MSI Journal of AI and Technology

Raman Scattering with Artificial Intelligence in Biomedicine: A New Paradigm for Molecular Diagnostics

Dr. Nadir Omar Massoud Driza[1*], Ola Mohammed Ibrahim[1], Rafa Saad Abdulsalam Hamad[1], Hanan Mohammed Abdulsalam Ali[1,2]

[1*] Department of Physics, Faculty of Arts and Sciences Elmarj, University of Benghazi, Elmarj, Libya.

[2] Higher institute of science and Technology Elmarj, Libya.

*** Correspondence:** Dr. Nadir Omar Massoud Driza

The authors declare that no funding was received for this work.

OPEN ACCESS

Received: 15-November-2025
Accepted: 10-December-2025
Published: 14-December-2025

This article is published in the **MSI Journal of AI and Technology**,

ISSN 3107-6181 (Online)

Volume: 1, Issue: 3 (Oct-Dec) 2025

ABSTRACT: Raman scattering is an inelastic light-scattering process that yields a highly specific molecular fingerprint derived from the vibrational modes of biological components, including lipids, proteins, nucleic acids, and metabolic compounds. While Raman spectroscopy offers unparalleled chemical specificity and label-free analysis, the high dimensionality and complexity of biological spectral data have historically limited its clinical throughput. The integration of Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), is transforming this paradigm. This paper reviews the foundational principles of Raman spectroscopy and details how AI-driven analysis is enabling the rapid, automated, and accurate interpretation of complex biological spectra, establishing a new frontier for molecular diagnostics in biomedicine.

Keyword: *Raman Scattering, Artificial Intelligence, Molecular Diagnostics in Biomedicine*

1. Introduction

Traditional biomedical diagnostics rely on macroscopic, morphological, or labor-intensive biochemical assays. In

contrast, Raman spectroscopy (RS) provides a molecular-level, non-destructive assessment of a sample's chemical composition [1]. The process involves the inelastic scattering of incident photons by molecular vibrations, resulting in a spectrum of shifted wavelengths. This spectrum is an intrinsic signature of the sample, sensitive to the subtle changes in lipids, proteins, nucleic acids, and metabolites characteristic of disease states.

The clinical transition of RS, however, faces a significant hurdle: the complexity and noise inherent in biological spectra [2]. Variations due to instrumental factors, fluorescence background, and the subtle nature of pathological changes within a high-dimensional dataset make manual or traditional statistical analysis challenging and time-consuming. The emergence of powerful AI techniques offers a solution, automating feature extraction and classification with unprecedented accuracy and speed [3].

2. Methods of Raman Scattering Fundamentals in Biomedicine and AI Analysis

2.1 Raman Scattering Principles

Raman scattering provides chemical specificity based on the principle of inelastic interaction:

$$\Delta\bar{\nu} = \frac{1}{\lambda_i} - \frac{1}{\lambda_s}$$

where $\Delta\bar{\nu}$ is the Raman shift (in cm^{-1}), λ_i is the incident wavelength, and λ_s is the scattered wavelength (see Fig. 1).

The spectral regions correspond to major biological components, generating a unique molecular fingerprint for each tissue or cell type. This phenomenon is critical for understanding the molecular changes associated with disease.

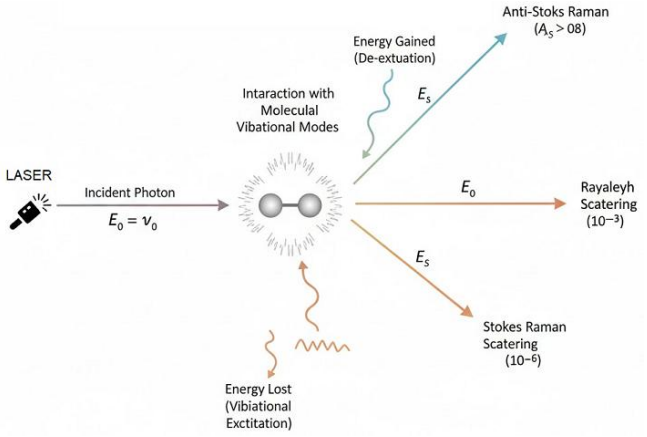

Figure 1: Illustration of Raman Scattering. A diagram illustrating the basic principle of Raman scattering, showing an incident photon interacting with a molecule, leading to Rayleigh scattering (elastic) and Stokes/Anti-Stokes Raman scattering (inelastic) with energy shifts corresponding to molecular vibrational states. These shifts create the unique molecular finegrnint.

