Babylonian Journal of Machine Learning
Vol.2023, **pp**. 35–41
DOI: https://doi.org/10.58496/BJML/2023/007; ISSN: xxxxx
https://mesopotamian.press/journals/index.php/BJML

Research Article

An Effective Deep Learning Model for Surface-Enhanced Raman Spectroscopy Detection Using Artificial Neural Network

Aythem Khairi Kareem[1,], Meaad Ali Khalaf[2,], Mohammed M AL-Ani[3,], Ahmed Adil Nafea[4,*,]

1 Department of Heet Education, General Directorate of Education in Anbar, Ministry of Education, Heet, 31007 Anbar, Iraq
2 Department of computer science, AUL University, Beirut, Lebanon
3 Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia
4 Department of Artificial Intelligence, College of Computer Science and IT, University of Anbar, Ramadi, Iraq

ARTICLE INFO

Article History
Received 27 Apr 2023
Accepted 11 Jun 2023
Published 23 Jul 2023

Keywords
Artificial Neural Network,
Surface-enhanced Raman Spectroscopy,
Deep Learning,
Detection,
Classification

ABSTRACT

Surface-enhanced Raman spectroscopy (SERS) is a powerful technique for molecular sensing and has gained significant attention due to its high sensitivity and selectivity. SERS based on deep learning technology have been used in this study of materials, biological recognition, food safety, and intelligence. Deep learning techniques have shown tremendous potential in various scientific fields, including spectroscopy-based detection methodologies. In this study, we propose an effective deep learning approach for SERS detection using an artificial neural network (ANN). Then, the results are compared using two datasets batch1 and batch2. This study used Rhumamine 6G (R6G) as an aim molecule in this study. The experimental develops show the effectiveness of proposed ANN.

1. INTRODUCTION

This SERS spectroscopy is a reasonable mechanism to determine and represent biomolecules and their construction blocks [1]. It thoroughly represents their structure, composition, effect of ligand binding, and aggregation state. Once adsorbed on the surface of a SERS substrate, molecules experience a significant enhancement in their Raman signal sufficiently to be organized even when they are current in trace levels. The opportunity to inspect samplings in the medium utilizing short accession times and, according to a label-free approach, designates further individual muscle points of SERS delivered different analytical, biotechnological, and biomedical applications [2].

SERS manipulates the ability of metallic nanostructures to consolidate electromagnetic energy via optical modes called surface Plasmon (SP). The first observance of SERS appeared with the unexpected expansion of the Raman signal from adsorbed pyridine on an electrochemically roughened silver electrode. The signal increase was initially attributed to more adsorbed molecules with the expanded surface area, but it was subsequently discovered to derive from irregular surface enhancement. This finding is of paramount significance for analytical objectives since it delivers a means of confounding the inherent low efficiency of the normal Raman Scattering (RS) processes corresponding with fluorescence emission and infrared absorption. Therefore, the approach invented based on the phenomenon of SERS allows the investigation of smallish numbers of molecules. Additionally, due to its increased sensitivity, SERS inherits from RS attributes like chemical characteristics and label-free analytical nature. SERS has grabbed the concentration of various study districts by relishing these characteristics, and the past four decades have noticed a dramatic growth in applicability [3].

Corresponding author. Email: ahmed.a.n@uoanbar.edu.iq

SERS has evolved as a widely used spectroscopic process, utilizing the advances of current optics, nanodevices, and laser technologies. The cause is very simple: it can deliver structural data at the level of a single molecule due to the extensive expansion of the electromagnetic region generated by the local surface platon excitation (LSP). Furthermore, SERS is non-invasive, water-free, and can investigate conformational differences because it requires vibration data for targeted analysis. Several studies have shown that SERS can be used for the monitoring of therapeutic drugs (TDMs), in the detection of cancer cells and tumor margins in in vivo conditions, or in the biosensing of DNA. SERS can replace traditional processes currently routinely used for DNA-based diagnostics [4].

