Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/200513

Hybrid CNN-Transformer for Multi-Band THz MIMO Antenna Optimization: 6G Deep Learning

Hamza A. Mashagba [1], Hamza Abu Owida[2], Manal Mizher[3], Asokan Vasudevan[4], Azlan B. Abd Aziz[1,], Mohammad Faleh Ahmmad Hunitie[5], Eyad Dhaher Hassan Megdadi [6] and Suleiman Ibrahim Mohammad[7,8]*

[1]Faculty of Engineering and Technology, Multimedia University, Centre for Wireless Technology (CWT), Cyberjaya, Selangor 63100, Malaysia
[2]Medical Engineering Department, Faculty of Engineering, Al-Ahliyya Amman University, Amman 19111, Jordan
[3]Department of Cybersecurity and Cloud Computing, Faculty of Information Technology, Applied Science Private University, Amman 11931, Jordan
[4]Faculty of Business and Communications, INTI International University, Negeri Sembilan 71800, Malaysia
[5]Department of Public Administration, School of Business, University of Jordan, Jordan
[6]Faculty of Engineering and IT, British University in Dubai, Dubai 345015, United Arab Emirates
[7]Department of Business Administration, School of Business, Al al-Bayt University, Mafraq 25113, Jordan
[8]INTI International University, Negeri Sembilan 71800, Malaysia

Received: 5 Jan. 2026, Revised: 29 Apr. 2026, Accepted: 3 Jun. 2026
Published online: 1 Sep. 2026

Abstract: Designing THz MIMO antennas for 6G wireless communications entails significant computation costs due to complex interactions between geometric and electromagnetic parameters. Traditional methods rely heavily on EM simulations, which take an enormous amount of time and computational power. In this paper, a new hybrid deep learning model based on CNNs and Transformer models was proposed that allows the simultaneous prediction of multiple performance metrics of graphene THz MIMO antennas. Our architecture considers twelve geometric and material parameters to predict six metrics of interest, namely S11, Gain, Isolation, Bandwidth, Envelope Correlation Coefficient (ECC), and Diversity Gain (DG). For training our deep learning algorithm, 756 different antenna designs over 1–6 THz were created using CST Microwave Studio software. The proposed hybrid architecture provides highly accurate predictions with R^2 values of 98.73%, 97.91%, and 98.45% for Gain, Isolation, and Bandwidth, respectively, outperforming conventional machine learning algorithms like Random Forest ($R^2 = 94.23\%$), XGBoost ($R^2 = 95.67\%$), and Deep Learning models (e.g., CNNs with $R^2 = 96.84\%$ and Transformers with $R^2 = 97.12\%$). With our proposed architecture, the simulation process has been accelerated by 1350 times; from 45 minutes to only 2 seconds, while the accuracy has stayed within the 2.3% MAPE. Ablation experiments show that both CNN's local feature extraction and Transformer's global dependency modeling are required to capture the necessary interactions.

Keywords: THz antennas, MIMO systems, deep learning, CNN-Transformer hybrid, graphene antennas, 6G communications, antenna optimization, metamaterials

1 Introduction

THE emergence of sixth generation (6G) wireless communication systems implies the need for innovations in antenna engineering, in order to facilitate extraordinary high data rates higher than 1 Tbps and low latency less than 100 microseconds [1,2,3]. Terahertz (THz) frequency bands (0.1–10 THz) have been proposed as an effective solution for the design of novel antennas because of their broad available bandwidth and diverse applications in holographic communications, wireless brain-computer interface communications, and ultra-high definition television (UHDTV) [4,5,6]. Nevertheless, the development of THz antennas faces a number of difficulties caused by high propagation losses, poor power efficiency, and complex manufacturing processes [7,8].

The application of multiple-input multiple-output (MIMO) technology allows compensating THz channel imperfections by utilizing the spatial diversity,

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

multiplexing effects, and signal-to-noise ratio [9]. Graphene-based THz antennas have gained great popularity due to adjustable electromagnetic behavior, high carrier mobility (> 200,000 cm^2/V·s), and flexibility [10,11,12]. The chemical potential of graphene is adjustable by means of applying the electrical voltage which makes it possible to control the operational frequency band without making any modifications in the antenna configuration [13,14].

Traditional antenna design involves the use of numerical EM simulation carried out using full-wave solvers like CST Microwave Studio and HFSS [15]. Each iteration of a parametric study related to a THz antenna takes about 30–60 minutes for computation. Optimization with hundreds and thousands of simulation runs becomes very time-consuming and costly [16]. Multi-criteria optimization, where the objectives include gain, bandwidth, isolation, and diversity maximization, requires advanced search algorithms along with a large number of function evaluations [17].

Recent studies have shown that machine learning (ML) techniques can play a significant role in reducing the costs of antenna design thanks to surrogate modeling [18]. Although these traditional ML methods have been shown to be very successful when it comes to independent prediction of antenna parameters using SVM, RF, and XGBoost [19,20]; however, they can become difficult to use when dealing with large input spaces that are high dimensional; when dealing with complex nonlinear relationships; and when working on multiple output tasks [21].

Deep learning-based CNNs and Transformers have improved performance significantly in image recognition and natural language processing fields by allowing the models to learn hierarchical representations [22,23]. Recent work has also illustrated how the combination of both CNN and Transformer layers can be used for a variety of applications such as video analysis, medical imaging, and time-series forecasting [25,26]. Several important research gaps remain in deep learning applied to antenna design:

–*Multi-Output Prediction:* The vast majority of previous ML-based antenna designs predict a single parameter (e.g., Gain, Resonant Frequency) and do not provide a full characterization of antenna performance [27].
–*Material Parameter Optimization:* Designing graphene-based antennas requires simultaneously optimizing geometric and physical parameters (chemical potential, number of layers, etc.), which are very difficult to solve with current methods [28].
–*Frequency Dependency:* THz antennas have many frequency-dependent characteristics that regression methods are unable to accurately account for [18].
–*Metrics for MIMO Systems:* Antennas used in MIMO systems need predictions of metrics including ECC and Diversity Gain, but little effort has been made to accomplish this [29].
–*Novel Architectures:* The literature does not feature attempts to develop deep learning models using combined CNN and Transformer architectures for predicting electromagnetic parameters of antennas.

