



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Driven Design and Optimization of Smart Materials for Advanced Chemical Engineering Applications

Yuao Zhao

University of Nottingham Ningbo China, Zhejiang, 315100. China. E-mail: ssyyz87@163.com

Abstract

Artificial intelligence (AI) is revolutionizing the design and optimization of smart materials for chemical engineering applications by fast predicting, designing and optimizing adaptive materials. Smart materials adapt intelligently to thermal, chemical, mechanical, electrical, magnetic or optical stimuli. These materials can be used in catalysis, separations, sensing, energy conversion and environmental remediation. Existing methods rely primarily on expensive and time-consuming trial-and-error experiments. Limited, imbalanced and noisy data and black box predictions limit the reliability, interpretability and sustainable decision making. Therefore, to tackle these issues, a Hybrid Interpretable Optimization Decomposition Neural Network (HIOD-NN) framework has been proposed for predicting, interpreting and multi-objective optimization of material properties. The Conditional Tabular Generative Adversarial Network (GAN) generates realistic synthetic samples the Physics-Guided Neural Network (PGNN) predicts material properties with scientific constraints and Local Interpretable Model-Agnostic Explanations (LIME) find influential composition and processing features. Particle Swarm Optimization (PSO) finds optimal synthesis parameters, whereas the Multi-Objective Evolutionary Algorithm Based on Decomposition (MOEA/D) balances performance, cost, energy consumption, toxicity, scalability and sustainability. Feedback from the experiments keeps improving the dataset and subsequent predictions and optimizations. The overall performance of HIOD-NN resulted in accuracy of 95.8%, R^2 of 0.96, RMSE of 0.061 and 2.18% experimental error. Thus, HIOD-NN explains clear reliable and sustainable way for fast-paced smart materials discovery and implementation in industry.

Keywords: Artificial Intelligence, Chemical Engineering, Machine Learning, Materials Design, Materials Optimization, Smart Materials

1. Introduction

Artificial intelligence (AI) is revolutionizing the field of materials research through efficient prediction, design and optimization of materials for certain functionalities. Different from existing trial-and-error technique, AI evaluates complex relationships between materials composition, structure, processing, properties thus saving time, costs and experiments. Machine Learning (ML) for Materials Design (MLMD) platform is one example of ML application to predict materials properties and designs regardless of high programming skills required [1]. The combination of active ML and Bayesian optimization has successfully found the optimal composition of magnesium alloys in a large and complex space of experimental designs, greatly cutting down the need for experimentation [2]. Smart materials are capable of exhibiting changes in the physical, chemical, thermal, mechanical, electrical or catalytic properties as a result of changes in temperature, pressure, Potential of Hydrogen (pH), light and electromagnetic fields. The properties enable applications in catalysis, adsorption, separation, sensing, energy conversion, environmental cleaning and process intensification. ML techniques can extract material features that influence properties [3] and graph neural networks learn atomic interactions in structures [4]. In the balancing predictive performance, computational efficiency and material design constraints is no easy task [5]. Diverse and robust data are needed to build an AI system that designs materials reliably. Incomplete data, noise, inconsistent units of measurement, imbalanced classes and the discrepancy between simulation and experiment can impair model performance. Domain adaptation uses in the prediction of the properties of materials in different domains [6]. While data-efficient learning and enhanced sampling make catalytic simulations and experiments less costly [7]. AI systems like ChatMOF have used in predicting and designing metal organic frameworks [8]. Materials optimization needs efficient exploration of large compositional and processing space. The exploration of stable multi-component alloys has been faster due to Bayesian optimization [9].

The thermoelectric properties of semiconductor thin films have been optimized through ML [10]. ML models have also been useful in research high entropy alloy catalysts economically [11]. Conditional generative models aid the design of polymer-electrolytes [12]. Human-in-the-loop generative models are used to discover novel ternary materials [13]. Efficient Bayesian optimization has investigated the design space of functionalized

nanoporous materials [14]. Active learning uses select experiments on materials acceleration platforms [15]. As well as inverse design of functional materials [16]. Symbolic regression uses identify material descriptors and discover acid stable oxides for electrocatalysis [17]. Data-efficient graph learning uses build accurate interatomic potentials [18]. Hybrid Transformer-Graph models learn higher-order atomic interactions to improve property prediction [19]. However, there is a need for more interpretability and optimization of performance, cost, energy, toxicity, scalability and sustainability. Data-driven models capture complex material relationships, but could produce scientifically implausible predictions in the absence of sufficient data. To overcome this limitation, the proposed Hybrid Interpretable Optimization Decomposition Neural Network (HIOD-NN) framework combines multisource data pre-processing, a Conditional Tabular Generative Adversarial Network (CTGAN) for data augmentation, an Explainable Autoencoder for feature extraction and a Decomposed Feedforward Neural Network, Physics-Guided Neural Networks (PGNNs) to predict material properties and Local Interpretable Model-Agnostic Explanations (LIME) for analyzing features importance. Particle Swarm Optimization (PSO) performs optimization of materials compositions and process parameters and MOEA/D multi-objective optimization algorithm is used to obtain Pareto-optimal solutions considering performance, cost, energy consumption, toxicity, scalability and sustainability criteria. Selected materials are tested experimentally and obtained measurements from the lab and sensors are fed back into the framework to continue improving the predictive models.

- A systematic research of AI approaches for the design and optimization of smart materials for chemical engineering applications is provided.
- A unified HIOD-NN approach combining CTGAN data augmentation, PGNN property prediction and LIME explanations is proposed.
- Powerful PSO and MOEA/D algorithms are used to optimize materials and manufacturing process parameters under performance, cost, energy, toxicity, scalability and sustainability criteria.
- Experimental validation and feedback loop mechanism is added for comparison of predictions with actual material properties.

The structure of this research is organized as follows. The background literature is mentioned in Section 2. Section 3 describes the proposed HIOD-NN model. Section 4 presents the results and validation of proposed. Section 5 addresses the challenges and future work and Section 6 concludes the research.

2. Background Study

Rahman et al. (2025) [20] designed Supervised Pretraining for Material Property Prediction (SPMat) to mitigate the reliance on a large number of labelled materials datasets. The supervised contrastive pretraining and graph augmentation enhanced six property-prediction tasks despite the need for surrogate labels and high training costs.

Wang et al. (2025) [21] focused on inefficiencies of multi-type hydrogen evolution catalysts screening through the application of feature engineering and Extremely Randomized Trees.

Pisal, Krejčí & Rinke (2025) [22] used adsorption energy distribution and ML to screen for catalysts in Carbon dioxide (CO₂) methanol production. This technique found a number of potential catalysts however it could be used for only some metal-based catalysts.

Nishimura et al. (2025) [23] used ML for the prediction of photocurrent for multielement-doped hematite photoelectrodes. This approach used identify significant dopants and effective material compositions however, its generalizability was constrained due to the lack of experimental data.

Yu, Chng and Sholl (2025) [24] combined ML with the Maxwell model for screening Metal Organic Framework (MOF) for mixed-matrix membranes. Although this approach enhanced the candidate identification process, simplified assumptions cannot effectively capture interfacial defects and transport in real conditions.

Ignacz et al. (2025) [25] used ML together with experimental data to predict thin-film composite membranes. Such an approach allowed the researchers to identify important polymerization parameters, which were limited by the narrow scope of the training data set.

Liu, Guo and Su (2025) [26] proposed a convolutional neural network for predicting the elasticity of porous materials. This model yielded and operated approximately 2,400 times faster than Finite Element Method (FEA) however, it took into account only two-dimensional structures.

Ma and Zhao (2025) [27] used machine-learning potentials in atomistic simulations to study heat transfer in high entropy alloys. It was found out that the chemical short-range order could regulate thermal conductivity, but the validation was possible only in simulation alloy systems.

Liu et al. (2025) [28] proposed domain knowledge aided anomaly detection techniques for enhancing data quality of materials. The method enhanced anomaly detection and material-property prediction however, its effectiveness is contingent upon the availability of complete expert-based rules.

Nitturi et al. (2026) [29] used physics-based modelling, time-series prediction and inverse generative learning to develop battery electrode microstructure designs. The technique supported performance driven generation, but the design needed further experimental validation.

Table 1: Summary of AI-Based Methods for Smart-Material Discovery, Property Prediction and Optimization

Author (Year)	Concept	Method	Limitation	Result
Alverson et al. (2024) [30]	Crystal generation	GAN and diffusion models	Requires physical validation	Diffusion models produced more realistic crystals
Masui et al. (2025) [31]	Experimental optimization	Maximum Partial Dependence Effect (MPDE)-based Bayesian optimization	Threshold-dependent	Reduced optimization trials
Liu et al. (2025) [32]	Property prediction	Self-Adaptable Graph Attention Networks with Symbolic Regression (SA-GAT-SR)	High feature complexity	Achieved $R^2 \geq 0.93$
Allec and Ziatdinov (2025) [33]	Uncertainty-aware learning	Partially Bayesian neural network	Prior-dependent	Reduced computational cost
Arendt et al. (2026) [34]	Sustainable glass discovery	MD, ML and robotic synthesis	Limited glass systems	Achieved R^2 up to 0.98

Table 1 illustrates some of the latest developments in AI that can be used to discover new materials and predict the properties. Generative models, Bayesian optimization, graph networks, Bayesian learning and autonomous experimentation use in enhancing efficiency and accuracy. Nevertheless, experimental verification, computation cost, parameter choice and model generalizability are significant drawbacks.