Due to the complexity of extensive data that are regularly obtained during SERS investigations, which cause the classical techniques are not acceptable for data processing. ML techniques concentrate on building models using learning patterns from large datasets and enhancing the performance of models, which belongs to the area of AI. Machine learning techniques have been successfully involved in clustering, regression, classification, and estimate tasks over high-dimensional and large datasets [5]. However, it still has issues with the accuracy of enhancing detection so this study proposed an ANN model of deep learning for the detection of SERS.

2. RELATED WORK

Shizhuang Weng et al. (2019) proposed DL networks were utilized as a fully convolutional network (FCN), fully connected network, Principal Component Analysis (PCA), network and Convolution Neural Network (CNN) to predict their capabilities to recognize medications in human urine excerpt in the two input constructions of two-dimensional matrix or one-dimensional vector. The most suitable distinction development for drugs in urine by an accuracy of 98.05% in the detection set was learned utilizing CNN with scopes as input in the matrix form. The best quantitation was accepted employing FCN with scopes in matrix form, which recognized the investigation with a coefficient of 0.9997. These networks achieved more value than the traditional ML techniques. Overall, the DL networks deliver feasible options for recognizing and quantifying SERS [6].

Wonil Nam et al. (2021) this proposed electronic Raman Scattering (ERS)-based SERS calibration is achieved by applying improved supervised ML techniques to classify SERS areas of labelless living cells the first subtyping breast cancer cells through different tumor widths, and second evaluating the drug reactions of cancer cells at different doses. In particular, the SERS calibration based on ERS has the benefits of exceptional photostability under laser excitation, no spectral interference beside Raman signatures of biomolecules, and no matched using biomolecules at hot spots. Consequently, the proposed approach predicts that SERS calibration based on ERS can significantly increase the implementation of multivariate investigation in SERS dimensions without labeling in biological methods [7].

Wei Hu et al. (2019) proposed a practical ML Random Forest (RF) technique that detection SERS signals of a trans-1,2-bis (4-pyridyl) ethylene molecule adsorbed on a gold substratum. The geometric descriptive characteristics extracted from quantum chemistry simulations. The ML approach predicts vibrational commonness and Raman powers in molecular dynamics conformations. The resulting scores agree with density practical theory experiments and calculations. Detection SERS answers of the molecule on various surfaces or beneath exterior areas of electric specializations and solvent environment display useful transferability of the protocol [8].

Fatma Uysal Ciloglu et al. (2020) presented SERS combined with ML techniques for quickly identifying Gram positive Staphylococcus aureus methicillin-resistant and exposed stresses and Gram-negative Legionella pneumophila. K-nearest neighbors (kNN) algorithm classification performed best with 97.8% accuracy among the traditional ML classifiers, including NB, SVM, and DT. The proposed approach results show that SERS integrated with ML can detect susceptible and antibiotic-resistant bacteria [9].

Jingyu Ding et al. (2021) developed a methodical technique integrating SERS with multi-scale CNN. The proposed approach organized the 34 nm gold nanoparticle (AuNP) as a label-free Raman substrate, calculated the SERS ranges of these three salmonella serovars in 1854, and presented a multi-dimensional CNN model with three parallel CNNs to extract SERS spectrum attributes in multiple dimensions. The present study changed the percentage of training and test sets to take into account the impact of the number of repetitions and training instances on recognition accuracy by approximating the estimated data to experimental data, the model has demonstrated that it can be accurate at more than 97%. Experienced results show the effectiveness of combining SERS spectroscopy with CNN for the Salmonella serotype designation [10].

Jia-Wei Tang et al. (2022) investigated SERS through ML techniques to classify bacterial pathogens accurately and quickly. This study proposed two unsupervised ML techniques, Agglomerative Nesting (AGNES), and K-means Clustering were conducted for clustering investigation. Also, eight supervised ML techniques were approximated in terminology of bacterial detection through Raman spectra, which demonstrated that CNN reached the most acceptable accuracy (99.86%) including the highest area (0.9996) under the curve ROC [11].

