DOI: https://doi.org/10.57233/ijsgs.v12i1.1031 ISSNp: 2488-9229; ISSNe: 3027-1118

IJSGS FUGUSAU VOL. 12(1) WEBSITE: https://fugus-ijsgs.com.ng

INTERNATIONAL JOURNAL OF SCIENCE FOR GLOBAL SUSTAINABILITY
(A PUBLICATION OF FACULTY OF SCIENCE, FEDERAL UNIVERSITY GUSAU, NIGERIA)

Deep Learning-Optimized Design of Asymmetric Optical Directional Couplers for High-Efficiency Coupling

Ibrahim Nasidi[1, *], Hilal Bello[2]

[1]Department of Electrical and Electronics Engineering, Usmanu DanFodiyo University, Sokoto, Nigeria.
[2]Department of Information and Communication Engineering, Usmanu DanFodiyo University, Sokoto, Nigeria,
[*] **Corresponding Author's Email & Phone No.:** ibronasidi@yahoo.com, hilal.bello@udusok.edu.ng, +234-8065498184)

ABSTRACT

This study investigates the application of a deep neural network to optimize the design of asymmetric optical directional couplers, addressing the limitations of traditional iterative methods, which are often resource-intensive and time-consuming. A feed-forward neural network was developed to predict the coupling length of a coupler based on its structural parameters. A large dataset of 30,000 samples was used to train, validate, and test the model. The network achieved a minimal mean square error (MSE) of 1.07×10−6, demonstrating high precision. The trained model exhibited a remarkable prediction accuracy of 99.88% for the coupling length. To validate these results, finite-difference time-domain (FDTD) simulations were conducted using the network's predictions. The simulations confirmed that injecting a Transverse Electric fundamental mode (TE0) into the access waveguide successfully excited a double mode (TE1) in the bus waveguide, achieving a high coupling efficiency of 99.5%. This research demonstrates that a deep learning approach can provide a rapid and accurate method for designing efficient mode-conversion-based optical asymmetric directional couplers.

Keywords: Deep learning, Neural network, Nanophotonics, Asymmetric directional coupler, Optics

1.0. INTRODUCTION

The ability to control light-matter interactions at the nanoscale has significantly advanced the field of optics, largely driven by progress in nanophotonics (Rivera & Kaminer, 2020). As the demand for sophisticated functionalities in nanophotonics components increases, the design process becomes increasingly complex and challenging. Such complex structures have enabled designs and applications in metasurfaces (metamaterials and plasmonics) for colour displays (Yang et al., 2021), detectors (Nordin et al., 2021), lasers (Azzam et al., 2020), and photonic crystals for slow light (Nasidi et al., 2022).

Although substantial progress has been made, the design and optimization of these components often rely on iterative processes. Conventional approaches typically begin by numerically solving Maxwell's equations for a proposed structure using specialized software. The results obtained are subsequently compared against the target, which may or may not match well, necessitating multiple iterations to meet the target. Despite its proven success, this iterative, trial-and-error methodology is frequently resource-intensive and time-consuming, particularly for complex designs (Yao et al., 2018). Other techniques, such as the reduced basis method for finite element method (FEM) (Hesthaven et al., 2022), the integrated evolutionary strategy (ES) (Ferraz et al., 2024), and inverse design (Nasidi et al., 2023), have been used to enhance the design and optimisation processes. However, these techniques are computationally expensive because of intensive power and memory requirements.

Data-driven machine learning techniques have recently emerged as powerful tools to augment these traditional design and optimization workflows. As a cornerstone of deep learning, neural networks (NN) process raw data to identify autonomously underlying patterns and representations essential for tasks like prediction or classification, operating without pre-defined rules (Barth & Becker, 2018). Machine learning/deep learning has been applied in health analysis (Aru et al., 2021; Habehh & Gohel, 2021; Okandeji et al., 2022), environmental sciences (Liu et al., 2022), material informatics (Sha et al., 2021), strategy decision-making (Lu, 2025), speech recognition (Vashisht et al., 2021), path loss prediction (Shaibu et al., 2024) and security (Liang et al., 2019). In nanophotonics, deep-learning NN has been applied to design optical waveguides (Alagappan & Png, 2019), photonic coupler devices (Tu et al., 2021), gratings (Hammond & Camacho, 2019), power splitters (Tahersima et al., 2019), diffractive networks (Liu et al., 2023), chiral devices (Li et al., 2019), metasurfaces (Ghorbani et al., 2021), slow light (Chen et al., 2023) and nanocavities (Li et al., 2021; Li et al., 2022).