3. Proposed Methodology

The proposed HIOD-NN is a unified approach towards the development of smart materials in the field of chemical engineering which includes multisource data pre-processing, generation of synthetic data, physics informed prediction, explainability, parameter tuning, multi-objective decision making and experiments.

3.1 Datasets description

The benchmark Data of materials are obtained from various sources, including the Materials Research, Joint Automated Repository for Various Integrated Simulations–Density Functional Theory dataset (JARVIS-DFT), Open Quantum Materials Database (OQMD), Novel Materials Discovery (NOMAD) Repository, Catalysis-Hub and CoRE MOF. Such data include material compositions, crystal structure, processing conditions, adsorption energy, as well as properties such as thermal, mechanical, catalytic, electronic and transport. The collected data are then preprocessed and partitioned into training, validation and test sets.

3.2 Hybrid Interpretable Optimization Decomposition Neural Network (HIOD-NN)

The Proposed HIOD-NN is an integrated AI-driven system for the design and optimization of smart materials with high performance and sustainability of applications in chemical engineering. At first, material components, structure, manufacturing process details, physical characteristics, cost, toxicity and environmental impact information is gathered through experimental work, simulations, databases and real-time sensors. CTGAN synthesizes the data to solve the problems of insufficient data and imbalance between different data points, whereas PGNN uses the available data and physical and chemical laws to provide precise and physically meaningful material property predictions. LIME explains the prediction made by the system by revealing the effect of features including composition, temperature, pressure, particle size and porosity on the output of PGNN. After that, PSO navigates the complicated design space to find the optimum material compositions and synthesis process parameters. MOEA/D solves the Multi-Objective optimization problem of balancing the performance, cost, energy consumption, toxicity, scalability and environmental sustainability. In applications involving real-time data, the newly generated data from experiments or sensors are constantly fed back into the system to update the prediction and optimization process.

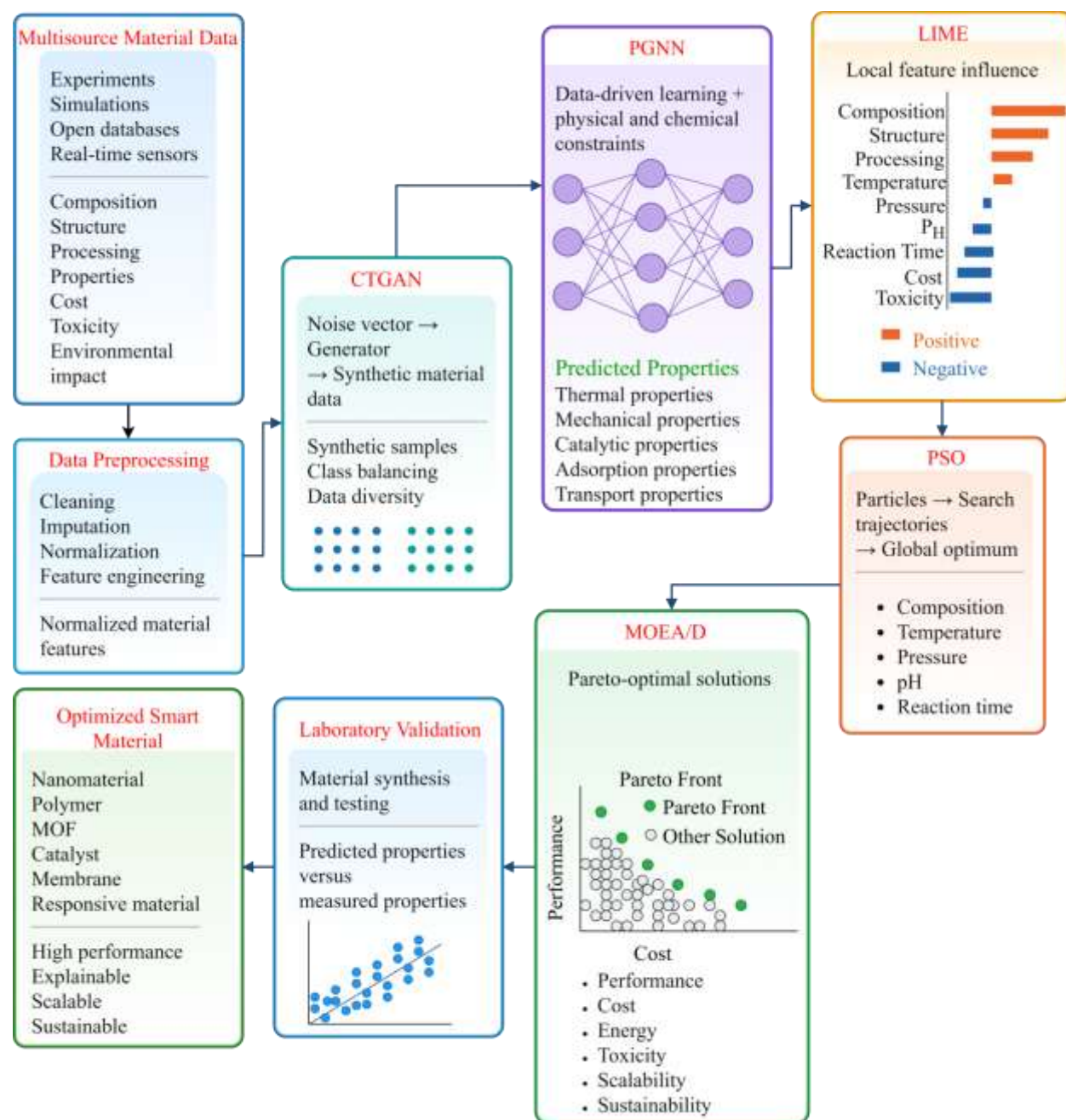


Figure 1: Systematic Architecture of the Proposed HIOD-NN Framework for Smart-Material Optimization

The figure 1 consists of multisource material data, preprocessing, CTGAN data augmentation, PGNN property prediction and LIME explainability techniques. The PSO model determines the optimal composition and fabrication conditions, whereas the MOEA/D model creates Pareto-optimal material solutions by compromising performance, cost, energy consumption, toxicity, scalability and sustainability. Candidate materials are synthesized experimentally and verified via comparisons of the predicted and experimental properties.

3.2.1 CTGAN-Based Material Data Augmentation

In CTGAN can be used to overcome the problems of scarcity and imbalance in smart material data. CTGAN learns about the relationships between materials and the performance. After that, it creates real-looking data points that have both continuous and categorical attributes.

$$X = [X_{str}, X_{chem}, X_{th}, X_{mech}, X_{cat}, X_{proc}] \quad (1)$$

Equation (1) describes the multimodal material feature representation, where diverse sets of smart material data are represented in one input vector. These X is the material representation as a whole, while X_{str} , X_{chem} , X_{th} , X_{mech} , X_{cat} and X_{proc} are the structural, chemical, thermal, mechanical, catalytic and processing properties, respectively.

$$x'_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \quad (2)$$

Equation (2) describes feature normalization for bringing all material descriptors to the same scale. These, x_{ij} is the j feature of i material, while $x_j^{\min}$ and $x_j^{\max}$ represent the minimum and maximum values of that feature. The feature normalization provides numerical stability and excludes domination of features with large magnitude during the training.

$$X_w = W \odot X \quad (3)$$

Equation (3) presents the adaptive feature weighting for assigning different importance scores to the material descriptors. These, X is the feature matrix without any transformation, while W are the learned feature weights.

3.2.2 PGNN-Based Material Property Prediction

The PGNN integrates information on the materials with well-known physics and chemistry laws. Properties predicted include the materials strength, conductivity, thermal stability, adsorption ability and catalytic performance. Physical limitations ensure that no unphysical predictions are made. Composition, structure, property relations are thus established.

$$H^{(l)} = \sigma(W^{(l)}H^{(l-1)} + b^{(l)}) \quad (4)$$

Equation (4) presents the process of transformation of deep features in a neural network. $H^{(l-1)}$ is the previous layers feature, $W^{(l)}$ is the learnable weight and $b^{(l)}$ is the bias. The activation function σ converts the data, using to learn the hierarchical nature of material features.

$$\sigma(z) = \max(0, z) \quad (5)$$

Equation (5) the definition of the Rectified Linear Unit (ReLU) activation function for adding nonlinearity to the AI model. Z is the weighted input to the neuron, these is retained in case of positive value and turned to be zero for negative value. The network learns complex nonlinear relations between material composition, structure, processing and performance.

$$\hat{Y} = f_{\theta}(X) \quad (6)$$

Equation (6) describes the AI prediction of material properties based on the input features. X is the material feature vector, f_{θ} is the trained AI model and θ is the learnable parameters. The output $\hat{Y}$ is the property predictions such as strength, conductivity, stability and catalytic performance.

$$\theta^* = \operatorname{argmin}_{\theta} \mathcal{L}(\theta) \quad (7)$$

Equation (7) provides an expression for finding the optimum parameters of the AI model by minimizing its loss function. In the formula above, θ refers to the learnable model parameters, $\mathcal{L}(\theta)$ stands for the loss during training, while θ^* refers to the set of optimum parameters. The optimization allows learning material-property relationship accurately.

