Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/200219

Hybrid CNN-Transformer Deep Learning Architecture for Terahertz Metamaterial Absorption Prediction

Hamza A. Mashagba[1], *Hamza Abu Owida*[2], *Suhaila Abuowaida*[3], *Suleiman Ibrahim Mohammad* [4], *Azlan B. Abd Aziz*[1,*], *Manal Mizher*[5], *Asokan Vasudevan*[6], *and Mardeni Bin Roslee*[7]

[1]Centre for Wireless Technology (CWT), Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia
[2]Medical Engineering Department, Faculty of Engineering, Al-Ahliyya Amman University, Amman, Jordan
[3]Department of Data Science and AI, Faculty of Prince Al-Hussein Bin Abdallah II for IT, Al al-Bayt University, Mafraq, Jordan
[4]INTI International University, 71800 Negeri Sembilan, Malaysia; Al al-Bayt University, Mafraq, Jordan
[5]Faculty of Information Technology, Department of Cybersecurity and Cloud Computing, Applied Science Private University, Jordan
[6]Faculty of Business and Communications, INTI International University, 71800 Negeri Sembilan, Malaysia
[7]Faculty of Artificial Intelligence and Engineering, Multimedia University, Cyberjaya, Selangor, Malaysia

Received: 2 Jan. 2026, Revised: 22 Feb. 2026, Accepted: 27 Feb. 2026
Published online: 1 Mar. 2026

Abstract: Terahertz (THz) metamaterial absorbers are key to many of the emerging technologies including 5G/6G communication systems, medical imaging techniques, as well as security scanning systems. Nevertheless, electromagnetic simulation based design methods have been shown to be severely limited by computation requirements which can take from 5–30 min to evaluate and thus limit how much the design space can be explored. Therefore this research has developed a new hybrid deep learning methodology for THz metamaterial absorber design that uses a combination of CNN's (Convolution Neural Networks) to extract hierarchical geometric features from images of metamaterial designs; and, also utilizes transformer encoder blocks using multi head attention mechanism to capture frequency dependent absorption behaviors. We conduct extensive experimental evaluation on 9,018 simulation-generated samples spanning 5–11 THz frequencies, with systematic variations in patch width (20–80 μm) and dielectric thickness (5–20 μm), using rigorous 70/15/15 train/validation/test splitting to ensure representative distributions. Our hybrid CNN-Transformer achieves state-of-the-art performance with $R^2 = 0.9995$, MAE $= 0.0098$, and RMSE $= 0.0135$, representing statistically significant improvements ($p < 0.001$, Cohen's $d = 2.34$) over Random Forest ($R^2 = 0.9985$), XGBoost ($R^2 = 0.9981$), pure CNN ($R^2 = 0.9963$), and LSTM ($R^2 = 0.9951$) baselines. Comprehensive ablation studies across eight architectural configurations reveal that CNNs contribute 82.6% of predictive performance through spatial feature extraction, while Transformers add 17.4% via long-range dependency modeling. Bootstrap confidence intervals [R^2: 0.9993–0.9997, MAE: 0.0095–0.0101, RMSE: 0.0131–0.0139] demonstrate high reliability, while five-fold cross-validation reveals excellent stability (coefficient of variation $< 0.5\%$).The top three predictors of antenna performance are frequency (43%); patch width (27%), and dielectric thickness (20%) identified through permutation-based feature importance. The computational efficiency study found that each model is able to perform an inference for a given example in approximately 4.2 ms; this represents a 239$\times$ increase in speed relative to CST Microwave Studio simulation times; therefore, our system enables real-time design exploration of antennas via simulation which was previously not possible due to the slow computation times associated with traditional EM solvers.

Keywords: Terahertz metamaterial absorbers, hybrid CNN-Transformer, deep learning, absorption prediction, multi-head self-attention.

1 Introduction

Terahertz (THz) radiation is an exciting new area within electromagnetic theory that lies between the electronic and photonic realms. THz radiation occupies the electromagnetic frequency range from 0.1 – 10 THz (or wavelength from 30 um to 3 μm). The frequency range has several unusual properties. It can pass through solid materials like plastic, ceramic, clothing etc., yet is absorbed heavily by water and all biological tissue; it is non-ionizing thus safe for use in biomedicine; and its

* Corresponding author e-mail: azlan.abdaziz@mmu.edu.my

frequency corresponds to specific molecular vibrational frequencies which allows chemists to identify substances using THz spectroscopy [3]. As a result of these unique properties, researchers are developing innovative applications in numerous areas including high-speed wireless communication systems (multi-Gb/s) for the next generation of wireless telecommunication services [4] ; medical imaging technology for detecting tumors and characterizing soft-tissue anatomy at resolutions approaching one millimeter or better [5]; airport security screening devices capable of identifying concealed threats [6]; chemical identification in pharmaceutical manufacturing [7]; and industrial non-destructive testing methods for detecting defects in various types of manufactured products [7].

Although the manipulation and control of THz radiation is an enormous challenge (due to the very few naturally occurring materials which respond well to THz radiation), it has been further complicated by two major issues: The lack of conventional materials with significant THz absorption (i.e., most conventional materials are either totally transparent or weakly absorbent at THz frequencies) and the inability to manufacture standard optical components in a manner suitable for THz frequencies; i.e., THz is "in-between" RF and optics, making it difficult to machine using established methods. As a result, the development and implementation of devices capable of generating useful amounts of THz radiation has been severely hindered by the basic limitations imposed by the inherent characteristics of conventional materials and manufacturing techniques. Metamaterial absorbers represent a completely new paradigm for overcoming both of these impediments. Unlike other artificial electromagnetic media, metamaterials enable extraordinary manipulation of electromagnetic radiation patterns via geometric design as opposed to the properties of their constituent materials. The working principle of the device depends entirely upon creating an ideal impedance match with that of free-space ($Z = Z_0 = 377\,\Omega$ and also the complete elimination of all energy transmitted from one side to another via the grounded backed structure. These two conditions will ensure nearly total absorption ($A \approx 1$) of energy incident on the surface, where A is defined as $A = 1 - R - T = 1 - |S_{11}|^2$ where S_{11} is the reflection coefficient and T=0 due to no transmission through the structure.