Motivated by these advancements, this study will investigate a deep-learning approach to designing an asymmetric optical directional coupler. This approach aims to streamline the design process and provide a time-efficient and accurate design value for optimal performance.

2.0 ASYMMETRIC DIRECTIONAL COUPLER DESIGN

The proposed asymmetric directional coupler designed on a silicon-on-insulator (SOI) platform is illustrated in Figure 1(a). It comprises two closely located waveguides, named access and bus waveguides, forming an evanescent coupling region. The device geometry is defined by four key parameters: the access waveguide width (w_a), the bus waveguide width (w_b), the gap (g), and the coupler length (l_c). To support the fundamental transverse electric (TE0) mode, w_a and w_b are set to 400nm and 800nm, respectively.

When the TE0 mode is launched into the access waveguide, as depicted in the inset of Figure 1(b), it couples to the bus

DOI: https://doi.org/10.57233/ijsgs.v12i1.1031 ISSNp: 2488-9229; ISSNe: 3027-1118

IJSGS FUGUSAU VOL. 12(1) WEBSITE: https://fugus-ijsgs.com.ng

waveguide over a specific length, known as the coupling length. This length is determined by varying the gap between the waveguides while maintaining fixed waveguide widths.

Figure 1(b) displays the electric field distribution within the directional coupler. For a specific gap $g = 430nm$, and coupling length $l_c = 425nm$, the mode is effectively transferred to the bus waveguide, achieving a coupling efficiency of 96%.

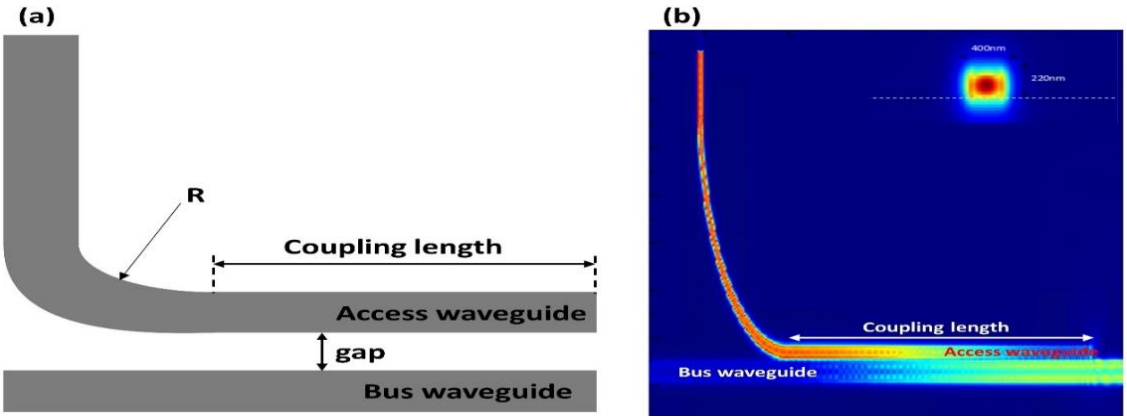

Figure 1: (a) Schematic diagram of an asymmetric directional coupler. (b) Electric field distribution in the asymmetric directional coupler. The inset shows the TE0 mode launched into the access waveguide.

3.0 DATA GENERATION

The asymmetric directional coupler, shown in Figure 1(b), was designed using Lumerical MODE Solution software. A substantial dataset of 30,000 discrete samples was generated from this design to serve as the basis for model training. Each sample in the dataset is defined by four key geometric parameters of the directional coupler.

The datasets were subsequently exported to MATLAB and Excel spreadsheets for post-processing. During the post-processing, the parameters were categorized as inputs and outputs for the model. Three parameters, namely, w_a, w_b, and g were defined as the model's inputs, while the coupling length (l_c) was set as the corresponding output. Moreover, the total dataset of 30,000 samples was subsequently partitioned into two distinct sets: 25,000 samples for model training and the remaining 5,000 for independent testing.

4.0 FEED-FORWARD NEURAL NETWORK

A feed-forward neural network (NN) is a fundamental deep learning architecture characterized by its layered structure. This multi-layered design allows the network to process and extract intricate, abstract patterns from input data. Through a series of interconnected neurons, the network learns to map complex, non-linear functions to a single, continuous output. This capability makes it suitable for tasks such as prediction, classification, and optimization of design parameters to achieve a desired performance.

4.1 NEURAL NETWORK ARCHITECTURE

Figure 2 presents a schematic diagram of the feed-forward deep NN model. This architecture comprises an input layer with three neurons, followed by two fully connected hidden layers, each containing 256 neurons, and a single-neuron output layer. Data processing proceeds in a forward direction, where each neuron receives input from all neurons in the previous layer and sends its output to all neurons in the following layer. This process is governed by a series of computational operations, including the application of weights, biases, and an activation function at each neuron.

