11: The subplots show the selected initial measurement areas (red crosses) overlaid on the resistance map of the library (342 measurement areas in total). The baseline fixed grid initialization (left) (Thelen et al., 2023). The minimum composition strategy (second) and the visual modality (third). The multimodal initialization (right) combines both approaches, producing a more diverse and representative set of initial points that capture complementary aspects of the library.

Table 2: Effectiveness of mixed initialization strategies with different data modalities priors across eight materials libraries using uncertainty sampling. The table reports the number and percentage of mixed strategies that reduced the required number of measurements compared to four-point probe measurements. The resulting improvement rate is also reported.

Material ID	**Mixed Strategies (Count)**	**Improvement Rate (%)**
10272	2	3.03
10275	30	44.70
10304	48	72.73
10311	53	80.30
10403	24	36.36
10402	18	27.27
10399	46	69.70
10374	13	18.94

Figure 12 further confirms this reduction by directly comparing the fixed grid baseline (Thelen et al., 2023) with the best-performing mixed strategies across eight materials libraries. Each bar pair corresponds to one library, with green bars representing the number of measurements required by the baseline and orange bars showing the number of measurements by the best mixed strategies. In all cases, the orange bars are consistently shorter, indicating faster convergence using multimodal cold-start. The magnitude of the savings varies by materials library, ranging from modest reductions of around 10% to substantial reductions exceeding 50%. For example, the materials library with ID 10311 shows one of the largest improvements, where the mixed strategy nearly halves the required number of measurements compared to the baseline. These results visually highlight that proposed initialization strategies consistently outperform uniform composition-based initialization, accelerating the learning process while reducing experimental effort.

Summarizing, the combination of different strategies improves data efficiency and robustness in our AL pipelines: our approach outperforms random sampling and constitutes a significant improvement over previous AL approaches for materials libraries (Thelen et al., 2023; Stricker et al., 2025).

A closely related study is presented by (Hastings et al., 2025), where the authors systematically investigate the effect of initialization strategies on Bayesian optimization performance in materials discovery. In that work, domain-informed criteria are used to select diverse and representative starting points, and the impact of different initialization schemes is evaluated in a controlled setting. Their analysis highlights that expert-guided initialization can substantially improve optimization efficiency compared to purely random selection.

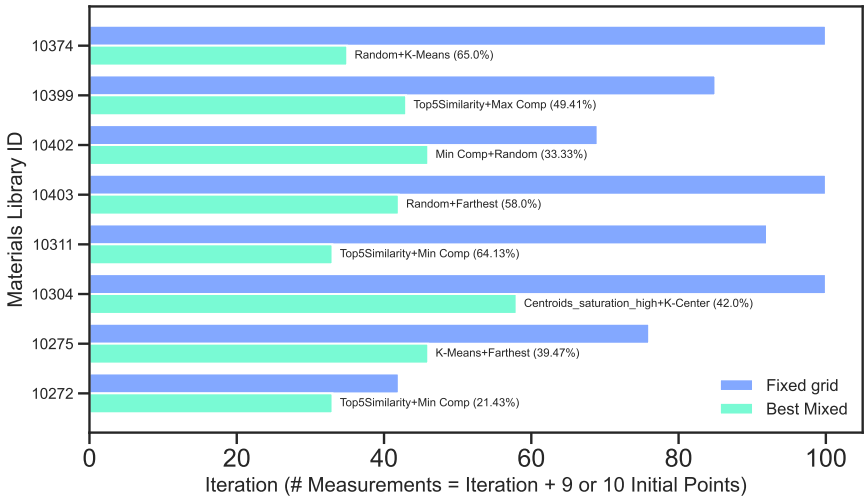


Figure 12: Comparison of the number of measurements needed between the evenly fixed grid baseline (Thelen et al., 2023) and the best-performing mixed strategies across eight materials libraries using the SAWEI acquisition function. Each bar pair corresponds to a specific materials library ID.

While conceptually related, our study differs in several important aspects. First, (Hastings et al., 2025) primarily evaluates initialization strategies in simulated or model-driven settings, where the underlying response surface is known or computationally generated. In contrast, we investigate multimodal initialization strategies directly within a real experimental high-throughput measurement framework. Second, our approach explicitly integrates heterogeneous prior information into the initialization process like numerical composition data, visual features extracted from material library images, and textual knowledge from literature.