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}(\theta_t) \quad (8)$$

Equation (8) gives an expression for the update of the parameters during training using gradient descent. In the formula above, θ_t refers to the parameters at iteration t , η to the learning rate and $\nabla_{\theta} \mathcal{L}$ to the loss gradient.

$$\mathcal{L}_{total} = \mathcal{L}_{data} + \lambda \mathcal{L}_{physics} \quad (9)$$

Equation (9) combines experimental data fitting with physics based understanding in AI training. Here, $\mathcal{L}_{data}$ is the prediction loss while $\lambda \mathcal{L}_{physics}$ represents physics constraint loss and λ is a coefficient representing the weighting of physics-based knowledge.

$$\mathcal{L}_{physics} = \frac{1}{N} \sum_i 1N |\mathcal{R}(\hat{Y}_i, X_i)|^2 \quad (10)$$

Equation (10) evaluates the deviations from known physical or chemical rules for AI predictions. These, R is the physical residual, $\hat{Y}_i$ is the predicted property of the material and X_i is the input of the material. Loss minimization uses derive physically and scientifically consistent predictions.

3.2.3 LIME-Based Interpretable Material Analysis

The technique for explaining LIME is used to explain the predictions made by the PGNN. The process uses factors such as composition, temperature, pressure, porosity and particle size that influence the performance of materials. This technique improves clarity in understanding the models decision process.

$$I_j = \frac{1}{N} \sum_i 1N |\phi_j^{(i)}| \quad (11)$$

Equation (11) shows how much weight each material property holds in AI-based predictions. Here, I_j refers to the importance of feature j , N refers to the total number of samples and $\phi_j^{(i)}$ is the contribution made by the j feature in sample i . A high value of I_j means more important features in terms of predicting material properties.

3.2.4 PSO-Based Synthesis Parameter Optimization

The PSO searches for suitable material compositions and synthesis conditions within a complex design space. It evaluates variables such as precursor concentration, temperature, pressure, reaction time and additive ratio. PSO efficiently identifies high-performing candidate configurations. This reduces conventional trial-and-error experiments, cost and development time.

$$z^* = \operatorname{argmin}_{z \in \Omega} J(f_\theta(X, z)) \quad (12)$$

Equation (12) is a formula for the optimal synthesis conditions needed to obtain the desired material performance characteristics. In this case, z stands for controllable process variables like temperature, pressure, concentration and time of reaction. Ω Determines possible conditions while J is the optimization criterion. Consequently, this z^* is the optimal synthesis setup.

3.2.5 MOEA/D-Based Sustainable Multi-Objective Optimization

The MOEA/D subdivides a large-scale optimization problem into smaller problems. The algorithm simultaneously evaluates the performance of the material, cost, toxicity, energy consumption, scalability and environmental impact. The algorithm comes up with several Pareto-optimal material solutions. This allows the engineers to come up with a balanced smart-material solution.

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_i i = 1N(Y_i - \hat{Y}_i)^2 \quad (13)$$

Equation (13) gives the Mean Squared Error (MSE) for experimental and calculated material properties. These, N are stands for the number of samples Y_i stands for the experimental value $\hat{Y}_i$ stands for the predicted value. The reduction in MSE leads to reduced prediction errors and enhanced learning performance of the AI model.

$$\mathcal{L}_{MAE} = \frac{1}{N} \sum_i i = 1N|Y_i - \hat{Y}_i| \quad (14)$$

Equation (14) gives the Mean Absolute Error (MAE) which gives the average of the absolute values of the error in prediction of experimental and calculated material properties. Y_i Stands for the measured value and $\hat{Y}_i$ stands for the predicted value. MAE is easy to interpret and less affected by large values of error unlike MSE .

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (15)$$

Equation (15) gives the determination coefficient R^2 , which gives the efficiency of the AI model in explaining variations in material properties. Y_i Stands for the measured value, $\hat{Y}_i$ stands for this predicted value $\bar{Y}$ stands for the mean measured value.

$$RMSE = \sqrt{\frac{1}{N} \sum_i i = 1N(Y_i - \hat{Y}_i)^2} \quad (16)$$

Equation (16) estimates the Root Mean Squared Error ($RMSE$), which quantifies the difference between predictions and the actual material property in its original units. In the formula above, N refers to the number of data points, while Y_i and $\hat{Y}_i$ refer to the observed and estimated material properties respectively. The smaller $RMSE$ proposed the predicted material properties to be close to the experimental observations.

$$J(z) = \sum_k k = 1Kw_k \quad (17)$$

Equation (17) is a weighted optimization criterion that integrates various material design criteria into one equation. In this case, K stands for the number of objectives, $J(z)$ stands for objective k and w_k for the weight of objective k . Objectives may include strength, conductivity, thermal stability, durability, costs and environmental performance.

$$z^* \in \mathcal{P} = \{z: \nexists z' \text{ such that } F_k(z') \geq F_k(z), \forall k\} \quad (18)$$

Equation (18) provides a definition of Pareto-optimal material solutions that can only achieve improvement in one objective by worsening the other objectives. Here, z represents an arbitrary material solution or synthesis conditions, z' is another solution and F_k represents objective k . The set $\mathcal{P}$ is comprised of balanced optimal solutions.

$$E_{val} = \frac{1}{N} \sum_i i = 1N \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \quad (19)$$

Equation (19) computes the percent difference between the experimentally measured values and the AI-estimated material property values during validation. In the equation, E_{val} is the validation error, Y_i is the experimental value and $\hat{Y}_i$ is the predicted value. Validation error is low when there is high similarity between the AI predictions and the experimentally obtained material properties.

$$S_{overall} = \sum_k k = 1Kw_k \left(\frac{P_k - P_k^{min}}{P_k^{max} - P_k^{min}} \right) \quad (20)$$

Equation (20) generates an overall score for material performance based on normalized material properties. In this case, P_k is the value attained for property k , $P_k^{\min}$ and $P_k^{\max}$ is the minima and maxima of that property and w_k is the weight of important

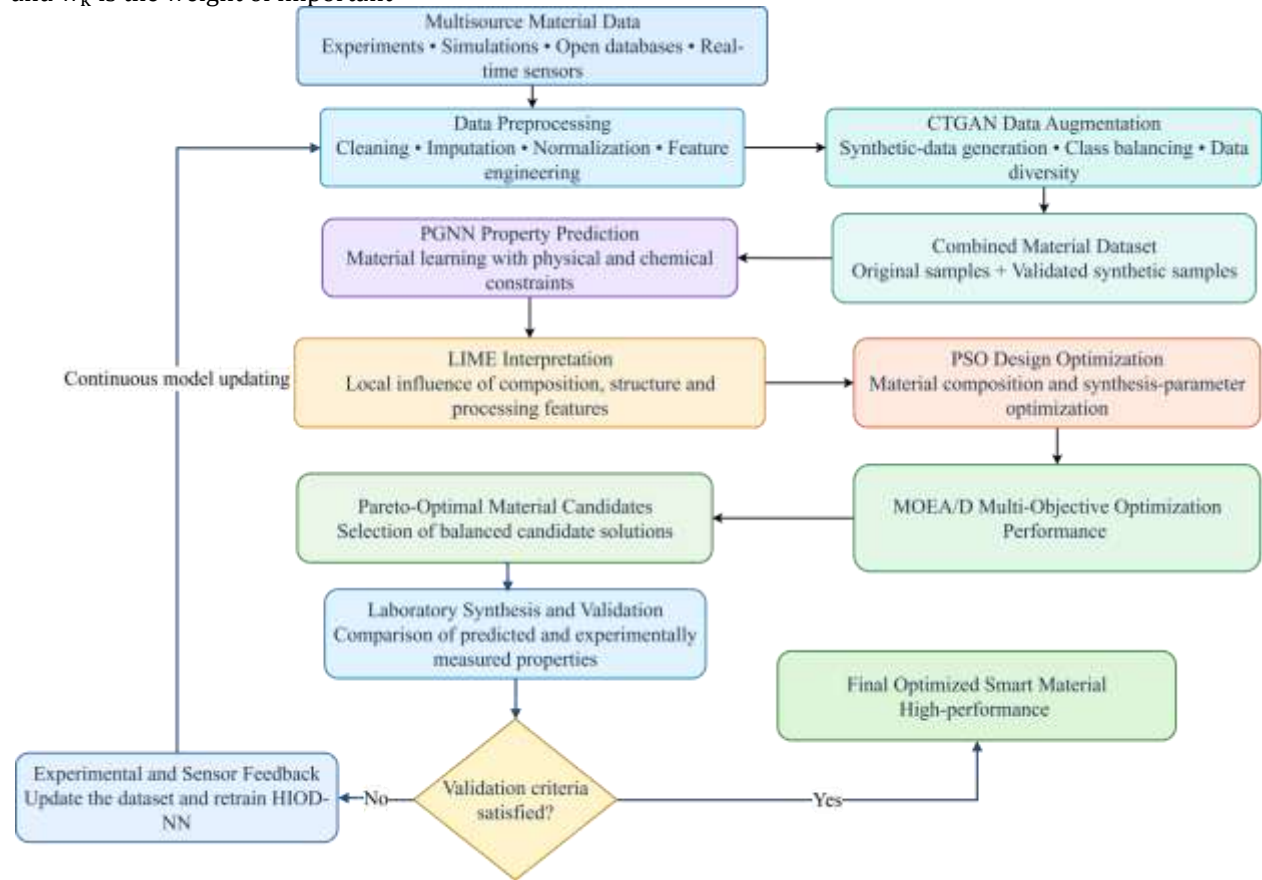


Figure 2: Workflow of the HIOD-NN

This figure 2 represents the multi-source material data, pre-processing, CTGAN enhancement, property prediction using PGNNs and explanation using LIME. PSO and MOEA/D then optimize the material composition, manufacturing process parameters, performance, cost, toxicity, energy consumption, scalability and sustainability. The selected Pareto-optimal solutions will be subjected to lab synthesis and validation. If this meet the validation criteria, this can become the final optimized material otherwise, these are fed back to retrain the models.