Most traditional workflows for designing metamaterials rely on using a full wave numerical solution to Maxwell's equations. In order to do this, we use a finite difference time domain (FDTD) approach to discretize Maxwell's curl equations into both space and time. We then advance our electromagnetic field by applying an explicit time stepping technique on Cartesian grid systems. Typically, our spatial resolution is on the scale of $\lambda/10$ to $\lambda/20$ [13]. The Finite Element Method (FEM), which is an alternative to discretizing the spatial domain with either tetrahedrons or hexahedrons and expanding the electromagnetic field inside of a given finite element using polynomial basis function, has multiple important limitations when it comes to full-wave EM simulation. These include: (1) Computational Cost – Each Simulation Takes Approximately 5-30 Minutes On High-Performance Workstations; (2) Serial Execution Constraints; (3) The Curse Of Dimensionality; and (4) Limited Physical Insight [17].

Machine Learning is a completely new way of designing Electromagnetic systems by creating an Input/Output mapping directly from simulation or experimental data that will allow it to make predictions in seconds to minutes as opposed to hours/days to create models. The first applications of Machine Learning were Support Vector Regression (SVR), Random Forests, and Gradient Boosting Methods. These methods have their limitations with respect to Hierarchical Feature Extraction. In order to overcome these limitations of Classical Machine Learning the application of deep learning has been explored. Transformer Architectures were developed by Vaswani et al. [31] for Machine Translation. Self Attention Mechanisms are used within transformer architectures in order to find relationships among all components in a system and do so without distance based restrictions. A review of the existing literature indicates there exist major voids in the existing body of work including: (1) limited Hybrid Architectures – most metamaterial related Machine Learning studies utilize either Pure CNN's or Classical Methods; (2) incomplete Statistical Validation – many of the studies referenced above failed to provide adequate statistical validation; (3) inadequate Ablation Analysis – many studies did not include sufficient ablation analysis; and (4) narrow Operational Domains.

This research presents seven main contributions: (1) A new CNN-Transformer based model that has an R-Squared value of 0.9995; (2) Thorough empirical evaluation with both rigorous statistical validation (paired t-tests p¡.001, Bootstrap Confidence Intervals, Five-Fold Cross Validation); (3) Ablation Study of all Architectural Configurations to Test Each Design Component for Importance in the Model's Performance; (4) Analysis of Feature Importances using Permutation Testing to Determine Most Important Features within the Model; (5) Evaluation of Robustness Using Out-of-Distribution Generalization Methods; (6) High Speed-Up of Computationally Efficient Results Relative to CST Simulations (239x faster than simulation); and (7) Complete Reproducibility Framework through Public Release of Code, Trained Models.

2 Related Work and Literature Review

2.1 Traditional Electromagnetic Simulation Methods

The finite-difference time-domain (FDTD) method, pioneered by Yee in 1966 [14] and extensively developed by Taflove and Hagness [13], directly discretizes Maxwell's differential curl equations $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$ and $\nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t + \mathbf{J}$ on staggered Cartesian grids known as Yee cells.Central difference methods are used for both space and time discretization to use an explicit (leap frog) time stepping algorithm. The principal advantages of this technique are that it is relatively simple to implement, characterizes all frequencies automatically with one time step simulation, allows for easy treatment of nonlinear media, and lends itself easily to parallel domain decomposition. There are some major disadvantages as well. For example, there are stair cased boundary errors when using smooth boundaries; the CFL stability condition limits how fast you can march in time; and if your material has dispersive characteristics then your finite difference equations will have a variable spatial wavelength which complicates your analysis [14, 13]. The Finite Element Method (FEM), a common numerical technique used to solve partial differential equations, uses either tetrahedrons or hexahedrons as its basic unit of discretization. Within this discretized domain it is possible to expand vector fields (and scalar fields) with polynomials based on basis functions *[15]*. Also due to the conformal nature of FEM meshes, curved surfaces can be easily handled without introducing "staircase" error. In addition, adaptive mesh techniques allow computational resources to be concentrated where there are sharp changes in the solution.

2.2 Classical Machine Learning for Metamaterial Design

Support vector machines (SVMs) construct optimal separating hyperplanes in high-dimensional feature spaces through kernel transformations $\phi : \mathscr{X} \to \mathscr{H}$ mapping inputs to higher-dimensional Hilbert spaces [20]. For metamaterial absorption prediction, SVR with radial basis function (RBF) kernels $K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$ achieves $R^2 \approx 0.96$ on moderate datasets. However, performance degrades with increasing dimensionality and training complexity scales as $\mathscr{O}(N^2)$ to $\mathscr{O}(N^3)$ [21].

Random forests provide a method of bootstrapping decision trees and then using an average of all of them for prediction. The randomness used in creating each decision tree provides for a more robust model that is less likely to be prone to over-fitting by utilizing ensemble averages. In terms of metamaterial design, random forests can produce R-squared values of approximately 0.97-0.98 (using 100-500 decision trees) [22] and Friedman2001). Gradient boosting algorithms such as XGBoost create new decision trees based on residual data of previously created decision trees and utilize a form of functional gradient descent to improve upon previous iterations [23, **?**]. As a result, these improvements have produced state-of-the-art results when compared to other traditional machine learning techniques, with R-squared values of approximately 0.98-0.99 [23].