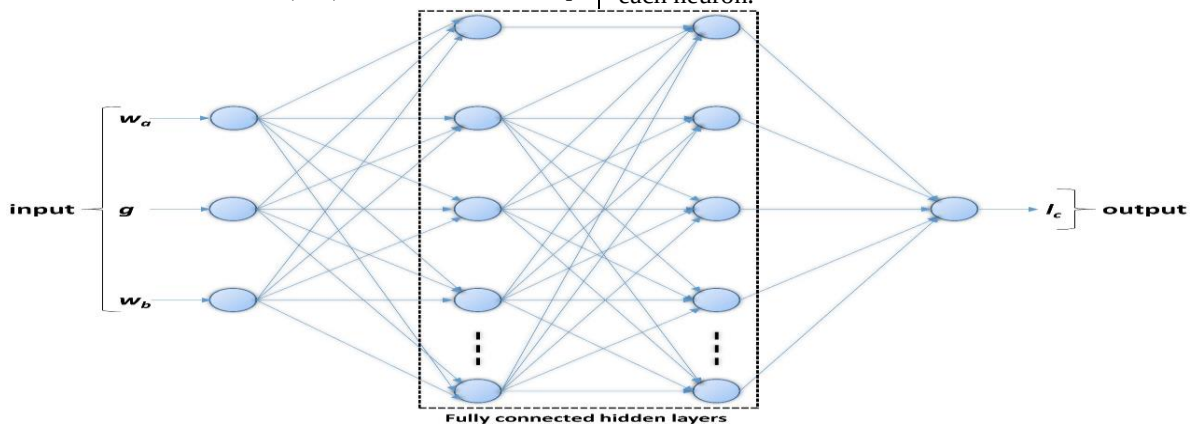

Figure 2: Schematic diagram of a feed-forward deep-learning model.

DOI: https://doi.org/10.57233/ijsgs.v12i1.1031 ISSNp: 2488-9229; ISSNe: 3027-1118

5.0 METHODOLOGY

5.1 DEEP-LEARNING NEURAL NETWORK IMPLEMENTATION

The NN was implemented, trained, and tested using the TensorFlow open-source framework within the Python programming environment. The neural network architecture comprises three (3) neurons at the input layer, one (1) neuron in the output layer, and two (2) hidden fully connected layers, each containing 256 neurons. All neurons, with the exception of those in the output layer, employed the rectified linear unit (ReLU) activation function. The mean square error (MSE) was used as the loss function. The root-mean-square backpropagation (RMSProp) algorithm was chosen for optimization, with a learning rate of 0.001. Given the large training dataset and the absence of overly complex patterns to learn, the batch size was set to 100, and the number of epochs was set to 30.

6.0 RESULT AND DISCUSSION

6.1 LOSS FUNCTION AND TRAINING

The NN was trained on a dataset split into an 80% training set and a 20% validation set. The validation set was used to monitor overfitting. Figure 3 illustrates the mean squared error (MSE) as a function of epochs (loss function) for both the training and validation datasets. Initially, the MSE was high for both sets but progressively decreased with an increasing number of epochs, demonstrating the model's learning capability. The MSE converged after 15 epochs, reaching a minimum of 1.07×10^{-6} for the training set and 2.75×10^{-6} for the validation set. This convergence indicates that the model successfully identified the underlying patterns within the data. A comparison of the curves reveals that the training MSE curve is smooth and remains slightly below the validation MSE curve, which is indicative of a good model fit. The minor difference in magnitude and the slight fluctuations in the validation error curve is consistent with the validation set being smaller than the training set.