Figure 2 declares representative Raman spectra of biological components. It clarifies the typical Raman spectra with distinct peaks for major biological components like lipids, proteins, and nucleic acids, highlighting their characteristic molecular fingerprints.

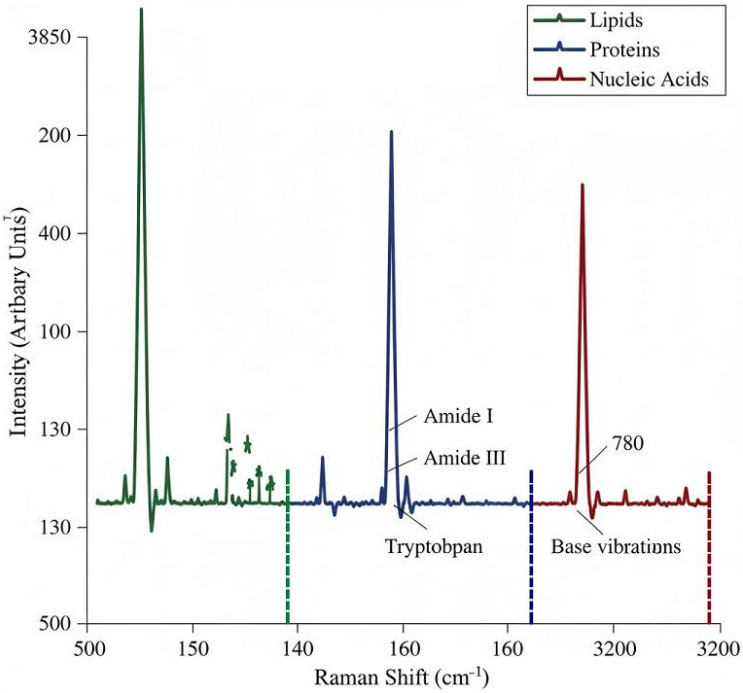

Figure 2: Representative Raman Spectra of Biological Components.

Table I: It illustrates that the [4] Mol2Raman model is an advanced AI tool that predicts the chemical fingerprint of a molecule (the Raman spectrum). It achieves this by directly analyzing the molecule's structural identity, which is provided as a simple line of code called SMILES. This specialized analysis is performed by a Graph Neural Network (GNN), allowing the system to quickly and accurately forecast spectral results without requiring any physical laboratory testing.

Biomolecule Class	**Example Peak Region (cm^{-1})**	**Chemical Significance**
Proteins	1600-1700 (Amide I)	Secondary structure (e.g.,α-helix, β-sheet)
Lipids	2800-3000 (C–H stretch)	Lipid-to-protein ratio, fatty acid saturation
Nucleic Acids	700-1000 (Phosphate backbone)	Cell proliferation rate, DNA damage
Carbohydrates	1000-1150 (C–O stretch)	Glycosylation changes, cell wall components

Pathological conditions, such as cancer or infection, induce quantifiable shifts in the intensity and position of these peaks, which are captured as large datasets of spectral information [5].

2.2 The Role of Artificial Intelligence in Spectral Analysis

AI, encompassing ML and DL, provides the necessary computational framework to handle the volume and complexity of Raman data, enabling the effective translation of spectral shifts into clinical decisions [6] [7] [8] [9].

2.2.1 Machine Learning for Classification and Regression

Traditional Machine Learning [10] (ML02) models are frequently employed for classification tasks, which involve differentiating between biological states, such as healthy versus diseased. Several common algorithms are utilized for this purpose. The Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA)

approach is a standard method for tissue type differentiation [11] [12] [13] [14]. In this combined technique, PCA is initially used for dimensionality reduction and noise filtering, which is then followed by LDA to maximize the separation between predefined classes. Another effective algorithm, Support Vector Machines (SVM), is particularly well-suited for high-dimensional data as it works by finding the optimal hyperplane to separate classes within the spectral space; this technique is commonly applied in areas such as single-cell sorting and microbial identification. Finally, Random Forests (RF) algorithms, which rely on using multiple decision trees, provide robust and interpretable classification models and are often leveraged to identify the most important spectral features, or biomarkers.

