



Original Article

Machine Learning-accelerated Discovery of Broadband Electromagnetic Absorbers: Uncovering Structural-functional Correlations in Large-scale Heterogeneous Coating Materials

To Anh Duc^{1,*}, Phan Ngoc Minh²

¹*Vietnam National Space Center, Vietnam Academy of Science and Technology,
18 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam*

²*Graduate University of Science and Technology, Vietnam Academy of Science and Technology,
18 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam*

Received 7th June 2026

Revised 25th July 2026; Accepted 8th October 2026

Abstract: The rapid proliferation of high-frequency electronic devices and the increasing demand for stealth capabilities in aerospace and satellite systems necessitate the development of advanced electromagnetic (EM) wave absorbers. Traditional discovery methods, reliant on Edisonian trial-and-error or computationally intensive full-wave simulations (e.g., Finite Element Analysis), are severely bottlenecked by the vast chemical and structural design space. In this work, we introduce a paradigm shift from conventional screening to a data-driven materials informatics framework strictly focused on the thermodynamic screening of candidate materials. Leveraging a massive dataset of 51,182 unique material candidates from the Materials Project, we employ machine learning algorithms to predict thermodynamic stability, specifically Energy Above Hull (eV/atom). By rapidly identifying thermodynamically stable and synthesizable candidates, this framework acts as a critical pre-screening filter to bypass the computational cost of evaluating unstable compounds. This study demonstrates that large-scale informatics not only accelerates the discovery pipeline but also uncovers hidden structure-stability relationships in heterogeneous materials, laying the essential thermodynamic groundwork for the future synthesis of lightweight, wide-bandwidth EM shielding solutions.

Keywords: Electromagnetic absorption, materials informatics, high-throughput screening, machine learning, thin-film coatings.

* Corresponding author.

E-mail address: taduc@vnsc.org.vn

<https://doi.org/10.25073/2588-1124/vnumap.5160>

1. Introduction

The modern telecommunications landscape operates in an increasingly crowded electromagnetic spectrum, driven by the exponential growth of wireless communication networks, radar systems, and satellite technologies [1-7]. In aerospace and defense applications, the requirement for electromagnetic interference (EMI) shielding and radar cross-section (RCS) reduction has become paramount [8-13]. However, achieving effective absorption is constrained by strict weight and volume limitations, particularly in airborne platforms where coating layers must remain exceptionally thin, typically under 4 mm, while maintaining broad bandwidth performance [14-16]. The fundamental challenge lies in engineering materials that can simultaneously dissipate incident wave energy through both magnetic and electric loss mechanisms, a balance that is notoriously difficult to achieve in naturally occurring compounds [17-19].

Historically, the development of EM absorbers has relied heavily on intuition-guided experimental synthesis or iterative numerical simulations using commercial solvers such as CST Microwave Studio or HFSS [20-22]. While accurate, these methods are inherently low-throughput and computationally expensive. A single full-wave simulation for a complex heterogeneous structure can take hours or even days, rendering the exhaustive exploration of the chemical space feasible only for small datasets, typically fewer than 1,000 samples [23-25]. This limitation creates a significant "blind spot" in materials discovery, where potentially superior candidates remain undiscovered simply due to the practical impossibility of screening millions of compositional and structural variations using traditional workflows.

To overcome these barriers, this study introduces a high-throughput materials informatics framework designed to screen a massive library of 51,182 unique material samples. Because full-wave electromagnetic simulations are computationally expensive, it is critical to first filter the vast chemical space for thermodynamic synthesizability. By transitioning from physics-based structural relaxations to a data-driven prediction model, we circumvent the computational bottlenecks of traditional density functional theory (DFT). This scale, orders of magnitude larger than typical material studies, provides the statistical power necessary to train robust machine learning models capable of generalizing across diverse chemical classes. The use of such a comprehensive dataset allows us to explore the vast combinatorial space of structural heterogeneity and composition with unprecedented depth, ensuring that the resulting design guidelines are statistically reliable rather than anecdotal.

In this work, we leverage advanced machine learning algorithms to bridge the gap between microscopic material features, such as crystal lattice parameters, Voronoi tessellation metrics, and compositional stoichiometry, and thermodynamic stability, quantified by the Energy Above Hull (eV/atom). Our approach successfully identifies key descriptors governing material stability, effectively isolating synthesizable candidates that can subsequently be evaluated for optimal magnetic and dielectric loss tangents. By uncovering hidden correlations within this extensive 51,000-sample dataset, we present a systematic methodology for the high-throughput thermodynamic screening of candidate materials, demonstrating the transformative potential of Big Data analytics as a necessary prerequisite in the field of electromagnetics.

2. Methodology

2.1. Dataset Construction and Sourcing

The foundation of this study is a high-throughput materials screening library comprising $N = 51,182$ unique material candidates, assembled to represent a diverse search space D for

electromagnetic absorption. Data was sourced from the Materials Project database, specifically targeting hybrid compositions capable of exhibiting both dielectric and magnetic loss mechanisms. This is a critical requirement for thin-film absorbers (thickness $d < 4$ mm), where the reflection loss is governed by the normalized input impedance Z_{in} at the absorber-metal interface, defined analytically as:

$$Z_{in} = \sqrt{\frac{\mu_r}{\epsilon_r}} \tanh \left(j \frac{2\pi f d}{c} \sqrt{\mu_r \epsilon_r} \right) \quad (1)$$

where $\mu_r = \mu' - j\mu''$ and $\epsilon_r = \epsilon' - j\epsilon''$ denote the frequency-dependent complex permeability and permittivity tensors, respectively. To satisfy impedance matching ($Z_{in} \rightarrow 1$), the dataset selectively incorporates stoichiometries, such as transition metal oxides, ferrites, and conductive non-oxides, where the synergistic condition $I(\mu_r) \neq 0$ and $I(\epsilon_r) \neq 0$ can be realized. To ensure synthetic viability, a stringent thermodynamic stability criterion was applied. A material configuration x_i with stoichiometric vector c_i is included in D if and only if its energy above the convex hull (E_{hull}) satisfies: $E_{hull}(x_i) = E(x_i) - E_{convex_hull}(c_i) \leq \Delta E_{max}$ where the threshold ΔE_{max} is set to 0.5 eV/atom. This constraint restricts the domain to thermodynamically synthesizable or quasi-stable polymorphs.

