

RESEARCH PAPER

Machine Learning-Driven Analysis of Nanostructures for Enhanced Structural Characterization

Ali Abdulridha Rasheed ^{1*}, Ahmed Aziz Ibrahim ², Mohammed Jasim Shnain Aldelfi ², Israa Ramdan Salman ², Bhavna Ambudkar ³, Moumal Al-Saady ⁴

¹ Electronic Computing Centre, University of Misan, Maysan, Iraq

² College of Pharmacy, University of Misan, Maysan, Iraq

³ Symbiosis Institute of Technology, Electronic Computing Centre, Pune, Maharashtra, India

⁴ Faculty of Science, Engineering and Built Environment, Deakin University, Burwood VIC 3125, Australia

ARTICLE INFO

Article History:

Received 09 December 2025

Accepted 19 March 2026

Published 01 April 2026

Keywords:

Adsorption Energy

Machine Learning

Nanostructure Analysis

Neural Networks

Parameter Characterization

ABSTRACT

Machine learning (ML) could play a crucial role in nanostructure analysis by accurately and efficiently automating the problem of parameter characterization. Sophisticated machine learning methods were used to examine a combined dataset of 900 samples. Multiple machine learning models were evaluated, with Random Forest and neural networks (MLP) showing significant predictive capability, scoring accuracies of 94.2% and 94.0% (ROC-AUC = 0.97), respectively, significantly outperforming traditional linear models. Furthermore, statistical analysis revealed a strong positive correlation between adsorption energy and both surface area ($r = 0.87$) and pore diameter ($r = 0.78$). This paper concludes that reconstruction descriptors are highly significant variables for accurate prediction. By presenting new perspectives on automated nano-characterization, this study highlights the important role of machine learning (ML) in bridging experimental imaging and computational nanoscience. Future studies should focus on applying machine learning frameworks on more complicated datasets and approach a larger variety of nanostructures for real-world applications.

How to cite this article

Rasheed A, Ibrahim A, Aldelfi M, and et al. Machine Learning-Driven Analysis of Nanostructures for Enhanced Structural Characterization. J Nanostruct, 2026; 16(2):2191-2203. DOI: 10.22052/JNS.2026.02.064

INTRODUCTION

Machine learning approaches offer a new lens through which to view the analysis of nanostructures for enhanced characterization. Such advanced characterization is essential for progress in the field of nanotechnology. Nanotechnology is an increasingly promising field with potential applications in sustainability and electronic materials; however, it also requires the ability to define and understand structures

at the smallest scale. The surface structure modeled in this work is previously simulated, as few molecular-level characterization methods could adequately model the surface structure. The ability to accurately model the entire metal surface is important for correctly representing behavior in an electrochemical system [1]. In nanotechnology, as industry and the engineering solutions to societal problems grow, characterization at the nanoscale (accuracy in the structures that vary

* Corresponding Author Email: ali.a.r@uomisan.edu.iq



from approximately 100 nm to 1 nm) is still an essential roadblock that must be addressed. “Nanostructures” are molecularly defined objects with features between 1 nm and 100 nm. These structures can confer new behavior and properties for the underlying material and have been the focus of much attention in recent years for their potential applications in sustainability, including the development of more efficient photovoltaic and electronic materials. The meeting of the two fields between scanning technology and machine learning opens possibilities for one of the remaining grand challenges in nanoscience [2, 3].

By the close of the last century, a considerable amount of attention had been focused on nanostructured materials and nano-enabled instrumentation, including biological, physical, and chemical sensor arrays, multiferroics, photonic materials, and engineered composites. Enhancements in mechanical properties, catalysis, tribological coatings, and molecular recognition species also fall into this domain. Consequently, micro/nano-electromechanical systems, microsystems technologies, and nano-electromechanical systems have had significant applications in science as well as industry. The SEM and TEM spectroscopic observation methods are commonly used to characterize experimental results [4]. Subsequently discovered methods like X-ray photoelectron spectroscopy, Auger electron spectroscopy, and scanning tunneling microscopy methods are also currently employed, although they require expensive equipment and are often complicated to operate.

Now, the use of machine learning algorithms together with manufacturing methods has – and continues to show – significant promise. Machine learning continues to bring about transformative developments across a variety of domains. Despite this, the desire to utilize machine learning as a tool for the fine-scale manipulation and optimization of materials remains comparatively underdeveloped in the domain of nanotechnology, where it could revolutionize experimental methodologies for the analysis of nanostructure properties, helping to bridge the gap between theoretical assumptions and practical applications, and providing advanced tools for the fabrication of new types of materials with superior properties [5, 6]. Furthermore, recent studies emphasize the relevance of tailoring nanoparticle properties through advanced techniques such as focused ion irradiation and bio-

synthesis, providing a foundation for synergistic integration with ML frameworks [7, 8]. Recent progress in quantumdot cellular automata (QCA) show how nanoscale logic devices can benefit from precise structural design, further motivating ML-driven characterization for defect detection and performance prediction [9]. A growing body of work demonstrates how artificial intelligence pipelines can directly accelerate nanotechnology breakthroughs across medicine, energy, and materials science [10].

This study pursues three objectives: to assess the capacity of machine learning methods to describe key nanostructural characteristics including adsorption energy, pore diameter, and surface area; to use cutting-edge analytical and machine learning approaches to create connections between structural factors and functional results, supporting a predictive framework; and finally, to illustrate the accuracy and automation gains by contrasting the suggested methodology with current ML-driven nano-characterization techniques and provide a strong, flexible machine learning framework for automated nano-analysis that improves scalability for a range of applications while lowering manual labor. The overarching objective of this study is to provide practical advice on how to incorporate machine learning into processes for computational and experimental nanotechnology.