This experimental setting enables us to assess the practical impact of multimodal priors on early-stage active learning performance under realistic measurement noise, spatial inhomogeneities, and experimental constraints. Therefore, beyond demonstrating improved optimization efficiency, our work provides empirical evidence that multimodal, data-informed initialization strategies lead to systematic and reproducible gains in experimental active learning workflows.

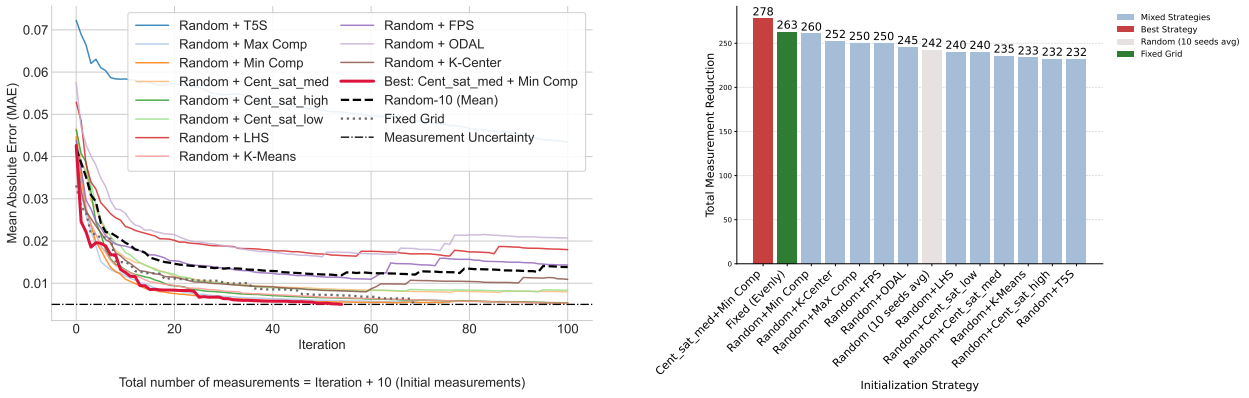
4.5 Robustness Across Random Initializations

To evaluate the robustness of the proposed mixed cold-start strategies, we analyze their performance across multiple random seeds. For each materials library, we select $K = 10$ initial measurement areas using ten different random seeds.

For the purely random baseline, all K initial points are sampled uniformly at random from the available measurement areas without replacement. In contrast, the mixed initialization strategies are constructed by selecting $K = 5$ random points without replacement and combined with $K = 5$ points obtained from one of the three modalities (numerical, visual, or textual), using the same set of ten random seeds for fair comparison. We additionally compare these mixed strategies to (i) the best-performing initialization strategy identified from the heatmap analysis in the previous section and (ii) the fixed grid baseline.

For each configuration, we run the AL loop under identical conditions and report: (i) the MAE averaged across seeds as a function of the active learning iteration, and (ii) the total reduction in required measurements until convergence. Averaging across seeds ensures that the reported performance does not arise from a favorable random initialization, but instead reflects systematic and reproducible behavior.

Figure 13 evaluates the robustness of the proposed multimodal initialization strategies under the Uncertainty Sampling (US) acquisition function for materials library 10402. In Figure 13(a), we report the MAE trajectories averaged over ten random seeds for each mixed initialization strategy. Variations across seeds arise because some runs converge earlier while others continue until the full iteration budget, leading to small



(a) (b)

Figure 13: Performance across multiple random seeds for data set 10402 using the US acquisition function. (a) Mean absolute error (MAE) as a function of active learning iteration. Thin coloured curves correspond to mixed initialization strategies (averaged over 10 seeds) with random sampling; the thick red curve highlights the best-performing strategy, the dashed black curve shows the mean over 10 purely random initializations, and the dotted curve represents the fixed grid baseline. (b) Total measurement reduction achieved by each strategy. Multimodal initialization consistently outperforms the averaged random baseline.

fluctuations at later iterations. The majority of mixed initialization strategies achieve lower MAE than the averaged random baseline throughout most of the learning process. The best-performing multimodal strategy combining a visualization-based prior with a numerical composition-based strategy (highlighted in red) consistently achieves the fastest error reduction and maintains the lowest MAE compared to both the averaged random baseline and the fixed grid initialization.