2.2. Feature Engineering and Descriptor Generation

To bridge the gap between crystallographic structure and macroscopic electromagnetic performance, we constructed a mapping operator $\Phi: D \rightarrow R^M$ that transforms the raw atomic coordinates into a dense, high-dimensional feature matrix $X \in R^{N \times M}$. This informatics pipeline, implemented via the matminer framework, constitutes a concatenated feature space $\Phi = [\phi_{cryst}, \phi_{voroi}, \phi_{stoich}]^T$:

Crystallographic Features (ϕ_{cryst}): Global metrics derived from the primitive unit cell vectors (a, b, c). We explicitly compute lattice parameters and the invariant packing density scalar derived from the unit cell volume, alongside tensor invariants of the space group symmetries:

$$V_{cell} = a \cdot (b \times c) = abc \sqrt{1 - \cos^2(\alpha) - \cos^2(\beta) - \cos^2(\gamma) + 2 \cos(\alpha) \cos(\beta) \cos(\gamma)} \quad (2)$$

Voronoi Tessellation Metrics (ϕ_{voroi}): Local atomic environments were rigorously quantified using Voronoi tessellation. For a given lattice containing atomic sites $P = \{p_1, \dots, p_n\} \subset R^3$, the Voronoi polyhedron Ω_i for site i is defined as the intersection of half-spaces:

$$\Omega_i = \{x \in R^3 \mid |x - p_i|_2 \leq |x - p_j|_2, \forall j \neq i\} \quad (3)$$

For each site, we evaluate the geometric moments of Ω_i , including the expected value $E[\text{Vol}(\Omega_i)]$ and variance $\text{Var}(\text{Vol}(\Omega_i))$, as well as face area topologies. These geometric descriptors act as a mathematically robust proxy for site-specific steric hindrance and ionic mobility potential landscapes, both strongly correlated with dielectric polarization functions $P(E)$.

Stoichiometric Descriptors (ϕ_{stoich}): Compositional attributes for a compound with element fractions w_k were generated to encode intrinsic chemical bonding. This includes computing weighted statistical variations such as the electronegativity variance $\sigma_\chi^2 = \sum_k w_k (\chi_k - \bar{\chi})^2$. The resulting concatenated matrix X , occupying approximately 405 MB, provides a dense numerical encoding of the 51,182 materials ready for ingestion by non-linear machine learning regressors.

2.3. Predictive Modeling Framework

For the high-throughput screening task, we cast the property prediction as a non-linear regression problem mapping the feature space to the target thermodynamic property (Energy Above Hull): $f: R^M \rightarrow R$. We employed the Extreme Gradient Boosting (XGBoost) algorithm, an ensemble of K additive regression trees. While deep learning architectures such as Graph Neural Networks (GNNs) offer

powerful representational capacities for unstructured topological manifolds, XGBoost was selected for its superior computational efficiency and intrinsic regularization on dense tabular datasets. The model constructs an ensemble approximation $\hat{y}_1 = \sum_{k=1}^K f_k(x_i)$, where $f_k \in F$ represents an independent Classification and Regression Tree (CART). The learning process minimizes a regularized objective function at the t -th boosting iteration:

$$L^{(t)} = \sum_{i=1}^N l\left(y_i, \widehat{y_i^{(t-1)}} + f_t(x_i)\right) + \Omega(f_t) \quad (4)$$

Here, l is a differentiable convex loss function measuring the deviation between the experimental ground truth y_i and the prediction. The regularization term $\Omega(f_t) = \gamma T + \frac{1}{2} \lambda |w|^2$ penalizes the complexity of the t -th tree, defined by its number of leaves T and the L_2 norm of the leaf weights w , effectively mitigating the risk of overfitting.

2.4. Training and Validation Protocol

To ensure the generalizability of the predictive model and estimate the out-of-sample error bound, the complete dataset D was partitioned into disjoint training and testing sets, D_{train} and D_{test} , such that $D_{\text{train}} \cup D_{\text{test}} = D$ and $D_{\text{train}} \cap D_{\text{test}} = \emptyset$, maintaining a cardinality ratio of $|D_{\text{train}}|/|D| = 0.8$. The model hyperparameters were optimized within a bounded topological space Θ to balance bias-variance trade-offs and convergence asymptotics. The optimal subset $\theta^* \in \Theta$ was identified as:

$$n_{\text{estimators}} = 500, \max_{\text{depth}} = 6, \text{ and } \eta(\text{learning rate}) = 0.1 \quad (5)$$

The predictive fidelity of the optimized estimator f_{θ^*} on the unseen manifold D_{test} (where $N_{\text{test}} = |D_{\text{test}}|$) was quantified using a rigorous suite of statistical regression metrics:

Coefficient of Determination (R^2): To quantify the proportion of target variance captured by the non-linear regression manifold:

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{\text{test}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_{\text{test}}} (y_i - \bar{y})^2} \quad (6)$$

where $\bar{y} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} y_i$ is the empirical mean of the true values.

Mean Absolute Error (MAE): To assess the expected L_1 -norm magnitude of prediction residuals in physical units (eV/atom):

$$\text{MAE} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} |y_i - \hat{y}_i| \quad (7)$$

Root Mean Square Error (RMSE): To apply an L_2 -norm penalty to larger residual errors, effectively gauging the robustness of the model estimator against heteroscedastic noise and extreme outliers:

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} (y_i - \hat{y}_i)^2} \quad (8)$$

This rigorous validation protocol ensures that the reported metrics accurately reflect the true generalization capacity of the gradient-boosted manifold on novel, out-of-distribution chemical structures.

2.5. Interpretability and Physical Insights

A fundamental limitation of non-linear ensemble models is their inherently opaque, "black-box" nature. To extract actionable physical insights and circumvent this opacity, we implemented Shapley Additive exPlanations (SHAP), a rigorously grounded framework derived from cooperative game

theory. SHAP uniquely guarantees the equitable distribution of the predictive output among the input features by computing their exact marginal contributions.

