Machine Learning-Assisted Optimization of Photonic Crystal Fiber-Surface Plasmon Resonance (PCF-SPR) Sensors for Microplastic Detection

Marinah Othman
Faculty of Engineering and Built Environment,
Universiti Sains Islam Malaysia,
Negeri Sembilan, Malaysia.
marinah@usim.edu.my

Muhammad Hasif Saiful Ashik
Faculty of Engineering and Built Environment,
Universiti Sains Islam Malaysia,
Negeri Sembilan, Malaysia.
hasyff@raudah.usim.edu.my

Muhammad Firdaus Mohd Fadhir
Faculty of Engineering and Built Environment,
Universiti Sains Islam Malaysia,
Negeri Sembilan, Malaysia.
firdauz1308@gmail.com

Sarina Mansor
Faculty of Engineering,
Multimedia University,
Selangor, Malaysia.
sarina.mansor@mmu.edu.my

Siti Nur Aisyah Binti Mohd Robi
Faculty of Engineering and Built Environment,
Universiti Sains Islam Malaysia,
Negeri Sembilan, Malaysia.
aisyah.robi@lne.st

Juliza Jamaludin
Faculty of Engineering and Built Environment,
Universiti Sains Islam Malaysia,
Negeri Sembilan, Malaysia.
juliza@usim.edu.my

Khairul Nabilah Zainul Ariffin
Faculty of Engineering and Built Environment,
Universiti Sains Islam Malaysia,
Negeri Sembilan, Malaysia.
nabilahzainul@usim.edu.my

2026 International Conference on Frontiers of Engineering and Emerging Technologies (FET) | 979-8-3195-1886-6/26/$31.00 ©2026 IEEE | DOI: 10.1109/FET68771.2026.11601643

Abstract **-Microplastic pollution poses a significant threat to aquatic ecosystems; however, the development of high-sensitivity Photonic Crystal Fiber Surface Plasmon Resonance (PCF-SPR) sensors remains limited by computationally intensive trial-and-error simulations. This study investigates the effectiveness of machine learning (ML) algorithms in optimizing dual sliced PCF-SPR sensor parameters to accelerate the design process. A comprehensive dataset of geometric parameters and optical performance metrics is generated using COMSOL Multiphysics and used to train Artificial Neural Network (ANN) and Extreme Gradient Boosting (XGBoost) regression models. Results show that XGBoost achieves better predictive accuracy and identifies an optimal sensor configuration with a sensitivity of 5000 nm/RIU. The proposed ML-assisted framework significantly reduces computational cost while providing deeper insight into parameter–performance relationships, offering a scalable approach for real-time microplastic monitoring system development.**

Keywords - Photonic crystal fiber, surface plasmon resonance, machine learning, COMSOL Multiphysics

I. INTRODUCTION

The proliferation of microplastics (MPs) in aquatic ecosystems has evolved into a critical environmental engineering challenge, posing severe risks to marine biodiversity and human health through the trophic transfer of contaminants [1], [2]. The identification of MPs is critical for effective environmental management; however, conventional characterization techniques such as Fourier Transform Infrared (FTIR) spectroscopy, Raman spectroscopy, and electrochemical voltammetry are limited by their offline nature, complex sample preparation, and inability to support real-time, in-situ monitoring [3]-[5]. Consequently, there is an urgent demand for advanced sensing technologies capable of high-sensitivity, label-free, and rapid detection of MPs in aqueous environments [7].

PCF-SPR sensors represent an advanced optical sensing solution, offering high sensitivity to small variations in refractive index induced by pollutant concentrations [8]. Specifically, D-shaped PCF-SPR architectures have garnered significant attention due to their exposed core structure, which facilitates direct analyte interaction and simplifies the coating process of plasmonic materials such as gold or silver compared to closed-hole structures [9], [10].

However, the structural optimization of these nanophotonic devices presents a significant computational bottleneck. Conventional design workflows rely heavily on Finite Element Method (FEM) solvers, such as COMSOL Multiphysics, to solve Maxwell's equations iteratively. This trial-and-error approach is computationally expensive and time-consuming, particularly when exploring multi-dimensional design spaces involving multiple geometric parameters. The lack of efficient optimization frameworks often results in sub-optimal designs or prohibitive development cycles [11].