Importantly, even when mixed strategies incorporate $K = 5$ randomly selected initial points (using ten different random seeds) together with modality-informed priors, their averaged performance remains competitive with and often superior to purely random initialization and the fixed grid baseline. This indicates that the observed improvements are not driven by favorable seed selection but instead arise from the informative nature of the prior-based initialization.

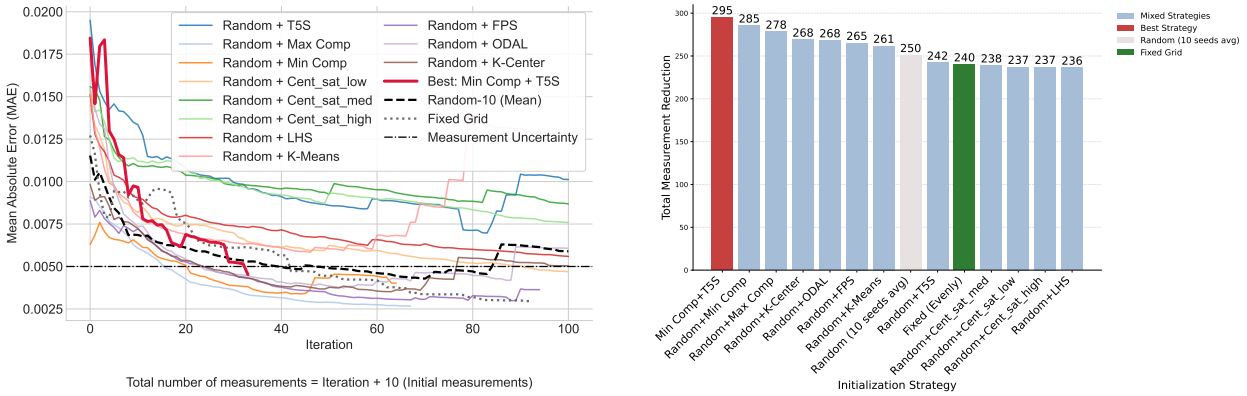
Figure 13(b) further quantifies this behavior by reporting the total measurement reduction until convergence. The multimodal strategy achieves larger reductions compared to the average over ten purely random initializations, confirming that the improvements are systematic rather than stochastic. While individual random seeds mixed with prior-based initialization strategies may occasionally perform competitively compared to fixed grid initialization, their averaged performance remains significantly inferior to the best multimodal configurations.

Figure 14 presents the results for materials library 10311 under the SAWEI acquisition function. In this case, the strongest performance is achieved by combining numerical and textual priors.

Overall, these results confirm that combining modality-informed priors with random initialization yields robust and reproducible improvements in active learning efficiency under uncertainty-driven exploration.

4.6 Effect of the number of initial points

We mention that a single-modality strategy using just five initial measurement areas can reduce the number of required measurements by up to 87% for certain materials libraries. Mixed modality cold-start strategies where the number of initial points is effectively doubled—achieve over 85% reduction across all libraries. This naturally raises the question: Are the improvements in mixed strategies simply due to having more initial points? To answer that, we run a comparison using all the base initialization strategies, except for minimum (Min Comp) and maximum (Max Comp) composition (since they depend on the number of elements in the materials library), and fix the number of initial points to eight, which is approximately equivalent to



(a) (b)

Figure 14: Performance across multiple random seeds for data set 10311 using the SAWEI acquisition function. (a) Mean absolute error (MAE) as a function of active learning iteration. Thin coloured curves correspond to mixed initialization strategies (averaged over 10 seeds) with random sampling; the thick red curve highlights the best-performing strategy, the dashed black curve shows the mean over 10 purely random initializations, and the dotted curve represents the fixed grid baseline. (b) Total measurement reduction achieved by each strategy. Multimodal initialization consistently outperforms the averaged random baseline.

the mixed modality strategies. This provides a fair comparison to assess if the improvements by using two modalities are an affect of the initialization quality or the quantity of initialization points.

Table 3 summarizes the improvement rates of mixed strategies compared to their single-strategy baselines. The improvement rate represents the percentage of cases (across all materials libraries) in which a mixed strategy converges earlier than the baseline strategy. For example, an improvement rate of 66% means that in two out of three materials libraries tested, the mixed strategy converges faster than the single-strategy baseline. Across all base strategies, mixed initialization is faster in 35–67 % of the comparisons, confirming that the benefits are the results of a more informative initialization rather than simply increasing the number of starting points. Combining modalities captures more diverse and informative regions of the composition space-property space enabling faster and more efficient model convergence.