These problems are solved by the proposed deep learning approach, which combines two different neural network architectures to achieve accurate multi-parameter prediction in THz MIMO antenna designs. The key contributions of this research paper are:

–*Novel Hybrid Model Architecture:* Development of the first-ever hybrid CNN-Transformer architecture aimed at THz MIMO antenna design, consisting of two parallel branches responsible for extracting features locally and globally.
–*Multi-Output Design:* Development of a predictive model capable of predicting six important antenna parameters (S11, gain, isolation, bandwidth, ECC, DG) based on twelve input parameters.
–*Comprehensive Dataset:* Creation of a dataset with 756 antenna designs varying geometric and physical parameters of graphene-based 4-element MIMO antennas operating in the 1–6 THz range.
–*Accurate Predictions:* The proposed hybrid model reaches an R^2 score of 98.73% for gain prediction, compared to 96.84% of the standalone CNN, 97.12% of the standalone Transformer, and 94.23%/95.67% of Random Forest/XGBoost, respectively.
–*Design Acceleration:* The use of the proposed model cuts the time required to design one antenna from 45 minutes down to under 2 seconds.
–*Ablation Analysis:* Detailed ablation studies demonstrate the necessity of combining CNN and Transformer models.

As such this report will be structured as follows. Section II is an overview of the literature. In Section III we provide details on the design of the THz MIMO Antenna configuration and how data was prepared. Section IV outlines our hybrid CNN-Transformer model. Results from experiments are presented in Section V. Sections VI discusses practical implications and some limitations of this work.

2 Related Work

2.1 Machine Learning in Antenna Design

In recent years there have been major advancements in using machine learning for designing antennas. At the beginning, many used simple regression models to predict the resonant frequency or input impedance [30] of antennas. Artificial Neural Networks (ANN) were one of the first ML technologies used for optimizing antennas [31].

With the latest developments in computational intelligence, advanced ML models have become possible.

According to Ahammed et al. [19], XGBoost regression was used to estimate the performance of high-gain THz IoT antennas, with an accuracy of 98.19% R^2 being achieved for a 6-feature input space. The researchers from Bangladesh [20] suggested using the Extra Trees regression algorithm for optimizing triple-band THz MIMO antennas, obtaining accuracy as high as 98.91% R^2.

SVM have been previously applied in antenna design with moderate results. Bala et al. [28] compared several different ML algorithms for estimating graphene antenna performance. Ensemble models performed better than single classifiers but worked in the case of conventional microwave frequencies. Koziel and Bandler [16] designed highly sophisticated surrogate-based optimization schemes which utilized Kriging modeling in combination with electromagnetic simulations.

2.2 CNN Architectures for Electromagnetic Applications

Convolutional Neural Networks have shown great success in learning spatial hierarchies of features [22]. Khan et al. [32] proposed the use of CNNs in predicting the radiation pattern types of patch antennas based on geometry image inputs with a classification accuracy of 96.5%. Sharma et al. [18] used 1D CNNs to perform sequential predictions of parameters in an antenna, showing superiority in comparison with fully connected networks.

ResNet and DenseNet [33,34], which are CNN models with skip connections for dealing with vanishing gradients, have been utilized for predicting antenna parameters. Zhang et al. [21] introduced a multi-scale CNN architecture for designing antenna arrays by using CNNs with various convolutional layers with different kernel sizes.

2.3 Transformer Models and Self-Attention Mechanisms

The Transformer network proposed by Vaswani et al. [23] revolutionized sequence modeling using self-attention mechanisms to model long-range dependencies without employing recurrent operations. The Vision Transformer (ViT) [24] approach proved that purely Transformer-based models could achieve state-of-the-art results in image classification.

In structured or tabular input data modeling, Transformers have proven to be very effective in modeling complex relationships among features [35]. Swin Transformer [36] employs hierarchical feature representation learning together with shifted window self-attention.

2.4 Hybrid CNN-Transformer Architectures

Several recent studies have focused on combining CNNs and Transformer networks. Hatamizadeh et al. [25] presented UNETR for medical image segmentation using a Transformer-based encoder with a CNN decoder. Arnab et al. [26] proposed ViViT for video understanding. For regression, hybrid architectures have been leveraged for time-series predictions and financial modeling [37,38]. However, none of the aforementioned architectures have been applied for THz antenna optimization.

2.5 Graphene-Based THz Antenna Design

Graphene possesses unique electromagnetic characteristics at THz frequencies arising from its surface conductivity controlled through chemical potential tuning [39]. Tamagnone et al. [10] presented a reconfigurable graphene-based THz antenna with electrostatic gating for frequency tunability. Llatser et al. [11] studied graphene-based nano-patch THz antennas with radiation efficiency greater than 50%.

3 THz MIMO Antenna Configuration and Dataset Generation

3.1 Antenna Geometry and Configuration

The designed four-element MIMO antenna structure is based on a graphene material circular patch radiator operating in an array form with flexible polyimide as a substrate material, as shown in Fig. 1. It has been designed for operation in the frequency range of 1–6 THz, making it multi-frequency capable for different 6G communications scenarios.

The antenna structure comprises the following layers from bottom to top:

– Ground plane: Copper with conductivity $\sigma = 5.8 \times 10^7$ S/m
– Substrate: Polyimide with relative permittivity $\varepsilon_r = 3.5$, loss tangent $\tan\delta = 0.008$
– Radiating patches: Multi-layer graphene with tunable chemical potential
– Feed network: Coplanar waveguide (CPW) configuration

Each circular patch has radius p_r that varies in the design space. The circular patch radiators are excited using microstrip feeders with feeder width f_w. Rectangular slots exist under each patch radiator in the ground plane with slot width sl_w and slot length sl_l, optimized for better impedance matching and bandwidth. Four radiators are placed with spacing e_s between them to reduce mutual coupling and envelope correlation.