To bridge this gap, the integration of artificial intelligence (AI) in photonic design, often termed *inverse design*, has emerged as a transformative strategy. Machine learning (ML) algorithms can function as surrogate models, rapidly mapping the non-linear relationships between geometric parameters and optical performance without the need for repetitive FEM simulations [11]. Recent studies have demonstrated the potential of algorithms such as ANN and XGBoost in predicting sensor characteristics with high accuracy [9]. For instance, an R^2 value of 99.64% has been reported using XGBoost for a D-shaped sensor configuration; however, a direct comparative analysis of such algorithms specifically for microplastic detection optimization remains limited in the existing literature [9].

A comparative study between Artificial Neural Networks (ANN) and XGBoost is conducted to evaluate their efficacy as surrogate models for nonlinear regression tasks in photonic applications. By leveraging a dataset generated via rigorous FEM simulations, this study aims to identify the optimal geometric parameters that yield maximum wavelength sensitivity. The dataset size is considered sufficient due to the use of a full factorial design strategy, which ensures complete coverage of the parameter space and captures all significant interactions between design variables. This study focuses on establishing a machine learning–assisted optimization framework, with experimental validation planned as part of future work.

II. RELATED WORK

Microplastic detection research has predominantly focused on enhancing analytical techniques based on spectroscopy and microscopy [1]–[4]. More recently, machine learning (ML) has been integrated to automate spectral and image-based classification, improving identification accuracy and throughput. However, these approaches primarily address data analysis and do not contribute to the advancement of sensor hardware design.

Photonic crystal fiber–surface plasmon resonance (PCF-SPR) sensors have been widely investigated for biosensing and refractive index measurement applications. Advanced configurations, including D-shaped, dual-core, and elliptical-core structures, have demonstrated sensitivities exceeding 10,000 nm/RIU [8], [10]. Despite their strong potential for environmental sensing, most existing designs are optimized for biomedical refractive index ranges rather than microplastic-contaminated water.

The integration of ML into photonic device optimization has recently gained attention. Artificial neural network (ANN) models have achieved low prediction errors in optimizing plasmonic sensor geometries, while tree-based methods such as random forest (RF) and extreme gradient boosting (XGBoost) have shown strong robustness and predictive accuracy for complex PCF-SPR structures [9]. Nevertheless, there remains no clear consensus on the most suitable ML model for PCF-SPR optimization, and comparative studies are limited.

Recent ML-assisted optimization efforts have demonstrated notable progress in accelerating sensor design. For instance, Dogan *et al.* [9] employed XGBoost integrated with NSGA-II optimization, achieving a sensitivity of 4814.14 nm/RIU (R^2 = 99.64%), however, they used a D-shaped sensor. Similarly, Saha *et al.* [10] applied multiple regression techniques, including multiple linear regression (MLR), support vector regression (SVR), and random forest, to model their sensor, but they also used a different structure and material . Additionally, these studies primarily target general biosensing within conventional biomedical refractive index ranges and lack explicit model interpretability analysis.

In contrast, this work focuses on microplastic detection within the low refractive index range of 1.38–1.42, which is relevant to fluorinated polymer contaminants. In addition, SHAP-based interpretability is incorporated to identify dominant design parameters, enabling more efficient and physically informed optimization. A direct comparative evaluation between XGBoost and ANN is also conducted, providing practical insights into model selection for ML-assisted photonic sensor design in environmental sensing applications.

III. METHODOLOGY

The proposed methodology is structured as a generalizable engineering workflow that integrates numerical simulation, data preprocessing, machine learning, and optimization

A. PCF-SPR Sensor Dataset

A simulation-based dataset is employed to train and evaluate the machine learning models. The dataset is generated using COMSOL Multiphysics, where a dual-sliced shaped photonic crystal fiber surface plasmon resonance (PCF-SPR) sensor structure, as shown in Fig. 1, is modeled using the finite element method (FEM) [9]. This configuration is selected due to its strong evanescent field interaction with the analyte region, which is suitable for refractive-index-based environmental sensing [8]. Key structural parameters of the PCF-SPR sensor are systematically varied within physically realistic ranges to capture the influence of geometry and material properties on sensor performance. Each data sample corresponds to a unique sensor geometry and includes both input parameters and performance outputs.