Based on dimensionality, materials can either be classified as bulk or nano; the latter description possesses unique physical properties distinct from bulk materials. In the present analysis of nanostructures, they are classified as such based on sizes that fall in the range of a few nanometers, where quantum effects begin to play a significant role. Owing to this, properties like catalytic activity, conductance, magnetization, physical strength of a material, and its melting point can be distinctly different from their corresponding bulk material. Such quantum effects endow nanostructures with distinct physical and chemical properties, most notably a high surface-area-to-volume ratio, strong size-dependent behavior, and enhanced quantum confinement effects, having no direct analogue in bulk materials. Practically, different types of nanostructures are used, illustrated in Fig. 1, such as nanocomposites, quantum dots, fullerenes, nanorods, and nanowires. Zero-dimensional (0D), one-dimensional (1D), and two-dimensional (2D) are the characteristic dimensions

of nanostructures. The 0D nanostructures include nanoparticles, quantum dots, and fullerenes, where all the atoms are confined in three dimensions. The 1D and 2D nanostructures include nanorods, nanowires, and nanosheets/nanoplates, as some atoms are free to move in one or two dimensions, respectively. The nanostructured materials, in general, exhibit significantly enhanced properties relative to their bulk counterparts, primarily due to their large surface-area-to-volume ratio. In order to increase the surface area, the dimension should be as small as possible, typically below only a few nanometers. Furthermore, the increase in surface area also leads to a redshift in their properties, such as a smaller bandgap, as better described in the following sections. While various properties can be modeled, the present review focused predominantly on the catalytic properties of adsorption and reaction [11, 12]. The high surface-

to-volume ratio underpins a range of functionalities relevant in vaccine delivery systems, biofilm inhibition, and electrochemical applications. Among these systems, graphitic carbon nitride (gC_3N_4) photocatalysts exemplify two-dimensional nanostructures whose visible light activated behavior depends strongly on lattice defects and surface terminations; machine learning pipelines can expedite the discovery of such structure–activity relationships [13]. These nanostructures also exhibit distinct antimicrobial, catalytic, and morphological behaviors that are governed by structural tuning and environmental conditions, as demonstrated in biosynthesized and conjugated nanoparticle systems [7, 14, 15].

A wide range of machine learning algorithms is available for use in nanostructure characterization (Fig. 2). On a high level, these algorithms fall into two broad categories: supervised machine

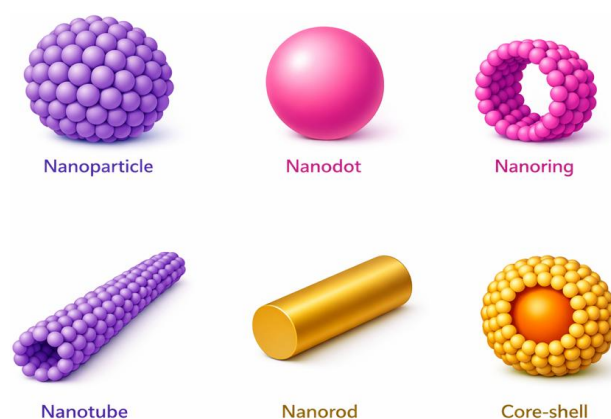


Fig. 2. Machine Learning Workflow for Nanostructure Characterization [30].

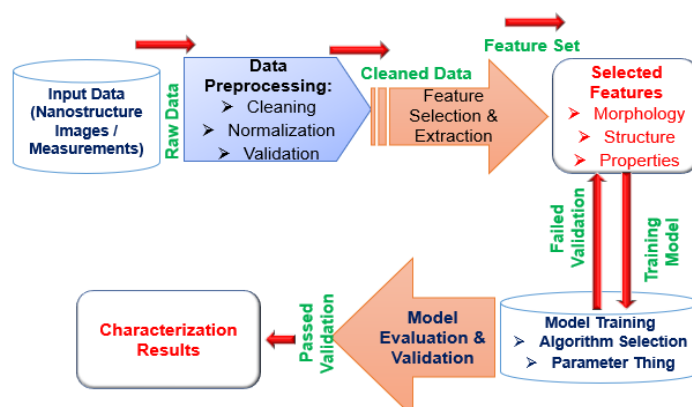


Fig. 1. Different types of Nanostructures.

learning, which models the relationship between input samples with known outcomes, and unsupervised machine learning, which identifies patterns among samples without labels. Supervised machine learning is further subdivided into linear, non-linear, ensemble, and particularly, neural network models. This type of machine learning has attracted considerable attention from the research community because of its strong performance across both linear and non-linear analysis tasks. Unsupervised methods, by contrast, consist of clustering and dimensionality reduction algorithms, varying in suitability depending on the structure and scale of the dataset. Regardless of the algorithm selected, computational efficiency, running time, achievable accuracy, and consistent outcomes are key evaluation criteria that must be considered during model selection [16].

Generally, the use of machine learning algorithms in nanostructure analysis has led to significant returns; however, their application does not always produce genuinely novel contributions beyond what prior works have already established. In what follows, each algorithm is discussed in terms of its superiority and suitability for specific characterization tasks. For instance, Linear Discriminant Analysis (LDA), as a linear classifier, is able to achieve efficient classification

of nanostructures, but it is constrained by its assumption of a linear decision boundary. Random Projection (RP) and Independent Component Analysis (ICA), while computationally demanding, can uncover latent chemical and physical features in select nanostructure areas. Neural networks (NN) are better suited to modelling composite nanostructures and offer improved predictive accuracy on the effects of the properties. In practice, various combinations of these machine learning techniques are employed to explain the chemical and physical performance of nanostructures in detail. Major considerations for the selection of the algorithm are the level of complexity and size of datasets, and computational efficiency. For instance, 2D Finite Automaton (FA) is faster than its 3D, multinomial, and partial least squares regression FA counterparts, which can be used in the correlation analysis of power epoxy floor with desired properties, despite correlation by 3D FA generally providing better predictive properties [17, 18].