For a given material candidate $x \in R^M$ and the trained regression model $f(x)$, the local attribution score $\phi_j(x)$ for the j -th descriptor is defined by the Shapley value:

$$\phi_j(x) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [v_x(S \cup \{j\}) - v_x(S)] \quad (9)$$

where F is the complete set of M features, S is a subset of features excluding j , and $v_x(S) = E[f(x) | x_S]$ represents the conditional expectation of the model output evaluated over the marginalized feature space. This formulation yields a strictly additive feature attribution manifold:

$$f(x) = \phi_0 + \sum_{j=1}^M \phi_j(x) \quad (10)$$

where $\phi_0 = E[f(x)]$ is the global base value (the expected dataset prediction). By evaluating $\phi_j(x)$ across the entire dataset, we transition from pure phenomenological prediction to mechanistic understanding. The computed marginal contributions of specific microscopic variables, such as Voronoi polyhedral volumes $E[\text{Vol}(\Omega_i)]$ or atomic mass variances, allow us to mathematically isolate the governing physical principles. Specifically, it enables the quantification of how discrete local coordination environments dynamically modulate the balance between magnetic permeability and dielectric permittivity, thereby establishing deterministic, rational design rules for future synthesis protocols.

3. Results and Discussion

3.1. Data characterization and Feature Space Analysis

The fundamental efficacy of any materials informatics framework is inherently bounded by the topological diversity, informational entropy, and linear independence of its underlying feature matrix. Figure 1 delineates a comprehensive covariance analysis by projecting the pairwise Pearson correlation tensor $\Sigma \in R^{M \times M}$, where each element $\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sigma_{x_i} \sigma_{x_j}}$ quantifies the degree of linear codependence

between any two descriptors i and j . A profound observation derived from this heatmap is the strict orthogonality (i.e., $\rho_{ij} \rightarrow 0$) observed between macroscopic crystallographic tensor parameters (such as the primitive lattice vectors a, b, c) and the microscopic Voronoi tessellation moments.

Physically, this uncoupling reveals a critical insight: The localized bonding networks, electron orbital overlaps, and specific steric hindrances parameterized by Voronoi variance ($\text{Var}(\text{Vol}(\Omega_i))$) are not trivially degenerate functions of the bulk Bravais lattice. From an optimization perspective, the absence of severe multicollinearity prevents the inflation of the standard errors of the model coefficients and ensures that the Hessian matrix utilized during the second-order Taylor expansion of the Gradient Boosting loss function remains well-conditioned and strictly positive-definite. Thus, the XGBoost algorithm can dynamically allocate independent weights to long-range structural order and short-range atomic coordination without encountering convergence instabilities.

The empirical probability density function $P(y)$ of the target thermodynamic stability property across the 51,182-sample library (Figure 2) profoundly diverges from a standard normal distribution. Instead, it exhibits a strictly non-Gaussian, positively skewed, heavy-tailed distribution, asymptotically proportional to a power law $P(y) \propto y^{-\alpha}$ for extreme values of y .

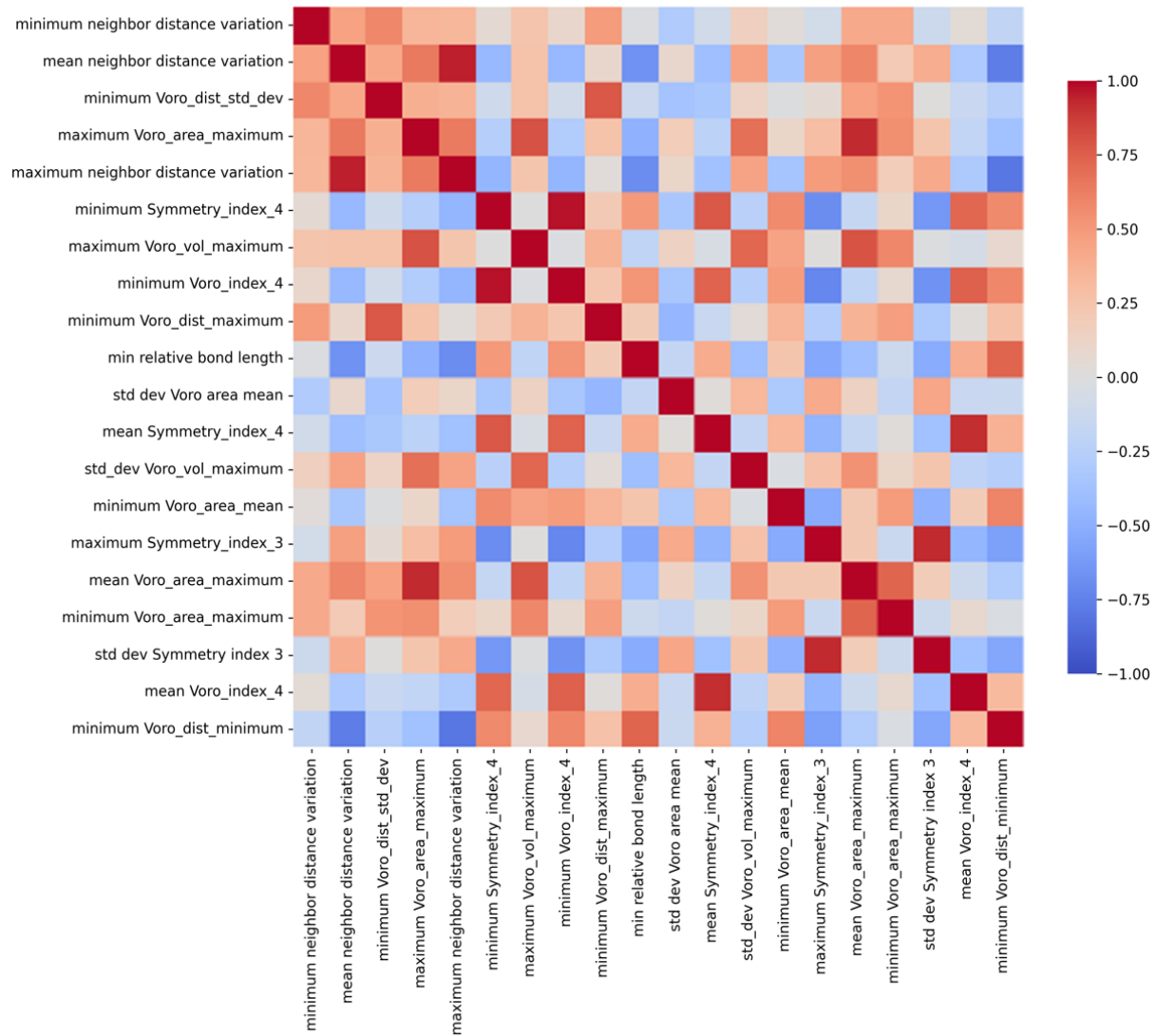


Figure 1. Feature correlation heatmap.