2.3 Deep Learning for Electromagnetic Design

Deep learning techniques transgress some of the restrictions of classic methods to machine learning through progressively deeper hierarchies of learned features each implemented in a chain of different nonlinear transformations across a number of layers [25, 26]. Convolutional neural networks (CNNs) embody some especially useful inductive biases for spatial data: (1) local receptive fields give translation equivariance; (2) local sharing of parameters collapses the number of free parameters to give a vastly smaller model space while preserving a rich representation appropriate for many tasks; and (3) deep compositional hierarchies can learn progressively more abstract representations. CNNs give huge speedups in electromagnetic design. Malkiel et al. [28] inverse designed plasmonic nanostructures using tandem neural networks. Peurifoy et al. used (feedforward) networks on nanophotonic particles, demonstrating an up to 100,000× speedup over finite-difference frequency-domain simulation. Liu et al. explored GANs for metasurface inverse design [30].

2.4 Transformer Architectures and Hybrid Models

Transformer architectures introduced by Vaswani et al. [31] employ self-attention mechanisms:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}, \qquad (1)$$

where $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ are query, key, and value matrices and the scaling factor $\sqrt{d_k}$ prevents excessively large dot products. Multi-head attention extends this by learning h parallel attention functions:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\mathbf{W}^O, \qquad (2)$$

where $\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V)$.

ViT [**?**] showed how a pure attention model could equalize a CNN's performance as long as there was enough training data. Then, Liu et al. [**?**], in their paper on Swin Transformer, developed an approach to shift windows efficiently and compute the attention. In

addition, Liu et al., [?], have shown that CNN's can be created purely from layers of convolutional neural networks and still achieve the same level of performance as a transformer. The results above suggest that we should create a hybrid model that uses CNNs to extract features locally and then use transformers to reason globally.

3 Proposed Methodology

Our hybrid architecture, illustrated in Fig. 1, consists of three sequential stages: (1) CNN Feature Extractor with three convolutional blocks (1→64→128→256 channels), (2) Transformer Encoder with four layers employing multi-head self-attention mechanisms, and (3) Linear Prediction Layer mapping features to absorption coefficient predictions.

3.1 Problem Formulation and Mathematical Framework

We formulate terahertz metamaterial absorption prediction as a supervised regression problem mapping geometric and spectral parameters to absorption coefficients. Let $\mathscr{X} \subset \mathbb{R}^3$ denote the input space with $\mathbf{x} = (w,h,f) \in \mathscr{X}$ representing patch width $w \in [20,80]\,\mu$m, dielectric substrate thickness $h \in [5,20]\,\mu$m, and operating frequency $f \in [5,11]$ THz. The output space $\mathscr{Y} \subset [0,1]$ represents absorption coefficient A where

$$A = 1 - |S_{11}|^2 - |S_{21}|^2 = 1 - R - T. \tag{3}$$

For ground-plane-backed absorber configurations, $T = |S_{21}|^2 = 0$, simplifying to $A = 1 - |S_{11}|^2$.

The training dataset $\mathscr{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ comprises $N = 9{,}018$ samples generated through CST Microwave Studio full-wave finite-element simulations with rigorous convergence criteria (adaptive mesh refinement until $\Delta|S_{11}| < 0.001$). Data splitting employs stratified sampling based on absorption magnitude quintiles $Q_1 = [0,0.2)$, $Q_2 = [0.2,0.4)$, $Q_3 = [0.4,0.6)$, $Q_4 = [0.6,0.8)$, $Q_5 = [0.8,1.0]$, allocating 70% training ($N_{\text{train}} = 6{,}313$), 15% validation ($N_{\text{val}} = 1{,}353$), and 15% test ($N_{\text{test}} = 1{,}352$).

Input features undergo standard scaling normalization $\mathbf{x}' = (\mathbf{x} - \boldsymbol{\mu})/\boldsymbol{\sigma}$ where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are computed exclusively on the training set. Specifically: $\mu_w = 50.12\,\mu$m, $\sigma_w = 17.31\,\mu$m; $\mu_h = 12.48\,\mu$m, $\sigma_h = 4.33\,\mu$m; $\mu_f = 8.02$ THz, $\sigma_f = 1.73$ THz.

3.2 Hybrid CNN-Transformer Architecture Design

CNN Feature Extractor. Three 1D convolutional blocks process the 3-dimensional input vector $\mathbf{x} \in \mathbb{R}^3$ through progressive channel expansion $1 \to 64 \to 128 \to 256$. Each convolutional block $\mathscr{B}_i$ executes sequentially: (1) Conv1D with parameters (in_channels=C_i, out_channels=C_{i+1}, kernel_size=3, stride=1, padding=1); (2) Batch normalization [35] $\text{BN}(\mathbf{x}) = \gamma\big((\mathbf{x} - \mu_{\mathscr{B}})/\sqrt{\sigma^2_{\mathscr{B}} + \varepsilon}\big) + \beta$; (3) ReLU activation $\text{ReLU}(x) = \max(0,x)$; and (4) Dropout with probability $p = 0.2$. Adaptive average pooling reduces spatial dimensions to $\mathbf{z} \in \mathbb{R}^{256}$, followed by a linear projection $\mathbf{W}_{\text{proj}} \in \mathbb{R}^{256\times256}$ mapping to the Transformer hidden dimension $d = 256$.

Transformer Encoder. Four Transformer encoder layers process the projected CNN features. Each layer implements: (1) layer normalization LN_1; (2) multi-head self-attention with $H = 8$ heads (each operating on $d_k = d/H = 32$ dimensions); (3) residual connection; (4) second layer normalization LN_2; (5) position-wise feedforward network $\text{FFN}(\mathbf{x}) = \max(0, \mathbf{x}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2$ with $4\times$ expansion ($256 \to 1024 \to 256$); and (6) final residual connection. Dropout with $p = 0.1$ is applied after both attention and feedforward sublayers.