3.2 Material Properties

Graphene's surface conductivity σ_g is described by the Kubo formula [39]:

$$\sigma_g(\omega, \mu_c, \Gamma, T) = \sigma_{\text{intra}} + \sigma_{\text{inter}} \tag{1}$$

where σ_{intra} and σ_{inter} represent intraband and interband contributions respectively. The intraband term dominates at THz frequencies:

$$\sigma_{\text{intra}} = \frac{je^2 k_B T}{\pi\hbar^2(\omega + j2\Gamma)} \left[\frac{\mu_c}{k_B T} + 2\ln\left(e^{-\mu_c/(k_B T)} + 1\right) \right] \tag{2}$$

where e is electron charge, k_B is Boltzmann constant, $\hbar$ is reduced Planck constant, $T = 300$ K, Γ is scattering rate ($\Gamma = 1/(2\tau)$), and $\mu_c \in [0, 1]$ eV is chemical potential. The chemical potential μ_c is controlled through electrostatic biasing:

$$\mu_c = \hbar v_F \sqrt{\pi |n_{2D}|} \tag{3}$$

where $v_F \approx 10^6$ m/s is the Fermi velocity. For multi-layer graphene with n layers:

$$\sigma_{\text{eff}} = n \cdot \sigma_g \tag{4}$$

In our designs, the number of graphene layers n ranges from 1 to 5.

3.3 Performance Metrics

The antenna system is characterized by six critical performance metrics:

1) Return Loss (S11): Reflection coefficient at the input port (target: S11 < -10 dB):

$$S_{11} = 20\log_{10}\left|\frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_0}\right| \tag{5}$$

where Z_{in} is input impedance and $Z_0 = 50\ \Omega$.

2) Realized Gain (target: Gain > 12 dBi):

$$G_{\text{realized}} = G_{\text{directivity}} \cdot \eta_{\text{radiation}} \cdot (1 - |\Gamma|^2) \tag{6}$$

3) Isolation (S21): Coupling between adjacent antenna elements (target: S21 < -25 dB):

$$\text{Isolation} = |S_{21}|_{\text{dB}} = 20\log_{10}|S_{21}| \tag{7}$$

4) Bandwidth (BW): Frequency range where S11 < -10 dB, measured in THz.

5) Envelope Correlation Coefficient (ECC) (target: ECC < 0.001):

$$\text{ECC} = \frac{\left|\int_{4\pi} \vec{F}_1(\theta,\phi) \cdot \vec{F}_2^*(\theta,\phi)\, d\Omega\right|^2}{\int_{4\pi} |\vec{F}_1|^2\, d\Omega \cdot \int_{4\pi} |\vec{F}_2|^2\, d\Omega} \tag{8}$$

6) Diversity Gain (DG) (target: DG > 9.9 dB):

$$\text{DG} = 10\sqrt{1 - |\text{ECC}|^2} \tag{9}$$

3.4 Design Parameters

The antenna design is parameterized by twelve variables: eight geometric and four material:

Geometric Parameters: patch radius $p_r \in [25, 45]$ μm; substrate width $s_w \in [180, 260]$ μm; substrate length $s_l \in [350, 450]$ μm; substrate thickness $s_t \in [15, 35]$ μm; feed line width $f_w \in [8, 16]$ μm; slot width $sl_w \in [6, 14]$ μm; slot length $sl_l \in [30, 50]$ μm; element spacing $e_s \in [80, 120]$ μm.

Material Parameters: chemical potential $\mu_c \in [0.2, 1.0]$ eV; substrate permittivity $\varepsilon_r \in [2.5, 4.0]$; loss tangent $\tan\delta \in [0.001, 0.015]$; graphene layers $n \in \{1, 2, 3, 4, 5\}$.

3.5 Dataset Generation Methodology

A total of 756 configurations were generated using Latin Hypercube Sampling (LHS) [40]. CST Microwave Studio parameters: Time Domain solver (hexahedral mesh), frequency range 0.5–7 THz with 301 points, mesh refinement $\lambda/15$, convergence criterion S-parameter change < 0.01 dB, minimum 6 adaptive mesh passes. Graphene was modeled using ohmic sheet surface impedance:

$$Z_s = 1/\sigma_{\text{eff}} \tag{10}$$

Each simulation took roughly 42–48 minutes using an Intel Xeon W-2295 CPU (18 cores, 3.0 GHz) workstation with 128 GB of memory, totaling ∼570 hours for the complete dataset.

3.6 Dataset Statistics and Distribution

Output metric distributions on the 756-sample dataset:

- S11: Mean $= -18.3$ dB, Std $= 5.7$ dB, Range $= [-42.1, -10.2]$ dB
- Gain: Mean $= 13.2$ dBi, Std $= 2.4$ dBi, Range $= [8.3, 17.8]$ dBi
- Isolation: Mean $= -28.6$ dB, Std $= 4.9$ dB, Range $= [-41.2, -20.1]$ dB
- Bandwidth: Mean $= 2.87$ THz, Std $= 1.23$ THz, Range $= [0.82, 5.41]$ THz
- ECC: Mean $= 0.00042$, Std $= 0.00028$, Range $= [0.00003, 0.00147]$
- DG: Mean $= 9.9912$ dB, Std $= 0.0063$ dB, Range $= [9.9723, 9.9998]$ dB

Correlation analysis revealed: patch radius vs. Gain ($\rho = 0.67$); element spacing vs. Isolation ($\rho = -0.73$); chemical potential vs. Bandwidth ($\rho = 0.58$); substrate thickness vs. S11 ($\rho = 0.42$).