2.2.2 Deep Learning for Feature Extraction and Automated Analysis

Deep Learning (DL) models [15], especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), represent the cutting edge in spectral analysis by learning directly from raw, high-dimensional spectral data, thus bypassing the need for traditional feature engineering. CNNs are particularly well-suited for processing spectral data, which can be treated as one-dimensional signals. Their convolutional layers automatically learn the most relevant spatial features (such as peak positions and widths) for subsequent classification or regression tasks. A typical CNN architecture for Raman analysis comprises several 1D convolutional layers followed by pooling layers for dimensionality reduction, and ultimately, fully connected layers for classification. In terms of application, CNNs have demonstrated high accuracy (often exceeding 95%) in critical tasks, including differentiating malignant from benign tissue and classifying single-cell bacterial strains based purely on their spectral signatures [16] [17] [18]. Furthermore, CNNs excel in Hyperspectral Imaging Analysis by processing Raman hyperspectral images (where each pixel contains a full Raman spectrum), which enables automated tissue segmentation and real-time visualization of pathological regions [19] [20] [21]. While less common than CNNs, Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, are also being explored for processing sequential spectral data, giving them the potential to capture long-range dependencies across the entire spectral range.

Figure 3 is conceptual diagram of an AI-Raman system for diagnostics a flowchart illustrating the typical workflow of an AI-Raman system, from sample acquisition and Raman spectroscopy data collection, through data preprocessing, to AI model training (e.g., CNN) and final diagnostic output.

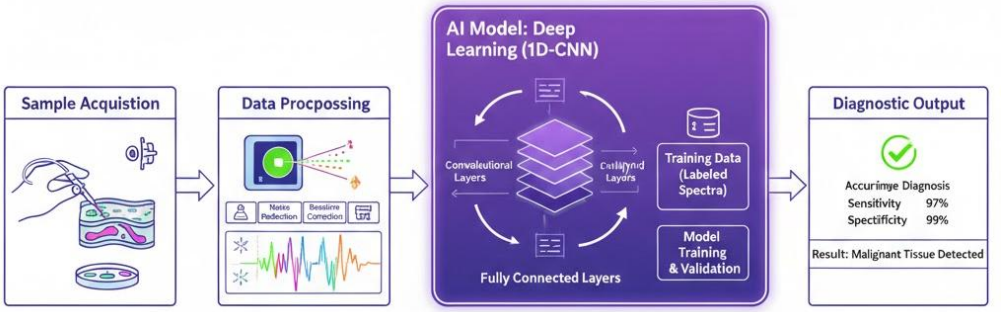

Figure 3: Concepual Diagram of Al-Raman System for Diagnostics.

Table II: Here is a table summarizing the specific differences between Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) as applied to spectral analysis.

Feature	**Convolutional Neural Network (CNN)**	**Recurrent Neural Network (RNN) (e.g., LSTM)**
Data Interpretation	Treats the spectrum as a 1D spatial signal (like a narrow image).	Treats the spectrum as a sequence of data points over time/wavenumber.
Core Mechanism	Convolutional Filters (Kernels) slide over small, local windows of the data.	Recurrent Connections and Hidden States process data sequentially.
Primary Strength	Captures local spectral features (peak position, width, shape) and is translation invariant.	Captures long-range dependencies and context across the entire spectral range.

Feature Focus	Local patterns (e.g., the specific signature of a single functional group).	Sequential context (e.g., the relationship between a peak at 400 cm^{-1} and one at 1800 cm^{-1}).
Architecture	Built from 1D Convolutional Layers, Pooling Layers, and Fully Connected Layers.	Built from Recurrent Cells (e.g., LSTM or GRU), which form a deep sequential structure.
Typical Application	Classification (e.g., tissue type, bacterial strain) and Hyperspectral Imaging (segmentation).	Time-series prediction (less common in general spectral classification, but explored for complex sequences).

3. Results and Discussion

3.1. Performance Metrics and Validation

The efficacy and clinical utility of these AI-Raman models are quantified using standard machine learning performance metrics, ensuring robust and generalizable results.

Table III: This table detailing the performance metrics (Accuracy, Sensitivity, Specificity, and AUC) is crucial because it provides the standardized, quantifiable evidence needed to confirm that the AI models are effective for clinical use.