Algorithm: Hybrid Interpretable Optimization and Decomposition Neural Network (HIOD-NN)	
Input:	
$D \leftarrow$ Multisource smart-material dataset	
$P \leftarrow$ Physical and chemical constraints	
$O \leftarrow$ Optimization objectives:	
performance, cost, energy, toxicity, scalability, sustainability	
Output:	
$Y \leftarrow$ Predicted material properties	
$X_{best} \leftarrow$ Optimized synthesis parameters	
$S^* \leftarrow$ Pareto-optimal smart-material solutions	
Begin	
1. Collect material composition, structure, processing, property, cost, toxicity and environmental-impact data.	
2. Preprocess dataset D :	
a. Remove duplicate and inconsistent records	
b. Handle missing values and noise	
c. Normalize numerical features	

- d. Encode categorical variables
 - e. Split data into training, validation and testing sets
 3. Apply CTGAN:
 - a. Train CTGAN using the training dataset
 - b. Generate realistic synthetic material samples
 - c. Combine original and synthetic samples to obtain Daug
 4. Train PGNN using D aug and constraints P :
 - a. Predict material properties Y
 - b. Calculate prediction loss
 - c. Calculate physics-constraint loss
 - d. Update model parameters until convergence
 5. Apply LIME to PGNN predictions:
 - a. Calculate the contribution of each material feature
 - b. Rank influential composition and processing parameters
 6. Apply PSO:
 - a. Initialize candidate material compositions and synthesis conditions
 - b. Evaluate each candidate using the trained PGNN
 - c. Update particle velocity and position
 - d. Return the best synthesis parameters X best
 7. Apply MOEA/D:
 - a. Decompose the multi-objective problem into subproblems
 - b. Optimize performance, cost, energy, toxicity, Scalability and sustainability
 - c. Generate Pareto-optimal solutions S^*
 8. Experimentally synthesize and test the selected materials.
 9. Compare predicted and experimental properties using Accuracy, RMSE, R^2 and experimental deviation.
 10. Perform descriptive statistics, hypothesis testing, 95% confidence interval analysis and ANOVA.
 11. Add experimental results to D and update the framework.
 12. Return Y , X best and S^* .
- End

The pseudo-code explained the entire process of the proposed HIOD-NN system. The process starts with the collection and preprocessing of multisource smart-material data. CTGAN synthesizes data in order to enhance the quality and balance and PGNN predicts material properties through the use of physical and chemical constraints. LIME analyzes the significance of individual material and process parameters. PSO discovers the best compositions and manufacturing conditions, whereas MOEA/D creates Pareto-optimal sustainable solutions by balancing performance, cost, energy usage, toxicity, scalability and environmental effects.

4. Experimental result

This section describes the performance of the HIOD-NN approach in the context of predicting, explaining, optimizing and validating smart materials experimentally. The performance evaluation includes data augmentation with CTGAN, prediction performance of PGNN, feature contribution explanation with LIME, synthesis parameter optimization with PSO and sustainability-oriented material solutions obtained by MOEA/D. The dataset is split into 70% for training, 15% for validation and 15% for testing. In the training phase, the training set is used to train the model, the validation set is employed for hyperparameters tuning and model selection, while the testing set evaluates the generalization performance of the trained model independently. The approach is implemented using Python 3.10, which employs TensorFlow, Keras, Scikit-learn, NumPy, Pandas, Matplotlib and LIME libraries. Evaluation metrics include accuracy, RMSE, coefficient of determination (R^2), experimental deviation, descriptive statistics, hypothesis testing, confidence intervals and ANOVA.

Table 2: Smart-Material Optimization Results Using HIOD-NN

Material category	Evaluated property	Initial value	Optimized value	Unit
Nanomaterials	Thermal conductivity	48.6	61.8	$W \cdot m^{-1} \cdot K^{-1}$

Functional polymers	Tensile strength	52.4	67.9	MPa
Metal–organic frameworks	Adsorption capacity	318.5	426.7	mg·g ⁻¹
Catalysts	Catalytic conversion	81.3	94.7	%
Smart membranes	Permeability	186.4	243.8	LMH
Responsive materials	Response efficiency	78.6	92.1	%

Table 2 indicates that the performance of all smart materials classes was improved after optimization. Metal-organic frameworks turned out to have the biggest improvement in performance among all smart materials classes because the adsorption ability was raised from 318.5 to 426.7 mg⁻¹ and the permeability of smart membranes was raised from 186.4 to 243.8 LMH. The thermal conductivity and tensile strength were optimized to over 92% for catalysts and responsive materials.

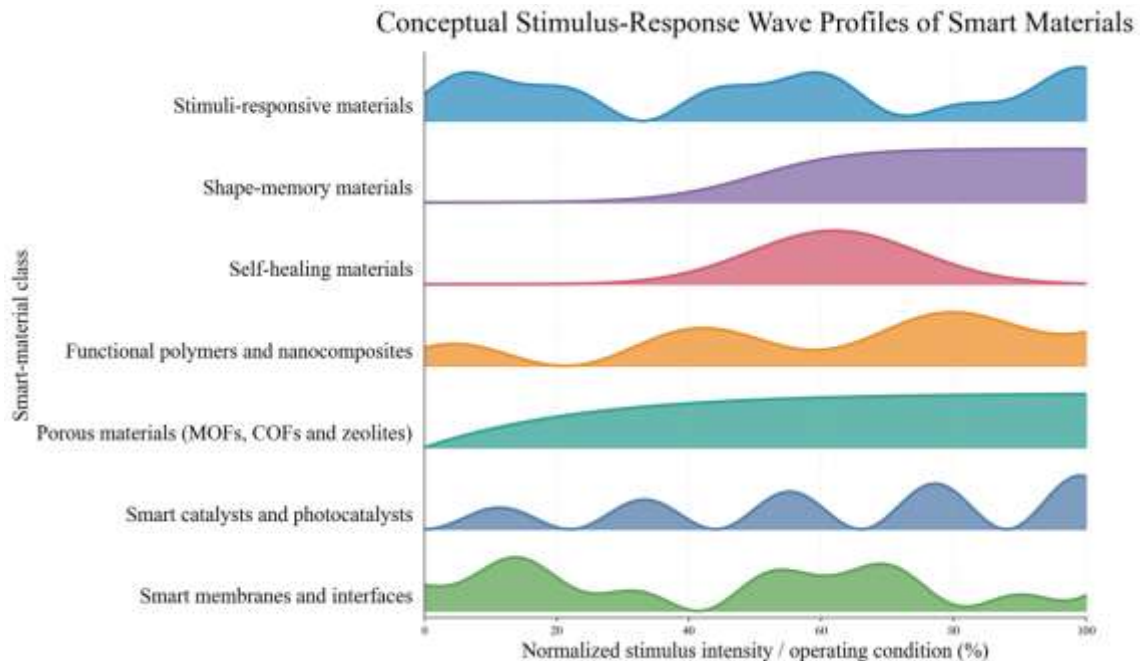


Figure 3: Response Curves of Smart Material Types Due to Different Stimuli

Figure 3 shows the response curve of seven smart material types due to different stimuli or operating conditions. Shape memory and porous materials show gradual response curves, whereas the maximum response of self-healing materials is obtained in a specific range of operating conditions. Stimuli-responsive materials, catalysts, polymers and membranes exhibit periodic behavior.

Table 3: Data Quality Evaluation Before and After CTGAN-Augmented Data

Data-quality indicator	Before CTGAN	After CTGAN	Outcome
Number of material samples	2,500	7,500	Dataset expanded
Missing values (%)	8.4	1.2	Data completeness improved
Minority-class samples	320	1,850	Class imbalance reduced
Class-balance ratio	0.38	0.91	Balanced distribution achieved
Feature coverage (%)	76.5	94.8	Data diversity increased
Duplicate records (%)	4.7	0.9	Redundancy reduced
Data-quality score	0.72	0.94	Overall quality improved

The table 3 shows the way in which CTGAN has enhanced the material dataset from 2,500 samples to 7,500 samples through the increase of the number of samples in the minority class from 320 to 1,850. The missing data and duplicate rates have been brought down from 8.4% to 1.2% and from 4.7% to 0.9%, respectively. This has led to an enhancement of the class balance ratio and feature coverage from 0.91 and 94.8%. The overall data quality has thus been improved from 0.72 to 0.94 after augmentation.

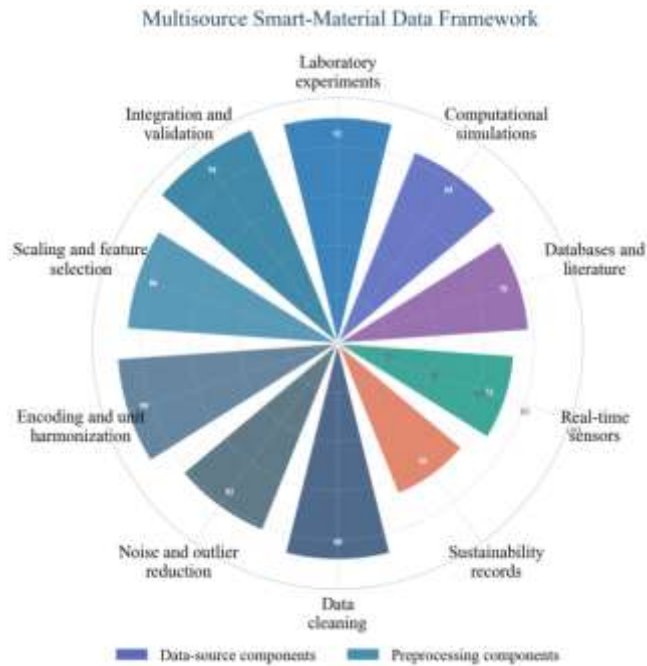


Figure 4: Analysis of the Multisource Smart-Material Data Framework

This figure 4 shows the normalized contributions of five different sources of data and five different methods of pre-processing used in the proposed system. The laboratory experiments have shown the greatest contribution by any data source (92), whereas the integration and validation have the greatest contribution by any pre-processing method (94). Nevertheless, the sustainability shows the smallest contribution (66) and thus reveals the lack of data in this category.