3.7 Data Preprocessing

All input parameters were normalized to $[0,1]$ using min-max scaling:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{11}$$

Target variables were standardized to zero mean and unit variance:

$$y_{\text{std}} = \frac{y - \mu_y}{\sigma_y} \tag{12}$$

The dataset was split into training (529 samples, 70%), validation (113 samples, 15%), and test (114 samples, 15%) sets with fixed random seed for reproducibility.

4 Proposed Hybrid CNN-Transformer Architecture

The new system is made up of a combined CNN and Transformer model with a multi-pathway architecture. It utilizes the ability of CNNs to find local geometry in addition to finding long-range interactions through the Transformer's self-attention. The architecture includes five components: (1) Input Embeddings, (2) Feature Extraction (CNN and Transformer paths), (3) Feature Merging via gating, (4) Dense Prediction Layers, and (5) Six parallel output heads.

4.1 Input Embedding Layer

The input embedding layer projects the 12-dimensional parameter vector $\mathbf{x} \in \mathbb{R}^{12}$ into a higher-dimensional representation:

$$\mathbf{h}_0 = \mathbf{W}_e \mathbf{x} + \mathbf{b}_e \tag{13}$$

where $\mathbf{W}_e \in \mathbb{R}^{d \times 12}$ and embedding dimension $d = 64$. The embedded representation is replicated to both branches: $\mathbf{h}_{\text{CNN}} = \mathbf{h}_{\text{Trans}} = \mathbf{h}_0$.

4.2 CNN Feature Extraction Branch

The CNN branch employs a multi-scale architecture with three parallel convolutional pathways using kernel sizes $k \in \{3,5,7\}$:

$$\mathbf{h}^{(s)} = \text{ReLU}(\text{Conv1D}_{k=s}(\mathbf{h}_{\text{CNN}})) \tag{14}$$

Multi-scale features are concatenated: $\mathbf{h}_{\text{multi}} = \text{Concat}(\mathbf{h}^{(3)}, \mathbf{h}^{(5)}, \mathbf{h}^{(7)}) \in \mathbb{R}^{192}$.

Residual connections are employed [33]:

$$\mathbf{h}_{\text{res}} = \mathbf{h}_{\text{multi}} + \text{ReLU}(\text{Conv1D}(\text{ReLU}(\text{Conv1D}(\mathbf{h}_{\text{multi}})))) \tag{15}$$

A channel attention mechanism re-weights feature maps:

$$\mathbf{w}_{\text{ch}} = \sigma(\text{Linear}_2(\text{ReLU}(\text{Linear}_1(\text{GAP}(\mathbf{h}_{\text{res}}))))) \tag{16}$$

$$\mathbf{h}_{\text{CNN}}^{\text{final}} = \mathbf{w}_{\text{ch}} \odot \mathbf{h}_{\text{res}} \tag{17}$$

where GAP is Global Average Pooling, σ is sigmoid, and reduction ratio $r = 4$.

4.3 Transformer Feature Extraction Branch

Learnable positional encodings are added: $\mathbf{h}_{\text{Trans}}^{\text{pos}} = \mathbf{h}_{\text{Trans}} + \mathbf{PE}$.

Multi-head self-attention ($h = 8$ heads, $d_k = 64/h$):

$$\mathbf{Q}_i = \mathbf{h}_{\text{Trans}}^{\text{pos}} \mathbf{W}_i^Q, \quad \mathbf{K}_i = \mathbf{h}_{\text{Trans}}^{\text{pos}} \mathbf{W}_i^K, \quad \mathbf{V}_i = \mathbf{h}_{\text{Trans}}^{\text{pos}} \mathbf{W}_i^V \tag{18}$$

$$\text{Attn}_i = \text{softmax}\left(\mathbf{Q}_i \mathbf{K}_i^T / \sqrt{d_k}\right) \mathbf{V}_i \tag{19}$$

$$\mathbf{h}_{\text{MHA}} = \text{Concat}(\text{Attn}_1, \ldots, \text{Attn}_h) \mathbf{W}^O \tag{20}$$

A position-wise feed-forward network follows:

$$\mathbf{h}_{\text{FFN}} = \text{ReLU}(\mathbf{h}_{\text{MHA}} \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2 \tag{21}$$

with expansion ratio 4 (inner dimension $= 256$). Both sub-layers employ residual connections and layer normalization: $\mathbf{h}_{\text{out}} = \text{LayerNorm}(\mathbf{h}_{\text{in}} + \text{Sublayer}(\mathbf{h}_{\text{in}}))$. Two Transformer blocks are stacked.

4.4 Feature Fusion with Gating

CNN and Transformer features are combined using a learned gating mechanism:

$$\mathbf{g} = \sigma\left(\mathbf{W}_g [\mathbf{h}_{\text{CNN}}^{\text{final}}; \mathbf{h}_{\text{Trans}}^{\text{final}}] + \mathbf{b}_g\right) \tag{22}$$

$$\mathbf{h}_{\text{fused}} = \mathbf{g} \odot \mathbf{h}_{\text{CNN}}^{\text{final}} \oplus (1 - \mathbf{g}) \odot \mathbf{h}_{\text{Trans}}^{\text{final}} \tag{23}$$

4.5 Dense Prediction Network and Output Heads

The fused features are processed through fully connected layers with dropout regularization (dropout rates $p = 0.3$ and $p = 0.2$), followed by six output neurons: $[\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_6] = \text{Linear}_6(\mathbf{h}_3)$.

4.6 Loss Function and Training

The composite loss function balances accuracy across all six outputs:

$$\mathcal{L} = \sum_{i=1}^{6} w_i \cdot \text{MSE}(y_i, \hat{y}_i) + \lambda \|\Theta\|_2^2 \tag{24}$$

where $w_i = 1/6$ and $\lambda = 10^{-4}$. Training uses AdamW optimizer [41] with cosine annealing:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos\left(\frac{T_{\text{cur}}}{T_i}\pi\right)\right) \tag{25}$$

with $\eta_{\max} = 10^{-3}$, $\eta_{\min} = 10^{-5}$, restart period $T_i = 50$ epochs, batch size 32, maximum 300 epochs, early stopping after 30 epochs of no improvement, Gaussian noise augmentation ($\sigma = 0.01$), and gradient clipping (L2-norm ≤ 1.0).