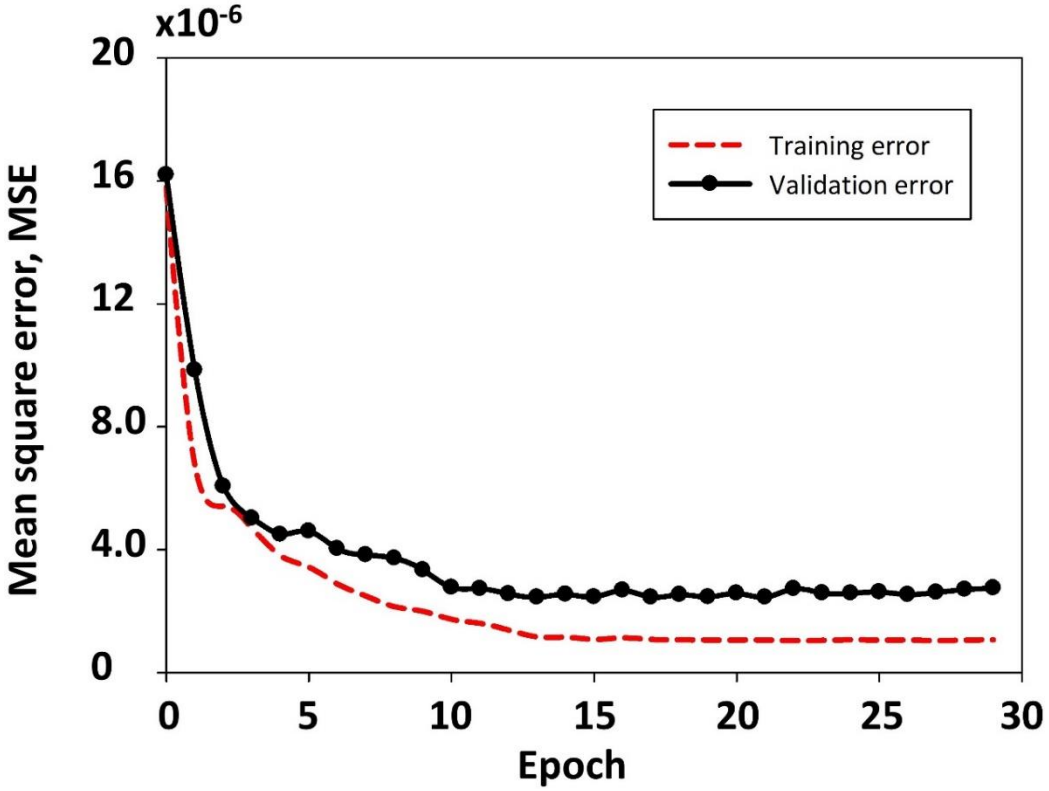

Figure 3: The learning curves for training and validation datasets. The mean squared error sharply declines, indicating that the network finds a pattern in the data.

6.2 HYPERPARAMETER OPTIMIZATION AND OVERFITTING MITIGATION

After an extensive hyperparameter search, it was found that the optimal model configuration involved adjusting the number of layers, learning rates, and number of neurons per hidden layer to minimize validation error. A notable observation was that increasing the number of neurons consistently led to a reduction in the MSE, thereby improving the model's performance. However, increasing the neuron count beyond 256 resulted in the vanishing gradient problem. Additionally, increasing the depth of the hidden layers did not significantly affect performance gains. A large learning rate, on the other hand, resulted in gradient explosion, an issue common in fully connected neural networks (Hesthaven et al., 2022). To mitigate overfitting, an early stopping callback was implemented. This mechanism monitored the validation error and automatically halted training if performance ceased to improve, ensuring that the model generalized well to unseen data.

6.3 PREDICTION ACCURACY

The coefficient of determination, or R-squared (R^2), is a metric used to evaluate how well a regression model

DOI: https://doi.org/10.57233/ijsgs.v12i1.1031 ISSNp: 2488-9229; ISSNe: 3027-1118

predicts outcomes. While most commonly used with linear regression, it can also be applied to a NN when the model is used for a regression task. Essentially, it measures the proportion of the variance in the dependent variable that the independent variables can explain. The value of R^2 ranges from 0 to 1, and it's often expressed as a percentage. A value of 1 (or 100%) indicates that the model perfectly predicts the outcome, while a value of 0 suggests the model doesn't explain any of the variability in the data. Therefore, a higher R^2 generally means a better-performing model. In this case, the output, or the dependent variable, is the coupling length l_c. The R^2 value ranges from 0 to 1, where 0 indicates no predictability (being the minimum) and 1 (being the maximum) represents a perfect prediction. R^2 is computed by comparing the actual output values (*y_actual*) with the NN predicted values (*y_predict*) using the following expressions:

$$R^2 = 1 - \frac{SSR}{SST} \quad (1)$$

$$SSR = \sum \left(y\_actual - y\_predict \right)^2 \quad (2)$$

$$SST = \sum \left(y\_actual - \overline{y\_actual} \right)^2 \quad (3)$$

$\overline{y\_actual}$ is the mean of *y_actual*. In statistical modeling, particularly linear regression, SST means Total Sum of Squares, and SSR means Regression Sum of Squares. They are key components used to measure the variability in a dataset.

