11th LIS International Conference
Liga Ilmu Serantau 2025

ISSN: 2985-4393

Comparative Analysis of Machine Learning Algorithms for Wide-range Refractive Index Classification in Multimode Interference No-Core Fiber Sensors

Nurul Farah Adilla Zaidi[1, b)], Muhammad Yusof Mohd Noor[1, a)], Nur Najahatul Huda Saris[1, c)], Sumiaty Ambran[2, d)], Azizul Azizan[3, e)] and Joshu Leonardy[1, f)]

[1]*Faculty of Electrical Engineering,*
Universiti Teknologi Malaysia (UTM), 81310, Johor, Malaysia
[2]*Malaysian-Japan International Institute of Technology,*
Universiti Teknologi Malaysia (UTM), 54100, Kuala Lumpur, Malaysia
[3]*Faculty of Artificial Intelligence,*
Universiti Teknologi Malaysia (UTM), 54100, Kuala Lumpur, Malaysia

[a)] *Corresponding author: yusofnor@utm.my*
[b)] *nfadilla5@graduate.utm.my*
[c)] *nurnajahatulhuda@utm.my*
[d)] *sumiaty.kl@utm.my*
[e)] *azizulazizan@utm.my*
[f)] *joshu@graduate.utm.my*

Abstract. Multimode interference (MMI) fiber sensors based on no-core fiber (NCF) are attracting growing interest due to their structural simplicity, cost-effectiveness, and ability to operate without cladding removal or chemical modifications. These sensors are widely utilized in applications requiring precise refractive index (RI) detection, such as environmental monitoring and biomedical diagnostics. However, accurate classification across a wide RI range is challenging, particularly due to the transition between guided modes (when the surrounding RI is below the fiber RI) and leaky modes (when the surrounding RI is above the fiber RI). This study proposes a machine learning (ML)-based approach to address these challenges and enhance RI classification accuracy. Transmission spectra from an NCF-based MMI sensor were acquired under both low-RI (LRI) and high-RI (HRI) conditions and analyzed using MATLAB. Three ML classifiers involved, Decision Tree (DT), Support Vector Machine (SVM), and Neural Network (NN) were evaluated using standard metrics: accuracy, precision, recall, and F1-score. The NN model achieved the highest classification accuracy at 80.0%, outperforming SVM (77.5%) and DT (42.5%). These results demonstrate the effectiveness of ML, particularly NN, in improving the performance and reliability of fiber optic sensors for broad-range RI detection.

Keywords: Multimode interference (MMI), no-core fiber (NCF), refractive index (RI), machine learning (ML), classification, fiber optic sensors.

1. INTRODUCTION

Multimode interference (MMI) fiber sensors have emerged as a versatile and promising class of optical sensors due to their simple configuration, ease of fabrication, and cost-effective production [1, 2]. Among these, no-core fiber (NCF) based structures are particularly notable for their ability to detect refractive index (RI) changes without requiring cladding removal or chemical modification [3, 4]. These properties make NCF-based MMI sensors highly suitable for real-time and in-situ applications across biomedical diagnostics [5], chemical sensing [6, 7], and environmental monitoring [8].

The core operating principle of MMI sensors is based on the interference of multiple higher-order modes excited within the NCF when light from a single-mode fiber (SMF) enters it. These interference patterns are highly sensitive to the surrounding RI [9, 10]. Depending on whether the surrounding medium has a lower or higher RI than the fiber itself, the sensor transitions between two operational regimes: guided mode in low-RI (LRI) environments and leaky mode in high-RI (HRI) environments [11]. This guided-to-leaky mode transition introduces classification challenges due to spectral complexity and RI mismatch. Traditional signal processing approaches often fall short in interpreting the non-linear variations associated with this transition [12]. In response, data-driven methods such as machine learning (ML) offer a compelling solution, capable of learning intricate spectral patterns for robust RI prediction [13, 14].

In this study, we explore the use of three ML classifiers which were Decision Tree (DT), Support Vector Machine (SVM), and Neural Network (NN) to classify RI values based on spectral data acquired from an MMI-NCF fiber sensor. Our objective is to assess the comparative performance of these models and determine the most effective approach for accurate, wide-range RI detection using ML-enhanced optical sensing.