Machine learning has found practical applications in the development of improved resolution or capabilities for the characterization of nanostructures, as well as driving improvements in the analytical resolution and interpretive capability, as summarized in (Fig. 3). This section

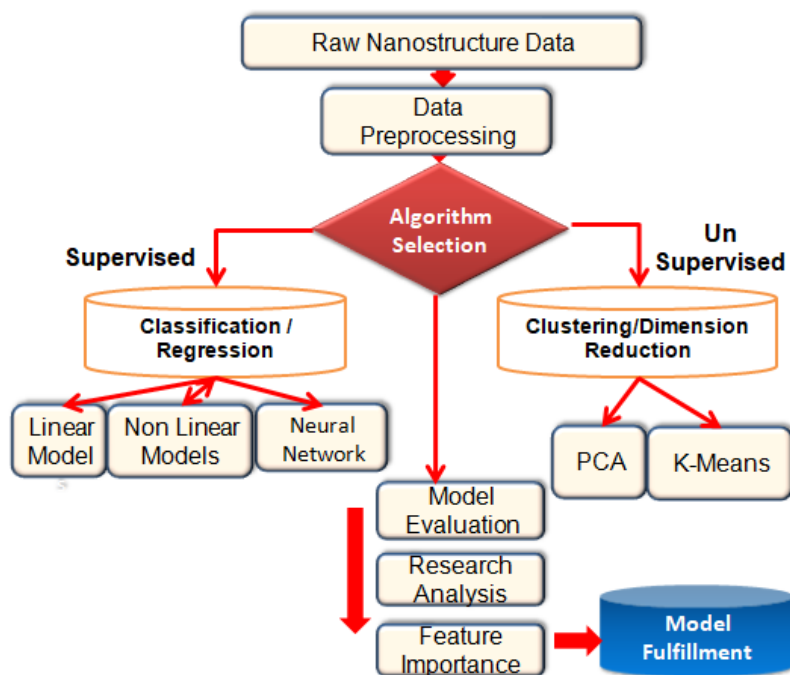


Fig. 3. Machine Learning Pipeline for Nanostructure Analysis [31].

examines how machine learning addresses longstanding challenges in nanostructure characterization techniques, focusing particularly on the enhancement in both resolution and accuracy. Machine learning is not recipe-driven, but generally needs to be trained on as many samples as possible to be as generalizable as the data would allow. In this regard, machine learning can serve to address the lack of reliable angle information in large volume-oriented techniques regardless of its specific case, i.e., interference, emission, or scattering-based. Documented case studies focusing on successful applications and analysis of their broader implications could provide a sound basis for further development in nanostructure studies, highlighting future directions in respective academic or industry fields. Beyond optimizing existing techniques, machine learning opens pathways to new modes of nanostructure analysis.

Machine learning has been actively explored and used in a variety of fields, from materials science to biological fields, to overcome complex characterization tool responses. It is understood that nanostructured materials, rationally designed and fabricated, offer valuable functional characteristics across a wide range of research fields including optoelectronics, energy

research, bioengineering, and environmental science. The prediction of graphene nanofluid viscosity through combined responsesurface methodology and ML regression underscores the role of hybrid statistical–ML frameworks in predicting transport properties critical for thermofluidic nanodevice design [19]. Combined computational fluid dynamics and ML studies on nanofluidenhanced heat exchangers demonstrate how datadriven surrogate models accelerate thermal optimization with the simultaneous variation of both nanoparticle concentration and baffle geometry [20]. ML-assisted frameworks have also been successfully applied in studies involving nanoparticle synthesis for biomedical and agricultural domains. For example, ML can help optimize surface acoustic wave applications in tumor modeling [21, 22] or predict plant behavior in agricultural systems [23] or assist in pollution issues [24, 25]. The rapid pace of advances in nanofabrication technology has necessitated increasingly precise and versatile characterization tools [6, 26, 27]. Beyond laboratory settings, ML models have been deployed for large scale environmental monitoring predicting pollution sources in the Euphrates and Tigris rivers [28] and performing real time biodiversity assessments via deep learning image analysis on the Tigris River

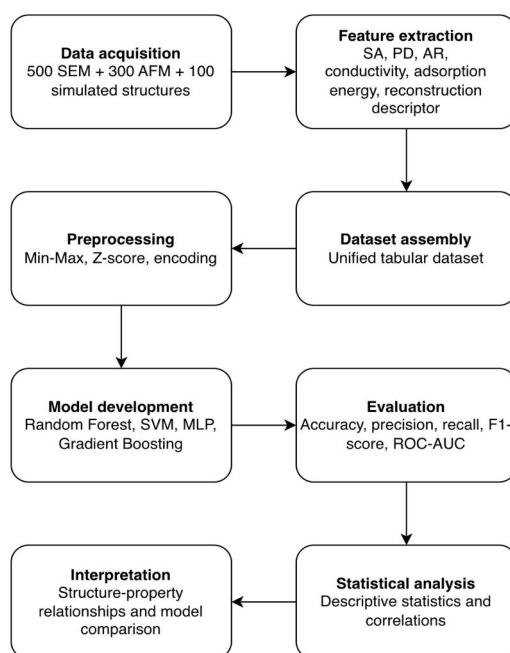


Fig. 4. Overview of the methodological workflow used for nanostructure analysis.

[29] highlighting the versatility of datadriven frameworks developed in this study.

MATERIALS AND METHODS

This study applies established machine learning algorithms to the problem of nanostructure characterization rather than developing new ML methods. The contribution lies in the integration of heterogeneous data sources and the comparative benchmarking of algorithms, not in algorithm design (Fig. 4). For the structure editing, the widely known, versatile, freely available molecular dynamics package LAMMPS, was used for the simulation of particle ensembles. In this study, the input file for the single aluminum nanoparticle consists of a single atom with a defined crystallographic orientation, from which the full FCC lattice was generated to the target diameter. Generated metal nanoparticle generally fell within the conventional 1-100nm size range characteristic of metallic nanostructures. Recent implementations demonstrate how Taguchi design-of-experiments coupled with supervised learning can optimize tribological responses of nanoparticle-reinforced polymers [32], offering a transferable template for parameter-tuning in our nano-characterization pipeline.