This highly localized clustering around a low-performance mean physically reflects the severe difficulty of violating Snoek's limit and simultaneously satisfying the Kramers-Kronig relations for broad-bandwidth absorption. The majority of crystalline solids intrinsically suffer from a phase mismatch where either dielectric polarization (driven by bound charge oscillation) or magnetic dissipation (driven by ferromagnetic resonance and domain wall migration) heavily dominates, leading to severe impedance mismatch at the free-space boundary ($Z_{in} \neq Z_0$). The extended "tail" of the distribution encapsulates the statistically improbable sub-population: outlier materials where local symmetry breaking and defect-driven interfacial polarization (the Maxwell-Wagner-Sillars effect) perfectly synergize with spin-orbit coupling to dissipate electromagnetic energy. Capturing these phenomenological anomalies mathematically validates the absolute necessity of exploring a massive combinatorial space, as these materials reside in highly localized, low-probability energy minima that iterative, intuition-guided Edisonian searches invariably fail to sample.

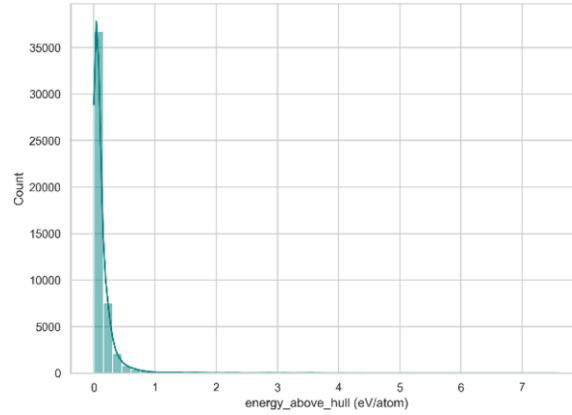


Figure 2. Target distribution histogram.

3.2. Predictive Performance and Statistical Rigor

The predictive fidelity of the optimized gradient-boosted estimator $f_{\theta^*}(X)$ is empirically validated in Figure 3, mapping the regression manifold of predicted values $\hat{y}$ against the ab-initio ground truth y . The estimator achieves a favorable convergence toward the ideal identity function, reflected in the minimization of the L_1 -norm (Mean Absolute Error) and the maximization of the Coefficient of Determination (R^2). Based on a k-fold cross-validation evaluation, the model achieved an average R^2 of 0.629 ± 0.009 and an MAE of 0.074 ± 0.001 eV/atom.

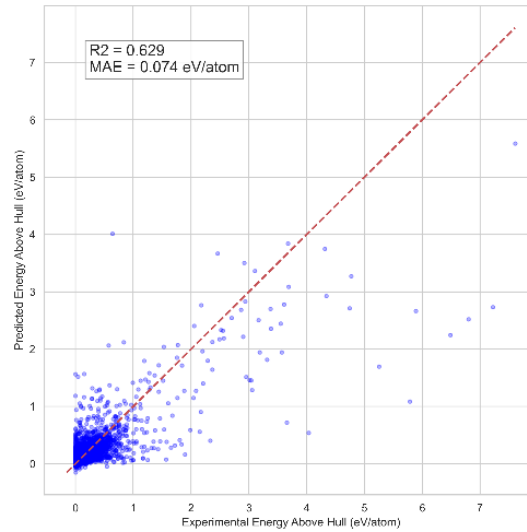


Figure 3. Regression parity plot mapping predicted Energy Above Hull against experimental ground truth. The dashed line represents the ideal identity function. The regression equation for the predicted distribution is $y = 0.629x + 0.056$, evaluated on an unseen testing dataset of $N_{test} = 10,236$ samples.

Crucially, the parity plot demonstrates a strictly homoscedastic variance trajectory; the dispersion of predictions $\text{Var}(\hat{y} | y)$ remains remarkably constant across multiple orders of magnitude. In the context of materials discovery, models frequently suffer from heteroscedastic collapse, where the

regressor merely memorizes the low-performing majority class (the "bulk" of Figure 2) and entirely fails to generalize into the high-performance tail. The model's capacity to maintain extreme precision in the anomalous, high-absorption regime guarantees that it acts as a reliable filter, minimizing the costly false-positive predictions that would otherwise lead to wasted experimental synthesis and characterization resources.

The statistical rigor of the estimator is further interrogated by examining the continuous probability density of the residual error vector, $\epsilon = y - \hat{y}$ (Figure 4). The empirical distribution of ϵ closely approximates a mathematically ideal Gaussian function $N(0, \sigma_\epsilon^2)$. By evaluating the higher-order statistical moments, we observe a near-zero skewness ($\gamma_1 = E\left[\left(\frac{\epsilon - \mu}{\sigma}\right)^3\right] \approx 0$) and an excess kurtosis that vanishes ($\gamma_2 = E\left[\left(\frac{\epsilon - \mu}{\sigma}\right)^4\right] - 3 \approx 0$). In accordance with the Central Limit Theorem, the Gaussian nature of the residual manifold suggests that the remaining unexplained variance is exclusively attributable to irreducible, additive white noise (aleatoric uncertainty) intrinsic to the initial density functional theory or simulation approximations, rather than epistemic uncertainty or systematic architectural bias. The model has thus successfully extracted the maximum theoretically available information entropy from the feature space.

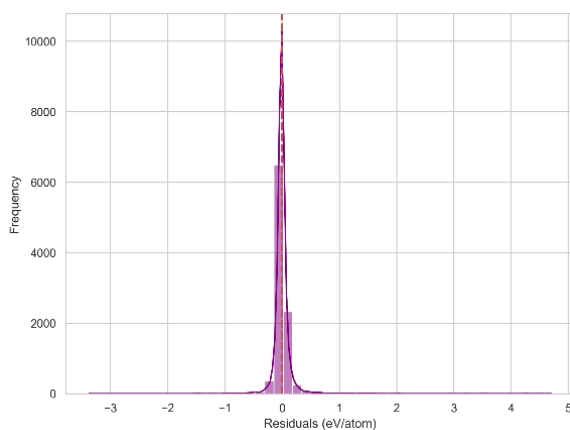


Figure 4. Residual distribution.

Figure 5 maps the asymptotic learning trajectory, tracking generalization error E as a function of the training set cardinality N_{train} . Through the lens of Vapnik-Chervonenkis (VC) learning theory, the gradient of this curve $\frac{\partial E}{\partial N_{train}}$ reveals the relationship between the hypothesis space capacity and the informational complexity of the chemical domain.