2. EXPERIMENTAL METHODOLOGY AND MODEL VALIDATION

2.1 NCF-based MMI Sensor Design and Experimental Setup

The sensor configuration utilized a 3.5 cm segment of 125 μm diameter NCF-F300, spliced between two SMF-28 fibers. Made of pure fused silica (RI ≈ 1.4440 at 1550 nm), the NCF has a numerical aperture (NA) of 0.47, supporting a broad range of modes. Spectral transmission was measured from 1520 to 1580 nm using a broadband Amplified Spontaneous Emission (ASE) source and optical spectrum analyzer (OSA). Certified Cargille RI liquids (Series A and AAA) were applied in two RI ranges: LRI (1.30 to 1.39 RIU) and HRI (1.46 to 1.55 RIU), with 0.01 RIU increments. Each RI class underwent 10 cycles of measurement, followed by thorough cleaning with isopropyl alcohol and tissue to avoid contamination.

2.2 Data Collection and Processing

A total of 200 transmission spectra were collected and uniformly sampled at 0.18 nm intervals, yielding 3340 data points per spectrum. The data was pre-processed using baseline correction and averaged to reduce noise. MATLAB was used for feature extraction and classification.

2.3 Dataset Splitting and Model Configuration

The dataset was split using an 80-20 train-test ratio with stratified sampling to maintain class distribution. Five-fold cross-validation was employed during training. MATLAB R2024b's Classification Learner toolbox was used for model training. DT used a maximum of 100 splits (Gini), SVM applied a linear kernel with box constraint 1, and NN used a single hidden layer with 10 neurons and ReLU activation.

2.4 Model Evaluation

Classification model performance was evaluated using standard metrics: accuracy, precision, recall, and F1-score. Accuracy reflects overall classification correctness, precision measures the proportion of true positive predictions, recall assesses the ability to identify actual positive cases, and F1-score provides a balanced metric that combines both precision and recall. MATLAB's visualization tools facilitated interpretation through confusion matrices, enabling a detailed examination of each model's ability to distinguish between closely spaced RI classes. This comprehensive methodology provides a solid foundation for comparing the classification effectiveness of DT, SVM, and NN in the context of NCF-based MMI sensor data, supporting the advancement of practical wide-range RI sensing applications.

3. RESULTS AND DISCUSSION

3.1 Spectral Analysis

The MMI-NCF sensor exhibited distinct spectral responses across both LRI and HRI ranges. In the LRI region (1.30-1.39 RIU), noticeable wavelength shifts, and intensity variations were observed as in Figure 1, attributed to enhanced light-medium interaction. Conversely, the HRI region (1.46-1.55 RIU) showed relatively stable wavelengths with dominant intensity fluctuations in Figure 2, characteristic of leaky mode propagation. These spectral features were extracted and processed using MATLAB. Specifically, the entire 3340 point pre-processed transmission spectrum vector was used as input to the ML models. This automated approach enables more consistent and precise RI prediction, addressing limitations commonly encountered with manual interpretation of overlapping peaks or minor intensity changes.