6.4 PREDICTION ACCURACY ANALYSIS

Figure 4 shows the R^2 against the epoch for both training and validation datasets. As can be seen, the prediction accuracy for both datasets starts quite low but steadily increases with each epoch. This improvement in accuracy largely stops after approximately five epochs, which suggests the data isn't overly complex and further increases in epochs offer no additional accuracy benefits. The model demonstrated exceptional predictive accuracy, attaining R^2 values of 99.88% on the training data and 99.08% on the validation data. These high values indicate that the model fits the data well and can accurately predict the output, which is the coupling length (l_c).

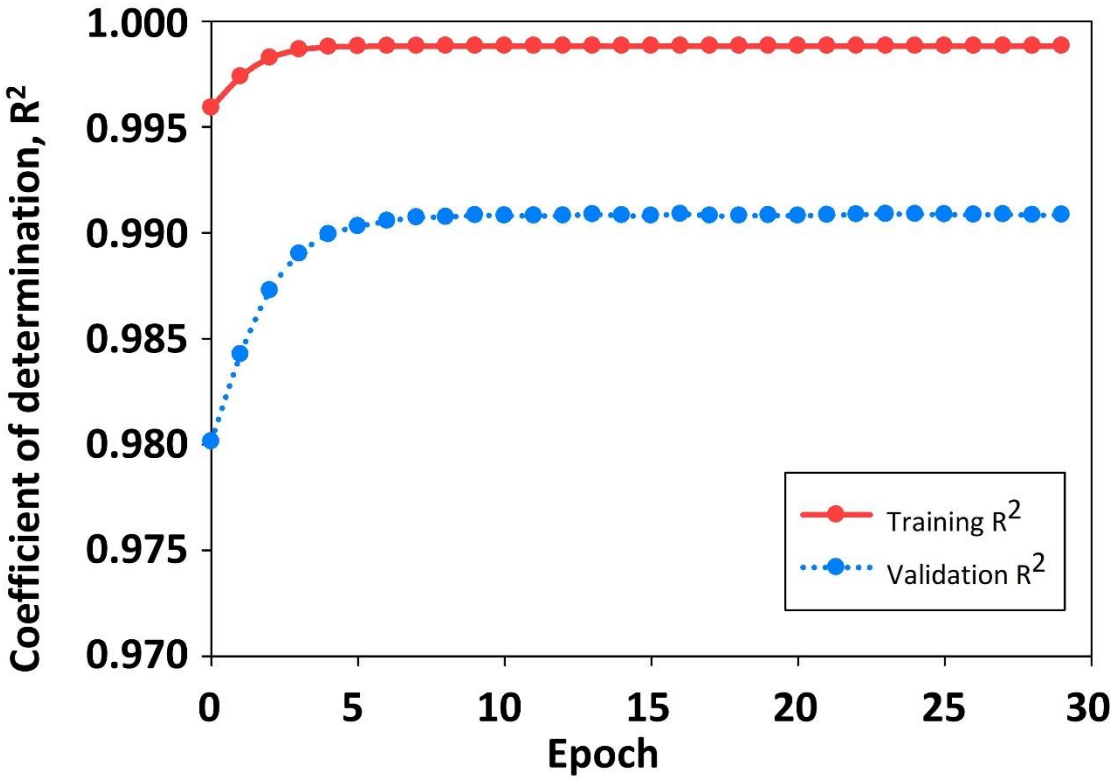

Figure 4: Coefficient of determination R^2 for training and validation data. The prediction accuracy R^2 for both datasets are above 99%.

7.0 EVALUATION OF THE MODEL

A test dataset of 5,000 samples was used to evaluate the model's performance. This dataset, taken from the larger pool of generated data, was specifically set aside for this purpose and was not used in the training process. This ensures an unbiased assessment of how well the model generalizes to new, unseen data. Crucially, these samples were unseen by the model during training. This evaluation

DOI: https://doi.org/10.57233/ijsgs.v12i1.1031 ISSNp: 2488-9229; ISSNe: 3027-1118

represents the initial test of the neural network's forward computation capabilities. The results are impressive: the test MSE is 1.105 x 10^{-6}, which agrees very well with the training MSE of 1.07 x 10^{-6}. The relative difference between the test and training MSE is 3.5 x 10^{-8}, which is also a very small value, indicating strong consistency. Moreover, the model achieved a test R^2 of over 90%. This shows that the NN has generalized exceptionally well, meaning it performs effectively on new, unseen data.

8.0 COUPLER LENGTH PREDICTION

With the deep learning NN model successfully established, it's ready for practical application. The next step is to confidently examine the network's forward computation in predicting the coupling length (l_c) of the asymmetric directional coupler. Given the input parameters of $w_a = 400nm$, $w_b = 800nm$ and $g = 180nm$, the model successfully predicted a coupling length of $l_c = 35.59\mu m$.