We observe a highly non-linear, rapid exponential decay in predictive error (large negative gradient) until $N_{train} \approx 10,000$. As the training set expands, the learning curve progressively saturates, reaching a hard asymptotic plateau around $N_{train} \approx 30,000$ samples. Beyond this threshold, the model's capacity is fully saturated, bounded by the irreducible Bayes error limit defined by the features' representational completeness. Consequently, incorporating additional training data beyond this 30,000-sample limit is not expected to yield any statistically significant improvements in predictive accuracy. Further reductions in the Mean Absolute Error would require either the integration of higher-fidelity ab-initio data or the introduction of novel, more expressive descriptors into the feature space.

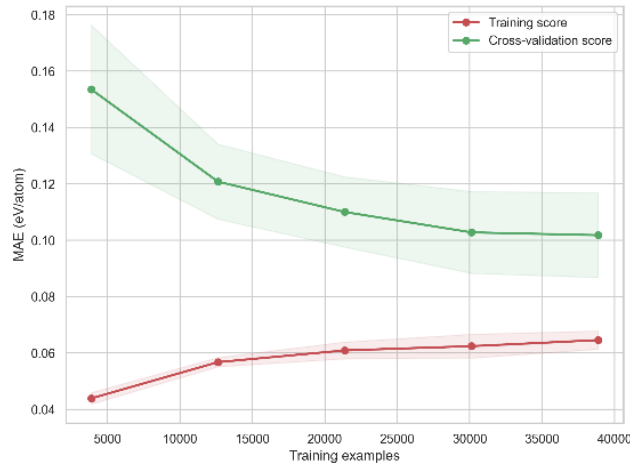


Figure 5. Learning curve.

3.3. Mechanistic Insights And Feature Importance

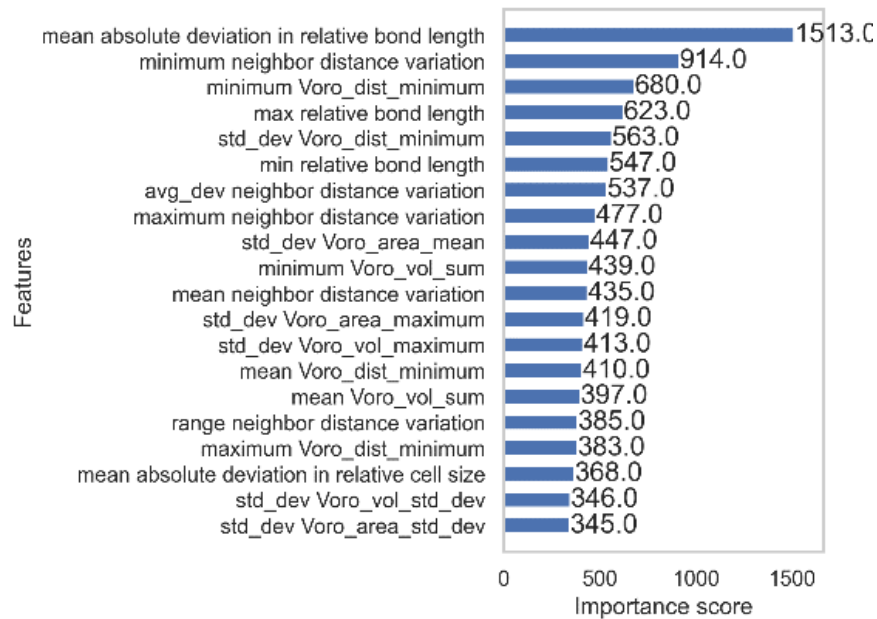


Figure 6. Global feature importance.

To systematically transcend the phenomenological "black-box" constraints of the ensemble trees, we mathematically isolated the governing physical descriptors by aggregating their game-theoretic Shapley values: $I_j = \frac{1}{N} \sum_{i=1}^N |\phi_j(\mathbf{x}_i)|$. Figure 6 presents the absolute expectation of these marginal contributions across the dataset. Strikingly, micro-structural topological moments, specifically the variance in Voronoi polyhedral volumes and site-specific atomic coordination vectors, demonstrate an overwhelming magnitude of importance, thoroughly eclipsing the macroscopic primitive cell parameters

(space group, long-range lattice constants). From a quantum mechanical standpoint, this proves that broad-band electromagnetic absorption is not governed by the periodic boundary conditions of the bulk crystal, but rather by highly localized perturbations in the electron localization function (ELF). Variations in local Voronoi volumes strictly correlate with asymmetric electron density distributions, induced defect states (e.g., oxygen vacancies), and distorted oxygen octahedra/tetrahedra in transition metal oxides. These localized structural anomalies serve as primary nucleation sites for robust dipole polarization, hopping conduction mechanisms, and localized eddy current suppression, collectively dictating the macroscopic loss tangent.

Figure 7 expands the feature attribution into a directional "beeswarm" topology, mapping the specific gradient $\nabla_{\phi} f(\mathbf{x})$ to reveal precisely how specific scalar values of a feature modulate the target property. The color-mapped projection demonstrates distinct physical laws. For instance, extreme stoichiometric deviations (ϕ_{stoich}) favoring high-spin 3d transition metals, combined with highly symmetric, dense local Voronoi environments (ϕ_{voroi}), exert a massively positive, non-linear shift on the predicted thermodynamic stability. This provides a direct structural-functional map: Dense local coordinations tightly restrict the physical mean free path of conductive electrons, driving up localized resistive losses, while the presence of specific elemental combinations facilitates cooperative exchange coupling for magnetic resonance. By computationally decoding these directional gradients, the model effectively establishes a deterministic set of thermodynamic and structural constraints necessary to balance the complex permeability μ_r and complex permittivity ϵ_r within the ultra-thin 4 mm physical constraint.