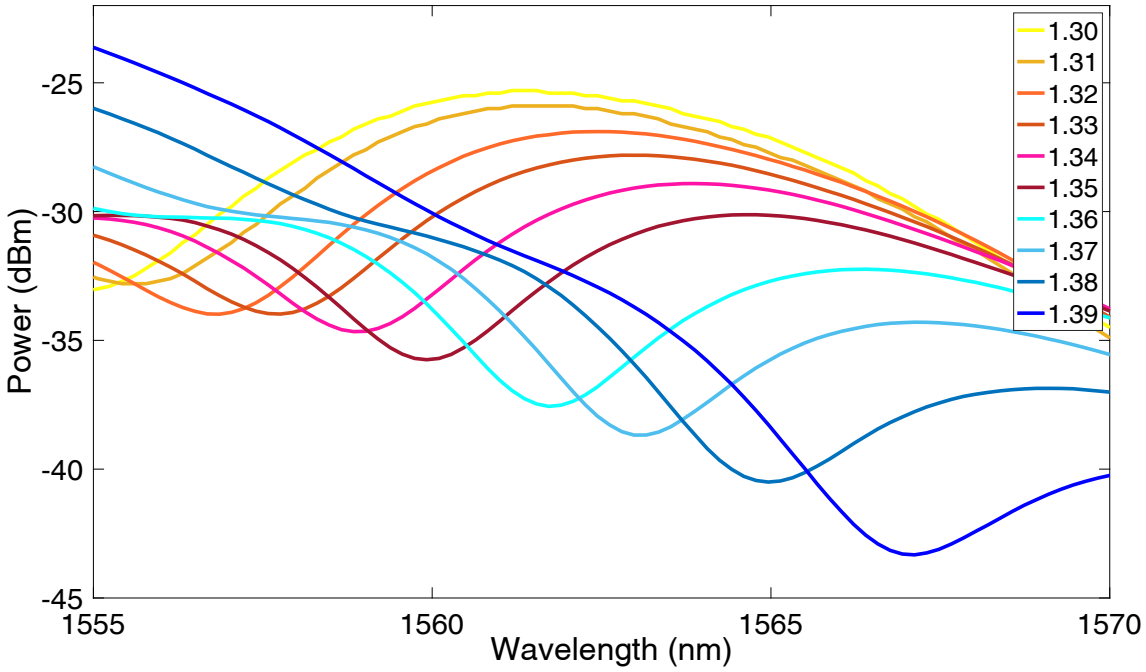

Figure 1. Spectral behavior for LRI region.

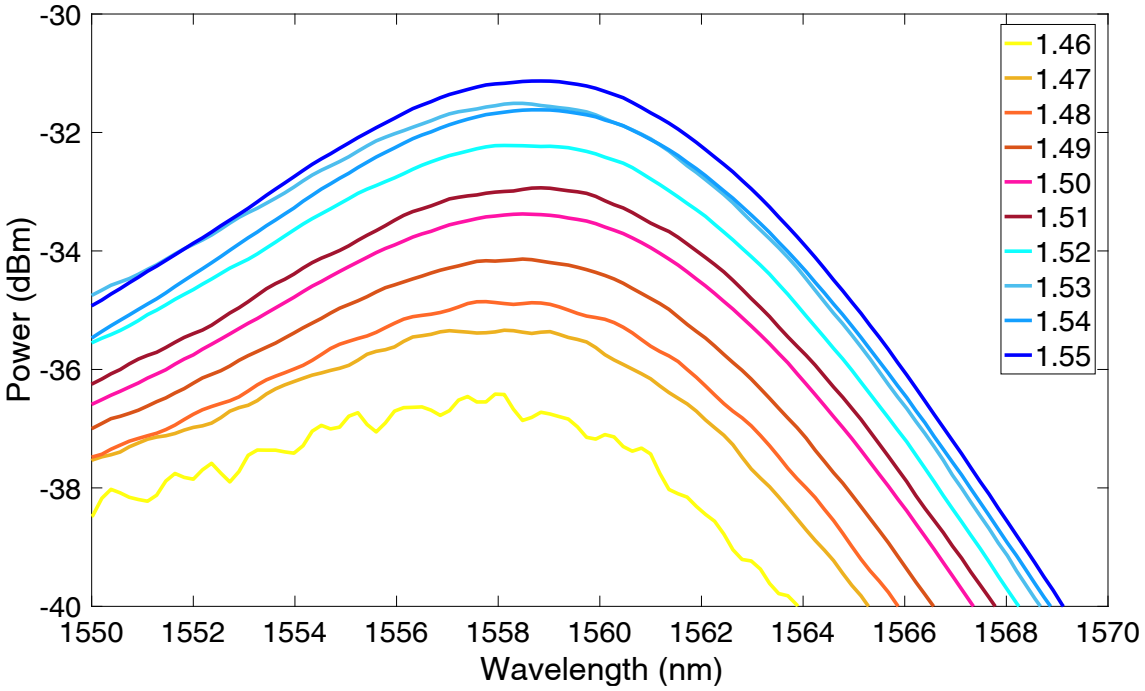

Figure 2. Spectral behavior for HRI region.