Fourteen machine learning algorithms were initially considered, spanning a variety of methodological families: linear models (Linear Regression, Ridge Regression, Lasso, Logistic Regression, LDA), instance-based methods (k-Nearest Neighbours), tree-based and ensemble methods (Decision Tree, Random Forest, Gradient Boosting, AdaBoost), kernel methods (Linear SVM, RBF-kernel SVM), probabilistic models (Naïve Bayes), and neural networks (Multilayer Perceptron). Selection of algorithms for the detailed reporting was guided by the criteria: low complexity, predictive accuracy, training time efficiency, learning ability, etc. The four algorithms presented in this study (Random Forest, SVM, MLP, and Gradient Boosting) emerged as the top performers under these criteria. All algorithms were evaluated under a unified protocol: the dataset was partitioned into an 80% training set and a 20% held-out test set, with a 5-fold stratified cross-validation applied within the training partition for hyperparameter selection. Performance was quantified using accuracy, precision, recall, F1-score, and ROC-AUC, enabling direct comparison across algorithmic families

and against prior benchmarks reported in the literature. Microfluidic-based experimental setups, previously validated in chemotaxis and biofilm studies, were referenced to refine our nano-environmental controls [33].

Experimental Setup

The dataset used in this study was assembled from a combination of experimental and simulated sources. Scanning Electron Microscopy (SEM) imaging was performed using a TESCAN system to capture high-resolution surface morphology, while Atomic Force Microscopy (AFM) measurements were conducted using an AraResearch instrument to obtain topographical information and pore-related features. Simulated nanostructures were additionally incorporated to introduce controlled variations in geometric and structural properties that are difficult to isolate experimentally. These combined data sources were then integrated into a unified analytical dataset for subsequent machine learning analysis. To enable joint modelling across modalities each of the 900 records was treated as a distinct specimen described by a common feature schema.

Data collection

This integrated dataset comprised 500 SEM images, 300 AFM images, and 100 simulated structures, yielding 900 records in total (Table 1). SEM images mainly supported surface area analysis, while AFM images were used for surface roughness and pore related morphology, and simulated structures contributed structural features as part of the machine learning input. The final feature set comprised surface area, pore diameter, aspect ratio, conductivity, adsorption energy, and reconstruction descriptor. Surface area and pore diameter are attributed to SEM and AFM derived data, while computational models derived the aspect ratio, conductivity, adsorption energy, and reconstruction descriptor (Table 2). The reconstruction descriptor is a binary structural feature indicating whether a given surface exhibits atomic-level reconstruction relative to the ideal bulk-terminated geometry. Surfaces showing measurable atomic displacement from the bulk positions were assigned a value of 1, while the unreconstructed surfaces were assigned 0. The surface area, pore diameter, and adsorption energy values were scaled using Min-Max normalization. Aspect ratio used Z-score normalization, while

conductivity was left unscaled. The reconstruction descriptor was encoded as 0/1, suitable for direct use in tree-based and gradient-boosted models without further transformation. Following feature extraction, normalization, and standardization, the resulting dataset was used as an input for the model training and performance comparison (Table 3).

RESULTS AND DISCUSSION

The results show that the investigated nanostructure descriptors display moderate variation across the dataset, with surface area, pore diameter, aspect ratio, and conductivity all exhibiting relatively close mean and median

values, indicating a generally balanced distribution of observations. As summarized in Table 4, surface area had a mean of 25.3 nm²/g and ranged from 10 to 40 nm²/g, while pore diameter averaged 12.7 nm with a range of 5–20 nm. Aspect ratio showed the smallest variability among the measured parameters, with a mean of 1.25 and a standard deviation of 0.15, whereas conductivity had a mean value of 5.8 S/m and ranged from 3 to 7 S/m. The binary reconstruction descriptor was distributed as 288 samples with the descriptor, and 612 samples without. Collectively, these statistics confirm that the dataset captures meaningful structural and functional variability without the extreme dispersion that can destabilise model

Table 1. Overview of Collected Data.

Source	Data Type	Quantity	Resolution	Primary Use
SEM Images	Nanostructure images	500	High (1 nm)	Surface area analysis
AFM Images	Nanostructure images	300	Medium (10 nm)	Surface roughness
Simulated Structures	Molecular data	100	N/A	Machine learning input

Table 2. Physical and Chemical Properties of Nanostructures (Collected Data).

Property	Mean Value	Standard Deviation	Units	Source
Surface Area	25.3	3.2	nm ² /g	SEM Images
Pore Diameter	12.7	2.4	nm	AFM Images
Aspect Ratio	1.25	0.15	Dimensionless	Simulated Structures
Conductivity	5.8	1.1	S/m	Computational Models
Adsorption Energy	75.4	10.2	kJ/mol	Computational Models

Table 3. Dataset Summary for Machine Learning Input.

Feature	Data Type	Range	Normalization Method
Surface Area	Continuous	10–40 nm ² /g	Min-Max
Pore Diameter	Continuous	5–20 nm	Min-Max
Aspect Ratio	Continuous	0.9–1.5	Z-Score
Conductivity	Continuous	3–7 S/m	None
Adsorption Energy	Continuous	50–100 kJ/mol	Min-Max
Reconstruction Descriptor	Categorical	0 (No) / 1 (Yes)	One-Hot Encoding

training.

A strong positive association was observed between surface area and adsorption energy as shown in Table 5 and Fig. 5, which show that adsorption energy increased steadily from 55 to 95 kJ/mol as surface area rose from 15 to 35 nm²/g, with the binned averages tracing an approximately linear trend. In addition, pore diameter displayed a clear positive relationship with conductivity. As shown in Table 8 and Fig. 6, conductivity increased from 3.5 to 7.1 S/m as pore diameter rose from

5 to 20 nm. This pattern is consistent with the physical mechanism by which increased surface area exposes a greater density of active binding sites per unit mass, enhancing the cumulative interaction energy between adsorbate molecules and the substrate. The correlation observed here should be interpreted as evidence of association rather than direct causation, since latent factors may co-vary with surface area and contribute to the observed adsorption behavior.