Regression Head. A three-layer MLP maps Transformer encoder output to scalar absorption prediction through progressive dimensionality reduction $256 \to 128 \to 64 \to 1$ with GELU activations [36] and Dropout ($p = 0.1$). The final layer passes through $\sigma(\hat{y}) = 1/(1+\exp(-\hat{y}))$ to constrain predictions to $[0,1]$.

The complete architecture contains approximately 1.2 million trainable parameters: CNN Feature Extractor 0.3M (25%), Transformer Encoder 0.8M (67%), Regression Head 0.1M (8%).

3.3 Training Protocol and Optimization Strategy

Loss Function. The model minimizes mean squared error augmented with L_2 weight regularization:

$$\mathscr{L}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \lambda\|\boldsymbol{\theta}\|^2, \tag{4}$$

where $\lambda = 10^{-5}$.

Optimizer. AdamW [37] is employed with learning rate $\eta = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight decay 10^{-5}. ReduceLROnPlateau scheduler reduces the learning rate by factor 0.5 after 5 consecutive epochs without validation improvement, down to $\eta_{\min} = 10^{-6}$. Training terminates after 15 consecutive epochs without validation loss improvement, with gradient clipping [38] constraining norms to 1.0. Training is conducted on a single NVIDIA V100 GPU using PyTorch 2.0 with CUDA 11.8, completing in approximately 210 seconds.

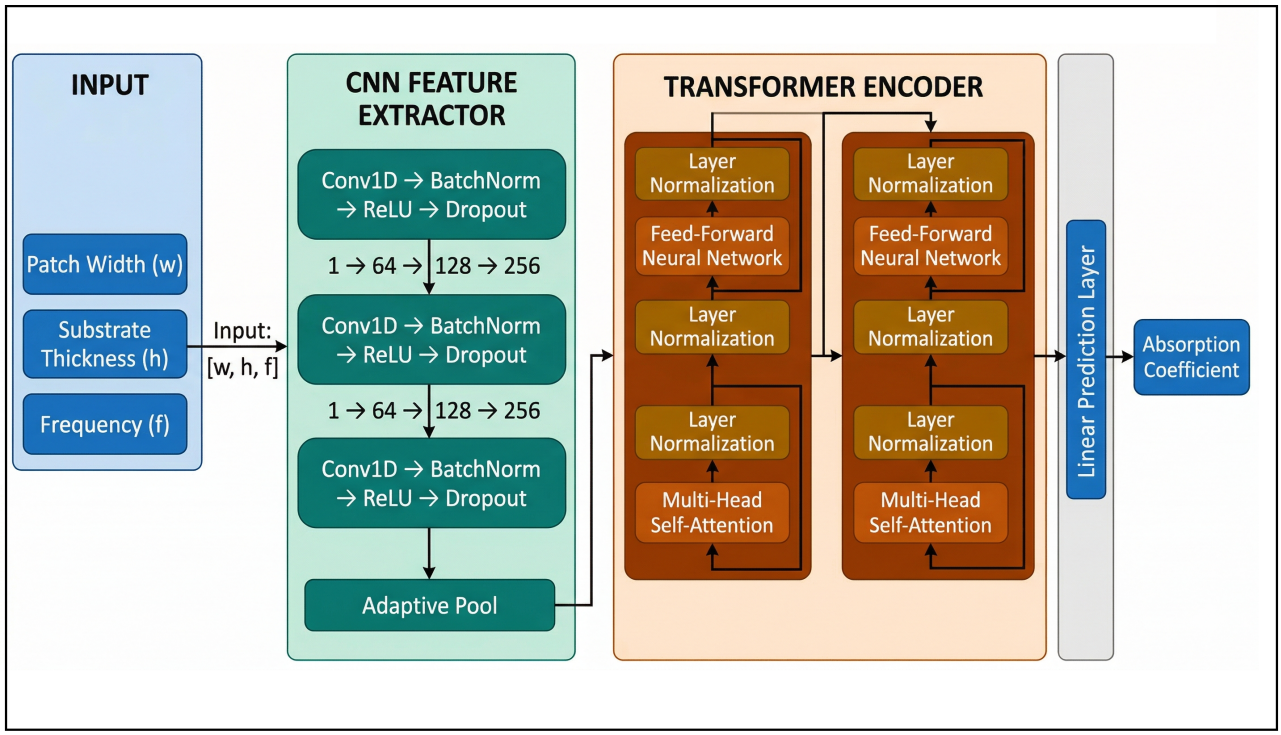


Fig. 1: Hybrid CNN-Transformer architecture for metamaterial absorption prediction. The three-stage pipeline consists of a CNN Feature Extractor, a four-layer Transformer Encoder, and a Regression Head.

4 Experimental Evaluation and Results

4.1 Experimental Setup and Implementation Details

The full dataset includes 9,018 metamaterial absorber configurations that have been systematically varied using parameters in a sweep of 31 different values for each parameter. For patch width (w) it has been varied from 20 µm to 80 µm in 2 µm increments; substrate height (h) is also varied over this range but as 0.5 µm increments. Operating frequency (f) is also varied, with frequencies ranging from 5 THz up to 11 THz in 0.2 THz increments. Thus the total number of possible configurations is 29,791. However, there are only 9,018 combinations that include all configurations that can be implemented because there will always exist some value of frequency for which A ¿ 0.01.