3.2 Classification Performance and Confusion Matrix Evaluation

Three machine learning classifiers (DT, SVM, and NN) were applied to classify RI values based on the processed spectral data. Their classification performances were assessed using overall accuracy, precision, recall, F1-score. Table 1 presents a summary of the results. Among the evaluated classifiers, the NN model demonstrated the highest accuracy at 80.0%, outperforming the SVM at 77.5% and DT at 42.5%. This confirms NN's strong suitability for high-precision RI classification over a wide sensing range, particularly in applications demanding sensitivity to subtle spectral differences.

Table 1. Summarizes the classification results for each ML model

Model	Accuracy	Precision	Recall	F1-score
DT	42.5%	Nan	42.5%	38.0%
SVM	77.5%	79.2%	77.5%	76.3%
NN	80.0%	79.2%	80.0%	78.2%

ML = Machine Learning, DT = Decision Tree, SVM = Support Vector Machine, NN = Neural Network, NaN = Not a number

To gain deeper insight into the classification behavior of each ML model, confusion matrices were analyzed for DT, SVM, and NN classifiers. These visualizations highlight the strengths and limitations of each model in distinguishing between closely spaced RI classes.

The DT model exhibited the lowest classification performance, with only a few RI classes achieving perfect True Positive Rates (TPR) as proved in Figure 3. While classes such as 1.36, 1.39, 1.47, and 1.49 RIU attained 100% TPR, a large number of classes showed substantial misclassification. For example, symmetrical confusion between adjacent RI classes such as 1.30 vs 1.31 RIU and 1.50 vs 1.51 RIU resulted in 50/50 split predictions, indicating the model's

inability to resolve small spectral differences. This is because the DT model is prone to overfitting and relies heavily on discrete decision boundaries, making it unsuitable for handling the nonlinear and overlapping spectral patterns seen in the MMI sensor data.

While the SVM model in Figure 4 provided a significant improvement over DT. The kernel-based nature of SVM allows for modelling of nonlinear spectral relationships, which is essential in distinguishing RI values affected by guided and leaky mode transitions. However, some misclassifications persisted, especially near RI class boundaries such as 1.30-1.32 RIU and 1.50-1.54 RIU, where spectral features are highly similar or noisy. SVM performed well overall but exhibited limited flexibility in capturing complex hierarchical spectral features compared to the NN model.

The NN model in Figure 5 outperformed both DT and SVM across nearly all metrics and classes. It achieved 100% TPR in most classes, including 1.30 to 1.32, 1.34 to 1.39, and several HRI classes such as 1.46, 1.47, 1.50, and 1.55. Minor misclassifications occurred only between adjacent RI classes such as 1.33-1.34 RIU, 1.48-1.49 RIU, 1.52-1.53 RIU, typically with 50/50 splits. These are likely due to spectral overlap, not model failure. The multilayer architecture of the NN allows it to extract deeper and more nuanced spectral features, making it well-suited for complex sensing environments.