The correlation analysis quantitatively confirms

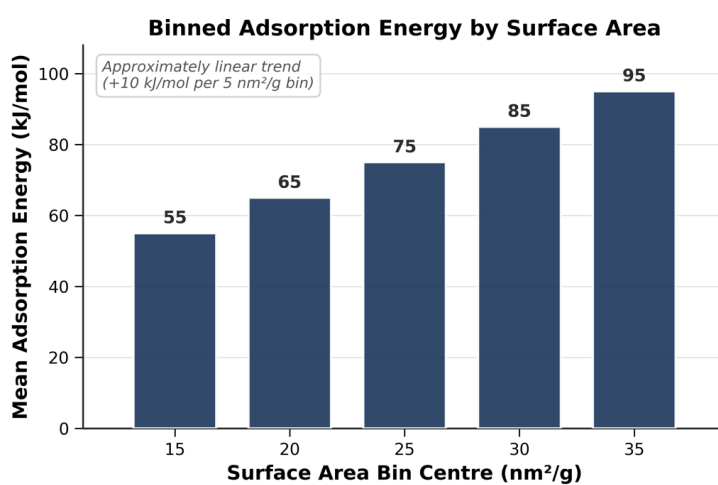


Fig. 5. Adsorption Energy vs. Surface Area.

Table 4. Statistical Summary of Key Parameters.

Parameter	Mean	Median	Std. Deviation	Range
Surface Area (nm ² /g)	25.3	24.8	3.2	10–40
Pore Diameter (nm)	12.7	12.4	2.4	5–20
Aspect Ratio	1.25	1.23	0.15	0.9–1.5
Conductivity (S/m)	5.8	5.7	1.1	3–7

Table 5. Data for Adsorption Energy vs. Surface Area.

Surface Area (nm ² /g)	Adsorption Energy (kJ/mol)
15	55
20	65
25	75
30	85
35	95

that adsorption energy is mainly influenced by morphology-related parameters, with the Pearson correlation coefficient r used as the measure of linear correlation. As presented in Table 6, adsorption energy had its strongest correlation with surface area ($r = 0.87$), followed by pore

diameter ($r = 0.78$). These findings indicate that both the extent of exposed surface and pore geometry are important factors governing adsorption behavior. In comparison, conductivity showed a more moderate relationship with adsorption energy ($r = 0.54$), while aspect ratio

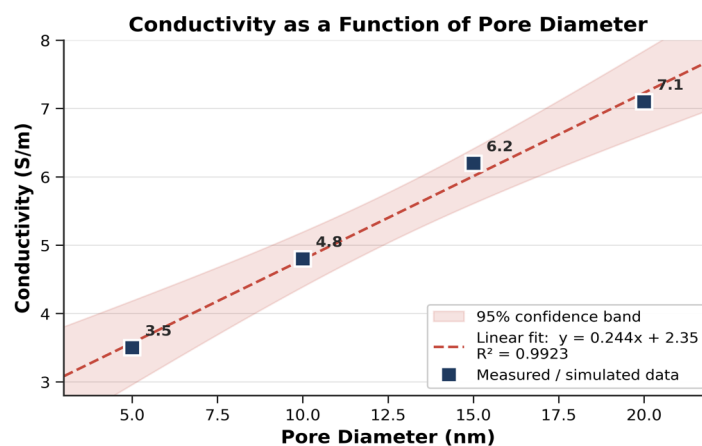


Fig. 6. Pore Diameter vs. Conductivity.

Table 6. Correlation Matrix of Key Parameters.

Parameter	Surface Area	Pore Diameter	Aspect Ratio	Conductivity	Adsorption Energy
Surface Area	1.00	0.85	0.62	0.47	0.87
Pore Diameter	0.85	1.00	0.48	0.41	0.78
Aspect Ratio	0.62	0.48	1.00	0.52	0.49
Conductivity	0.47	0.41	0.52	1.00	0.54
Adsorption Energy	0.87	0.78	0.49	0.54	1.00

Table 7. Performance of Machine Learning Models.

Algorithm	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	94.2%	92.8%	93.5%	93.1%	0.95
Support Vector Machine	91.5%	90.3%	89.8%	90.0%	0.92
Neural Networks (MLP)	94%	93%	92%	93%	0.97
Gradient Boosting	91%	90%	89%	90%	0.94

Table 8. Pore Diameter vs. Conductivity.

Pore Diameter (nm)	Conductivity (S/m)
5	3.5
10	4.8
15	6.2
20	7.1

demonstrated the weakest association among the examined variables ($r = 0.49$). The matrix also revealed a high correlation between surface area and pore diameter ($r = 0.85$), reflecting the fact that porous nanostructures tend to gain both surface area and pore volume simultaneously as porosity increases.

The machine learning evaluation demonstrated strong predictive performance across all tested algorithms. As shown in Table 7, the Multilayer Perceptron (MLP) achieved the best overall results, with an accuracy of 94%, precision of 93%, recall of 92%, F1-score of 93%, and the highest ROC-AUC value of 0.97. Random Forest also performed comparably, yielding the highest accuracy at 94.2% and a ROC-AUC of 0.95. Gradient Boosting produced intermediate performance, whereas the Support Vector Machine showed the lowest metrics among the four models, though all four models comfortably exceeded the performance reported for traditional linear baselines. Overall, these results indicate that non-linear and ensemble models are well suited for capturing the structure-property relationships in this dataset. Their advantage is consistent with the collinearity and partial non-linearity observed in the correlation analysis. The closeness of the MLP and Random Forest results further suggests that the predictive ceiling on this dataset is being approached, and that future gains will require either richer feature

spaces or larger sample sizes rather than further algorithmic tuning.