Table 4: LIME-Based Interpretation of Material and Processing Features

Material/processing feature	Input value	LIME contribution	Effect on performance	Interpretation
Surface area	$685 \pm 12 \text{ m}^2/\text{g}$	+0.24	Positive	Improves active-site availability
Porosity	$72 \pm 2\%$	+0.19	Positive	Enhances adsorption and mass transport
Synthesis temperature	$180 \pm 5^\circ\text{C}$	+0.15	Positive	Supports stable material formation
Additive concentration	$4.5 \pm 0.3 \text{ wt.}\%$	+0.12	Positive	Improves functional performance
Particle size	48 nm	+0.09	Positive	Increases surface reactivity
Processing time	$8 \pm 0.5 \text{ h}$	-0.06	Negative	Longer processing increases energy use
Material toxicity	0.18 ± 0.02	-0.11	Negative	Reduces sustainability and safety
Production cost	$\text{₹}1,250 \pm \text{₹}75/\text{kg}$	-0.08	Negative	Limits large-scale implementation

Table 4 indicates the LIME contribution of material properties for predicting the performance. Surface area is the greatest contributor (+0.24), followed by porosity (+0.19) and synthesis temperature (+0.15). This is due to the significance of these properties in terms of active sites, mass transport and structural durability. Toxicity of the material (-0.11) is the most negative property, while production cost (-0.08) and processing time (-0.06) negatively affects the performance of materials.

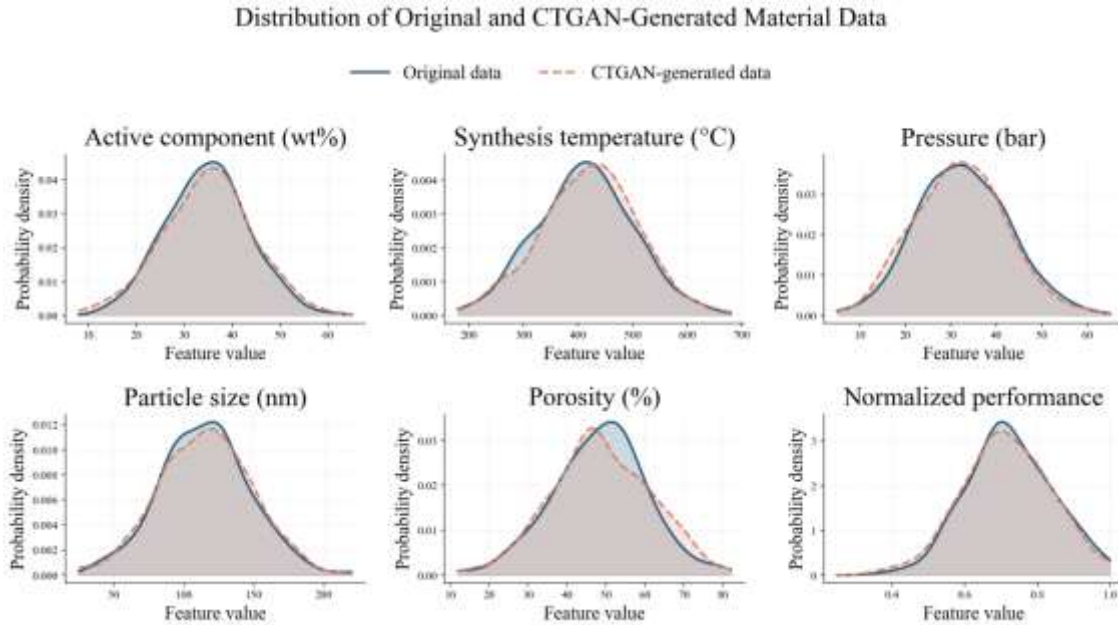


Figure 5: Distributional Comparison of Original and CTGAN-Generated Smart-Material Data

Figure 5 this graph demonstrates the comparative research of the probability density function of both the original data set and that of CTGAN generated data set with respect to the six important parameters of the smart material. Overlap in the above graph shows that the CTGAN model has been successful in simulating the statistics of the original data set. Minor variation in parameters like temperature, pressure and porosity shows the diversity of synthetic data set without any distortion in its distribution.

Table 5: Optimization of Material Composition and Synthesis Parameters

Optimization parameter	Initial value	PSO-optimized value	Permissible range	Influence
Active material content	60 ± 2 wt. %	72 ± 1.5 wt. %	50–80 wt. %	Improves functional performance
Additive concentration	3.0 ± 0.3 wt. %	4.6 ± 0.2 wt. %	1–6 wt. %	Enhances material stability
Synthesis temperature	$150 \pm 5^\circ\text{C}$	$182 \pm 3^\circ\text{C}$	120–220°C	Improves structural formation
Operating pressure	2.5 ± 0.2 bar	3.2 ± 0.1 bar	1–5 bar	Enhances reaction efficiency
Reaction time	10 ± 0.5 h	7.8 ± 0.3 h	5–12 h	Reduces processing time
Solution pH	6.5 ± 0.2	7.2 ± 0.1	5–9	Improves chemical stability
Particle size	65 ± 4 nm	48 ± 3 nm	30–100 nm	Increases active surface area

The table 5 presents the optimal composition and synthesis parameters by using PSO in the range of possible values. The amount of active substance, the amount of an additive, temperature, pressure and pH are increased to improve the performance, stability, structural formation and effectiveness of reactions. Reaction time is reduced from 10 to 7.8 hours to achieve effective production. Particle size is reduced from 65 to 48 nm.

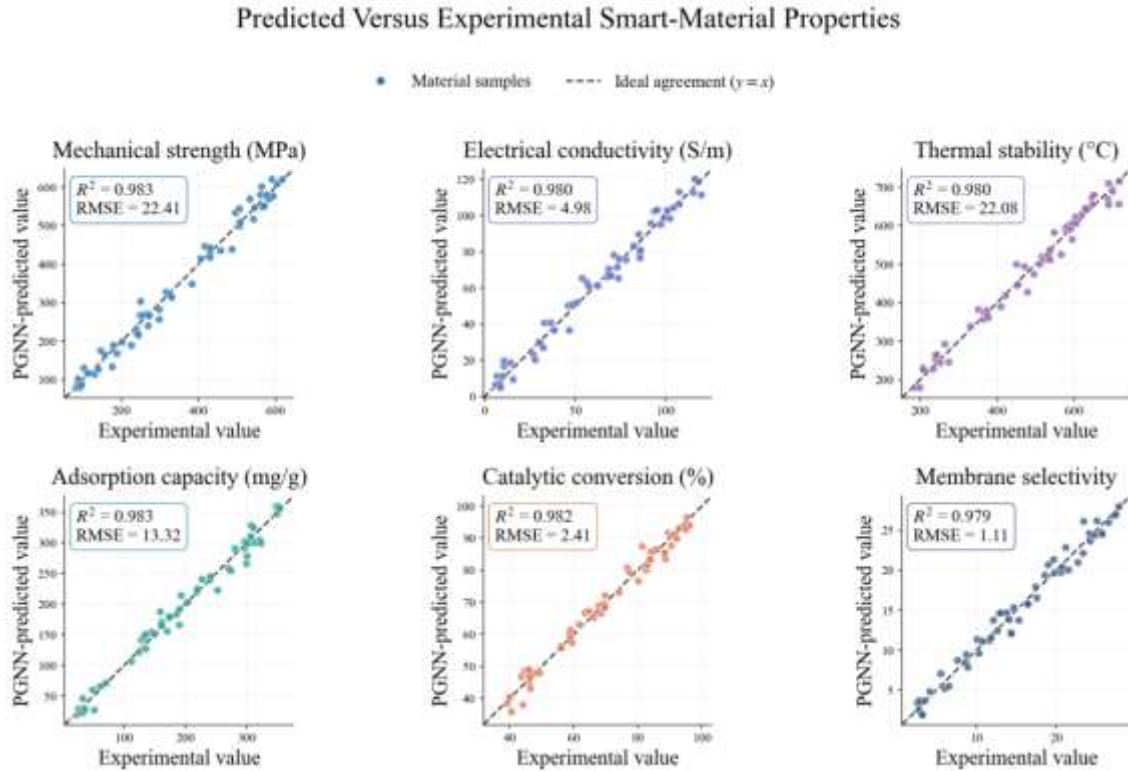


Figure 6: Predicted Versus Experimental Smart-Material Properties

This figure 6 represents the comparison between the predicted values obtained from the PGNN model and the experimental values of six different smart materials' properties. The values are close to the perfect correlation line and it have very high R^2 values of 0.979–0.983 and low RMSE values.