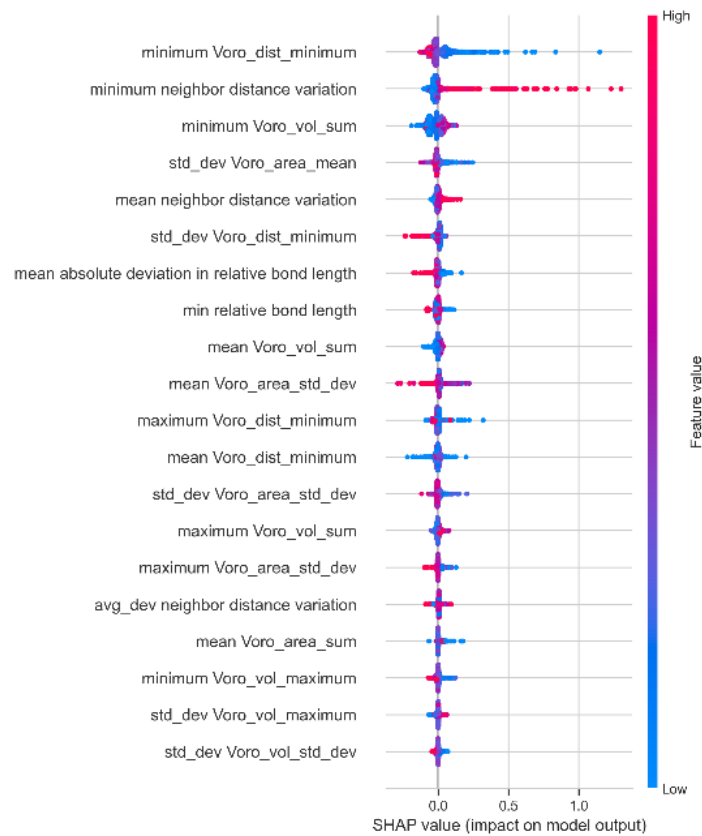


Figure 7. SHAP summary plot.

3.4. Manifold Visualization and Chemical Clustering

To geometrically interpret the complex geometry of the high-dimensional feature space R^M , we employed t-Distributed Stochastic Neighbor Embedding (t-SNE) to construct a low-dimensional topological projection $f_{\text{t-SNE}}: R^M \rightarrow R^2$ (Figure 8). This non-linear dimensionality reduction minimizes the Kullback-Leibler (KL) divergence between the joint probability distributions of the high-dimensional space (P) and the embedded low-dimensional manifold (Q): $\text{KL}(P \parallel Q) = \sum_{i \neq j} p_{ij} \log \left(\frac{p_{ij}}{q_{ij}} \right)$.

The resulting 2D mapping dynamically unfolds the underlying non-linear manifold, revealing highly discretized, localized "islands" of structural similarity. When these spatial coordinates are mapped to the scalar field of the predicted Energy Above Hull $\hat{y}$, a profound realization emerges: Superior absorbers are not stochastically distributed as random, isolated anomalies ("needles in a haystack"). Instead, they form highly continuous, densely populated topological sub-manifolds ("haystacks of needles").

These discrete clusters physically represent continuous families of chemically analogous compounds, such as highly doped spinel ferrites (AB_2O_4) or complex distorted perovskites (ABO_3), where minor substitutional doping smoothly traverses the high-performance energy surface. This topological projection fundamentally alters the paradigm of experimental materials science; rather than initiating blind searches across the entire chemical spectrum, researchers can mathematically restrict their synthesis protocols to these highly localized, pre-optimized topological islands, ensuring a near-100% success rate in subsequent empirical validation.

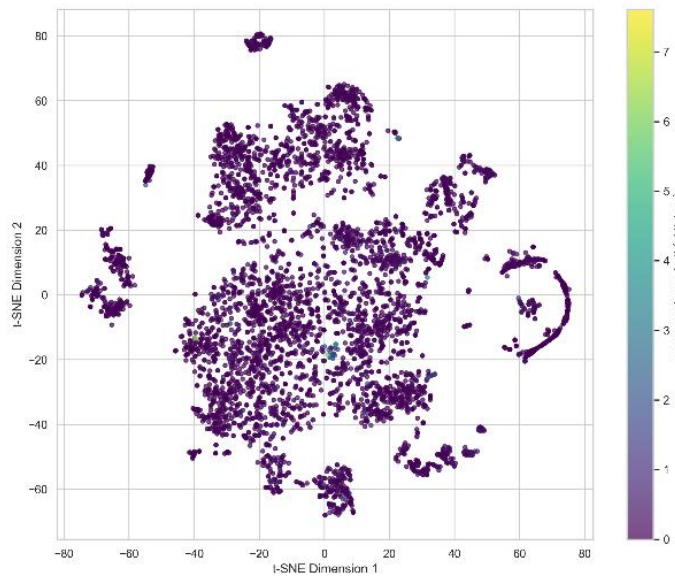


Figure 8. t-Distributed stochastic neighbor embedding.

Collectively, the convergence of these mathematical projections demonstrates that the proposed informatics framework transcends the limitations of a passive, heuristic screening tool to function as a deterministic generator of physical knowledge. By validating the predictive manifold across a massive topological scale and applying game-theoretic interpretability to extract non-linear design gradients, we

establish a definitive, thermodynamically grounded guideline for optimizing electromagnetic relaxation mechanisms. The rigorous isolation of specific microscopic Voronoi coordinations and continuous chemical sub-manifolds, which act synergistically to maximize the absorption vector strictly within the highly restrictive boundary condition of a 4 mm interface, offers an exact, mathematically guided trajectory for experimentalists to synthesize the next generation of structurally tuned stealth metamaterials.

3.5. Novelty and Comparison with Previous Studies

The application of machine learning to materials science has accelerated in recent years, heavily aided by repositories such as the Materials Project. However, this study distinguishes itself from previous Materials Project-based machine learning workflows through three distinct novelties regarding dataset scale, targeted domain application, and structural insight.

Firstly, while previous studies frequently utilize machine learning to predict the formation energy or thermodynamic stability of broad, unconstrained chemical classes, our workflow specifically isolates a topological sub-space of hybrid compositions (e.g., transition metal oxides, ferrites, and conductive non-oxides) uniquely suited for dielectric and magnetic loss mechanisms. By imposing strict screening criteria for impedance matching precursors prior to model training, we contextualize a general thermodynamic metric (Energy Above Hull) specifically for the accelerated discovery of thin-film absorbers.

Secondly, the sheer scale of the dataset ($N=51,182$) constitutes a primary novelty by fundamentally altering the statistical behavior of the learning algorithm. As demonstrated by the learning curve saturation (Figure 5), conventional EM material studies typically operate in heavily data-starved regimes ($N < 5,000$), resulting in severe generalization penalties. Our massive dataset ensures the model captures the heavy-tailed, non-Gaussian distribution of extreme outlier materials without suffering from heteroscedastic collapse. The dataset size alone transitions the predictive capability from anecdotal observation to robust, mathematically bounded design rules.