Ryosuke Nishitsuji et al. (2023) investigate the multivariate analysis of SERS scopes using ML prototypes. Substrates with gold nanostructures whose surfaces include Self-Assembled Monolayers (SAMs) of eclectic benzene thiol products are

simulated, and their SERS scopes are obtained. After dataset preprocessing, the scopes are investigated by employing additional ML techniques to approximate the prediction accurateness of the mixing arrangements of the SAM reagents. The experience results demonstrate that the LDA technique reaches the highest accuracy (0.996).

3. RESEARCH METHODOLOGY

This methodology discusses the steps of using deep learning and shows guidance on data preprocessing, splitting the data set, PCA, and evaluation as shown in "Figure 1".

3.1 Dataset Description

This study used R6G datasets as a target molecule [12]. The RS wavelength is 785 nm, and the power and exposure duration of the laser are set at 200 mW and 500 ms, separately. To acquire the scopes of SERS, this study descended a 2.5 μL selection on the SERS substratum and dried it at a temperature (27 °C). Every SERS amount is registered with 10 s periods to miscalculate signal degradation. Each sampling DS∈R×2000, which represents The SER dataset, contains 2000 samples.

The dataset employed is two individual files, bacth1, and batch2, including data from distinct focuses. In a individual batch, it's have 500 negative and 1500 positive samples. The concentration of ≥0.01 μM is the threshold for positive, which conforms with the transmitted boundary of recognition of the R6G molecule. In batch 1, we calculate (10000, 0, and 10) μM. Batch 2 (100, 0.1, and 0.01) μM are calculated. Table I displays an entire explanation of the selection division of the R6G dataset.

TABLE I. THE ENTIRE EXPLANATION OF R6G DATASET.

Batch No.	Positive			Negative		
	10	100	10000	0	0.01	0.1
Bach2	0	500	0	0	500	500
Batch1	500	0	500	500	0	0

3.2 Convert Target Variables to Categorical

The target variables, representing the classes or categories of the SERS spectra, are converted to categorical format. This step is essential for classification tasks where the ANN model needs to predict the class labels accurately. Categorical conversion ensures that the target variables are represented as binary vectors, with each element corresponding to a specific class.

3.3 Principal Component Analysis

PCA is applied to reduce the dimensionality of the input features. This step helps in capturing the most significant variations in the dataset while reducing noise and computational complexity. The number of components is set to 19 in this case, but it can be adjusted based on the desired level of dimensionality reduction and information retention.

3.4 Splitting

This study splits the dataset into 30% for testing and 70% for training.

3.5 Proposed Model ANN

This study proposed an ANN model for SERS detection. The architecture of the model by adding layers sequentially. In this case, the model architecture contains three layers the first is a dense layer with 512 units and a ReLU activation function, followed by the second layer is a dense layer with 256 units and a ReLU activation function, and finally, a dense layer with 2 units and a SoftMax activation function for binary classification of SERS.

The main structure of an artificial neural network (ANN) is shown in Fig. 1. Stage 3 This is the main building block on which the proposed approach is based. A neural network consists of 3 types of layers [13].

- Input layer: This layer is reliable for transferring the input dataset to the afterward layer.
- Hidden layers: It is responsible for the identification of the connection among output values and input values.
- Output layer: This layer generates ANN detection results based on inputs.

The input layer of an ANN begins with input data and random weights. The feedforward process produces prediction values, and the loss score is evaluated using the loss function. Backpropagation updates weights by computing gradients, a gradient-based optimization mechanism. This allows the loss score to be reduced through weight updates, aiming to find the best weights for the least error in output prediction.