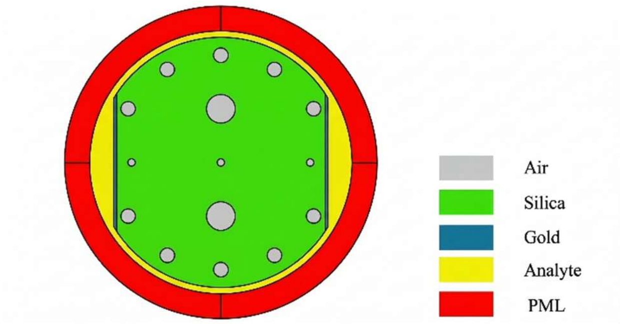


Fig. 1. Cross section of the dual-side polished PCF.

The parameter ranges used in this study are listed in Table 1. The dataset includes five discrete levels of analyte refractive index ranging from 1.38 to 1.42 with a step size of 0.01, representing the effective index of water–microplastic mixtures associated with advanced industrial pollutants. This low-index range is primarily attributed to fluorinated polymers, which are extensively used in coatings, lubricants, and high-technology applications, and can be introduced into the environment as primary or secondary microplastics. Additionally, three levels were established for each of the five geometric parameters, specifically gold layer thickness (40, 50, and 60 nm), lattice pitch (3.0, 3.5, and 4.0 µm), air hole diameter (1.6, 1.8, and 2.0 μm), core diameter (400, 600, and 800 nm), and analyte thickness (0.35, 0.40, and 0.45 μm).

Table 1. Input Parameters for COMSOL Multiphysics.

Parameter	Range/Values	Unit	Step size
Operating wavelength	0.6-1.0	μm	0.05
Refractive index (Analyte)	1.38-1.42	-	0.01
Gold layer thickness	40, 50, 60	nm	10
Pitch distances	3.0, 3.5, 4.0	μm	0.5
Air-hole diameter	1.6, 1.8, 2.0	μm	0.2
Core diameter	400, 600, 800	nm	200
Analyte thickness	0.35, 0.40, 0.45	μm	0.05

The generation of a systematic dataset using COMSOL Multiphysics based on the FEM was conducted to ensure high model credibility and comprehensive design space coverage. A Full Factorial Design (FFD) sampling strategy was implemented, involving a total of 1,215 unique sensor configurations. This approach evaluates every possible combination of the design variables to ensure that no critical interaction between parameters is overlooked, distinguishing it from random sampling techniques.

Each of the 1,215 configurations underwent a wavelength sweep from 0.6 μm to 1.0 μm with a discrete step size of 0.05 μm to capture the complete spectral response. The simulation process generated a raw dataset exceeding 10,000 data points before the feature extraction procedure was applied to identify the resonance peak for each configuration. Only the critical resonance parameters, specifically the resonance wavelength and the corresponding peak confinement loss, were extracted for the purpose of model training. This rigorous preprocessing resulted in a high-quality final dataset consisting of 1,215 samples. The dataset size is considered sufficient due to the use of a full factorial design strategy, which ensures complete coverage of the parameter space and captures all significant interactions between design variables. These samples provide a robust foundation for the ANN and XGBoost surrogate models to accurately map complex nonlinear optical responses for precise microplastic detection in industrial effluents.

Prior to model training, the dataset will undergo preprocessing to improve learning stability and prediction reliability. Invalid or non-convergent simulation samples are removed to ensure data consistency. Continuous input parameters are normalized using Min–Max scaling to prevent bias toward parameters with larger numerical ranges and to facilitate efficient convergence of the ANN model [11]. The dataset is then randomly divided into training, validation, and testing subsets, using an 80:10:10 ratio. The training set is used to fit the machine learning models, the validation set supports hyperparameter tuning and early stopping, and the testing set provides an unbiased evaluation of model generalization. This separation ensures that model performance reflects true predictive capability rather than overfitting to simulation data.

B. Machine Learning Models

The optimization problem is formulated as a supervised regression task in which sensor structural parameters are mapped to key performance indicators. Two machine learning models are employed as surrogate predictors:

Artificial Neural Network (ANN): ANN is employed as a deep learning–based surrogate model to approximate the nonlinear relationship between PCF-SPR design parameters and sensor performance [11]. The ANN architecture consists of an input layer corresponding to the sensor parameters, multiple hidden layers with nonlinear activation functions, and an output layer predicting wavelength sensitivity and modal loss.