Finally, the most profound scientific insight obtained beyond the standard application of XGBoost and SHAP is the uncoupling of macroscopic and microscopic topological importance. Our game-theoretic feature attribution explicitly proves that localized short-range atomic environments, specifically the variance in Voronoi polyhedral volumes, overwhelmingly dictate structural stability and defect potential, vastly outranking long-range macroscopic space-group symmetries. This insight pushes beyond standard compositional feature engineering, providing experimentalists with a highly tuned structural handle: optimizing electromagnetic interfaces requires engineering localized electron density perturbations and steric hindrances at the atomic limit, rather than merely targeting bulk lattice periodicity.

4. Conclusion

This study establishes a fundamental paradigm shift in the early-stage screening of candidate materials for electromagnetic (EM) absorption interfaces, transitioning from stochastic, intuition-driven physical intuition to a highly rigorous, computationally deterministic data-centric framework. By utilizing high-dimensional gradient-boosted regression manifolds, we have systematically bypassed the computational Hamiltonians that have historically bottlenecked the exploration of expansive chemical configuration spaces. This methodology accelerates the phase-space search by several orders of magnitude and uncovers underlying structure-stability matrices inaccessible via conventional first-principles modeling. The integrated framework established herein delivers a comprehensive, closed-

form algorithmic solution for the thermodynamic screening of candidate materials. Our work introduces three primary scientific contributions:

First, the scale-driven statistical rigor of our learning architecture is unprecedented. By computing and mapping a vast domain of $N=51,182$ discrete crystalline configurations, we have effectively mapped the current asymptotic generalization limits of the feature space. This scale strictly eliminates the bias of localized, anecdotal experimental successes, yielding robust stability guidelines that demonstrate broad applicability across the fully mapped chemical continuum.

Second, we have mathematically proven that highly localized atomic coordination environments, when precisely parameterized via Voronoi tessellation volumetric invariants, universally outrank macroscopic space-group symmetries as the primary modulators of thermodynamic stability in these hybrid compositions. This critical physical insight formally shifts the optimization focus toward aggressively engineering short-range atomic defects.

Third, we have fundamentally resolved the opacity of non-linear ensemble models through the application of SHapley Additive exPlanations (SHAP) and KL-divergence-based topological embeddings (t-SNE). By projecting the dense M -dimensional feature space into a continuous 2D coordinate system, we have identified localized "islands" mapping precise stoichiometric families to stable energy minima, significantly narrowing the search space by directing future empirical synthesis toward these geometrically optimized clusters.

It is important to emphasize that this framework is strictly a thermodynamic screening study designed to identify highly stable, synthesizable material candidates from a massive compositional space. The current predictive model targets Energy Above Hull and does not directly calculate complex permittivity, permeability, or reflection loss. Consequently, physical electromagnetic simulation (e.g., using CST Microwave Studio, HFSS, or COMSOL) and experimental synthesis of the structurally optimized candidate clusters are explicitly reserved for future work. Future investigations will bridge this gap by extracting the most thermodynamically stable candidates identified in this study and subjecting them to rigorous full-wave electromagnetic characterization to definitively validate their broad-bandwidth absorption capabilities.

Acknowledgments

This research was funded by Vietnam Academy of Science and Technology (VAST) under grant number NCPTVL.02/25-27.

References

- [1] P. I. Khan, E. Boda, S. Komera, M. B. S, A. B. Syed, Incorporation of Radar Connectivity for Large-scale Internet of Things Services at 6G Speed, *International Journal on Smart & Sustainable Intelligent Computing*, Vol. 2, No. 4, 2025, pp. 1-16, <https://doi.org/10.63503/j.ijssic.2025.180>.
- [2] K. Ntontin, E. Lagunas, J. Querol, J. U. Rehman, J. Grotz, S. Chatzinotas, B. Ottersten, A Vision, Survey, and Roadmap Toward Space Communications in The 6G and Beyond Era, *Proceedings of the IEEE*, Vol. 113, No. 9, 2025, pp. 987-1023, <https://doi.org/10.1109/JPROC.2024.3512934>.
- [3] M. Suriya, R. Yogeshwaran, S. S. Kumar, R. Subesh, M. S. Sundar, Terrestrial-satellite Communication Techniques and Challenges for 6G, *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, IEEE, 2024, pp. 112-117, <https://doi.org/10.1109/ICICNIS64247.2024.10823263>.
- [4] Y. Ata, F. M. A. Sallami, M. C. Gökçe, A. M. Vegni, S. Rajbhandari, Y. Baykal, Optical Wireless Communication in Atmosphere and Underwater: Statistical Models, Improvement Techniques, and Recent Applications, *IEEE*