To validate the NN prediction accuracy and evaluate its performance, the predicted values are employed in Lumerical FDTD software, which allows for the design and simulation of the asymmetric directional coupler in three-dimensional (3D). The design parameters for the simulation are $n_{si} = 3.45$, $n_{SiO_2} = 3.45$, $w_a = 400nm$, $w_b = 800nm$, $g = 180nm$, and $l_c = 35.59\mu m$. A mode source, with a centre wavelength of $\lambda = 1550nm$ injects light into the structure.

9.0 ENHANCED COUPLING EFFICIENCY

Figure 5 shows the transmission or coupling efficiency recorded with a monitor at the bus waveguide. As is evident, the coupling efficiency at $\lambda = 1550nm$ is an impressive 99.5%. This simulation yielded a coupling efficiency of 99.5%, marking a considerable enhancement over the 96% efficiency obtained through pre-optimization methods.

This high efficiency shows that the TE0 single mode (as displayed in the inset of Figure 1b) in the access waveguide is efficiently coupled to the TE1 double/multi-mode in the bus waveguide. Specifically, the fundamental TE0 single mode from the access waveguide is effectively converted and coupled into a TE1 double/multi-mode within the bus waveguide, as depicted in the inset of Figure 5. Therefore, an asymmetric directional coupler can be effectively utilized in mode conversion steps within optical switches to selectively couple to specific modes (Zhu et al., 2015).

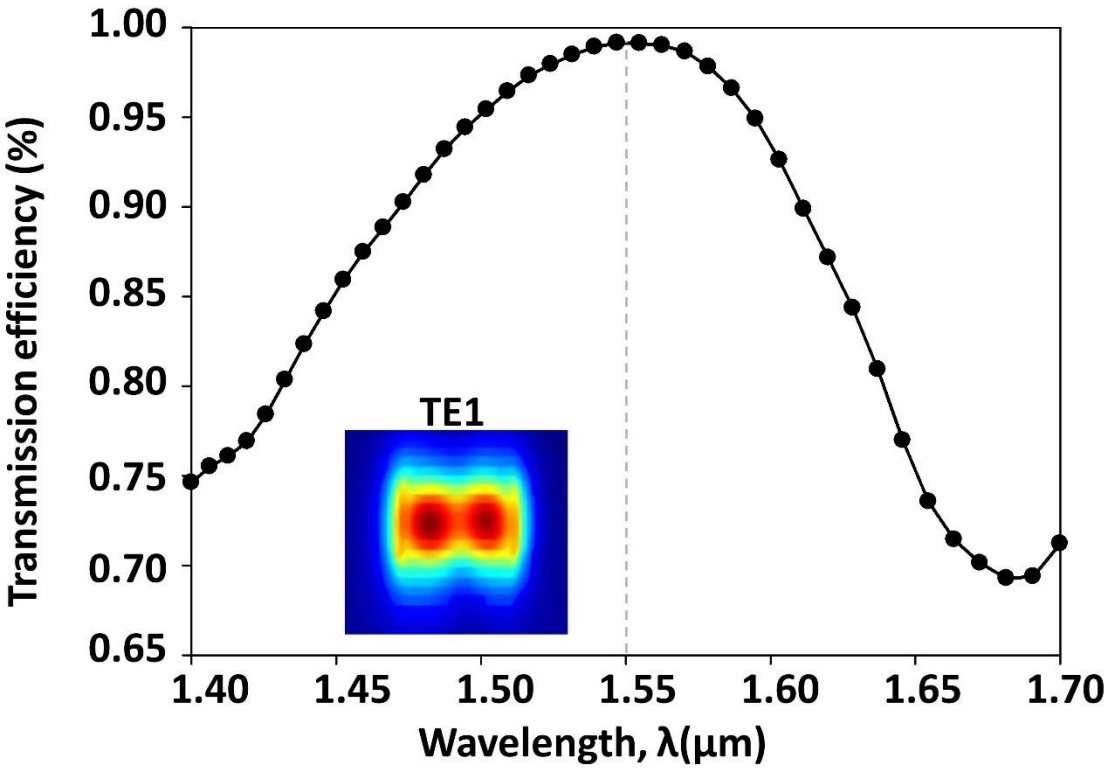

Figure 5: Transmission efficiency of an asymmetric directional coupler. At $\lambda = 1550nm$, the transmission efficiency is 99.5%. The inset depicts a TE1 mode in the bus waveguide at $\lambda = 1550nm$.