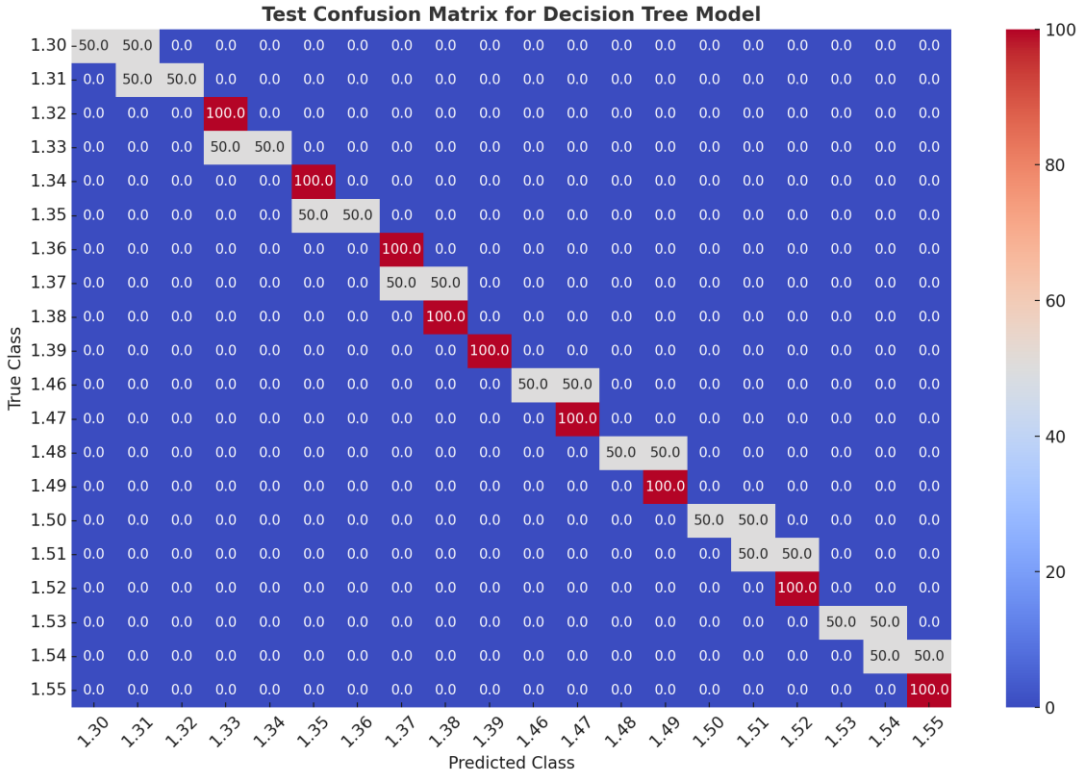

Figure 3. Confusion matrix for DT model.

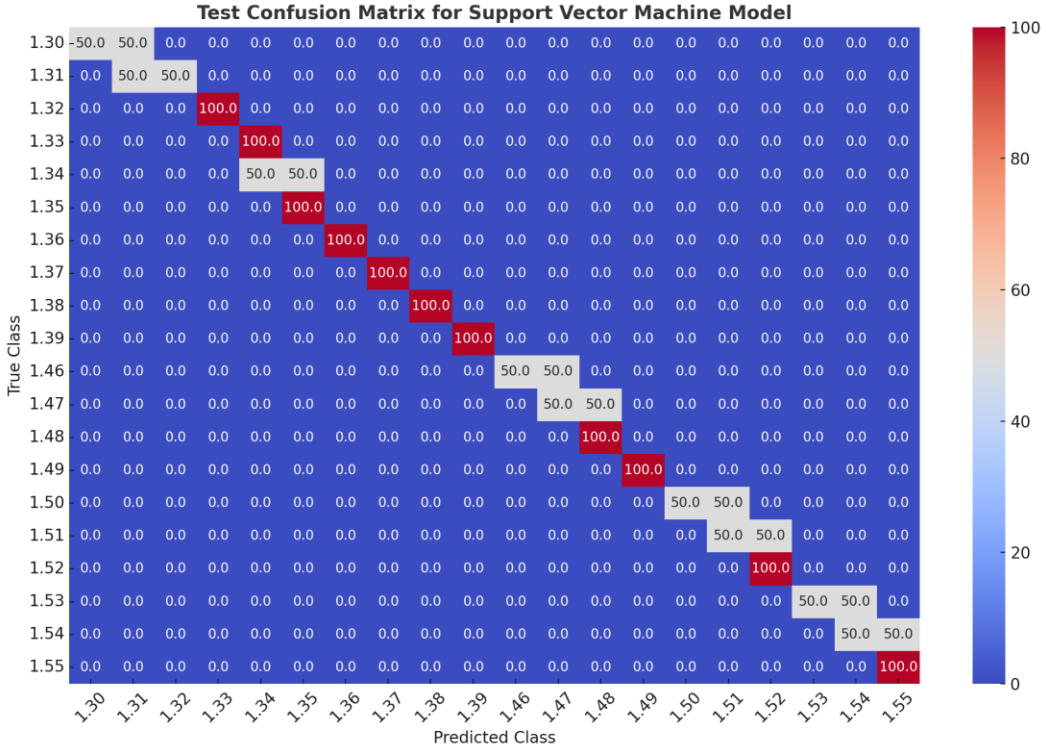

Figure 4. Confusion matrix for SVM model.