Metric	**Definition**	**Clinical Interpretation**
Accuracy	Ratio of correctly classified instances to the total instances.	Overall correct diagnosis rate.

Sensitivity (Recall)	True Positives / (True Positives + False Negatives).	Ability to correctly identify diseased samples.
Specificity	True Negatives / (True Negatives + False Positives).	Ability to correctly identify healthy (negative) samples.
Area Under the Curve (AUC)	Area under the Receiver Operating Characteristic (ROC) curve.	Evaluates how well the model can distinguish between classes at all potential classification thresholds.
F1 Score	Harmonic mean of precision and sensitivity.	A measure of a model's accuracy on a specific class, especially useful when datasets are imbalanced.

Validation typically employs techniques like k-fold cross-validation to ensure that the model's performance is not specific to the training subset. A high AUC value (close to 1.0) is often cited as proof of strong diagnostic capability, demonstrating the AI system's ability to reliably translate complex spectral changes into a simple, binary diagnostic output [22] [23].

3.2. AI-Driven Applications in Biomedical Diagnostics

The synergy between high-specificity Raman spectroscopy and high-efficiency Artificial Intelligence (AI) analysis is creating transformative applications across medicine and diagnostics. Specifically, AI-Raman systems are under development for Intraoperative Margin Assessment, providing real-time tissue [24] classification—such as distinguishing between tumor and normal brain or breast tissue—orders of magnitude faster than conventional frozen section analysis, thereby significantly decreasing operation time and enhancing positive margin rates [25]

[26]. Similarly, AI models are used for Rapid Antimicrobial Susceptibility Testing (AST) by analyzing subtle spectral shifts in bacteria in response to antibiotics within minutes, drastically cutting the time required for AST from days to hours; this allows clinicians to prescribe targeted antibiotics earlier, which is critical in the fight against Antimicrobial Resistance (AMR) [27]. Furthermore, the integration of AI with Surface-Enhanced Raman Scattering [10] [28] (SERS) technology enables Liquid Biopsy and Biofluid Analysis through the ultra-sensitive detection and classification of low-concentration biomarkers, such as circulating tumor cells or exosomes, found in biofluids, thus opening up new pathways for non-invasive diagnostics [28].

4. Conclusion

In this article, we made it clear that Raman scattering offers an invaluable molecular window into biological systems. However, its potential is fully realized only when paired with sophisticated computational tools. The synergy between Raman spectroscopy and Artificial Intelligence addresses the inherent challenges of biological data complexity, enabling rapid, automated, and highly accurate molecular diagnostics. This convergence marks a significant advancement in biomedicine, promising to revolutionize diagnostics from the pathology lab to the point-of-care setting by providing clinicians with instant, high-specificity molecular information.

References

1. Balan, V., Mihai, C.-T., Cojocaru, F.-D., Uritu, C.-M., Dodi, G., Botezat, D., & Gardikiotis, I. (2019). Vibrational spectroscopy fingerprinting in medicine: from molecular to clinical practice. *Materials*, *12*(18), 2884. $https://doi.org/10.3390/ma12182884$

2. Chen, X., Shen, J., Liu, C., Shi, X., Feng, W., Sun, H., Zhang, W., Zhang, S., Jiao, Y., Chen, J., Hao, K., Gao, Q., Li, Y., Hong, W., Wang, P., Feng, L., & Yue, S. (2024). Applications of data characteristic AI-assisted Raman spectroscopy in pathological classification. *Analytical Chemistry*, *96*(16), 6158–6169. $https://doi.org/10.1021/acs.analchem.3c04930$

3. Coca-Lopez, N., Alcolea-Rodriguez, V., Bañares, M. A., Brockhauser, S., Gorenflot, J., Henderson, A., Hildebrandt, R., Jeliazkova, N., Kochev, N., Lozano Diz, E., Pilat, Z., Polli, D., Strömert, P., Sturm, C., Vanna, R., & Portela, R. (2025). Artificial intelligence-powered Raman spectroscopy through open science and FAIR principles. *ACS Nano*, *19*(44), 38189–38218. $https://doi.org/10.1021/acsnano.5c09165$