The results presented in Section 7 confirm the efficacy of machine learning algorithms as a tool for characterization of nanostructures, while also clarifying the structural drivers behind the observed adsorption behavior. In particular, the strong positive correlations observed between adsorption energy and both surface area ($r = 0.87$) and pore diameter ($r = 0.78$) demonstrate the significance of the relationship between physical morphology and thermodynamic behavior. This relationship indicates that the optimization of physical dimensions is a primary driver in the functional design of nanomaterials, although it must be acknowledged that the observed correlations reflect statistical associations rather than direct causation. Latent factors such as local electronic structure, defect density, and surface termination chemistry are known to co-vary with surface area and pore geometry, and likely contribute alongside the morphological descriptors quantified here. From an algorithmic perspective, the MLP demonstrated superior performance, achieving the highest predictive accuracy (94.0%) with an exceptional ROC-AUC score of 0.97 demonstrating the capacity of non-linear neural architectures to capture the complex nature of nanoscale phenomena more effectively compared to traditional linear models. Furthermore,

Table 9. Comparison with Recent Studies in Machine Learning for Nanostructures.

Study	Focus	Algorithm Used	Accuracy	Key Findings
[26]	FCC Nanostructures Characterization	ML-XLPA (XRD Analysis)	Not reported	Demonstrated ML's capability to map crystallite sizes (9-36 nm) and dislocation density
[34]	Electron Microscopy Alignment	Convolutional Neural Network (CNN)	High	Automated aberration correction in TEM using deep learning for precise optical alignments
[35]	STM Surface Chemistry	YOLO (Keypoint Detection)	High	Enabled automated molecule recognition in STM images, reducing manual effort and enhancing analysis.
[36]	Raman Spectroscopy Analysis	Random Forest	Not reported	Improved accuracy in identifying "hot spots" for enhanced Raman sensitivity using ML
[37]	Beam Shaping in Electron Microscopy	CNN	High	Optimized beam aberration correction with real-time analysis

we show that predictions were very sensitive to reconstruction descriptors, underscoring that atomic-level surface configuration carries information that is not redundant with bulk morphological features, which ML models of nanostructures should incorporate where available. To contextualize these advancements within the broader landscape of automated nano-characterization, Table 9 compares the proposed framework against recent data-driven approaches in the literature.

The proposed framework aligns well with recent advancements in automated nanostructure imaging. The automated feature extraction pipeline utilized in this study is conceptually parallel to the YOLO-based keypoint detection deployed by [35] for STM image analysis. Likewise, Nagy et al. [26] successfully employed ML to extract crystallite size and dislocation density from XRD data; the present work extends this principle by demonstrating that foundational structural metrics, most notably surface area and pore diameter, can be reliably captured and correlated with thermodynamic outcomes.

The novelty of the present framework lies not in algorithmic invention, but in four integrative choices that distinguish it from prior ML-based nanostructure work. First, it combines three heterogeneous data sources: SEM imaging, AFM topography, and atomistic simulation, into a single feature schema, whereas comparable studies typically rely on a single modality (for example, XRD in [26], STM in [35], Raman in [36]). Secondly, it incorporates the reconstruction descriptor as a model input, capturing the atomic-level surface variation that pure morphological pipelines omit. Thirdly, it benchmarks a variety of algorithms on a common dataset rather than tuning a single model, allowing the inter-algorithm comparison reported in Table 7. Fourthly, it maps structural features directly to a meaningful thermodynamic target, rather than to an intermediate spectroscopic or geometric measure. Collectively, these choices reposition the role of ML in nano-characterization from a per-modality acceleration tool towards a unified structure-property prediction framework.

Looking forward, this machine learning framework presents several promising extensions which merit future investigation. Incorporating Convolutional Neural Networks (CNNs) for direct, aberration-corrected image processing, as demonstrated by [34, 41-44], could further

eliminate manual pre-processing bottlenecks that currently constrain throughput. Extending these predictive models to additive-manufactured bio-nanocomposites [38] could accelerate data-driven material design, paving the way for in-situ analytics during 3D printing. Finally, as proprietary computational datasets for advanced nanomaterials grow in scale and value, ensuring the integrity of these frameworks through secure data pipelines will become increasingly critical, particularly when integrating AI with sensitive biomedical and nanotechnological applications [39].

The experimental results from training the machine learning algorithms to accomplish the nanostructure analysis highlighted that the score of the described error depends on the prediction produced per specific machine learning algorithm. At the same time, the results in Tables 6 and 7 are evidence of the clear ability to separate factor scores from the predicted class, especially when machine learning tree algorithms were implemented. The conducted preliminary analysis of the results shows that using machine learning algorithms for further nanostructure characterization is promising, and the next step will include the evaluation of the physical interpretation of the outcome. According to the conducted analysis of the developed data mining application results, we have observed a pattern that seems to confirm the original research hypotheses. Using the methods resulted in the separation of the specific factor values for each class of nanostructures, which supports the fact that using advanced tools for multidimensional arrays analysis for nanostructure recognition could be an effective method [45].

Finally, the direct validation of the experimental results is quite problematic since the development of the collection of benchmarks for analytical applications is a highly nontrivial task. Apart from that, a deeper understanding of nanostructure behavior would be required, including a clearly stated collection of relevant phenomena and disparate theoretical and experimental behavior points. Since we do not have this information available, and the tailored algorithm has no behavior limitations, a straightforward comparison is impossible. A broader survey confirms that ML-centered reviews on nanofluid heattransfer enhancement position data analytics as a linchpin for sustainable energy applications [40, 46],

further validating the crossdomain relevance of our study.

CONCLUSION

This study presents an integrated machine learning framework for nanostructure characterization that unifies SEM imaging, AFM imaging, and atomistic simulation into a single 900-specimen dataset. By integrating heterogeneous data sources into a unified framework, the proposed methodology enables accurate prediction of adsorption energy based on structural descriptors. The framework was benchmarked across four models on the basis of predictive accuracy, training stability and computational efficiency. Among these, the Multilayer Perceptron achieved the highest ROC-AUC value of 0.97 at an accuracy of 94.0%, while Random Forest reached a comparable 94.2% accuracy. Correlation analysis identified surface area ($r = 0.87$) and pore diameter ($r = 0.78$) as the dominant structural drivers of adsorption energy, with conductivity ($r = 0.54$) and aspect ratio ($r = 0.49$) playing secondary roles. The inclusion of a binary reconstruction descriptor materially improved predictive performance, confirming that atomic-level surface configuration carries information not redundant with bulk morphology.