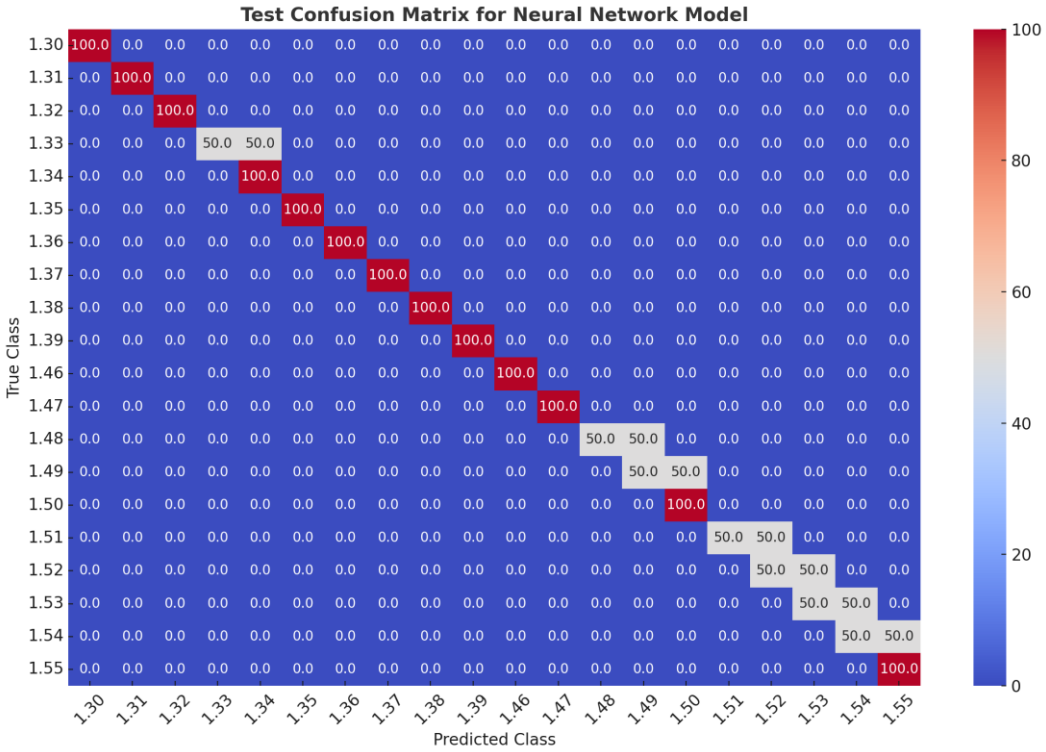

Figure 5. Confusion matrix for NN model.

Therefore, this comparative evaluation clearly demonstrates that the NN model offers superior classification performance for wide-range RI detection using an MMI based-NCF sensor. While SVM provides a strong alternative, especially in handling nonlinear transitions, its performance declines slightly in overlapping regions. The DT model, though simple and interpretable, lacks the capacity to resolve complex and overlapping spectral patterns effectively. In conclusion, the NN model stands out as the most reliable and accurate ML classifier for this application, with the potential for deployment in real-time, high-sensitivity optical sensing platforms. Future work could explore deeper network architectures or hybrid models to further enhance classification accuracy in RI transition zones.

Although no traditional baseline model was explicitly implemented, it is important to note that simpler rule-based approaches such as fixed-threshold peak tracking were evaluated during preliminary analysis. However, due to the overlapping and nonlinear nature of the spectra, particularly across guided-to-leaky mode transitions such methods were unable to produce meaningful classification results. Prior studies have similarly reported the limitations of traditional spectral interpretation when faced with complex, noisy, or highly coupled optical responses[12]. Therefore, the use of ML, and particularly NN models, is well-justified as they offer superior capability to model non-linear spectral behavior and extract nuanced features for accurate RI classification.