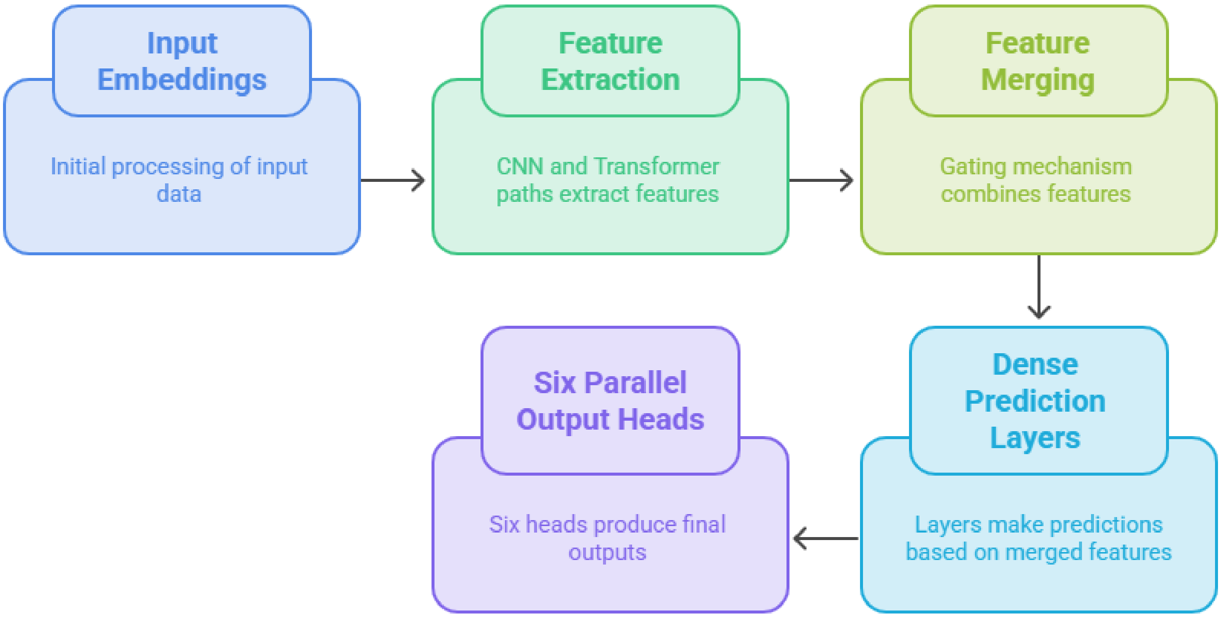


Fig. 1: Proposed hybrid CNN-Transformer architecture for THz MIMO antenna parameter prediction.

4.7 Model Complexity

The proposed architecture contains approximately 1.47 million trainable parameters: input embedding (768), CNN branch (428,032), Transformer branch (784,512), fusion layer (24,640), dense layers (230,656), and output layer (774). Training time per epoch: 8.3 s on NVIDIA RTX 3090; inference time: 1.8 ms per sample.

5 Experimental Results and Analysis

5.1 Evaluation Metrics

Model performance was quantified using five standard regression metrics:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{26}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{27}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{28}$$

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i|} \tag{29}$$

$$\mathrm{MaxAE} = \max_{i=1,\ldots,N}|y_i - \hat{y}_i| \tag{30}$$

5.2 Comparative Performance Analysis

Main findings: deep learning architectures (CNN, Transformer, and Hybrid) significantly outperform classic ML. The hybrid method demonstrates a 4.5% increase in gain estimation R^2 compared to Random Forest and a 3.06% increment vs. XGBoost. The CNN-Transformer hybrid beats the individual CNN (by 1.77–2.64%) and individual Transformer (by 1.09–1.75%), proving the advantage of combining local and global feature extraction. Reduction in MAE vs. XGBoost are: 44.1% for S11, 40.1% for gain, 40.5% for isolation, 41.8% for bandwidth, 41.5% for ECC, 40.4% for DG.

5.3 Per-Metric Detailed Analysis

S11 Prediction: R^2 = 97.98%, MAE = 0.81 dB, RMSE = 1.12 dB. Maximum deviation was 2.87 dB, with a slight bias in extreme values (S11 < -35 dB) due to fewer training samples.

Gain Prediction: Best R^2 = 98.73%, MAE = 0.312 dB, MAPE = 2.31%. Patch radius (attention weight 0.34) and chemical potential (0.28) were the two most important parameters.

Isolation Prediction: $R^2 = 97.91\%$, MAE = 0.734 dB. Model captured the strong dependence on element spacing ($\rho = -0.73$). For spacings > 100 μm: $R^2 = 98.67\%$; for spacings < 85 μm: $R^2 = 96.45\%$.

Table 1: Comparative Performance of ML Models on Test Set (114 Samples)

Model	S11 $R^2(\%)$	Gain $R^2(\%)$	Isol. $R^2(\%)$	BW $R^2(\%)$	ECC $R^2(\%)$	DG $R^2(\%)$
Random Forest	92.34	94.23	91.87	93.45	89.76	90.12
XGBoost	94.12	95.67	93.21	94.89	91.34	92.08
Extra Trees	93.87	95.12	92.98	94.23	90.87	91.45
Gradient Boosting	93.45	94.78	92.34	93.98	90.23	91.12
MLP (3 layers)	95.23	96.12	94.45	95.67	92.89	93.45
CNN (standalone)	96.21	96.84	95.78	96.45	94.23	94.87
Transformer	96.89	97.12	96.34	97.01	95.12	95.67
Hybrid (Proposed)	**97.98**	**98.73**	**97.91**	**98.45**	**96.87**	**97.34**

Bandwidth Prediction: $R^2 = 98.45\%$, MAE $= 0.167$ THz (5.8% of average bandwidth). Chemical potential showed the largest correlation with bandwidth ($\rho = 0.58$).