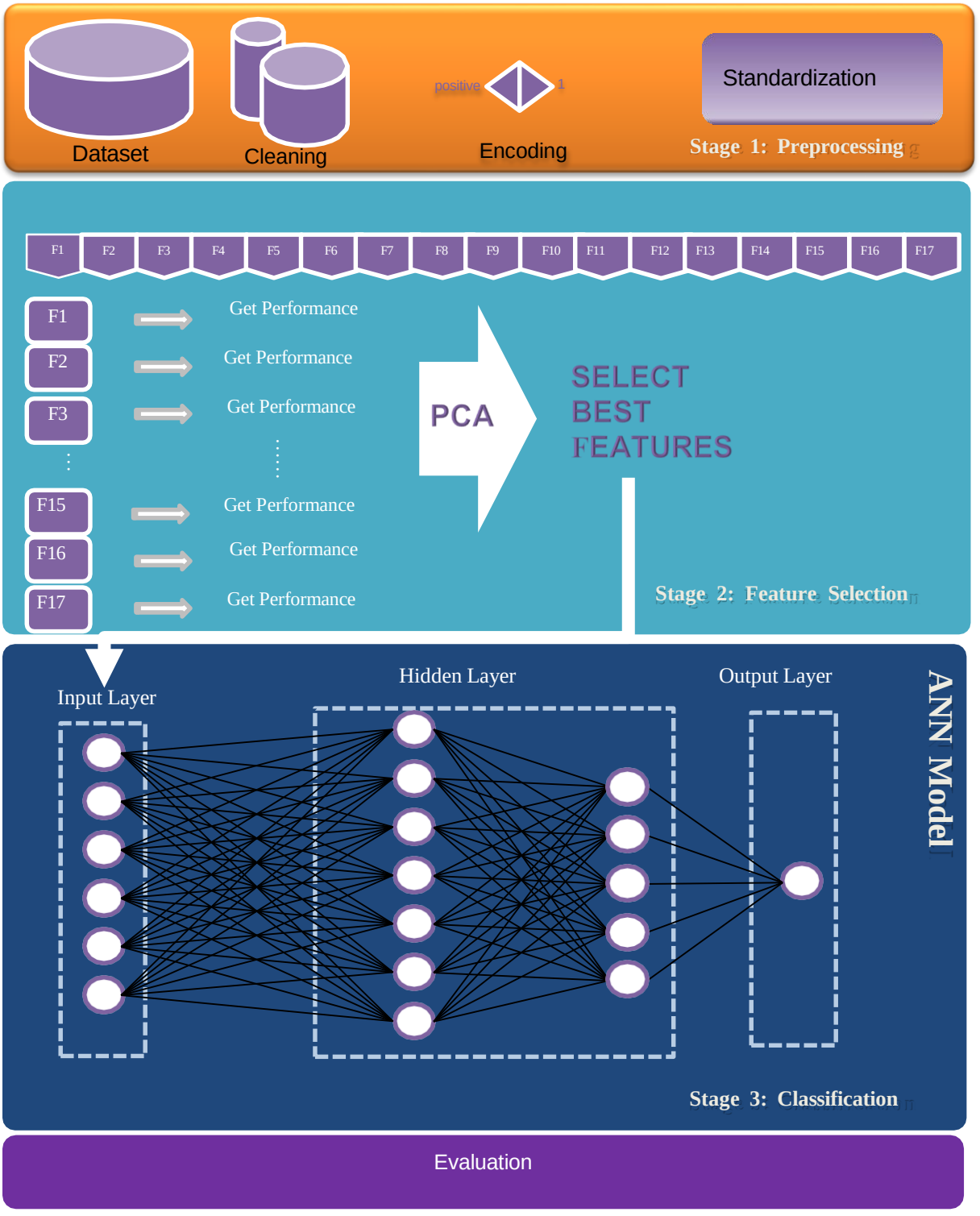

Fig. 1. Research Methodology

3.6 Evaluation

In the following steps explanation of accuracy, precision, recall, and F1 score along with their equations:

Accuracy: measures the overall correctness of the model's predictions [14].

Accuracy = (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives)

Precision: focuses on the quality of positive predictions, measuring how many correctly predicted positives were out of all instances predicted as positive [15].

Precision = True Positives / (True Positives + False Positives)

Recall (Sensitivity): true positive rate, measures the proportion of actual positive instances correctly identified by the model.

Recall = True Positives / (True Positives + False Negatives)

F1 Score: is the harmonic mean of precision and recall, providing a balanced evaluation of the proposed performance[16].