Table 3: Improvement rate (% of cases where the mixed strategy required fewer measurements than its single-modality baseline) using the *SAWEI* acquisition function.

Base Strategy	**Improvement (%)**
Outlier-based sampling (ODAL)	67.35
Farthest Point Sampling (FPS)	64.15
Greedy k-Center Sampling	64.00
Random Sampling	61.36
Textual Modality (T5S)	61.19
K-Means Sampling	56.25
Latin Hypercube Sampling (LHS)	41.46
Visual Modality (Centroids medium saturation)	40.48
Visual Modality (Centroids low saturation)	38.46
Visual Modality (Centroids high saturation)	37.21

When comparing mixed strategies with baselines with a similar number of starting points, we observe that the most consistent improvements come from multimodal combinations. In particular, the top ten strategies listed in Table 4 are a combination of a visual modality (from materials library photos) with a numerical modality (like composition-based initialization). They outperform their corresponding single strategies in 6 out of the 8 materials libraries, showing that it is not just about how many starting points you have, but

which ones you choose. This aligns with our main findings, where multimodal priors consistently converge faster than using only one modality.

Table 4: Top mixed strategies under the *uncertainty sampling* acquisition function. The table reports the number of materials libraries in which the mixed strategy outperformed its corresponding single-strategy baseline with approximately equal initialization points.

Mixed Strategy	**No. of outperformed libraries**
Visual Modality (Cent_sat_high) + Farthest Point Sampling (FPS)	6
Farthest Point Sampling (FPS) + K-Means sampling	6
Visual Modality (Cent_sat_high) + K-Means Sampling	6
Visual Modality (Cent_sat_medium) + Farthest Point Sampling (FPS)	6
Latin Hypercube Sampling (LHS) + Outlier-based Sampling (ODAL)	6
K-Means sampling + Top5Similarity	5
Visual Modality (Cent_sat_low) + K-Center Sampling	5
K-Means sampling + Latin Hypercube Sampling (LHS)	5
Visual Modality (Cent_sat_medium) + Outlier-based Sampling (ODAL)	5
Visual Modality (Cent_sat_medium) + K-Center Sampling	5

Cent_sat_high = high saturation centroids; Cent_sat_medium = medium saturation centroids; Cent_sat_low = low saturation centroids.

4.7 Summary of Empirical Findings

Across eight heterogeneous composition-spread materials libraries, our empirical results demonstrate that the choice of cold-start initialization has a decisive impact on the efficiency of active learning–guided characterization. Even without combining modalities, sample-specific single-modality cold-start strategies already yield substantial benefits, reducing the number of required measurements by approximately 20–50% on average compared to uniform, fixed grid initialization (Thelen et al., 2023; Stricker et al., 2025).

Multimodal cold-start strategies consistently provide larger gains. By combining two complementary priors from numerical (composition-based), visual (photographic), and textual (literature-derived) modalities, the required number of measurements is frequently reduced by more than 70%, reaching up to 85% in favorable cases. These improvements are observed across materials libraries with varying compositional complexity, indicating that the benefits of multimodal initialization are not limited to a specific alloy system.

Importantly, comparisons using matched numbers of initial measurement points confirm that the observed improvements are not merely a consequence of increasing the size of the initial training set. Instead, multimodal strategies provide higher-quality cold-start information by guiding early sampling toward more informative and diverse regions of the composition space. This finding is further supported by systematic analyses across different base strategies and modality pairings.

Robustness analyses show that the performance trends remain consistent across random seeds and across both acquisition functions considered. While the Self-Adjusting Weighted Expected Improvement (SAWEI) acquisition function generally achieves faster convergence than uncertainty sampling, no single acquisition function is universally optimal. Rather, the effectiveness of the acquisition strategy clearly depends on the chosen cold-start initialization.

Taken together, these empirical findings demonstrate that informed, multimodal cold-start initialization substantially improves both data efficiency and robustness in experimental active learning pipelines for materials characterization. By exploiting cheap, readily available priors before any costly measurements are performed, the proposed strategies enable faster convergence and reduced experimental effort without modifying the core active learning loop.