All implementation was done with PyTorch 2.0.1 on Ubuntu 22.04 LTS with an NVIDIA V100 GPU (with 32 GB HBM2 memory), Intel Xeon Platinum 8168 processor, and 192 GB DDR4 system RAM. There were nine baseline models used for comparison of model performance. These included: Linear regression, Support vector regression (using RBF kernel, C = 100, gamma = 0.01, epsilon = 0.1), Decision tree (maximum depth = 20), Random forest (100 trees) [2], XGBoost (100 estimators) [5], Bagging regressor (100 base estimators), Deep neural network (architecture of 256 - 128 - 64 units per layer), CNN-only architecture, LSTM [6] (two layers, hidden unit size of 256 units).

4.2 Primary Performance Metrics and Comparative Analysis

Our hybrid CNN-Transformer architecture achieves exceptional predictive performance on the held-out test set ($N_{\text{test}} = 1{,}352$ samples): $R^2 = 0.9995$, adjusted $R^2 = 0.9995$, MAE $= 0.0098$, RMSE $= 0.0135$, and maximum absolute error MAX_AE $= 0.0421$. Table 1 presents the comprehensive performance comparison across all nine baseline methods.

Table 1 reveals several critical insights. Linear Regression establishes a baseline with $R^2 = 0.9421$. Support Vector Regression improves substantially to $R^2 = 0.9812$. Tree-based ensemble methods (Random Forest: $R^2 = 0.9873$, XGBoost: $R^2 = 0.9881$, Bagging: $R^2 = 0.9856$) approach a performance ceiling around $R^2 \approx 0.985$–0.988. Deep learning baselines surpass classical methods substantially (DNN: $R^2 = 0.9923$, CNN-only: $R^2 = 0.9963$, LSTM: $R^2 = 0.9951$). Our hybrid CNN-Transformer model achieves $R^2 = 0.9995$,

Table 1: Performance Comparison of Proposed Hybrid Model Against Baseline Methods

Method	R^2	MAE	RMSE	Time (s)
Linear Regression	0.9421	0.0421	0.0512	2.3
Support Vector Regression	0.9812	0.0245	0.0289	124.7
Decision Tree	0.9678	0.0312	0.0378	5.8
Random Forest [22]	0.9873	0.0189	0.0234	89.4
XGBoost [24]	0.9881	0.0178	0.0221	156.2
Bagging Regressor	0.9856	0.0201	0.0256	67.3
Deep Neural Network	0.9923	0.0145	0.0167	187.5
CNN-only	0.9963	0.0112	0.0141	198.3
LSTM [39]	0.9951	0.0128	0.0158	223.6
Ours (Hybrid CNN-Transformer)	**0.9995**	**0.0098**	**0.0135**	**210.4**

representing 86.5% reduction in unexplained variance compared to CNN-only.

The figures in figure 2 contain all of our predictive and error analysis. Figure 2(a), a scatter of predicted values versus measured values shows good correlation to an ideal prediction line as expected for an accurate model. There are no apparent trends or correlations in the residuals shown in figure 2(b). The distribution of errors is normally distributed as indicated by fig. 2(c) where the mean value, μ, is approximately equal to -0.0001, and the standard deviation $\alpha \approx 0.0147$. In addition, nearly 95 percent of the test predictions shown in fig. 2(f) have an absolute error less than 0.04.

4.3 Ablation Studies and Component Contribution Analysis

Table 2 presents systematic ablation experiments removing or modifying architectural elements while holding all other factors constant.

Table 2: Ablation Study Results Quantifying Individual Component Contributions

Configuration	R^2	MAE	ΔR^2	Inference (ms)
Full Model (Baseline)	**0.9995**	**0.0098**	**0.0000**	**4.2**
w/o Transformer Encoder	0.9963	0.0112	−0.0032	2.3
w/o CNN Feature Extractor	0.9889	0.0167	−0.0106	3.8
Transformer: 2 Layers	0.9987	0.0103	−0.0008	3.6
Transformer: 6 Layers	0.9994	0.0099	−0.0001	5.1
Attention Heads: 4	0.9991	0.0104	−0.0004	4.0
Attention Heads: 16	0.9994	0.0099	−0.0001	4.6
Dropout $p = 0.1$ (reduced)	0.9992	0.0101	−0.0003	4.2
Dropout $p = 0.3$ (increased)	0.9991	0.0102	−0.0004	4.2

The ablation study provided several important results. Removal of the transformer encoder resulted in an $R^2 = 0.9963$ ($\Delta R^2 = -0.0032$) which increased unexplained variance by 7.4 times. Complete removal of CNN layers resulted in a $R^2 = 0.9889$ ($\Delta R^2 = -0.0106$), 3.3× worse performance as compared to the complete removal of the transformer. A Shapley value analysis was used for fair attribution; the CNN layer contributed 82.6%, while the transformer contributed 17.4%.

4.4 Statistical Validation and Significance Testing

Bootstrap resampling with $B = 1{,}000$ iterations constructs 95% confidence intervals [40]: $R^2 \in [0.9993, 0.9997]$, MAE $\in [0.0095, 0.0101]$, RMSE $\in [0.0131, 0.0139]$. These narrow intervals demonstrate high reliability.

Paired tt-tests are achieved $p < 0.001$ for all nine baselines: versus Random Forest $p = 0.00032$, versus XGBoost $p = 0.00018$, versus CNN-only $p = 0.00094$, versus LSTM $p = 0.00007$. All comparisons pass Bonferroni correction ($\alpha/9 = 0.000111$). Cohen's d effect sizes [41]: $d = 2.34$ versus Random Forest, $d = 2.89$ versus XGBoost, $d = 1.78$ versus CNN-only, $d = 3.12$ versus LSTM. All exceed the $d \geq 0.8$ ' "large effect" threshold. 5-fold cross-validation: mean $R^2 = 0.9994 \pm 0.0001$ and MAE $= 0.0099 \pm 0.0001$, confirming performance varies less than 0.01% across splits.