*F1 Score = 2 * (Precision * Recall) / (Precision + Recall)*

4. RESULTS AND DISCUSSION

SERS was classified into two types which are positive and negative. This model was trained using the ANN methods. The proposed model achieved the highest classification accuracy and outperformed classifiers in all testing scenarios. The accuracy of the proposed model score of 100%, which is the highest accuracy score compared with other proposed studies. Table II presents the results for two datasets Batch1 and Batch2.

TABLE II. PROPOSED RESULTS.

	Precision	***Recall***	***F1-score***	***accuracy***
ANN-Batch1	100	100	100	100
ANN-Batch2	100	1.00	100	100

As shown in "Figure 2" the First Figure (a) is a confusion matrix, and the second Figure (b) is actual with predicted values for the batch 1 dataset.

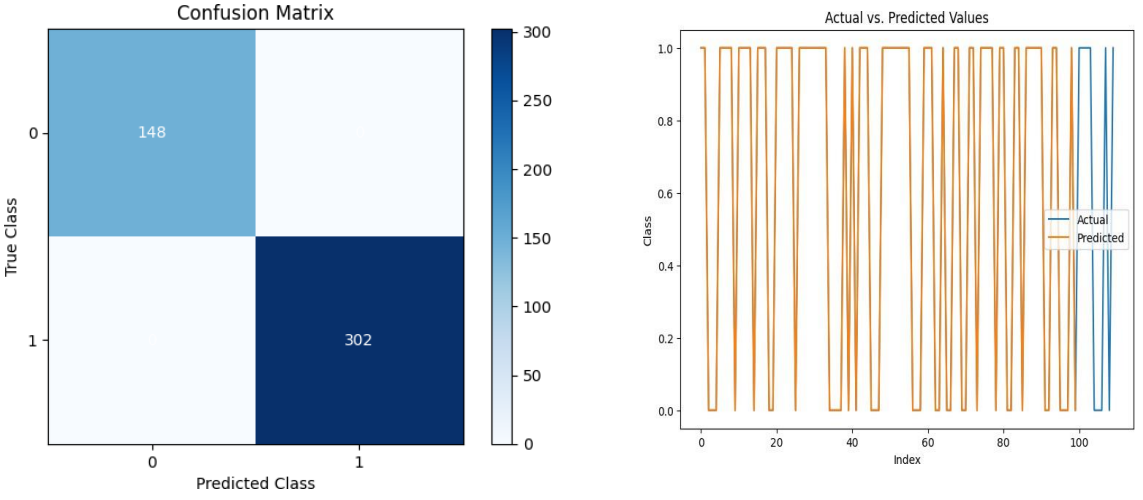

Fig. 2. (a) confusion matrix; (b) actual with predicted values

As shown in "Figure 3" the first Figure (a) is a confusion matrix, and the second Figure (b) is actual with predicted values for the batch 2 dataset.