The ANN model is trained using a backpropagation algorithm with a mean squared error loss function. Gradient-based optimization is applied to update network weights iteratively until convergence criteria are met. The validation dataset is used to adjust network depth, neuron count, and learning rate to balance model complexity and generalization performance. ANNs are particularly suitable for this application due to their ability to model complex, high-dimensional physical interactions that arise in photonic structures [11]. Once trained, the ANN provides fast performance prediction without requiring repeated FEM simulations.

Extreme Gradient Boosting (XGBoost): In parallel with the ANN, an XGBoost model is developed as a tree-based ensemble surrogate. XGBoost constructs an ensemble of regression trees through an iterative boosting process, where each new tree minimizes the prediction error of the previous ensemble. Key hyperparameters such as tree depth, learning rate, and number of estimators are optimized using the validation dataset. Regularization techniques inherent to XGBoost help prevent overfitting and improve robustness, particularly when dealing with nonlinear interactions and moderate-sized datasets [12].

Compared to ANN, XGBoost offers strong interpretability and stability, making it well suited for engineering applications where reliable prediction and sensitivity to design parameters are important. The inclusion of both ANN and XGBoost enables comparison between deep learning–based and boosting-based surrogate modeling approaches.

C. Model Evaluation and Validation

The predictive performance of the ANN and XGBoost models is evaluated using the testing dataset. Standard regression metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), are used to quantify prediction accuracy [12]. Beyond numerical accuracy, the models are assessed from an engineering perspective by examining their ability to capture physically meaningful trends in sensor behavior. Agreement between ML predictions and known PCF-SPR performance characteristics serves as qualitative validation of the surrogate models.

After training and validation, the ANN and XGBoost models are employed as surrogate predictors for rapid design exploration. Instead of performing computationally expensive FEM simulations, new PCF-SPR configurations are evaluated using the trained ML models to identify designs that maximize wavelength sensitivity while maintaining acceptable confinement loss. This ML-assisted optimization approach enables efficient scanning of the design space and supports rapid iteration, making it suitable for emerging engineering applications that demand fast adaptation and scalability. Microplastic detection is used as a representative environmental sensing scenario to demonstrate the practical relevance of the proposed framework.

IV. Results and Discussion

A. *Performance Analysis of ML Models*

Table 2 summarizes the comparative performance of the ANN and XGBoost models, highlighting significant improvements in prediction accuracy and error reduction achieved by XGBoost. The XGBoost model achieved a near-perfect coefficient of determination, $R^2 = 0.9995$, with significantly lower prediction errors, including approximately 2.6 times and 6.6 times reductions in MAE and MSE, respectively, compared to the ANN model. This result confirms the predictive accuracy and robustness of the XGBoost model for nonlinear regression in nanophotonic design.

Table 2. Comparative performance metrics of ML models.

Evaluation Metrics	ANN	XGBoost
Training time (s)	134.5	158.4
Coefficient of determination (R^2)	0.9968	0.9995
Mean Absolute Error (MAE)	0.3470	0.0522
Mean Squared Error (MSE)	0.4399	0.1689
Overfitting behavior	Mild (dropout asymmetry)	Minimal (L1+L2 via NSGA-II)

B. *Analysis of Overfitting Behavior*

The ANN exhibited a mild train-validation performance gap, characteristic of the regularization effect of Dropout. In contrast, the XGBoost model demonstrated minimal overfitting risk. This robust generalization capability is attributed to the optimal L1 and L2 regularization terms, tuned via the NSGA-II multi-objective optimization, which effectively managed model complexity and improved predictive stability on the unseen test data. As illustrated in Fig. 2, the XGBoost model consistently outperforms the ANN across key evaluation metrics, demonstrating superior predictive accuracy and stability.