Table 6: Laboratory Validation of HIOD-NN-Predicted Smart Materials

Smart-material category	Evaluated property	HIOD-NN prediction	Laboratory result	Relative deviation (%)
Nanomaterials	Thermal conductivity ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	61.8	60.9 ± 1.4	1.46
Functional polymers	Tensile strength (MPa)	67.9	66.8 ± 1.7	1.62
Metal-organic frameworks	Adsorption capacity ($\text{mg}\cdot\text{g}^{-1}$)	426.7	419.5 ± 6.8	1.69
Catalytic materials	Catalytic conversion (%)	94.7	92.9 ± 1.2	1.90
Smart membranes	Permeability (LMH)	243.8	238.6 ± 4.3	2.13
Responsive materials	Response efficiency (%)	92.1	90.4 ± 1.5	1.85

The Table 6 demonstrates the comparison of HIOD-NN prediction results with the experimental data of six classes of smart materials. Relative deviations are within the range of 1.46%–2.13%, meaning that there is a good match between the theoretical prediction and experimental data. Nanomaterials have the lowest deviation of 1.46%, whereas smart membranes have the highest deviation of 2.13% indicating consistency.

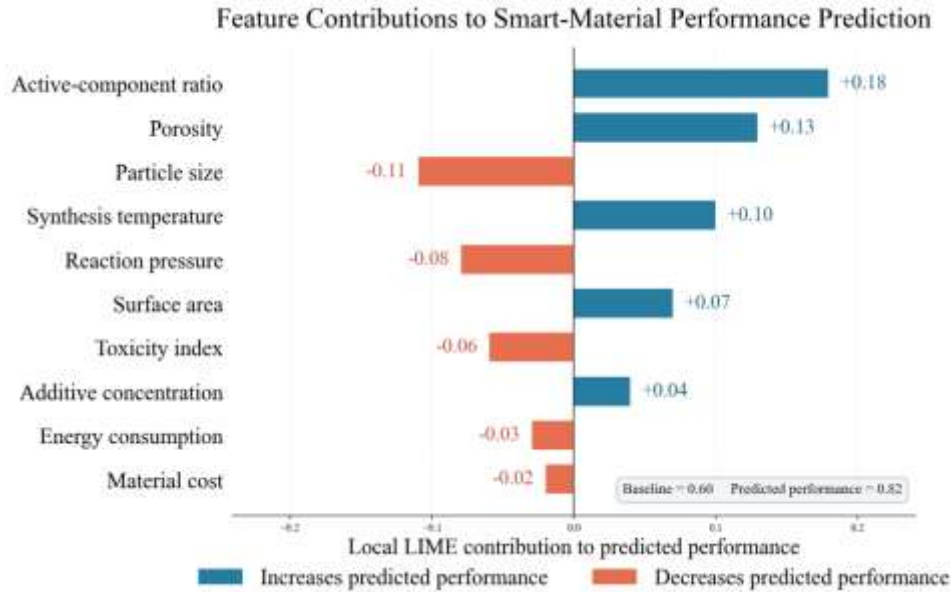


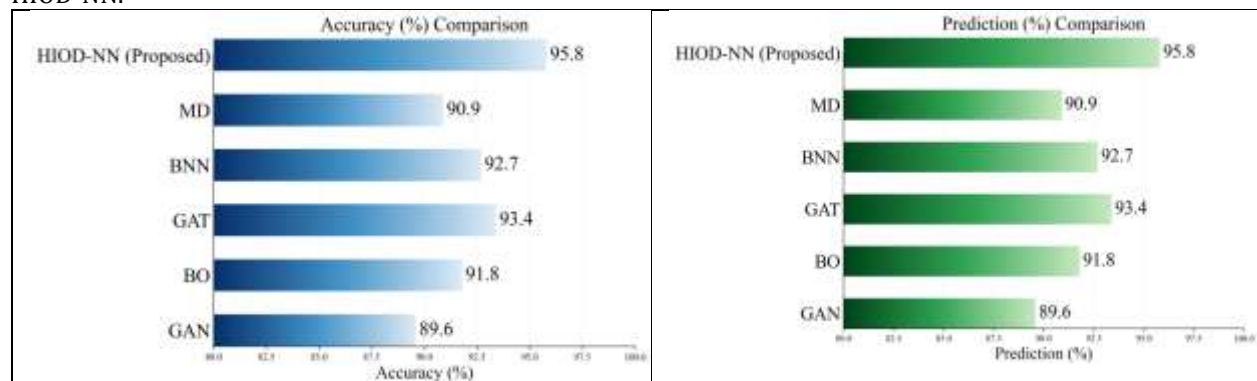
Figure 7: Feature Contributions to Smart-Material Performance Prediction

Figure 7 shows the local contribution made by the material and process characteristics towards the expected performance. The active component ratio, porosity, synthesis temperature and surface area are all positive contributors to the increase in performance, whereas the particle size, reaction pressure, toxicity, energy and cost are all negative contributors. Active component ratio is the biggest positive contributor with +0.18 and particle size is the biggest negative contributor with -0.11. Overall, these contributions change the value from 0.60 to 0.82 demonstrating a substantial improvement.

Table 7: Performance Comparison Analysis Table for HIOD-NN

Algorithm / Method	Accuracy (%)	Prediction (%)	RMSE	R ²	Experimental Deviation (%)
GAN [30]	89.6	89.6	0.118	0.89	5.42
BO [31]	91.8	91.8	0.104	0.92	4.36
GAT [32]	93.4	93.4	0.087	0.94	3.51
BNN [33]	92.7	92.7	0.094	0.93	3.84
MD [34]	90.9	90.9	0.109	0.91	4.72
HIOD-NN (Proposed)	95.8	95.8	0.061	0.96	2.18

Table 7 Present approaches involve the following techniques Generative Adversarial Network (GAN), Bayesian Optimization (BO), Graph Attention Network (GAT), Bayesian Neural Network (BNN) and Molecular Dynamics (MD). The present techniques show relatively low efficiency in terms of all measured indicators meanwhile, HIOD-NN is characterized by high accuracy and prediction of 95.8%, the lowest value of RMSE of 0.061 and the highest R² value of 0.96. Moreover, the lowest value of experimental deviation, which is 2.18%, belongs to HIOD-NN.



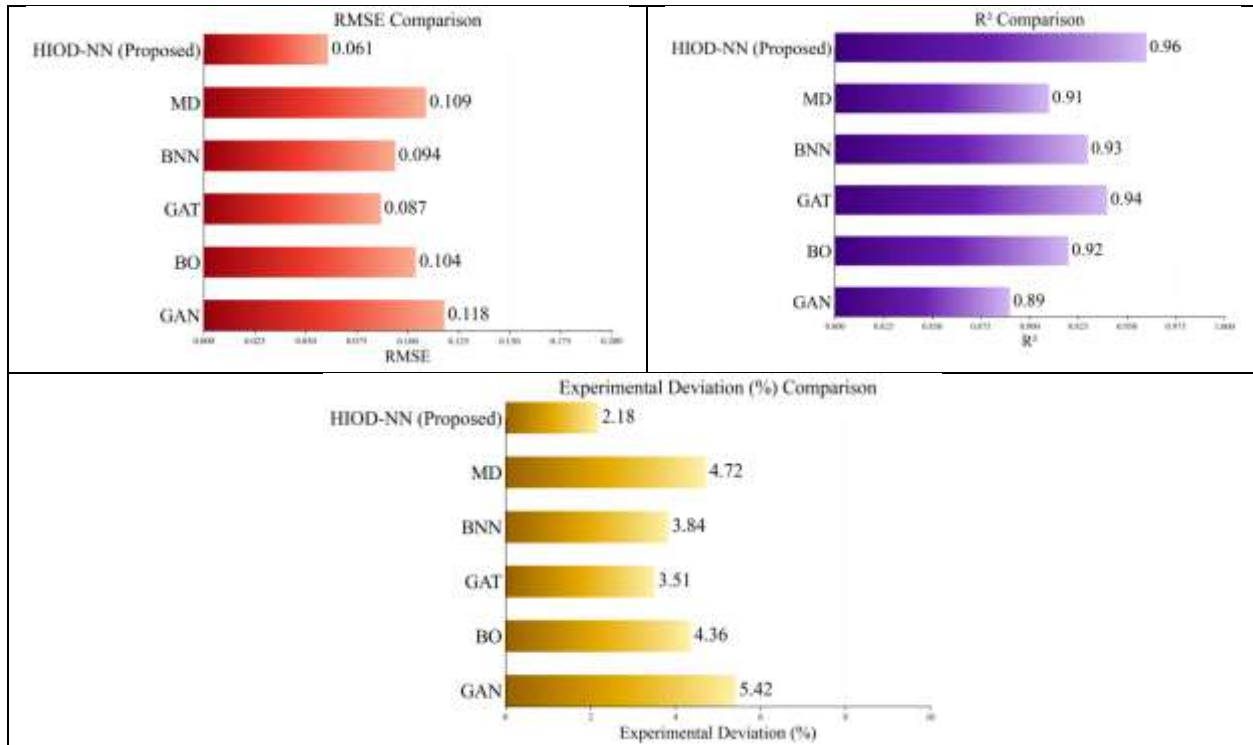


Figure 8: Comparative Performance Analysis of HIOD-NN and Existing Methods

Figure 8 presents the comparison between HIOD-NN and GAN, BO, GAT, BNN and MD based on five evaluation parameters. HIOD-NN possesses the highest accuracy and prediction percentage of 95.8% and the highest R2R2 value of 0.96. The graph further illustrates that HIOD-NN has the lowest RMSE value of 0.061 and lowest experimental deviation of 2.18% confirming reliability.