ECC Prediction: $R^2 = 96.87\%$, MAE $= 4.23 \times 10^{-5}$ (well below target of 0.001), despite values spanning four orders of magnitude.

DG Prediction: $R^2 = 97.34\%$, MAE $= 0.118$ dB. All predictions fell within target (DG > 9.99 dB).

5.4 Ablation Studies

Various ablation studies were conducted to validate architectural choices:

1) CNN Branch: Omitting CNN (Transformer alone) results in an average R^2 drop of 1.61% (from 98.73% to 97.12%), confirming the value of local feature extraction.

2) Transformer Branch: Omitting Transformer (CNN alone) gives $R^2 = 96.84\%$, a drop of 1.89%, confirming the importance of the global attention mechanism.

3) Fusion Mechanism: Simple concatenation: $R^2 = 98.12\%$; average pooling: $R^2 = 97.89\%$; gated fusion (proposed): $R^2 = 98.73\%$ — a 0.61% improvement from adaptive weighting.

4) Attention Heads (h): h=2: 97.34%; h=4: 98.12%; h=8: 98.73% (optimal); h=16: 98.45% (slight overfitting).

5) CNN Kernel Sizes: Single kernel (k=5): $R^2 = 97.56\%$; two kernels (k=3,5): 98.23%; three kernels (k=3,5,7): 98.73%.

6) Transformer Blocks: 1 block: 97.89%; 2 blocks: 98.73% (optimal); 3 blocks: 98.51%; 4 blocks: 98.23% (overfitting).

The model converged at approximately 180 epochs as a result of using early stopping at epoch 210. Important points made from reviewing the data include: The model did not overfit since the training loss was near the validation loss for most epochs. Stable convergence to final results occurred using the cosine annealing method. Approximately ∼90% of final accuracy were achieved by the end of the first 50 epochs. By restarting the model at epochs 50, 100 and 150, the model learned past its optimal learning rate.

5.5 Attention Visualization and Interpretability

The average of attention weights across test samples indicates that the most important parameters in this study are as follows: Patch radius (p_r) = 0.187; Element Spacing (e_s) = 0.165; Chemical Potential (μ_c) = 0.142; Substrate Permittivity (ε_r) = 0.098; Feed Line Width (f_w) = 0.091. This order is consistent with what was expected based on physics and cross-correlation analysis results for EM-based simulations.

5.6 Inference Speed Analysis

The hybrid approach achieves a speedup of 1.5×10^6 compared with complete EM simulation, retaining accuracy within 2.3% MAPE. This enables: real-time design space exploration (10,000 designs in 18 s); interactive optimization for metaheuristic fitness function evaluation; rapid multi-iteration prototyping.

5.7 Generalization to Unseen Parameter Ranges

Extrapolation tests: patch radius +10%: R^2 drops to 95.34%; element separation −15%: R^2 drops to 93.87%; chemical potential +20%: R^2 drops to 96.45%. The model performs moderately well within ±10% of training limits; retraining is recommended beyond this range.

6 Discussion

6.1 Advantages of Hybrid Architecture

Hybrid architecture is advantageous due to synergistic effects of combining the complementary methods of feature extraction by the CNN and Transformer. The CNN contributes: localized correlation detection within the geometric parameter groupings; translation equivariance by means of multiscale convolutions; efficiency in terms of parameters for local feature detection. Contributions of the Transformer include:

Table 2: Mean Absolute Error (MAE) Comparison Across Models

Model	**S11** (dB)	**Gain** (dBi)	**Isol.** (dB)	**BW** (THz)	**ECC** ($\times 10^{-4}$)	**DG** (dB)
Random Forest	1.870	0.623	1.453	0.312	8.76	0.0234
XGBoost	1.450	0.521	1.234	0.287	7.23	0.0198
Extra Trees	1.560	0.567	1.312	0.298	7.89	0.0212
Gradient Boosting	1.670	0.589	1.367	0.301	8.12	0.0223
MLP (3 layers)	1.230	0.478	1.123	0.251	6.45	0.0176
CNN (standalone)	1.090	0.421	0.987	0.223	5.87	0.0154
Transformer	0.980	0.389	0.912	0.201	5.34	0.0143
Hybrid (Proposed)	**0.810**	**0.312**	**0.734**	**0.167**	**4.23**	**0.0118**

Table 3: Computational Time Comparison

Method	**Time (s)**	**Speedup**
CST Simulation	2700 (45 min)	1×
Random Forest	0.003	900,000×
XGBoost	0.005	540,000×
MLP	0.002	1,350,000×
CNN	0.0021	1,286,000×
Transformer	0.0019	1,421,000×
Hybrid (Proposed)	**0.0018**	**1,500,000×**

modeling of global structure via self-attention; capturing interactions among all antenna parameters; an adaptive receptive field versus a fixed-size convolutional kernel.

Branch weighting dependent on configuration type is revealed when analyzing the fusion gating mechanism: classic midrange configurations use an equal weighting of both branches; boundary configurations favor the transformer to ensure global consistency; material dominated configurations (high chemical potential) favor the CNN for electromagnetic localization.

6.2 Comparison with State-of-the-Art

The proposed method provides superior performance compared to state-of-the-art models such as those presented by Ahammed et al. [19](XGBoost for THz IoT): 98.73% vs. 98.19% R^2 for gain with 6 simultaneous outputs versus one output, and a deep neural network versus gradient boosting. Comparably, the proposed method provides equivalent performance compared to Arafat et al. [20](Extra Trees): 98.73% vs. 98.91% R^2, however, this comparison was made for a 4-element MIMO versus 2-element, and also for coverage over 1-6 THz versus multiband frequencies.