5 Conclusion and outlook

We present a practical, experimentally validated cold-start strategy based on multimodal priors in AL frameworks to accelerate materials characterization of composition-spread materials libraries. By integrating diverse single-modality and multimodal initialization strategies we address one of the central challenges in AL: selecting informative starting points from *cheap* priors without labeled data, i.e. cold start. Our results across eight materials libraries indicate that informed initialization significantly and consistently accelerates surrogate model convergence. Single-modality approaches like minimum composition and Greedy k-Center sampling prove effective. Multimodal strategies such as the combination of Top 5 Similarity (textual) with visual clustering or composition-based methods further accelerate convergence.

Our proposed initialization techniques, when combined with a dynamic acquisition function such as SAWEI and a robust stopping criterion based on uncertainty flattening, achieve up to an 85% reduction in the number of measurements required for convergence, averaged across all eight materials libraries. In comparison, the fixed-grid baseline achieves an average reduction of approximately 73%, demonstrating that multimodal initialization substantially accelerates the characterization process without compromising accuracy w.r.t. experimental measurement uncertainty. Extensive validation based on experimental materials data suggests that the proposed approach is reliable, indicating potential for broader applicability.

In the future, we will deploy this framework in the loop with semi-automated experimental equipment. Further, we will extend this framework and use it for the AL-based characterization of other relevant material properties. Note, the proposed framework is directly deployable in existing high-throughput laboratories, as it relies exclusively on data modalities that are already part of standard experimental workflows. However, proprietary software and/or missing application programming interfaces make implementation unnecessarily complex or even impossible with existing hardware.

Our study contributes to the ongoing development of autonomous accelerated characterization, where existing cheap priors can be used to cold-start an AL loop with minimal human intervention and maximal efficiency. The code used in this study is available at: GitHub)

References

R. Barata, M. Leite, R. Pacheco, M. O. P. Sampaio, J. T. Ascensão, and P. Bizarro. Active learning for imbalanced data under cold start. In *Proceedings of the 2nd ACM International Conference on AI in Finance*, pp. 1–9, 2021.

Thomas A.A. Batchelor, Jack K. Pedersen, Simon H. Winther, Ivano E. Castelli, Karsten W. Jacobsen, and Jan Rossmeisl. High-entropy alloys as a discovery platform for electrocatalysis. *Joule*, 3(3):834–845, 2019. ISSN 2542-4351. doi: https://doi.org/10.1016/j.joule.2018.12.015. URL `https://www.sciencedirect.com/science/article/pii/S2542435118306214`.

J. Bejarano, K. Bose, T. Brannan, A. Thomas, K. Adragni, N. Neerchal, and G. Ostrouchov. Sampling within k-means algorithm to cluster large datasets. Technical report, Oak Ridge National Laboratory, Oak Ridge, TN, 2011.

C. Benjamins, E. Raponi, A. Jankovic, C. Doerr, and M. Lindauer. Self-adjusting weighted expected improvement for bayesian optimization, 2023. arXiv preprint arXiv:2306.04262.

Y. Eldar, M. Lindenbaum, M. Porat, and Y. Y. Zeevi. The farthest point strategy for progressive image sampling. *IEEE Transactions on Image Processing*, 6(9):1305–1315, 1997.

Trevor Hastings, James David Paramore, Brady Butler, and Raymundo Arroyave. Leveraging domain knowledge for optimal initialization in bayesian materials optimization. *Digital Discovery*, pp. –, 2025. doi: 10.1039/D5DD00361J. URL `http://dx.doi.org/10.1039/D5DD00361J`.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

A Gilad Kusne, Heshan Yu, Changming Wu, Huairuo Zhang, Jason Hattrick-Simpers, Brian DeCost, Suchismita Sarker, Corey Oses, Cormac Toher, Stefano Curtarolo, et al. On-the-fly closed-loop materials discovery via bayesian active learning. *Nature communications*, 11(1):5966, 2020.

David D Lewis and Jason Catlett. Heterogeneous uncertainty sampling for supervised learning. In *Machine learning proceedings 1994*, pp. 148–156. Elsevier, 1994.