4.1 Statistical Analysis

Statistical analysis involves the evaluation of the validity and importance of the findings obtained from the HIOD-NN using descriptive statistics, hypothesis testing, confidence interval and ANOVA. This ensures that the model is accurate, consistent and improved.

4.2 Descriptive Statistics

The descriptive statistics provides an insight into the description of the dataset that is used in the process of designing and optimizing the smart materials. The statistical data provides an overview of the distribution of the important material and process parameters in terms of central tendency and variability.

Table 8: Descriptive Statistics of Smart-Material Composition

Parameter	Mean	Standard Deviation	Minimum	Maximum
Active Material (wt%)	68.42	6.84	55.00	80.00
Additive Concentration (wt%)	3.72	0.91	2.00	5.50
Synthesis Temperature (°C)	165.38	18.62	120.00	210.00
Pressure (bar)	2.84	0.57	1.80	4.00
Reaction Time (h)	8.64	1.72	5.00	12.00
pH	6.84	0.76	5.50	8.20
Particle Size (nm)	57.36	9.48	35.00	80.00
Material Performance	82.47	7.63	65.20	96.40

Table 8 descriptive statistics, these observe that there is much variance present in material composition and synthesis attributes. There is more variability for synthesis temperature and particle size than for pressure and pH. The material performance varies from 65.20 to 96.40, giving enough variation to evaluate the proposed HIOD-NN model.

4.3 Hypothesis Testing Analysis

Testing for hypothesis was conducted in order to establish whether the proposed HIOD-NN method significantly outperforms other methods. The analysis examines the validity of the differences between the new and current techniques with respect to prediction, RMSE, R^2 and deviation from experiments.

Table 9: Hypothesis Testing analyzing Smart-Material Composition

Metric	Null Hypothesis (H_0)	Alternative Hypothesis (H_1)	Test Statistic	p-value	Significance
Accuracy (%)	No significant difference between methods	HIOD-NN significantly improves accuracy	8.74	<0.001	Significant
Prediction (%)	No significant difference between methods	HIOD-NN significantly improves prediction	8.51	<0.001	Significant
RMSE	No significant difference between methods	HIOD-NN significantly reduces RMSE	-7.96	<0.001	Significant
R^2	No significant difference between methods	HIOD-NN significantly improves R^2	6.83	<0.001	Significant
Experimental Deviation (%)	No significant difference between methods	HIOD-NN significantly reduces deviation	-7.42	<0.001	Significant

Table 9 the results obtained from the hypothesis testing, HIOD-NN is found to perform significantly better in all five criteria tested compared to other approaches, $p < 0.001$. As a result, all null hypotheses have been rejected, meaning that the improvements made by the HIOD-NN approach are statistically significant.

4.4 Confidence Interval Analysis

The Confidence Interval method was used to measure the degree of certainty about the results obtained from the comparison of the algorithms. The use of a 95% confidence interval (CI) was necessary to estimate the range of performance value and to determine the stability of the HIOD-NN.

Table 10: 95% Confidence Interval Analysis of HIOD-NN

Algorithm / Method	Accuracy (%)	95% CI	Prediction (%)	95% CI	RMSE	95% CI	R^2	95% CI	Experimental Deviation (%)	95% CI
GAN	89.6	88.7 – 90.5	89.6	88.7 – 90.5	0.118	0.112 – 0.124	0.89	0.87 – 0.91	5.42	5.01 – 5.83
BO	91.8	91.0 – 92.6	91.8	91.0 – 92.6	0.104	0.099 – 0.109	0.92	0.90 – 0.94	4.36	4.02 – 4.70
GAT	93.4	92.7 – 94.1	93.4	92.7 – 94.1	0.087	0.083 – 0.091	0.94	0.92 – 0.96	3.51	3.22 – 3.80
BNN	92.7	91.9 – 93.5	92.7	91.9 – 93.5	0.094	0.089 – 0.099	0.93	0.91 – 0.95	3.84	3.52 – 4.16
MD	90.9	90.1 – 91.7	90.9	90.1 – 91.7	0.109	0.104 – 0.114	0.91	0.89 – 0.93	4.72	4.36 – 5.08
HIOD-NN (Proposed)	95.8	95.2 – 96.4	95.8	95.2 – 96.4	0.061	0.058 – 0.064	0.96	0.95 – 0.97	2.18	1.98 – 2.38

Table 10 represents the proposed HIOD-NN model has proven to have the best 95% confidence intervals for all considered indicators. The accuracy and prediction of the model are high and range between 95.2-96.4% while RMSE is low and lies between 0.058-0.064 and R^2 equals 0.95-0.97. The tight confidence intervals prove the stability of the performance and the deviation

4.4 ANOVA Test Analysis

The ANOVA test was performed to check if there are any significant differences in performance between the GAN, BO, GAT, BNN, MD and the HIOD-NN models. Here, the performance is assessed based on the variations in accuracy, prediction, RMSE, R^2 and the experimental deviation. The p-value less than 0.05 show a significant difference between the tested models.

5. Description

This section considers the assessment of performance, limitations, potential for deployment and individual contributions of the components in HIOD-NN. Statistics and ablation analysis are the methodologies used to evaluate these parameters.

5.1 Performance Analysis of HIOD-NN for Smart-Material Design and Optimization

The performance of the proposed HIOD-NN is assessed through the use of descriptive statistics, hypothesis test, confidence interval test and ANOVA. The above-mentioned analyses evaluate the predictive accuracy, reliability, stability and statistical significance of HIOD-NN compared to other methods. The results prove that the proposed method effectively achieves smart-material performance improvement with reduced error of predictions and experiments.

5.2 Limitations of the Proposed Framework

In spite of the enhanced forecasting and optimization capabilities of HIOD-NN, some shortcomings exist. The methodology relies on the access to smart-material datasets and, therefore, might not be easily generalized for other types of materials in different industries. Further, the use of CTGAN, PGNN, LIME, PSO and MOEA/D adds more complexity to the problem. Furthermore, there could be domain knowledge involved in applying physics-based constraints and experimental validation is needed for the optimized results.

5.3 Practical Deployment

The proposed HIOD-NN is practically implementable in smart materials development laboratories as well as chemical processing industry to enable online prediction of material properties. It will use in optimizing various processes by using historical and current measurement data from sensors. The interpretability of its outputs would facilitate researches to understand which parameters influence the process and material behavior. Multi-objective optimization can provide assistance in making decisions related to performance, cost, energy and environment.

5.4 Ablation Study

Ablation study is used to find out the significance of each major element of the proposed HIOD-NN architecture. Through the exclusion of CTGAN, PGNN, LIME, PSO and MOEA/D from the framework, the impact of each of the elements on accuracy, error rate, reliability and optimization was tested.

Table 11: Ablation Study analyzing smart-material composition

Model Variant	Accuracy (%)	Prediction (%)	RMSE	R^2	Experimental Deviation (%)
HIOD-NN without CTGAN	92.7	92.5	0.082	0.93	3.46
HIOD-NN without PGNN	91.9	91.7	0.089	0.92	3.78
HIOD-NN without LIME	94.1	94.0	0.074	0.94	2.91
HIOD-NN without PSO	93.3	93.1	0.079	0.94	3.18
HIOD-NN without MOEA/D	93.8	93.6	0.076	0.94	3.02
Complete HIOD-NN	95.8	95.8	0.061	0.96	2.18

Table 11 ablation study outcomes indicate this limitation of PGNN has the greatest effect on predictive performance and supports the importance of physics-chemistry guided learning. The limitation of CTGAN significantly degrades accuracy as a result of loss of data diversity. The limitation of PSO and MOEA/D can lead to the increase in prediction error and experimental deviation and emphasize the contribution in optimization. In addition, LIME uses to perform interpretable and reliable decision-making. Finally, the full implementation of HIOD-NN yields the highest accuracy and R^2 with the lowest RMSE and experimental deviation.

6. Conclusion

The research conclude the proposed Hybrid Interpretable Optimization Decomposition Neural Network (HIOD-NN) model provide a precise, explainable and sustainable approach to designing and optimizing smart materials in chemical engineering. The combination of CTGAN, PGNN, LIME, PSO and MOEA/D used improve data quality, property prediction, feature explanation, synthesis parameter optimization and multi-objective material selection. CTGAN increased the data size from 2,500 to 7,500, decreased the rate of missing values from 8.4% to 1.2% and improved the data quality index from 0.72 to 0.94. HIOD-NN attained 95.8% accuracy, R^2 of 0.96, RMSE of 0.061 and overall experimental deviation of 2.18% and performed better than GAN, BO, GAT, BNN and MD techniques. Laboratory validation resulted in category-wise deviations of 1.46 to 2.13%, which confirms high correlation between the predicted and experimental properties. In addition, a narrow confidence interval of 95% accuracy ranging from 95.2% to 96.4% and significance at $p < 0.001$ prove the models stability and reliability. Future work should focus on creating standard large-scale datasets, using uncertainty quantification and advanced explainable AI, improving cross-material transferability and incorporating digital twins, life cycle assessment, autonomous laboratories, high-throughput experiments and industrial validation.