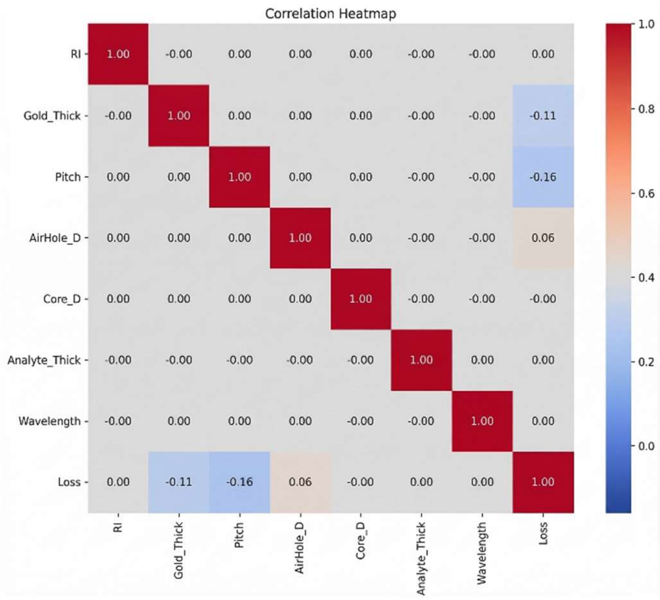
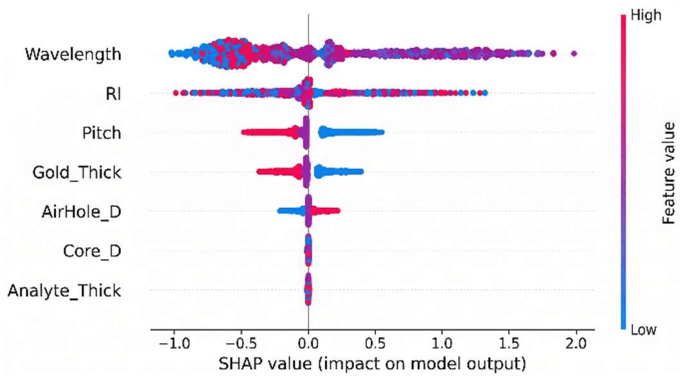


Fig. 2. Comparison of predictive performance between ANN and XGBoost models, illustrating differences in accuracy and generalization behavior.

Fig. 3 presents the training and validation performance of the ANN and XGBoost models, providing insight into their generalization behavior. The ANN model exhibits a noticeable gap between training and validation performance, indicating mild overfitting despite the application of regularization techniques such as dropout. This suggests that the ANN model is partially sensitive to the training dataset



and may not fully generalize to unseen configurations. In contrast, the XGBoost model demonstrates closely aligned training and validation performance, indicating strong generalization capability and minimal overfitting, confirming its robustness against overfitting.

Fig. 3. Training and validation performance comparison of ANN and XGBoost models, highlighting overfitting characteristics and model generalization capability.

C. Feature Importance and Sensitivity Analysis

The superiority of XGBoost enabled a reliable SHAP (SHapley Additive exPlanations) analysis to interpret the complex relationships between the input geometric parameters and the optical performance. This analysis established that wavelength (mean |SHAP| = 59.28) is the dominant predictor of confinement loss (CL), followed by the analyte refractive index (RI) (mean |SHAP| = 28.93). The physical basis for this dominance is rooted in SPR physics: longer wavelengths increase the evanescent field penetration depth, strengthening SPR coupling and increasing CL. Conversely, a higher analyte RI increases core-cladding contrast, which strengthens light confinement and suppresses CL. Crucially, parameters such as core diameter and analyte thickness were found to have a negligible influence on the final model output (mean |SHAP| ≈ 0.1), which directly supports the streamlining of future PCF-SPR design efforts by deprioritizing these less influential variables.

D. Optimized PCF-SPR Configuration

Table 3 presents the optimized PCF-SPR sensor configuration obtained through the ML-assisted optimization framework, along with the corresponding performance metrics. This optimal dual-sliced PCF-SPR configuration achieves a corrected maximum wavelength sensitivity of 5,000 nm/RIU. This sensitivity lies within the upper range reported for PCF-SPR sensors in the literature, demonstrating the power of the ML-assisted framework in efficiently navigating the design space to identify high-performance structures. The high sensitivity achieved, coupled with the rapid inference time of the XGBoost surrogate model, establishes a robust and scalable prototyping pathway for next-generation microplastic monitoring systems.