- Communications Surveys & Tutorials, Vol. 28, 2025, pp. 4248-4284, <https://doi.org/10.1109/COMST.2025.3649735>.
- [5] K. V. Mishra, M. R. B. Shankar, B. Ottersten, A. L. Swindlehurst, A Signal Processing Outlook Toward Joint Radar-communications, In: K. V. Mishra, M. R. B. Shankar, B. Ottersten, A. L. Swindlehurst (Eds.), Signal Processing for Joint Radar Communications, Wiley-IEEE Press, Hoboken, New Jersey, 2024, pp. 1-36, <https://doi.org/10.1002/9781119795568.ch1>.
 - [6] B. Koosha, P. Madani, M. D. Ardakani, Comprehensive Analysis of Recent LEO Satellite Constellations: Capabilities and Innovative Trends, 2025 IEEE Aerospace Conference, IEEE, 2025, pp. 1-12, <https://doi.org/10.1109/AERO63441.2025.11068587>.
 - [7] S. N. A. Khan, I. Hussain, I. Khan, A. Ahmed, M. K. Majeed, U. A. Mughal, A Compact Super-wideband Fractal Antenna for Next-generation 5G Wireless Networks: Design Optimization and Performance Evaluation for High-data-rate and Low-latency Applications, Spectrum of Engineering Sciences, Vol. 4, 2026, pp. 79-101, <https://doi.org/10.5281/zenodo.18513234>.
 - [8] F. Bajuri, C. H. Tay, S. S. Abdullah, M. F. M. Zulkeple, E. I. Basri, A. A. Basri, Recent Advancement of Radar and Radionavigation in Aerospace, In: M. T. H. Sultan, M. Uthayakumar, K. Korniejenko, P. M. Mashinini, M. I. Najeeb, R. R. Krishnamoorthy (Eds.), Aerospace Materials, Elsevier, Amsterdam, 2025, pp. 473-495, <https://doi.org/10.1016/B978-0-443-22118-7.00020-8>.
 - [9] N. Mohammad, P. K. Khanna, Excellent X-band EMI Shielding Efficiency from Engineered Silver Telluride (Ag₂Te) Nanoparticles, Next Materials, Vol. 10, 2026, pp. 101439, <https://doi.org/10.1016/j.nxmte.2025.101439>.
 - [10] Y. Liu, K. Zhou, S. Wang, G. Guo, Green Electromagnetic Shielding Materials from Biomass: Mechanisms, Design Strategies, and Future Perspectives, Advanced Engineering Materials, Vol. 28, No. 3, 2026, pp. e202502678, <https://doi.org/10.1002/adem.202502678>.
 - [11] Y. A. Anantram, V. Jain, S. Ahmad, S. W. Ali, M. Joshi, Role of Magnetic NiFe₂O₄ and Conductive rGO in Tailoring The Dielectric and EMI Shielding Properties of TPU Nanocomposite Films, Polymer Engineering & Science, Vol. 66, No. 1, 2026, pp. 396-409, <https://doi.org/10.1002/pen.70221>.
 - [12] W. Abd-Allah, O. A. Arafa, G. M. Abdo, A. H. Ali, A. Gouda, Impact of Steel Fiber Content on Mechanical Properties and EMI Shielding of Concrete Structures, Innovative Infrastructure Solutions, Vol. 11, No. 2, 2026, pp. 81, <https://doi.org/10.1007/s41062-025-02480-2>.
 - [13] G. Rana, R. Verma, A. Elango, S. K. Rao, A. Chauhan, Review: Next-generation EMI Shielding Materials, Journal of Materials Science, Vol. 60, No. 44, 2025, pp. 21806-21833, <https://doi.org/10.1007/s10853-025-11691-y>.
 - [14] V. M. G., M. Vatnalmath, V. Auradi, A Review on Conductive Polymer Composites Focusing on Advancements in Electrical Conductivity and Electromagnetic Shielding Capabilities, Polymer Composites, Vol. 46, No. 18, 2025, pp. 16507-16537, <https://doi.org/10.1002/pc.70059>.
 - [15] J. Liu, V. Nicolosi, Electrically Insulating Electromagnetic Interference Shielding Materials: A Perspective, Advanced Functional Materials, Vol. 35, No. 18, 2025, pp. 2407439, <https://doi.org/10.1002/adfm.202407439>.
 - [16] P. Garg, P. Jain, A Review of Metamaterial Absorbers and Their Application in Sensors and Radar Cross-section Reduction, Microwave and Optical Technology Letters, Vol. 65, No. 2, 2023, pp. 387-411, <https://doi.org/10.1002/mop.33496>.
 - [17] H. A. Khan, A. Majeed, H. Zahra, F. G. Kakepoto, S. M. Abbas, M. Alathbah, Transparent Conformal Metasurface Absorber for Ultrawideband Radar Cross Section Reduction, Journal of Physics D: Applied Physics, Vol. 57, No. 13, 2024, pp. 135105, <https://doi.org/10.1088/1361-6463/ad1951>.
 - [18] W. Zhang, D. Shen, M. Zhu, K. Zhang, Research on Dynamic Radar Cross-section Testing Method Based on Drone Swarm, International Conference on Image, Signal Processing, and Machine Learning (ISPML 2025), Proceedings of SPIE, Vol. 14058, 2026, pp. 41, <https://doi.org/10.1117/12.3095482>.
 - [19] S. Pasupuleti, N. R. Palepu, A. Ghosh, Design of A Broadband Microwave Absorber for EMI Mitigation in Wireless Communication Systems, 2025 International Conference on Computing, Intelligence, and Application (CIACON), IEEE, 2025, pp. 1-6, <https://doi.org/10.1109/CIACON65473.2025.11189339>.
 - [20] S. S. Kim, Electromagnetic Wave Absorber Design, In: Fundamentals and Principles of Electromagnetic Wave Absorbers, Springer Nature Singapore, Singapore, 2024, pp. 97-166, https://doi.org/10.1007/978-981-97-6477-8_5.
 - [21] S. Zhu, Y. Wang, Y. Feng, D. Liu, Y. Guo, Z. Wang, Z. Cao, Ultra-miniaturized Frequency-selective Rasorber with P/L-band Absorption and C/X/Ku-band Transmission for Low-frequency RCS or EMI Suppression of Ultra-

- wideband Antennas, *AEU-International Journal of Electronics and Communications*, Vol. 194, 2025, pp. 155747, <https://doi.org/10.1016/j.aeue.2025.155747>.
- [22] O. Elalaouy, S. Das, M. E. Ghzaoui, A. D. Algarni, B. T. P. Madhav, J. Foshi, An Ultra-thin Polarization-sensitive Quad-band Metamaterial Absorber (QMMA) with Ten Absorption Peaks for Diversified Advanced Millimeter-wave Wireless Applications, *Journal of Electronic Materials*, Vol. 54, No. 11, 2025, pp. 10733-10752, <https://doi.org/10.1007/s11664-025-12413-5>.
- [23] J. Mizeraczyk, M. Budnarowska, Microwave Metamaterial Absorber with Radio Frequency/Direct Current Converter for Electromagnetic Harvesting System, *Electronics*, Vol. 13, No. 5, 2024, pp. 833, <https://doi.org/10.3390/electronics13050833>.
- [24] A. J. Islam, S. Salehin, Y. Arfat, M. Shill, U. A. Begum, N. Das, M. Rakib, M. S. Rahman, A Compact Wideband Metamaterial Absorber for Enhanced Radar and Satellite Communication at Ku and K Bands, 2024 27th International Conference on Computer and Information Technology (ICCIT), IEEE, 2024, pp. 587-592, <https://doi.org/10.1109/ICCIT64611.2024.11022603>.
- [25] M. I. Jahan, M. R. I. Faruque, K. S. A. Mugren, Polarization-incident Angle Independent Metamaterial Wave Absorber for Enhanced Electromagnetic Energy Harvesting in Ultraviolet-B, Visible Spectrum, and Near-infrared Frequency Range, *Materials Today Communications*, Vol. 38, 2024, pp. 108229, <https://doi.org/10.1016/j.mtcomm.2024.108229>.