A. Ludwig. Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *npj Computational Materials*, 5(1): 70, 2019.

Tobias Löffler, Alfred Ludwig, Jan Rossmeisl, and Wolfgang Schuhmann. What makes high-entropy alloys exceptional electrocatalysts? *Angewandte Chemie International Edition*, 60(52):26894–26903, 2021. doi: https://doi.org/10.1002/anie.202109212. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202109212.

D. Maltz and K. Ehrlich. Pointing the way: Active collaborative filtering. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 202–209. ACM Press/Addison-Wesley, 1995.

E. W. McFarland and W. H. Weinberg. Combinatorial approaches to materials discovery. *Trends in Biotechnology*, 17(3):107–115, 1999.

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

Aditya Nandy, Chenru Duan, and Heather J Kulik. Audacity of huge: overcoming challenges of data scarcity and data quality for machine learning in computational materials discovery. *Current Opinion in Chemical Engineering*, 36:100778, 2022.

E. Pasolli and F. Melgani. Gaussian process regression within an active learning scheme. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pp. 3574–3577, 2011.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006.

K. Ravi, V. Fediukov, F. Dietrich, T. Neckel, F. Buse, M. Bergmann, and H.-J. Bungartz. Multi-fidelity gaussian process surrogate modeling for regression problems in physics. *Machine Learning: Science and Technology*, 5(4):045015, 2024.

A. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 253–260, 2002.

Burr Settles. Active learning literature survey. 2009.

M. Shi, K. Hua, J. Ren, and Y. Cao. Global optimization of k-center clustering. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 19956–19966, 2022.

Boris N Slautin, Yongtao Liu, Hiroshi Funakubo, Rama K Vasudevan, Maxim Ziatdinov, and Sergei V Kalinin. Bayesian conavigation: Dynamic designing of the material digital twins via active learning. *ACS nano*, 18(36):24898–24908, 2024.

Markus Stricker, Lars Banko, Nik Sarazin, Niklas Siemer, Jan Janssen, Lei Zhang, Jörg Neugebauer, and Alfred Ludwig. Computationally accelerated experimental materials characterization–drawing inspiration from high-throughput simulation workflows. *npj Computational Materials*, 12(1):2, December 2025. ISSN 2057-3960. URL https://doi.org/10.1038/s41524-025-01919-5.

Katrine L. Svane and Jan Rossmeisl. Theoretical optimization of compositions of high-entropy oxides for the oxygen evolution reaction**. *Angewandte Chemie International Edition*, 61(19):e202201146, 2022. doi: https://doi.org/10.1002/anie.202201146. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202201146.

F. Thelen, L. Banko, R. Zehl, S. Baha, and A. Ludwig. Speeding up high-throughput characterization of materials libraries by active learning: autonomous electrical resistance measurements. *Digital Discovery*, 2(5):1612–1619, 2023.

Rui Xin, Edirisuriya MD Siriwardane, Yuqi Song, Yong Zhao, Steph-Yves Louis, Alireza Nasiri, and Jianjun Hu. Active-learning-based generative design for the discovery of wide-band-gap materials. *The Journal of Physical Chemistry C*, 125(29):16118–16128, 2021.

Dezhen Xue, Prasanna V Balachandran, John Hogden, James Theiler, Deqing Xue, and Turab Lookman. Accelerated search for materials with targeted properties by adaptive design. *Nature communications*, 7 (1):1–9, 2016.

W. Yuan, Y. Han, D. Guan, S. Lee, and Y.-K. Lee. Initial training data selection for active learning. In *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, pp. 1–7, 2011.

L. Zhang and M. Stricker. Matnexus: A comprehensive text mining and analysis suite for materials discovery, 2023. arXiv preprint arXiv:2311.06303.

Lei Zhang, Lars Banko, Wolfgang Schuhmann, Alfred Ludwig, and Markus Stricker. Composition-property extrapolation for compositionally complex solid solutions based on word embeddings. *Digital Discovery*, 4:1578–1590, 2025. doi: 10.1039/D5DD00169B. URL `http://dx.doi.org/10.1039/D5DD00169B`.

Yu Zhu, Jinghao Lin, Shibi He, Beidou Wang, Ziyu Guan, Haifeng Liu, and Deng Cai. Addressing the item cold-start problem by attribute-driven active learning. *IEEE Transactions on Knowledge and Data Engineering*, 32(4):631–644, 2019.