Table 3. Optimized PCF-SPR sensor design and performance metrics.

Design Parameter	Optimized Value	Unit
Gold layer thickness	60.0	nm
Lattice pitch (Λ)	4.0	μm
Air-hole diameter	1.6	μm
Core diameter	400	nm
Analyte thickness	0.35	μm
RI interval	1.40–1.41	-
Peak wavelength shift (Δλ)	0.050 (50 nm)	μm
Wavelength sensitivity (S)	5,000	nm/RIU
Confinement loss at peak	172.997	dB/m

V. CONCLUSION

This paper presents a machine learning–driven framework for accelerating the optimization of PCF-SPR sensors for microplastic detection. By integrating FEM-based simulation data with ANN and XGBoost surrogate models, the proposed approach effectively mitigates the computational limitations of conventional design methods. The XGBoost model demonstrated superior performance, achieving $R^2 = 0.9995$ and lower prediction error compared to ANN. SHAP analysis further revealed wavelength and refractive index as dominant factors influencing sensor performance. The optimized dual-sliced PCF-SPR design achieved a maximum sensitivity of 5000 nm/RIU within the microplastic-relevant refractive index range. These results validate the effectiveness of ML-assisted optimization and highlight its potential as a scalable tool for next-generation environmental sensing applications.

ACKNOWLEDGMENT

This project is a collaborative effort between USIM, MMU, and UTM, funded by the Ministry of Higher Education Malaysia (MOHE) under the Fundamental Research Grant Scheme (FRGS/1/2025/WAS02/USIM/02/1).

REFERENCES

[1] J. Su *et al.*, "Machine learning: Next promising trend for microplastics study," *Journal of Environmental Management*, vol. 344, p. 118756, 2023.

[2] L. Xie, M. Ma, Q. Ge, Y. Liu, and L. Zhang, "Machine Learning Advancements and Strategies in Microplastic and Nanoplastic Detection," *Environmental Science & Technology*, vol. 59, pp. 8885–8899, 2025.

[3] H. de Medeiros Back, E. C. Vargas Junior, O. E. Alarcon, and D. Pottmaier, "Training and evaluating machine learning algorithms for ocean microplastics classification through vibrational spectroscopy," *Chemosphere*, vol. 287, p. 131903, 2022.

[4] J. Y. Lin, H. T. Liu, and J. Zhang, "Recent advances in the application of machine learning methods to improve identification of the microplastics in environment," *Chemosphere*, vol. 307, p. 136092, 2022.

[5] D. H. L. Kumar *et al.*, "Electrochemical approaches for detecting micro and nano-plastics in different environmental matrices," *International Journal of Electrochemical Science*, vol. 20, p. 101182, 2025.

[6] J. Sarmiento, M. Anaya, and D. Tibaduiza, "Microplastic Identification Using Impedance Spectroscopy and Machine Learning Algorithms," International Journal of Distributed Sensor Networks, vol. 20, 2024.

[7] M. M. Khanam, M. K. Uddin, and J. U. Kazi, "Advances in machine learning for the detection and characterization of microplastics in the environment," *Frontiers in Environmental Science*, vol. 13, p. 1573579, 2025.

[8] X. Luo *et al.*, "Low-loss and high-sensitivity surface plasmon resonance sensor composed of microstructured optical fibers for wide visible-to-infrared and refractive index ranges," *Optik*, vol. 295, p. 171389, 2023.

[9] Y. Dogan, R. Katirci, and I. Erdogan, "Machine learning-based optimization for D-shaped PCF SPR refractive index sensor," *Optics Communications*, vol. 575, p. 131304, 2025.

[10] C. Saha *et al.*, "Dual-core silver-coated plasmonic sensor modeling with machine learning," *Heliyon*, vol. 10, no. 8, p. e38175, 2024.

[11] P. Manfredi, A. Waqas, and D. Melati, "Stochastic and multi-objective design of photonic devices with machine learning," *Scientific Reports*, vol. 14, p. 7162, 2024.

[12] Z. Chen, X. Bai, Y. Chen, D. Xu, and W. Li, "Machine learning-driven prediction of biochar yield and fuel properties from biomass characteristics and pyrolysis conditions," *Biomass and Bioenergy*, vol. 206, p. 108594, 2026.