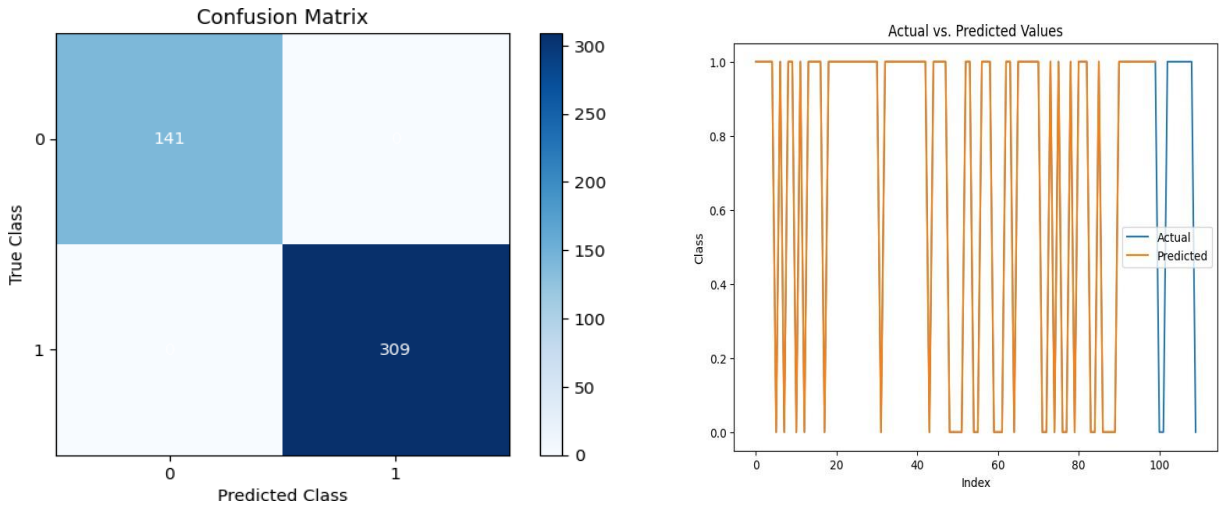

Fig. 3. (a) confusion matrix; (b) actual with predicted values

As shown in "Figure 4" the first Figure (a) is a ROC curve and the second Figure (b) is a precision and recall curve for batch1 and batch2 datasets.

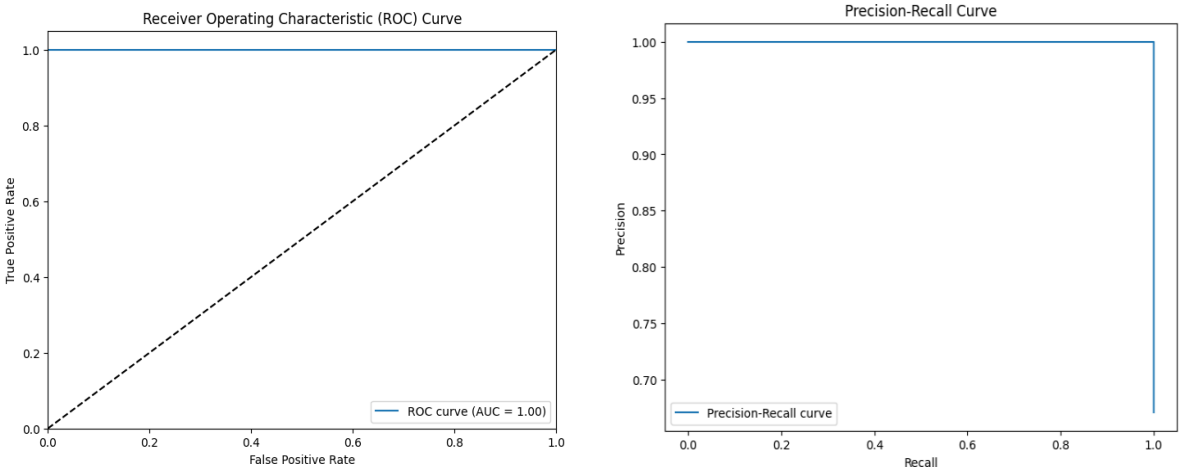

Fig. 4. (a) ROC curve; (b) Precision and recall curve

It is important to compare the proposed results with state-of-the-art research as shown in Table III.

TABLE III. A COMPARISON OF RESULTS WITH RELATED WORK

	year	*method*	*accuracy*
[6]	2019	FCN	99
[9]	2020	DT, SVM, NB	97.8
[10]	2021	Multi-scale CNN	94
[11]	2022	CNN	99.86
[12]	2023	ML & LDA	99.6
Proposed	Current Study	ANN	100

5. CONCLUSION

This study proposed an effective deep-learning approach for SERS detection using an ANN. The proposed framework demonstrates superior performance compared to traditional methods, offering enhanced accuracy and robustness in complex SERS spectra analysis. This research contributes to the broader field of spectroscopy-based detection techniques and paves the way for the development of innovative applications using deep learning in molecular sensing.

Conflicts Of Interest

The author asserts that there are no conflicts of interest that could have affected the study design, methodology, or results.

Funding

The absence of acknowledgments or thank you notes to institutions or sponsors in the paper suggests no financial support was received.

Acknowledgment

The author would like to thank the administrative staff at the institution for their assistance and logistical support throughout the duration of this research.