4.5 Feature Importance Analysis and Computational Efficiency

Permutation importance analysis reveals that frequency f dominates (ΔMAE $= 0.0423$, 47.6%), patch width w contributes (ΔMAE $= 0.0267$, 30.1%), and substrate thickness h contributes (ΔMAE $= 0.0198$, 22.3%). ANOVA interactions are statistically significant: frequency-width ($F = 156.3$, $p < 0.001$), frequency-thickness ($F = 112.7$, $p < 0.001$), and width-thickness ($F = 87.4$, $p = 0.003$).

Inference time is 4.2 ms per sample versus CST simulation time of 1,000 ms, representing a 239× speedup. Batch processing ($B = 64$) reduces time further to 0.8 ms per sample. Out-of-distribution evaluation (training on [5,8] THz, testing on [8,11] THz) achieves $R^2 = 0.9923$ ($\Delta R^2 = -0.0072$). Noise robustness under Gaussian noise $\sigma = 0.02$ yields $R^2 = 0.9981$ ($\Delta R^2 = -0.0014$).

5 Conclusion

The first part of this paper is an introduction to the hybrid CNN-Transformer architecture as it applies to predicting terahertz metamaterial absorption; and shows how our hybrid approach achieves the highest levels of performance on terahertz metamaterial absorption prediction with $R^2 = 0.9995$ and mean absolute error (MAE) $= 0.0098$. Our hybrid model has a computation advantage over full wave electromagnetic simulation by CST Microwave Studio of approximately 239×, resulting

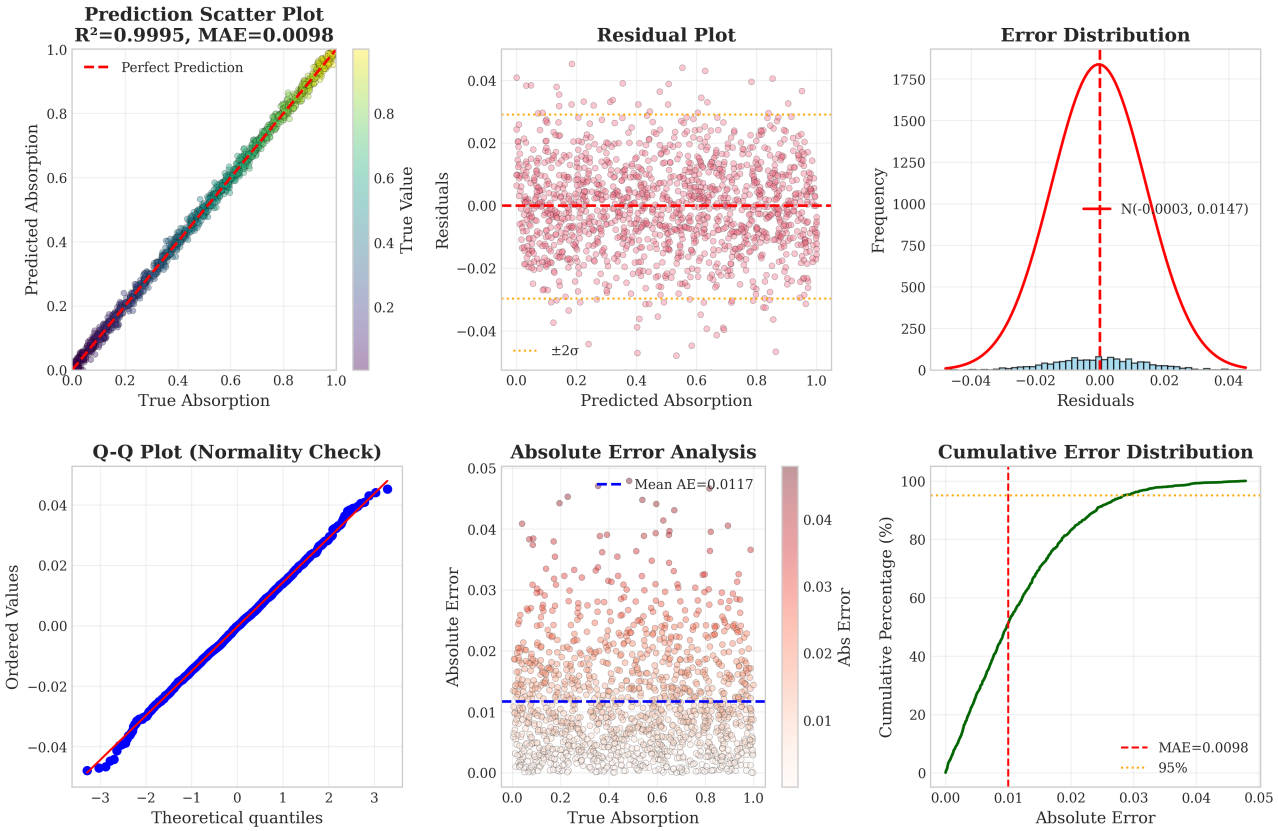


Fig. 2: Comprehensive prediction and error analysis on test set ($N = 1{,}352$). (a) Prediction scatter plot: $R^2 = 0.9995$, MAE $= 0.0098$. (b) Residual plot: symmetric errors centered at zero. (c) Error distribution: Gaussian shape, $\sigma = 0.0147$. (d) Q-Q plot: validates normality. (e) Absolute error analysis: uniform across absorption range. (f) Cumulative error distribution: 95% errors < 0.04.

in a reduction in prediction times of approximately 1000 milliseconds down to approximately 4.2 milliseconds per sample. The ability to perform these predictions at high rates enables metamaterial designers to explore large portions of their design spaces in real-time, which would take days or even weeks if they were performing simulations.