4. Cook, H., Crisford, A., Bourdakos, K., Dunlop, D., Oreffo, R. O. C., & Mahajan, S. (2024). Holistic vibrational spectromics assessment of human cartilage for osteoarthritis diagnosis. *Biomedical Optics Express*, *15*(7), 4264–4280. $https://doi.org/10.1364/BOE.520171$

5. Dai, R., Sun, M., Lu, M., & Deng, L. (2024). Interpretable machine learning models based on shear-wave elastography radiomics for predicting cardiovascular disease in diabetic kidney disease patients. *Journal of Diabetes Investigation*, *15*(9), 1637–1650. $https://doi.org/10.1111/jdi.14294$

6. Delrue, C., Speeckaert, M. M., & De Bruyne, S. (2025). Deep learning algorithms and Raman spectroscopy in the clinical laboratory setting. *Critical Reviews in Clinical Laboratory Sciences*. Advance online publication. $https://doi.org/10.1080/10408363.2025.2549305$

7. Fuentes, A. M., Narayan, A., Milligan, K., Lum, J. J., Brolo, A. G., Andrews, J. L., & Jirasek, A. (2023). Raman spectroscopy and convolutional neural networks for monitoring biochemical radiation response in breast tumour xenografts. *Scientific Reports*, *13*(1), Article 1530. $https://doi.org/10.1038/s41598-023-28479-2$

8. He, J., Li, H., Zhang, B., Liang, G., Zhang, L., Zhao, W., Zhao, W., Zhang, Y.-J., Wang, Z.-X., & Li, J.-F. (2025). Convolutional neural network-assisted Raman spectroscopy for high-precision diagnosis of glioblastoma. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *329*, 125615. $https://doi.org/10.1016/j.saa.2024.125615$

9. Hemanth, D. J. (2021). Automated feature extraction in deep learning models: A boon or a bane? *2021 8th International Conference on Electrical Engineering,*

Computer Science and Informatics (EECSI), 10–14. $https://doi.org/10.23919/EECSI53397.2021.9624287$

10. Ho, C.-S., Neal, J., Hogan, C. A., Blackmon, L., Jeffrey, S. S., Holodniy, M., Banaei, N., Saleh, A. A. E., Ermon, S., & Dionne, J. (2019). Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nature Communications*, *10*(1), Article 4927. $https://doi.org/10.1038/s41467-019-12898-9$

11. Humaran, D., Castillo, A., Blay, L., Pascual, I., Matute-Molina, K., Perez-Anker, J., Puig, S., Fernandez, P. L., & Julian, J. F. (2025). Towards a new standard: Prospective validation of *Ex Vivo* fusion confocal microscopy for intraoperative margin assessment in breast-conserving cancer surgery. *Cancers*, *17*(23), 3848. $https://doi.org/10.3390/cancers17233848$

12. Jong, L.-J. S., Veluponnar, D., Geldof, F., Sanders, J., Guimaraes, M. D. S., Peeters, M.-J. T. F. D. V., Duijnhoven, F. V., Sterenborg, H. J. C. M., Dashtbozorg, B., & Ruers, T. J. M. (2025). Toward real-time margin assessment in breast-conserving surgery with hyperspectral imaging. *Scientific Reports*, *15*(1), Article 9556. $https://doi.org/10.1038/s41598-025-94526-9$

13. Lee, S., Park, R., & Jung, H. S. (2025). AI-enhanced surface-enhanced Raman scattering for accurate and sensitive biomedical sensing. *Advanced Intelligent Discovery*. Advance online publication. $https://doi.org/10.1002/aidi.202500030A$

14. Liang, S., Singh, M., Dharmaraj, S., & Gam, L.-H. (2010). The PCA and LDA analysis on the differential expression of proteins in breast cancer. *Disease Markers*, *29*(4), 231–242. $https://doi.org/10.3233/DMA-2010-0753$

15. Liu, Y., Chen, S., Xiong, X., Wen, Z., Zhao, L., Xu, B., Guo, Q., Xia, J., & Pei, J. (2025). Artificial intelligence guided Raman spectroscopy in biomedicine: Applications and prospects. *Journal of Pharmaceutical Analysis*, *15*(2), 101271. $https://doi.org/10.1016/j.jpha.2025.101271$