4. CONCLUSION

This work demonstrated the successful application of ML techniques for wide-range RI classification using MMI-NCF fiber sensors. Among the three evaluated models, the NN outperformed SVM and DT, achieving the highest classification accuracy of 80.0%. Confusion matrix analysis confirmed the NN's strength in resolving closely spaced RI values due to its multilayer learning capability. These results underscore the potential of ML, particularly NN, to enhance the accuracy and robustness of optical fiber sensing systems for real-time, high-precision applications. Future work may explore deep learning architectures such as Convolutional Neural Networks (CNNs), which are capable of automatically extracting local and global spectral features. These models may further improve classification accuracy by capturing subtle spectral variations more effectively. Real-time embedded deployment could also be investigated to enhance practical usability.

REFERENCES

[1] M. Elsherif *et al.*, "Optical fiber sensors: Working principle, applications, and limitations," *Advanced Photonics Research,* vol. 3, no. 11, p. 2100371, 2022.

[2] X. Zhu *et al.*, "'Lab around fiber'humidity-enhanced ammonia sensor: Multimode interference functionalized by graphene oxide," *Optics & Laser Technology,* vol. 182, p. 112232, 2025.

[3] V. Bhardwaj and V. K. Singh, "Study of liquid sealed no-core fiber interferometer for sensing applications," *Sensors and Actuators A: Physical,* vol. 254, pp. 95-100, 2017.

[4] M. Shao, M. Wang, Y. Yuan, X. Zhang, Y. Zhang, and Y. Liu, "Wearable optical fiber sensor in no-core fiber for heart rate monitoring," *Optics & Laser Technology,* vol. 192, p. 113404, 2025.

[5] M. M. Hasan, H. J. Taher, and S. A. Mohammed, "Highly sensitive fiber-optic temperature sensor based on tapered no-core fiber for biomedical and biomechanical applications," *Periodicals of Engineering and Natural Sciences (PEN),* vol. 9, no. 2, pp. 762-773, 2021.

[6] V. Bhardwaj, A. K. Pathak, and V. K. Singh, "No-core fiber-based highly sensitive optical fiber pH sensor," *Journal of Biomedical Optics,* vol. 22, no. 5, pp. 057001-057001, 2017.

[7] Z. Mu, Z. Yang, X. Han, Y. Li, and X. Sun, "Tapered No-Core fiber sensor coated with graphene oxide for the detection of human hemoglobin," *Optical Fiber Technology,* vol. 94, p. 104351, 2025.

[8] D. Song *et al.*, "An ultrafine no-core fiber-based sensor for simultaneous liquid level and temperature measurement," *IEEE Sensors Journal,* vol. 23, no. 14, pp. 15628-15636, 2023.

[9] K. Wang, "Fiber-Optic Multimode Interference Sensing and Imaging," Technische Universität München, 2023.

[10] A. Guglielmelli *et al.*, "Plasmonic Coupling for High-Sensitivity Detection of Low Molecular Weight Molecules," *Small Science,* vol. 5, no. 1, p. 2400382, 2025.

[11] Q. Wang, C. Li, C. Zhao, and W. Li, "Guided-mode-leaky-mode-guided-mode fiber interferometer and its high sensitivity refractive index sensing technology," *Sensors,* vol. 16, no. 6, p. 801, 2016.

[12] P. Oliveri, C. Malegori, R. Simonetti, and M. Casale, "The impact of signal pre-processing on the final interpretation of analytical outcomes–A tutorial," *Analytica chimica acta,* vol. 1058, pp. 9-17, 2019.

[13] W. Strielkowski, A. Vlasov, K. Selivanov, K. Muraviev, and V. Shakhnov, "Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review," *Energies,* vol. 16, no. 10, p. 4025, 2023.

[14] S. Khani, P. Rezaei, and M. Rahmanimanesh, "Machine learning analysis of a Fano resonance based plasmonic refractive index sensor using U shaped resonators," *Scientific Reports,* vol. 15, no. 1, p. 23857, 2025.