10.0 CONCLUSION

This paper proposes a novel deep-learning approach for the design optimization of an asymmetric directional coupler. A feed-forward NN with two hidden layers was trained on a dataset of 30,000 samples to predict the device's coupling length based on its geometrical parameters. The model achieved a low MSE of 1.07×10^{-6}, demonstrating its ability to accurately learn the complex relationship between the

DOI: https://doi.org/10.57233/ijsgs.v12i1.1031 ISSNp: 2488-9229; ISSNe: 3027-1118

coupler's structure and its performance.

Validation of the NN's prediction in a separate FDTD simulation confirmed its efficacy, yielding a high coupling efficiency of 99.5%. This result represents a significant improvement over the 96% efficiency achieved with conventional methods. The findings affirm the viability of employing deep-learning models for the rapid and accurate design of complex nanophotonic components, which can accelerate the development of advanced devices such as multiplexers, de-multiplexers, and mode converters (Zhu et al., 2015).

REFERENCES

Rivera, N., & Kaminer, I. (2020). Light–matter interactions with photonic quasiparticles. *Nature Reviews Physics*, *2*(10), 538–561. https://doi.org/10.1038/s42254-020-0224-2

Yang, W., Qu, G., Lai, F., Liu, Y., Ji, Z., Xu, Y., Song, Q., Han, J., & Xiao, S. (2021). Dynamic bifunctional metasurfaces for holography and color display. *Advanced Materials*, *33*(36), Article 2101258. https://doi.org/10.1002/adma.202101258

Nordin, L., Petluru, P., Kamboj, A., Muhowski, A. J., & Wasserman, D. (2021). Ultra-thin plasmonic detectors. *Optica*, *8*(12), 1545–1551. https://doi.org/10.1364/OPTICA.438039

Azzam, S. I., Kildishev, A. V., Ma, R.-M., Ning, C.-Z., Oulton, R., Shalaev, V. M., Stockman, M. I., Xu, J.-L., & Zhang, X. (2020). Ten years of spasers and plasmonic nanolasers. *Light: Science & Applications*, *9*(1), Article 90. https://doi.org/10.1038/s41377-020-0319-7

Nasidi, I., Hao, R., Chen, J., Li, E., & Jin, S. (2022). Photonic moiré lattice waveguide with large slow light bandwidth and delay-bandwidth product. *Applied Optics*, *61*(19), 5776–5781. https://doi.org/10.1364/AO.462016

Yao, K., Unni, R., & Zheng, Y. (2018). Intelligent nanophotonics: Merging photonics and artificial intelligence at the nanoscale. *Nanophotonics*, *8*(3), 339–366. https://doi.org/10.1515/nanoph-2018-0183

Hesthaven, J. S., Pagliantini, C., & Rozza, G. (2022). Reduced basis methods for time-dependent problems. *Acta Numerica*, *31*, 265–345. https://doi.org/10.1017/S0962492922000058

Ferraz, W. M., Gonçalves, M. S., & Santos, C. H. S. (2024). Evolutionary strategy combined to Fuzzy Logic and Shannon's Entropy to maximize the optimization of optical coupler design. *SoftwareX*, *27*, Article 101835. https://doi.org/10.1016/j.softx.2024.101835

Nasidi, I., Hao, R., Jin, S., & Li, E. P. (2023). Inverse design of a photonic moiré lattice waveguide towards improved slow light performances. *Applied Optics*, *62*(10), 2651–2655. https://doi.org/10.1364/AO.485059

Barth, C., & Becker, C. (2018). Machine learning classification for field distributions of photonic modes. *Communications Physics*, *1*(1), Article 58. https://doi.org/10.1038/s42005-018-0060-1

Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. *Current Genomics*, *22*(4), 291–300. https://doi.org/10.2174/1389202922666210705124359

Aru, O. E., Adimora, K. C., & Nwankwo, F. J. (2021). Investigating the impact of 5G radiation on human health using machine learning. *Nigerian Journal of Technology*, *40*(4), 694–702. https://doi.org/10.4314/njt.v40i4.16

Okandeji, A. A., Odeyinka, O. F., Sogbesan, A. A., & Ogunye, N. O. (2022). A comparative analysis of haemoglobin variants using machine learning algorithms. *Nigerian Journal of Technology*, *41*(4), 789–796. https://doi.org/10.4314/njt.v41i4.16

Liu, X., Lu, D., Zhang, A., Liu, Q., & Jiang, G. (2022). Data-driven machine learning in environmental pollution: Gains and problems. *Environmental Science & Technology*, *56*(4), 2124–2133. https://doi.org/10.1021/acs.est.1c06157