References

[1] N. N. Jamil and A. K. Kareem, "Comparative Analysis on Machine Learning and One-Dimensional Convolutional Neural Network to Predict Surface Enhanced Raman Spectroscopy," in 2023 3rd International Conference on Computing and Information Technology (ICCIT), 2023, pp. 216–221.
[2] A. Barucci et al., "Label-free SERS detection of proteins based on machine learning classification of chemo- structural determinants," Analyst, vol. 146, no. 2, pp. 674–682, 2021.
[3] A. I. Pérez-Jiménez, D. Lyu, Z. Lu, G. Liu, and B. Ren, "Surface-enhanced Raman spectroscopy: benefits, trade-offs and future developments," Chem. Sci., vol. 11, no. 18, pp. 4563–4577, 2020.
[4] M. A. Tahir, N. E. Dina, H. Cheng, V. K. Valev, and L. Zhang, "Surface-enhanced Raman spectroscopy for bioanalysis and diagnosis," Nanoscale, vol. 13, no. 27, pp. 11593–11634, 2021.
[5] J.-W. Tang et al., "Comparative analysis of machine learning algorithms on surface enhanced Raman spectra of clinical Staphylococcus species," Front. Microbiol., vol. 12, p. 696921, 2021.
[6] M. Serhan et al., "Total iron measurement in human serum with a novel smartphone-based assay," IEEE J. Transl. Eng. Heal. Med., vol. 8, pp. 1–9, 2020.
[7] W. Nam, X. Ren, I. Kim, J. Strobl, M. Agah, and W. Zhou, "Plasmonically calibrated label-free surface-enhanced Raman spectroscopy for improved multivariate analysis of living cells in cancer subtyping and drug testing," Anal. Chem., vol. 93, no. 10, pp. 4601–4610, 2021.
[8] W. Hu et al., "Machine learning protocol for surface-enhanced raman spectroscopy," J. Phys. Chem. Lett., vol. 10, no. 20, pp. 6026–6031, 2019.
[9] F. U. Ciloglu, A. M. Saridag, I. H. Kilic, M. Tokmakci, M. Kahraman, and O. Aydin, "Identification of methicillin-resistant Staphylococcus aureus bacteria using surface-enhanced Raman spectroscopy and machine learning techniques," Analyst, vol. 145, no. 23, pp. 7559–7570, 2020.
[10] J. Ding et al., "Rapid identification of pathogens by using surface-enhanced Raman spectroscopy and multi-scale convolutional neural network," Anal. Bioanal. Chem., vol. 413, no. 14, pp. 3801–3811, 2021.
[11] J.-W. Tang et al., "Rapid discrimination of clinically important pathogens through machine learning analysis of surface enhanced Raman spectra," Front. Microbiol., vol. 13, p. 843417, 2022.
[12] S. Park, J. Lee, S. Khan, A. Wahab, and M. Kim, "Sersnet: surface-enhanced Raman spectroscopy based biomolecule detection using deep neural network," Biosensors, vol. 11, no. 12, p. 490, 2021.
[13] A. A. Nafea, "Artificial Neural Network and Latent Semantic Analysis for Adverse Drug Reaction Detection Abstract :," no. May, 2023.
[14] M. S. I. Alsumaidaie, K. M. A. Alheeti, and A. K. Alaloosy, "An Assessment of Ensemble Voting Approaches, Random Forest, and Decision Tree Techniques in Detecting Distributed Denial of Service (DDoS) Attacks," 2023.
[15] A. A. Nafea, M. Mishlish, A. M. S. Shaban, M. M. AL-Ani, K. M. A. Alheeti, and H. J. Mohammed, "Enhancing Student's Performance Classification Using Ensemble Modeling," Iraqi J. Comput. Sci. Math. vol. 4, no. 4, pp. 204–214, 2023.
[16] A. D. Sallibi and K. M. A. Alheeti, "Detection of Autism Spectrum Disorder in Children Using Efficient Pre-Processing and Machine Learning Techniques," in 2023 International Conference on Decision Aid Sciences and Applications (DASA), 2023, pp. 505–514.