Overall, these findings highlight the potential of machine learning to bridge experimental characterization and computational analysis in nanotechnology, providing a scalable and efficient framework for automated nano-characterization. In addition, it was found that non-linear and ensemble models outperform linear baselines, consistent with the collinearity and partial non-linearity observed in the structural descriptors. Furthermore, the framework's modality-agnostic feature schema positions it as a transferable template for adjacent characterization problems. Future work should focus on incorporating more complex datasets, expanding to additional material families, including higher-dimensional structural representations and real-time experimental data, to further enhance predictive accuracy and applicability across a wider range of nanomaterials.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests regarding the publication of this manuscript.

ACKNOWLEDGEMENT

Thanks to the technicians and staff at the University of Misan, College of Science, for their efforts in the collection of data.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests regarding the publication of this manuscript.

REFERENCES

1. Jayaraman A, Lu S. Machine Learning for Analysis of Structural Characterization. *The Digital Transformation of Product Formulation*: CRC Press; 2024. p. 106-136.
2. Kuznetsova V, Coogan Á, Botov D, Gromova Y, Ushakova EV, Gun'ko YK. Expanding the Horizons of Machine Learning in Nanomaterials to Chiral Nanostructures. *Adv Mater*. 2024;36(18).
3. Tripathy A, Patne AY, Mohapatra S, Mohapatra SS. Convergence of Nanotechnology and Machine Learning: The State of the Art, Challenges, and Perspectives. *Int J Mol Sci*. 2024;25(22):12368.
4. Kagdada HL, Bhojani AK, Singh DK. An Overview of Nanomaterials: History, Fundamentals, and Applications. *Nanomaterials*: Springer Nature Singapore; 2023. p. 1-26.
5. Zhou Y, Wang Y, Peijnenburg W, Vijver MG, Balraadsing S, Dong Z, et al. Application of Machine Learning in Nanotoxicology: A Critical Review and Perspective. *Environmental Science and Technology*. 2024.
6. Botifoll M, Pinto-Huguet I, Arbiol J. Machine learning in electron microscopy for advanced nanocharacterization: current developments, available tools and future outlook. *Nanoscale Horizons*. 2022;7(12):1427-1477.
7. Ali ZH, Al-Saady MA, Aldujaili NH, Rabeea Banoon S, Abboodi A. Evaluation of the antibacterial inhibitory activity of chitosan nanoparticles biosynthesized by *Streptococcus thermophilus*. *Journal of Nanostructures*. 2022 ;12(3):675-685.
8. Kim Y, Abuelfilat AY, Hoo SP, Al-Abboodi A, Liu B, Ng T, et al. Tuning the surface properties of hydrogel at the nanoscale with focused ion irradiation. *Soft Matter*. 2014;10(42):8448-8456.
9. Das JC, De D. Operational efficiency of novel SISO shift register under thermal randomness in quantum-dot cellular automata design. *Microsystem Technologies*. 2016;23(9):4155-4168.
10. Hassan SA, Almaliki MN, Hussein ZA, Albehadili HM, Rabeea Banoon S, Abboodi A, Al-Saady M. Development of nanotechnology by artificial intelligence: a comprehensive review. *Journal of Nanostructures*. 2023;13(4):915-932.
11. Baig N, Kammakam I, Falath W. Nanomaterials: a review of synthesis methods, properties, recent progress, and challenges. *Materials Advances*. 2021;2(6):1821-1871.
12. Ghosh CK. Quantum Effect on Properties of Nanomaterials. *Engineering Materials*: Springer Berlin Heidelberg; 2015. p. 73-111.
13. Abed HH, Ammar SH. Construction, characterization, and visible light-assisted photocatalytic degradation behaviour of Graphitic carbon nitride (gC_3N_4) photocatalyst.
14. Al-Saady AJ, Aldujaili NH, Banoon SR, Al-Abboodi A. Antimicrobial properties of nanoparticles in biofilms. *Revis Bionatura* 2022; 7 (4) 71.
15. Al-Abboodi A, Falih IQ, Al-Asadi M, Hussein BA, Al-Saady MAAJ, Abdullah TA, et al. Iron-oxide nanoparticles (SPIONs)