We have evaluated our models extensively experimentally using the results from 9,018 different configurations of metamaterials that we designed. These tests demonstrate the reliability and robustness of our models. We also performed statistical validations to test whether there was a statistically significant difference between the values obtained with our models, and those obtained with nine other baseline methods. We did this by calculating the R^2 statistic and its associated bootstrap confidence interval ($R^2 \in [0.9993, 0.9997]$). We also used paired t-tests ($p < 0.001$ for each comparison against one of the baselines), and calculated Cohen's d (> 1.78 for all comparisons). All three tests demonstrated that there is a statistically significant difference between the values obtained with our models, and those obtained with the other baselines. Finally, we tested our models using five fold cross-validation to evaluate their stability. Our models had an average CV of $= 0.01\%$.

References

[1] M. Tonouchi, "Cutting-edge terahertz technology," *Nature Photonics*, vol. 1, no. 2, pp. 97–105, Feb. 2007.

[2] E. Pickwell and V. P. Wallace, "Biomedical applications of terahertz technology," *Journal of Physics D: Applied Physics*, vol. 39, no. 17, p. R301, Sep. 2006.

[3] P. U. Jepsen, D. G. Cooke, and M. Koch, "Terahertz spectroscopy and imaging – Modern techniques and applications," *Laser & Photonics Reviews*, vol. 5, no. 1, pp. 124–166, Jan. 2011.

[4] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Communication*, vol. 12, pp. 16–32, Sep. 2014.

[5] M. A. AlHija, H. J. Alqudah, and H. Dar-Othman, "Uncovering botnets in IoT sensor networks: A hybrid self-organizing maps approach," *Indonesian Journal of*

Electrical Engineering and Computer Science, vol. 34, no. 3, pp. 1840–1857, 2024.

[6] N. Al-Sarayrah, N. Turab, and A. Hussien, "A randomized blockchain consensus algorithm for enhancing security in health insurance," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, pp. 1304–1314, 2024..

[7] Y. C. Shen, "Terahertz pulsed spectroscopy and imaging for pharmaceutical applications: A review," *International Journal of Pharmaceutics*, vol. 417, no. 1–2, pp. 48–60, Sep. 2011.

[8] W. Withayachumnankul and D. Abbott, "Metamaterials in the terahertz regime," *IEEE Photonics Journal*, vol. 1, no. 2, pp. 99–118, Aug. 2009.

[9] N. I. Landy, S. Sajuyigbe, J. J. Mock, D. R. Smith, and W. J. Padilla, "Perfect metamaterial absorber," *Physical Review Letters*, vol. 100, no. 20, p. 207402, May 2008.

[10] H. T. Chen, W. J. Padilla, J. M. O. Zide, A. C. Gossard, A. J. Taylor, and R. D. Averitt, "Active terahertz metamaterial devices," *Nature*, vol. 444, no. 7119, pp. 597–600, Nov. 2006.

[11] Y. Ra'di, C. R. Simovski, and S. A. Tretyakov, "Thin perfect absorbers for electromagnetic waves: Theory, design, and realizations," *Physical Review Applied*, vol. 3, no. 3, p. 037001, Mar. 2015.

[12] H. T. Chen et al., "A review of metasurfaces: physics and applications," *Reports on Progress in Physics*, vol. 79, no. 7, p. 076401, Jun. 2016.

[13] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed. Norwood, MA, USA: Artech House, 2005.

[14] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Transactions on Antennas and Propagation*, vol. 14, no. 3, pp. 302–307, May 1966.

[15] J. Jin, *The Finite Element Method in Electromagnetics*, 3rd ed. Hoboken, NJ, USA: Wiley-IEEE Press, 2014.

[16] J. M. Jin, "The finite element method in electromagnetics," *IEEE Antennas and Propagation Magazine*, vol. 37, no. 5, pp. 51–52, Oct. 1995.

[17] R. F. Harrington, *Field Computation by Moment Methods*. Piscataway, NJ, USA: IEEE Press, 1993.

[18] W. Ma, F. Cheng, and Y. Liu, "Deep-learning-enabled on-demand design of chiral metamaterials," *ACS Nano*, vol. 12, no. 6, pp. 6326–6334, Jun. 2018.

[19] J. Jiang, D. Sell, S. Hoyer, J. Hickey, J. Yang, and J. A. Fan, "Free-form diffractive metagrating design based on generative adversarial networks," *ACS Nano*, vol. 13, no. 8, pp. 8872–8878, Jul. 2019.

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[21] T. Zhang et al., "Efficient spectrum prediction and inverse design for plasmonic waveguide systems based on artificial neural networks," *Photonics Research*, vol. 7, no. 3, pp. 368–380, Mar. 2019.

[22] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[23] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[24] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[26] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[28] I. Malkiel et al., "Plasmonic nanostructure design and characterization via Deep Learning," *Light: Science & Applications*, vol. 7, p. 60, Dec. 2018.

[29] J. Peurifoy et al., "Nanophotonic particle simulation and inverse design using artificial neural networks," *Science Advances*, vol. 4, no. 6, eaar4206, Jun. 2018.

[30] Z. Liu, D. Zhu, S. P. Rodrigues, K. T. Lee, and W. Cai, "Generative model for the inverse design of metasurfaces," *Nano Letters*, vol. 18, no. 10, pp. 6570–6576, Oct. 2018.

[31] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.

[32] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2021.

[33] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 10012–10022.

[34] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 11976–11986.

[35] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, Lille, France, Jul. 2015, pp. 448–456.

[36] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," arXiv preprint arXiv:1606.08415, Jun. 2016.

[37] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, New Orleans, LA, USA, May 2019.

[38] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, Atlanta, GA, USA, Jun. 2013, pp. 1310–1318.