16. Luo, R., Guo, S., Hniopek, J., & Bocklitz, T. (2025). 3D hyperspectral data analysis with spatially aware deep learning for diagnostic applications. *Analytical Chemistry*, *97*(14), 7729–7737. $https://doi.org/10.1021/acs.analchem.4c05549$

17. Matthies, L., Gebrekidan, M. T., Tegtmeyer, J. F., Oetter, N., Rohde, M., Vollkommer, T., Smeets, R., Wilczaks, W., Stelzle, F., Gosau, M., Braeuer, A. S., & Knipfer, C. (2021). Optical diagnosis of oral cavity lesions by label-free Raman spectroscopy. *Biomedical Optics Express*, *12*(2), 836–851. $https://doi.org/10.1364/BOE.409456$

18. Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., & Ilono, P. (2025). Deep convolutional neural networks in medical image analysis: A review. *Information*, *16*(3), 195. $https://doi.org/10.3390/info16030195$

19. Paraskevaidia, M., Baker, M. J., Butler, H., Byrne, H. J., Chakkumpulakkal P. V., T., Christie, L., Crean, S., Gardner, P., Gassner, C., Kazarian, S. G., Kochan, K., Kyrgiou, M., Lima, K. M. G., Martin-Hirsch, P. L., Paraskevaidis, E., Pebotuwa, S., Adegoke, J. A., Sala, A., Santos, M., Sule-Suso, J., Tyagi, G., Walsh, M., & Wood, B. (2021). Clinical applications of infrared and Raman spectroscopy in the fields of cancer and infectious diseases. *Applied Spectroscopy Reviews*, *56*(8–10), 804–868. $https://doi.org/10.1080/05704928.2021.1946076$

20. Pimenta, S., & Correia, J. H. (2025). Biomedical applications of Raman spectroscopy: A review. *Photochemistry*, *5*(4), 29. $https://doi.org/10.3390/photochem5040029$

21. Rahman, M. H.-U., Sikder, R., Tripathi, M., Zahan, M., Ye, T., Gnimpieba Z., E., Jasthi, B. K., Dalton, A. B., & Gadhamshetty, V. (2024). Machine learning-assisted Raman spectroscopy and SERS for bacterial pathogen detection: Clinical, food safety, and environmental applications. *Chemosensors*, *12*(7), 140. $https://doi.org/10.3390/chemosensors12070140$

22. Rasool, N., Wani, N. A., Bhat, J. I., Saharan, S., Sharma, V. K., Alsulami, B. S., Alsharif, H., & Lytras, M. D. (2025). CNN-TumorNet: Leveraging explainability in

deep learning for precise brain tumor diagnosis on MRI images. *Frontiers in Oncology*, *15*, 1554559. $https://doi.org/10.3389/fonc.2025.1554559$

23. Sorrentino, S., Gussoni, A., Calcagno, F., Pasotti, G., Avagliano, D., Rivalta, I., Garavelli, M., & Polli, D. (2025). Mol2Raman: A graph neural network model for predicting Raman spectra from SMILES representations. *Chemical Science*. Advance online publication. $https://doi.org/10.1039/D5DD00210A$

24. Sun, L., Xu, Z., Huang, W., Wu, S., Lin, X., Zhu, F., Liu, N., Huang, M., Chen, R., & Zeng, H. (2016). Preliminary study of differentiating smears from cancerous and non-cancerous nasopharyngeal tissue using confocal Raman spectroscopy. *Journal of Cancer Research and Clinical Oncology*, *142*(3), 823–831. $https://doi.org/10.1007/s00432-015-2082-3$

25. Wetzel, W. (2024). Optimizing AI models for Raman spectroscopy: Improving disease diagnosis. *Spectroscopy*.

26. Witjes, H., Rijpkema, M., van der Graaf, M., Melssen, W., Heerschap, A., & Buydens, L. (2003). Multispectral magnetic resonance image analysis using principal component and linear discriminant analysis. *Journal of Magnetic Resonance Imaging*, *18*(6), 724–731. $https://doi.org/10.1002/jmri.10237$

27. Yadav, S., & Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 78–83. $https://doi.org/10.1109/IACC.2016.25$

28. Abougarair, A. J., Abolaeha, O., & Aburakhis, M. K. (2025). Classification of malignant and benign skin lesions using CNN models. *IJEIT On Engineering and Information Technology*, *14*(1). $https://doi.org/10.36602/ijeit.v14i1.572$