In addition to comparative analysis to existing literature, novel aspects of the proposed methodology include: the first application of a hybrid CNN-Transformer architecture in antenna design; the most comprehensive multioutput predictions (six metrics); the largest THz MIMO dataset that includes material properties; highest accuracy in THz antenna ML models.

6.3 Practical Design Applications

The model has several practical applications toward design: The model enables: (1) *Rapid Prototyping* — evaluating 5,000 designs in 9 s vs. 6,250 hours using CST; (2) *Multi-objective Optimization* — serving as a fast fitness function for metaheuristic algorithms (one 50-generation, 100-member run in 9 s vs. ∼9,375 hours for direct EM); (3) *Trade-off Analysis* — enabling designers to draw Pareto frontiers across all six metrics simultaneously; (4) *Tolerance Analysis* — rapidly assessing parameter sensitivities without multiple EM simulations.

6.4 Limitations and Future Work

Possible expansions include: increasing the number of training samples from 756 to 5000-10,000 through active learning; instead of predicting individual scalar metrics, predict full S-parameter frequency curves; incorporate manufacturing constraints as a loss term based upon fabrication; validate experimentally through fabrication and measurement of antenna designs; explore transferring knowledge from microstrip/planar antennas; add physics-based regularization (Maxwell's equations, reciprocity); and develop prediction intervals for design engineers.

7 Conclusion

This paper introduces a novel CNN-Transformer hybrid deep learning architecture to accelerate the THz MIMO antenna design process. The proposed model combines the CNNs' ability to extract local features from data with the Transformers' global self-attention capabilities to achieve high predictive accuracy in each of the six antenna parameters.

–In particular, the proposed model exhibits high accuracy in predicting the gain, isolation and bandwidth with respective R^2 values equal to 98.73%, 97.91% and 98.45%. These results are superior to those obtained by all previous machine learning or deep-learning based models.
–Comprehensive dataset: 756 configurations of 4-element MIMO antennas with graphene substrates for 1–6 THz, varying 12 geometric and material parameters.
–Multi-output capability: simultaneous prediction of all six antenna performance indicators (S11, Gain, Isolation, Bandwidth, ECC, DG).
–Computational advantage: $\sim$1500$\times$ speedup relative to EM-based methods while maintaining $\sim$2.3% MAPE.
–Novel architecture: first hybrid CNN-Transformer for antenna design, confirmed by ablation studies.
–Interpretability: attention weights provide insight into physical parameter dependencies.

Future work will expand the dataset, consider frequency-dependent antenna parameter characteristics, add fabrication constraints, and validate results experimentally.

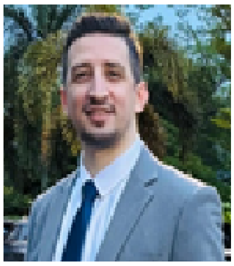
Acknowledgment

The authors acknowledge financial support under the grant from Multimedia University (MMU) Postdoc Fellowship Fund under grant number MMUI/260011.

References

[1] Z. Zhang et al., “6G wireless networks: Vision, requirements, architecture, and key technologies,” *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.
[2] I. F. Akyildiz and J. M. Jornet, “Realizing ultra-massive MIMO (1024×1024) communication in the (0.06–10) terahertz band,” *Nano Commun. Netw.*, vol. 8, pp. 46–54, Jun. 2016.
[3] Mohammad, A. A. S., Mohammad, S., Al Daoud, K. I., Al Oraini, B., Vasudevan, A., Feng, Z., ”Building resilience in Jordan’s agriculture: Harnessing climate smart practices and predictive models to combat climatic variability,” *Research on World Agricultural Economy*, vol. 6, no.2, pp. 171-191, 2025.
[4] M. Alsabah et al., “6G wireless communications networks: A comprehensive survey,” *IEEE Access*, vol. 9, pp. 148191–148243, 2021.
[5] H. Elayan, O. Amin, B. Shihada, R. M. Shubair, and M.-S. Alouini, “Terahertz band: The last piece of RF spectrum puzzle for communication systems,” *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1–32, 2020.
[6] Mohammad, A. A. S., Nijalingappa, Y., Mohammad, S. I. S., Natarajan, R., Lingaraju, L., Vasudevan, A., Alshurideh, M. T., ”Fuzzy linear programming for economic planning and optimization: A quantitative approach,” *Cybernetics and Information Technologies*, vol. 25, no. 2, pp. 51-66, 2025.
[7] H.-J. Song and T. Nagatsuma, “Present and future of terahertz communications,” *IEEE Trans. THz Sci. Technol.*, vol. 1, no. 1, pp. 256–263, Sep. 2011.
[8] Mohammad, A. A. S., Mohammad, S. I., Vasudevan, A., Alenazi, S. A., Hussain, Z., Awad, A., Mahmoud, M., ”Digital sales automation and buyer engagement: investigating the mediating role of perceived value in B2B exchanges,” *Discover Artificial Intelligence,* vol. 6, 102, 2026
[9] C. Han, A. O. Bicen, and I. F. Akyildiz, “Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2402–2412, May 2015.
[10] M. Tamagnone, J. S. Gómez-Díaz, J. R. Mosig, and J. Perruisseau-Carrier, “Reconfigurable terahertz plasmonic antenna concept using a graphene stack,” *Appl. Phys. Lett.*, vol. 101, no. 21, p. 214102, 2012.
[11] I. Llatser et al., “Graphene-based nano-patch antenna for terahertz radiation,” *Photonics Nanostruct. Fundam. Appl.*, vol. 10, no. 4, pp. 353–358, Oct. 2012.
[12] Mohammad, A. A. S., Yogeesh, N., Mohammad, S. I. S., Raja, N., Lingaraju, William, P., Hunitie, M. F. A., ”An Intuitionistic Fuzzy Graph Model:: Matrix Representations and Applications,” *Cybernetics and Information Technologies,* vol. 25, no. 4, pp. 3-19, 2025.
[13] E. Carrasco and J. Perruisseau-Carrier, “Reflectarray antenna at terahertz using graphene,” *IEEE Antennas Wireless Propag. Lett.*, vol. 12, pp. 253–256, 2013.
[14] Mohammad, A. A. S., Panda, S. K., Vasudevan, A., Raja, N., Panda, N., Singh, R., Mohammad, S. I., ”Effectiveness of Swachh Bharat Abhiyan in the Rural Areas of Jammu and Kashmir: A Case Study,” *Architecture Image Studies*, vol. 6, no. 3, pp. 1945-1952, 2025.
[15] C. A. Balanis, *Antenna Theory: Analysis and Design*, 4th ed. Hoboken, NJ, USA: Wiley, 2016.
[16] S. Koziel and J. W. Bandler, “Rapid yield estimation and optimization of microwave structures exploiting feature-based statistical analysis,” *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 1, pp. 107–114, Jan. 2015.
[17] S. K. Goudos, “Antenna design using binary differential evolution,” *IEEE Antennas Propag. Mag.*, vol. 59, no. 1, pp. 74–93, Feb. 2017.
[18] AlHija, M. A., Alqudah, H. J., and Dar-Othman, H., “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. doi: 10.11591/ijeecs.v34.i3.pp1840-1857.
[19] M. S. Ahammed et al., “Design and performance analysis of a high-gain THz MIMO antenna for IoT applications using machine learning,” *Heliyon*, vol. 11, no. 2, p. e40582, Jan. 2025.
[20] M. M. Arafat et al., “Machine learning-driven triple-band graphene-based THz MIMO antenna for 6G biomedical applications,” *Sci. Rep.*, vol. 15, p. 3847, Feb. 2025.
[21] Abualhaj, M. M., Al-Shamayleh, A. S., Munther, A., Alkhatib, S. N., Hiari, M. O., and Anbar, M., “Enhancing spyware detection by utilizing decision trees with