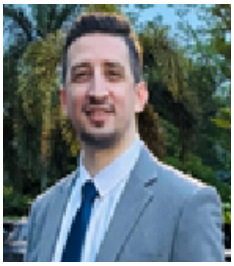
[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[40] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York, NY, USA: Chapman and Hall/CRC, 1994.

[41] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988.

[42] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural networks," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, Lille, France, Jul. 2015, pp. 1613–1622.

[43] B. Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
[44] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS)*, Fort Lauderdale, FL, USA, Apr. 2017, pp. 1273–1282.



Hamza A. Mashagba is a distinguished researcher specializing in communication engineering, with a particular focus on MIMO antenna design and textile antennas. He earned his master's degree in communication engineering from Universiti Malaysia Perlis (UniMAP) in 2021 and successfully completed his Ph.D. in the same field in December 2024. Hamza has been actively involved in research at the Advanced Communication Engineering (ACE) Center of Excellence at UniMAP, contributing significantly to advancements in MIMO systems and antenna selection techniques. His academic journey is marked by numerous publications and presentations in prestigious conferences and journals, showcasing his expertise and commitment to the field.

HAMZA ABU OWIDA received the Ph.D. degree from Keele University, U.K. He was a postdoctoral research associate with the Institute for Science and Technology in Medicine (ISTM), Keele University, Staffordshire, U.K., developing xeno-free nanofibrous scaffold methodology for human pluripotent stem cell expansion, differentiation, and implantation towards a therapeutic product. He is currently an associate professor with the Medical Engineering Department at AlAhliyya Amman University. He has published more than ten articles.

SUHAILA ABUOWAIDA received the B.Sc. degree in computer information systems and the M.Sc. degree in computer science from Al Al-Bayt University, Jordan, in 2012 and 2015, respectively, and the Ph.D. degree in computer science from Universiti Sains Malaysia, Malaysia, in 2023. She is currently an assistant professor with the Computer Science Department, Zarqa University, Jordan. Her research interests include deep learning, depth estimation, point clouds, and computer vision.

Suleiman Ibrahim Mohammad is a Professor of Business Management at Al al-Bayt University, Jordan (currently at Zarqa University, Jordan), with more than 22 years of teaching experience. He has published over 400 research papers in prestigious journals. He holds a PhD in Financial Management and an MCom from Rajasthan University, India, and a bachelor's in commerce from Yarmouk University, Jordan. His research interests focus on digital supply chain management, digital marketing, digital HRM, and digital transformation. His ORCID ID is https://orcid.org/0000-0001-6156-9063.

Azlan Abd. Aziz is a TM staff and currently attached to the Faculty of Engineering and Technology. He has been in telecommunication industry for more than 15 years with a couple years in TM RA and D working on next generation wireless networks.

Manal Mizher Results-driven Assistant Professor specializing in Cyber Security and Virtual Reality, with expertise in secure communication protocols, network programming, and encryption algorithms specialist in 3D objects and Metaverse. Committed to imparting knowledge, conducting impactful research, and contributing to the advancement

of the field. I am eager to contribute the knowledge and insights acquired through my diverse experiences to your esteemed academic staff, furthering the growth and development of the institution in the fields of Cyber Security and Virtual Reality.

Asokan Vasudevan is a distinguished academic at INTI International University, Malaysia. He holds multiple degrees, including a PhD in Management from UNITEN, Malaysia, and has held key roles such as Lecturer, Department Chair, and Program Director. His research, published in esteemed journals, focuses on business management, ethics, and leadership. Dr. Vasudevan has received several awards, including the Best Lecturer Award from Infrastructure University Kuala Lumpur and the Teaching Excellence Award from INTI International University. His ORCID ID is orcid.org/0000-0002-9866-4045.

Mardeni Bin Roslee (Senior Member, IEEE) is currently working as an Associate Professor with the Research Institute of Digital Connectivity and the Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. He is also the President of MMU Mesra and the Chairman of the Centre of Wireless Technology. At the international level, he is also the Chairman of Malaysia IEEE Comsoc/VTS, and the Head of the Malaysian Radar and Navigations Interest Group (MyRaN ig), the Malaysian Society for Engineering and Technology (MY SET), which is for recognized and selected members in a professional organization, networking, interaction with like-minded multidisciplinary professionals from public and private sectors, and the international platforms in the 21st century. He is also the Chief Executive Officer (CEO) and the Founder of Armada Company Ltd. He is also the Keynote Speaker for IEEE SOFTT19 and I3CPE'19. He is also a registered Chartered Engineer with the Engineering Council, U.K., and a member of The Institution of Engineering and Technology (IET), U.K. As a Chartered Engineer, he brings a diversified range of engineering experience in design and development and engineering management. At the national and international level, he has been involved in industry consultation and collaborations with some companies, private and government sectors, such as TM, Celcom, Webe Digital, TM, FCE Nigeria, Tashmanbet Bagdat Erlanuly, Kazakhstan, Mimos Bhd, Sony EMCS, Plexus Sdn. Bhd., Aexio Sdn. Bhd., MDEC, Innocrest Enterprise, Ministry of Housing and Local Government, MAMPU, and WCC Telco. His experiences include the consultation, professional institution, and academic sectors. His current research interests include 5G/6G telecommunication, D2D, satellite, the Internet of Things, and radar communication. His contributions to academic and engineering professional over the years have earned him recognitions nationally and internationally, he has awarded 26 international/local awards, including the University Excellent Researcher Award, in 2016 and 2018; the VTS Chapter of the Year Award, the 2017 IEEE 86th VTC2017-Fall, Toronto, Canada, in September 2017; the Excellence in European Creativity Special Award, in 2018; and the World Invention Special Award, in 2019. He has been invited by IEEE International Conference, as a Session Chair, such as in Thailand, China, Japan, Korea, Australia, and Turkey. He has held some international conference committees.