- enhance malaria vaccine antibody response. *Vaccine*. 2026;77:128353.
16. Timoshenko J, Lu D, Lin Y, Frenkel AI. Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles. *The Journal of Physical Chemistry Letters*. 2017;8(20):5091-5098.
17. Forni T, Baldoni M, Le Piane F, Mercuri F. GrapheNet: a deep learning framework for predicting the physical and electronic properties of nanographenes using images. *Sci Rep*. 2024;14(1).
18. Kaźmierczak S, Kasztelan R, Buczyński R, Mańdziuk J. Predicting optical parameters of nanostructured optical fibers using machine learning algorithms. *Engineering Applications of Artificial Intelligence*. 2024;132:107921.
19. Surakasi R, Subramani R, M J, V P, Rusho MA, Yishak S. Optimization of viscosity of propylene glycol and water (50:50)/Graphene nanofluid: A response surface methodology and machine learning approach. *Results in Engineering*. 2025;27:105692.
20. Al-darraj AR, Marzouk SA, Aljabr A, Almekhadi FA, Alqaed S, Kaood A. Enhancement of heat transfer in a vertical shell and tube heat exchanger using air injection and new baffles: Experimental and numerical approach. *Appl Therm Eng*. 2024;236:121493.
21. AlHasan L, Qi A, Al-Aboodi A, Rezk A, Shilton RR, Chan PPY, et al. Surface acoustic streaming in microfluidic system for rapid multicellular tumor spheroids generation. *SPIE Proceedings*; 2013/12/07: SPIE; 2013. p. 89235C.
22. Banoon SR, Ghasemian A. The Characters of Graphene Oxide Nanoparticles and Doxorubicin Against HCT-116 Colorectal Cancer Cells In Vitro. *J Gastrointest Cancer*. 2021;53(2):410-414.
23. Üstün R, Amjid M. Assessment of Germination and Seedling Development Factors of Soybean Cultivars in Different Salinity Levels. *Black Sea Journal of Agriculture*. 2024;7(5):477-485.
24. Sofo A, Elshafie HS, Camele I. Structural and Functional Organization of the Root System: A Comparative Study on Five Plant Species. *Plants*. 2020;9(10):1338.
25. Aziz ZS, Jazza SH, Dageem HN, Banoon SR, Balboul BA, Abdelzaher MA. Bacterial biodegradation of oil-contaminated soil for pollutant abatement contributing to achieve sustainable development goals: A comprehensive review. *Results in Engineering*. 2024;22:102083.
26. Nagy P, Kaszás B, Csabai I, Hegedűs Z, Michler J, Pethő L, et al. Machine Learning-Based Characterization of the Nanostructure in a Combinatorial Co-Cr-Fe-Ni Compositionally Complex Alloy Film. *Nanomaterials*. 2022;12(24):4407.
27. Angulo A, Yang L, Aydil ES, Modestino MA. Machine learning enhanced spectroscopic analysis: towards autonomous chemical mixture characterization for rapid process optimization. *Digital Discovery*. 2022;1(1):35-44.
28. Jawad LA, Al-Janabi MIG, Rutkayová J. Directional fluctuating asymmetry in certain morphological characters as a pollution indicator: Tigris catfish (*Silurus triostegus*) collected from the Euphrates, Tigris, and Shatt al-Arab Rivers in Iraq. *Fisheries and Aquatic Life*. 2020;28(1):18-32.
29. Hassan S, Al_Ezee AMM, Al Sulivany BSA, Hassan NE. Evaluating the Ecological Consequences of Heavy Metals Contamination on Aquatic Ecosystem Functioning in the Tigris River, Iraq. *Egyptian Journal of Aquatic Biology and Fisheries*. 2026;30(2):191-202.
30. Wang M, Wang T, Cai P, Chen X. Nanomaterials Discovery and Design through Machine Learning. *Small Methods*. 2019;3(5).
31. Margineanu MB, Julfakyan K, Sommer C, Perez JE, Contreras MF, Khashab N, et al. Semi-automated quantification of living cells with internalized nanostructures. *Journal of Nanobiotechnology*. 2016;14(1).
32. Sundarasetty H, Sahu SK. Tribological behavior of PLA reinforced with boron nitride nanoparticles using Taguchi and machine learning approaches. *Results in Engineering*. 2025;26:104772.
33. Al-Abboodi A, Tjeung R, Doran P, Yeo L, Friend J, Chan P. Microfluidic chip containing porous gradient for chemotaxis study. *SPIE Proceedings*; 2011/12/21: SPIE; 2011. p. 82041H.
34. Lobato I, Friedrich T, Van Aert S. Deep convolutional neural networks to restore single-shot electron microscopy images. *npj Computational Materials*. 2024;10(1).
35. Yuan S, Zhu Z, Lu J, Zheng F, Jiang H, Sun Q. Applying a Deep-Learning-Based Keypoint Detection in Analyzing Surface Nanostructures. *Molecules*. 2023;28(14):5387.
36. Rahman MH-U, Sikder R, Tripathi M, Zahan M, Ye T, Gnimpieba Z E, et al. Machine Learning-Assisted Raman Spectroscopy and SERS for Bacterial Pathogen Detection: Clinical, Food Safety, and Environmental Applications. *Chemosensors*. 2024;12(7):140.
37. Rotunno E, Tavabi AH, Rosi P, Frabboni S, Tiemeijer P, Dunin-Borkowski RE, et al. Alignment of electron optical beam shaping elements using a convolutional neural network. *Ultramicroscopy*. 2021;228:113338.
38. Maleki H, Asadi P, Akbari M, Peyghami M, Salarvand V, Bedir F. Incorporation of machine learning in the design of bio-nanocomposites produced by DLP 3D printer. *Results in Engineering*. 2025;27:105756.
39. Mousa AA, Hassan SA-DH, Rashid MK, Al-Saady M. Safeguarding Patient Data: Machine Learning for Phishing URL Detection in Healthcare Systems. *Journal of Advanced Research Design*. 2025;131(1):47-60.
40. Riyadi TWB, Herawan SG, Tirta A, Ee YJ, Hananto AL, Paristiawan PA, et al. Nanofluid heat transfer and machine learning: Insightful review of machine learning for nanofluid heat transfer enhancement in porous media and heat exchangers as sustainable and renewable energy solutions. *Results in Engineering*. 2024;24:103002.
41. Werdiningsih I, Puspitasari I, Hendradi R. Recognizing Daily Activities of Children with Autism Spectrum Disorder Using Convolutional Neural Network Based on Image Enhancement. *Cybernetics and Information Technologies*. 2025;25(1):78-96.
42. Bao J, Ye M. Head Pose Estimation Based on Robust Convolutional Neural Network. *Cybernetics and Information Technologies*. 2016;16(6):133-145.
43. Gyires-Tóth BP, Osváth M, Papp D, Szűcs G. Deep Learning for Plant Classification and Content-Based Image Retrieval. *Cybernetics and Information Technologies*. 2019;19(1):88-100.
44. Mohammed AJ, Al-Majidi SD, Al-Nussairi MK, Abbod MF, Al-Rawashidy HS. Design of a Load Frequency Controller based on Artificial Neural Network for Single-Area Power System. 2022 57th International Universities Power Engineering Conference (UPEC); 2022/08/30: IEEE; 2022. p. 1-5.
45. Long DT. Efficient DenseNet Model with Fusion of Channel and Spatial Attention for Facial Expression Recognition. *Cybernetics and Information Technologies*. 2024;24(1):171-189.
46. Rahmatulloh A, Haerani E, Gunawan R, Firdaus EA, Nugraha GF. ALEX: Automated Low-Light Enhancement eXpert for Intelligent Security Systems Using Vision Transformer. *Cybernetics and Information Technologies*. 2025;25(4):145-165.