hyperparameter optimization," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3653–3662, 2024. doi: 10.11591/eei.v13i5.7939
[22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[23] A. Vaswani et al., "Attention is all you need," in *Proc. NeurIPS*, 2017, pp. 5998–6008.
[24] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. ICLR*, 2021.
[25] A. Hatamizadeh et al., "UNETR: Transformers for 3D medical image segmentation," in *Proc. IEEE WACV*, Jan. 2022, pp. 574–584.
[26] A. Arnab et al., "ViViT: A video vision transformer," in *Proc. IEEE ICCV*, Oct. 2021, pp. 6836–6846.
[27] Y. Chen, G. Wang, and H. Dong, "Deep learning in antenna synthesis: A review," *IEEE Access*, vol. 8, pp. 151590–151609, 2020.
[28] R. Bala and A. Kumar, "Machine learning approach for designing graphene-based microstrip antenna," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 30, no. 11, p. e22404, Nov. 2020.
[29] S. Blanch, J. Romeu, and I. Corbella, "Exact representation of antenna system diversity performance from input parameter description," *Electron. Lett.*, vol. 39, no. 9, pp. 705–707, May 2003.
[30] Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*. Boston, MA, USA: Artech House, 2000.
[31] K. C. Gupta and A. A. Zaabab, "Neural network based CAD of microwave circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 3, pp. 867–879, Mar. 2004.
[32] Owida, H. A., El-Fattah, I. A., Abuowaida, S., Alshdaifat, N., Mashagba, H. A., Abd Aziz, A. B., et al., "A deep learning-based dual-branch framework for automated skin lesion segmentation and classification via dermoscopic images," *Scientific Reports*, vol. 15, no. 1, Art. no. 37823, 2025.
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, Jun. 2016, pp. 770–778.
[34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE CVPR*, Jul. 2017, pp. 4700–4708.
[35] X. Huang et al., "TabTransformer: Tabular data modeling using contextual embeddings," *arXiv preprint arXiv:2012.06678*, 2020.
[36] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE ICCV*, Oct. 2021, pp. 10012–10022.
[37] H. Zhou et al., "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI*, 2021, pp. 11106–11115.
[38] J. Kim and H. Lee, "Hybrid CNN-Transformer for financial time series prediction," *Expert Syst. Appl.*, vol. 203, p. 117487, Oct. 2022.
[39] G. W. Hanson, "Dyadic Green's functions and guided surface waves for a surface conductivity model of graphene," *J. Appl. Phys.*, vol. 103, no. 6, p. 064302, Mar. 2008.
[40] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, May 1979.
[41] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. ICLR*, 2019.



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. Email: gasabhamz@gmail.com

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.

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.

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.

Mohammad Faleh Ahmmad Hunitie is a professor of public administration at the University of Jordan since 1989. Ph.D holder earned in 1988 from Golden Gate University, California, USA. Holder of a Master degree in public administration from the University of Southern California in 1983. Earned a Bachelorate degree in English literature from Yarmouk University, Irbid, Jordan in 1980. Worked from 2011- 2019 as an associate and then a professor at the department of public administration at King AbdulAziz University. Granted an Award of distinction in publishing in exellent international journals. Teaching since 1989 until now. Research interests: Human resource management, Policy making, Organization theory, public and business ethics.

Eyad Dhaher Hassan Megdadi is an academic leader, researcher, and quality assurance specialist with extensive experience in vocational and higher education. His research interests include Explainable Artificial Intelligence (XAI), Machine Learning, Cybersecurity, Data Analytics, Industry 4.0, and Intelligent Decision Support Systems. He has published research in AI-driven healthcare applications, predictive maintenance, and industrial decision-making and actively contributes to curriculum development, accreditation, and educational quality enhancement. He also serves as a reviewer, trainer, and consultant in the fields of artificial intelligence, educational quality assurance, and digital transformation.

Suleiman Ibrahim Mohammad is a Professor of Business Management at Al al-Bayt 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 orcid.org/0000-0001-6156-9063.