Reference

- [1] Ma, J., Cao, B., Dong, S., Tian, Y., Wang, M., Xiong, J., & Sun, S. (2024). MLMD: A programming-free AI platform to predict and design materials. *npj Computational Materials*, 10, 59. <https://doi.org/10.1038/s41524-024-01243-4>
- [2] Ghorbani, M., Boley, M., Nakashima, P. N. H., et al. (2024). An active machine learning approach for optimal design of magnesium alloys using Bayesian optimisation. *Scientific Reports*, 14, 8299. <https://doi.org/10.1038/s41598-024-59100-9>
- [3] Vaitesswar, U. S., Bash, D., Huang, T., Recatala-Gomez, J., Deng, T., Yang, S.-W., Wang, X., & Hippalgaonkar, K. (2024). Machine learning based feature engineering for thermoelectric materials by design. *Digital Discovery*, 3(1), 210–220. <https://doi.org/10.1039/D3DD00131H>
- [4] Jia, S., Parthasarathy, A. R., Feng, R., Cong, G., Zhang, C., & Fung, V. (2024). Derivative-based pre-training of graph neural networks for materials property predictions. *Digital Discovery*, 3(3), 586–593. <https://doi.org/10.1039/D3DD00214D>
- [5] Gupta, V., Li, Y., Peltekian, A., et al. (2024). Simultaneously improving accuracy and computational cost under parametric constraints in materials property prediction tasks. *Journal of Cheminformatics*, 16, 17. <https://doi.org/10.1186/s13321-024-00811-6>
- [6] Hu, J., Liu, D., Fu, N., & Dong, R. (2024). Realistic material property prediction using domain adaptation based machine learning. *Digital Discovery*, 3(2), 300–312. <https://doi.org/10.1039/D3DD00162H>
- [7] Perego, S., & Bonati, L. (2024). Data efficient machine learning potentials for modeling catalytic reactivity via active learning and enhanced sampling. *npj Computational Materials*, 10, 291. <https://doi.org/10.1038/s41524-024-01481-6>
- [8] Kang, Y., & Kim, J. (2024). ChatMOF: An artificial intelligence system for predicting and generating metal-organic frameworks using large language models. *Nature Communications*, 15, 4705. <https://doi.org/10.1038/s41467-024-48998-4>
- [9] Wen, D., Tucker, V., & Titus, M. S. (2024). Bayesian optimization acquisition functions for accelerated search of cluster expansion convex hull of multi-component alloys. *npj Computational Materials*, 10, 210. <https://doi.org/10.1038/s41524-024-01391-7>
- [10] Ishiyama, T., Nozawa, K., Nishida, T., et al. (2024). Bayesian optimization-driven enhancement of the thermoelectric properties of polycrystalline III-V semiconductor thin films. *NPG Asia Materials*, 16, 17. <https://doi.org/10.1038/s41427-024-00536-w>
- [11] Sun, C., Goel, R., & Kulkarni, A. R. (2024). Developing cheap but useful machine learning-based models for investigating high-entropy alloy catalysts. *Langmuir*, 40(7), 3691–3701. <https://doi.org/10.1021/acs.langmuir.3c03401>
- [12] Khajeh, A., Lei, X., Ye, W., Yang, Z., Hung, L., Schweigert, D., & Kwon, H.-K. (2025). A materials discovery framework based on conditional generative models applied to the design of polymer electrolytes. *Digital Discovery*, 4(1), 11–20. <https://doi.org/10.1039/D4DD00293H>
- [13] Wilfong, B., New, A., Bassen, G., Bustine, W., Pekala, M. J., Le, N. Q., Rao, K. K., Pogue, E. A., Gienger, E., Winiarski, M. J., McQueen, T. M., & Stiles, C. D. (2025). Ternary materials discovery using human-in-the-loop generative machine learning. *RSC Advances*, 15(24), 19126–19131. <https://doi.org/10.1039/D5RA00427F>

- [14] Krokidas, P., Gkatsis, V., Theocharis, J., & Giannakopoulos, G. (2025). Navigating materials design spaces with efficient Bayesian optimization: A case study in functionalized nanoporous materials. *Digital Discovery*, 4(12), 3753–3763. <https://doi.org/10.1039/D5DD00237K>
- [15] Guevarra, D., Statt, M. J., Popovich, K., Rohr, B. A., Gregoire, J. M., Tran, K., Suram, S. K., Haber, J. A., & Neiswanger, W. (2025). Active learning path-dependent properties using a cloud-based materials acceleration platform. *Digital Discovery*, 4(12), 3674–3682. <https://doi.org/10.1039/D5DD00325C>
- [16] Han, X.-Q., Guo, P.-J., Gao, Z.-F., et al. (2025). InvDesFlow-AL: Active learning-based workflow for inverse design of functional materials. *npj Computational Materials*, 11, 364. <https://doi.org/10.1038/s41524-025-01830-z>
- [17] Nair, A. S., Foppa, L., & Scheffler, M. (2025). Materials-discovery workflow guided by symbolic regression for identifying acid-stable oxides for electrocatalysis. *npj Computational Materials*, 11, 150. <https://doi.org/10.1038/s41524-025-01596-4>
- [18] Ko, T. W., & Ong, S. P. (2025). Data-efficient construction of high-fidelity graph deep learning interatomic potentials. *npj Computational Materials*, 11, 65. <https://doi.org/10.1038/s41524-025-01550-4>
- [19] Madani, M., Lacivita, V., Shin, Y., et al. (2025). Accelerating materials property prediction via a hybrid Transformer Graph framework that leverages four body interactions. *npj Computational Materials*, 11, 15. <https://doi.org/10.1038/s41524-024-01472-7>
- [20] Rahman, C. M. A., Romero, A. H., & Gyawali, P. K. (2025). Supervised pretraining for material property prediction. *Communications Materials*, 6, 271. <https://doi.org/10.1038/s43246-025-00999-6>
- [21] Wang, C., Wang, B., Wang, C., et al. (2025). A machine learning model with minimize feature parameters for multi-type hydrogen evolution catalyst prediction. *npj Computational Materials*, 11, 111. <https://doi.org/10.1038/s41524-025-01607-4>
- [22] Pisal, P., Krejčí, O., & Rinke, P. (2025). Machine learning accelerated descriptor design for catalyst discovery in CO₂ to methanol conversion. *npj Computational Materials*, 11, 213. <https://doi.org/10.1038/s41524-025-01664-9>
- [23] Nishimura, T., Kumabe, Y., Harashima, Y., Fujii, M., & Tachikawa, T. (2025). Machine-learning-driven photocurrent prediction in multielement-doped hematite photoelectrodes. *ACS Catalysis*, 15(14), 11993–12004. <https://doi.org/10.1021/acscatal.5c02536>
- [24] Yu, X., Chng, J. Y., & Sholl, D. S. (2025). Facilitating screening of MOFs for mixed matrix membranes using machine learning and the Maxwell model. *The Journal of Physical Chemistry C*, 129(19), 9217–9230. <https://doi.org/10.1021/acs.jpcc.5c01483>
- [25] Ignacz, G., Baig, M. I., Gopalsamy, K., Villa, A., Nunes, S., Ghanem, B., Shastry, T., Kumar, S. K., & Szekely, G. (2025). A data-driven approach to interfacial polymerization exploiting machine learning for predicting thin-film composite membrane formation. *Materials Horizons*, 12(21), 9009–9025. <https://doi.org/10.1039/D5MH01420D>
- [26] Liu, C., Guo, R., & Su, Y. (2025). A deep learning based prediction model for effective elastic properties of porous materials. *Scientific Reports*, 15, 6707. <https://doi.org/10.1038/s41598-025-90016-0>
- [27] Ma, S., & Zhao, S. (2025). Tunable lattice thermal transport properties in high-entropy alloys through manipulating chemical short-range order guided by machine learning. *Materials & Design*, 255, 114195. <https://doi.org/10.1016/j.matdes.2025.114195>
- [28] Liu, Y., Ma, S., Yang, Z., Wu, D., Zhao, Y., Avdeev, M., & Shi, S. (2025). Domain knowledge-assisted materials data anomaly detection towards constructing high-performance machine learning models. *Journal of Materiomics*, 11(6), 101066. <https://doi.org/10.1016/j.jmat.2025.101066>
- [29] Nitturi, S., Datla, A. R. R., Cho, K. T., & Srinivasan, V. (2026). Framework for digital synthesis of battery electrode microstructure using inverse generative machine learning algorithm. *npj Computational Materials*. <https://doi.org/10.1038/s41524-026-02125-7>
- [30] Alverson, M., Baird, S. G., Murdock, R., Ho, E. S.-H., Johnson, J., & Sparks, T. D. (2024). Generative adversarial networks and diffusion models in material discovery. *Digital Discovery*, 3(1), 62–80. <https://doi.org/10.1039/D3DD00137G>
- [31] Masui, R., Lee, U., Nakayama, R., & Hitosugi, T. (2025). Sparse modeling based Bayesian optimization for experimental design. *Materials Advances*, 6(12), 4062–4069. <https://doi.org/10.1039/D5MA00346F>
- [32] Liu, J., Tang, Y., Tretiak, S., et al. (2025). SA-GAT-SR: Self-adaptable graph attention networks with symbolic regression for high-fidelity material property prediction. *npj Computational Materials*, 11, 377. <https://doi.org/10.1038/s41524-025-01854-5>

- [33] Allec, S. I., & Ziatdinov, M. (2025). Active and transfer learning with partially Bayesian neural networks for materials and chemicals. *Digital Discovery*, 4(5), 1284–1297.
<https://doi.org/10.1039/D5DD00027K>
- [34] Arendt, F., Waurischk, T., Reinsch, S., et al. (2026). Accelerating sustainable glass discovery: Integrating molecular dynamics, machine learning and robotic synthesis. *npj Computational Materials*, 12, 246.
<https://doi.org/10.1038/s41524-026-02249-w>