Sha, W., Li, Y., Tang, S., Tian, J., Zhao, Y., Guo, Y., Zhang, W., Zhang, X., Lu, S., Cao, Y.-C., & Cheng, S. (2021). Machine learning in polymer informatics. *InfoMat*, *3*(4), 353–361. https://doi.org/10.1002/inf2.12167

Lu, M. (2025). Automated machine learning and data-driven decision support system for strategy management in organizational activities. *CIT. Journal of Computing and Information Technology*, *33*(1), 57–71. https://doi.org/10.20532/cit.2025.1005877

Vashisht, V., Pandey, A. K., & Yadav, S. P. (2021). Speech recognition using machine learning. *IEIE Transactions on Smart Processing & Computing*, *10*(3), 233–239. https://doi.org/10.5573/IEIESPC.2021.10.3.233

Shaibu, F. E., Onwuka, E. N., Salawu, N., & Oyewobi, S. S. (2025). Evaluating the effectiveness of machine learning models for path loss prediction at 3.5 GHz with focus on feature prioritization. *Nigerian Journal of Technology*, *43*(4), 754–762. https://doi.org/10.4314/njt.v43i4.15

Liang, F., Hatcher, W. G., Liao, W., Gao, W., & Yu, W. (2019). Machine learning for security and the Internet of Things: The good, the bad, and the ugly. *IEEE Access*, *7*, 158126–158147. https://doi.org/10.1109/ACCESS.2019.2948912

Alagappan, G., & Png, C. E. (2019). Modal classification in optical waveguides using deep learning. *Journal of Modern Optics*, *66*(5), 557–561. https://doi.org/10.1080/09500340.2018.1552331

Tu, X., Xie, W., Chen, Z., Ge, M.-F., Huang, T., & Song, C. (2021). Analysis of deep neural network models for inverse design of silicon photonic grating coupler. *Journal of Lightwave Technology*, *39*(9), 2790–2799. https://doi.org/10.1109/JLT.2021.3057473

Hammond, A. M., & Camacho, R. M. (2019). Designing integrated photonic devices using artificial neural networks. *Optics Express*, *27*(21), 29620–29638. https://doi.org/10.1364/OE.27.029620

Tahersima, M. H., Kojima, K., Koike-Akino, T., Jha, D. K., Wang, B., Lin, C., & Parsons, K. (2019). Deep neural network inverse design of integrated photonic power splitters. *Scientific Reports*, *9*(1), Article 1368. https://doi.org/10.1038/s41598-018-37952-2

Liu, W., Fu, T., Huang, Y., Sun, R., Yang, S., & Chen, H. (2023). C-DONN: Compact diffractive optical neural network with deep learning regression. *Optics Express*, *31*(13), 22127–22143. https://doi.org/10.1364/OE.490072

Li, Y., Xu, Y., Jiang, M., Li, B., Han, T., Chi, C., Lin, F., Shen, B., Zhu, X., Lai, L., & Fang, Z. (2019). Self-learning perfect optical chirality via a deep neural network. *Physical Review Letters*, *123*(21), Article 213902. https://doi.org/10.1103/PhysRevLett.123.213902

Ghorbani, F., Beyraghi, S., Shabanpour, J., Oraizi, H., Soleimani, H., & Sonkusale, S. (2021). Deep neural network-based automatic metasurface design with a wide frequency range. *Scientific Reports*, *11*(1), Article 7102. https://doi.org/10.1038/s41598-021-86588-2

Chen, J., Hao, R., Nasidi, I., Zhang, H., Wang, X., & Jin, S. (2023). Deep learning-based modelling of complex photonic crystal slow light waveguides. *IEEE Journal of Selected Topics in Quantum Electronics*, *29*(6), 1–6. https://doi.org/10.1109/JSTQE.2023.3305450

Li, R., Gu, X., Li, K., Huang, Y., Li, Z., & Zhang, Z. (2021). Deep learning-based modeling of photonic crystal nanocavities. *Optical Materials Express*, *11*(7), 2122–2133. https://doi.org/10.1364/OME.425196

Li, R., Gu, X., Shen, Y., Li, K., Li, Z., & Zhang, Z. (2022). Smart and rapid design of nanophotonic structures by an adaptive and regularized deep neural network. *Nanomaterials*, *12*(8), Article 1372. https://doi.org/10.3390/nano12081372

Zhu, X., Stern, B., Chen, C. P., Tzuang, L. D., Cardenas, J., Bergman, K., & Lipson, M. (2015). On-chip mode-division multiplexing switch. *Optica*, *2*(6), 530–535. https://doi.org/10.1364/OPTICA.2.000530