The North African Journal of Scientific Publishing (NAJSP)
مجلة شمال إفريقيا للنشر العلمي (NAJSP)
E-ISSN: 2959-4820
Volume 4, Issue 2, 2026
Page No: 68-79
Website: https://najsp.com/index.php/home/index
SJIFactor 2024: 5.49 معامل التأثير العربي (AIF) 2025: 0.69 ISI 2024: 0.696

Raman Spectroscopy Integrated with Artificial Intelligence for Advanced Cancer Diagnosis

Nadir Omar Massoud Driza[1*], Rafa Saad Abdulsalam Hamad[2], Ola Mohammed Ibrahim[3], Hanan Mohammed Abdulsalam Ali[4]
[1,2,3]Department of Physics and Medical Physics, faculty of Arts and Sciences Elmarj, University of Benghazi, Elmarj, Libya
[4]Higher Institute of Science and Technology, Elmarj, Libya.

تقنية رامان الطيفية المدمجة مع الذكاء الاصطناعي لتشخيص السرطان المتقدم

نادر عمر مسعود ادريزه[1*]، رافع سعد عبدالسلام حمد[2]، علاء محمد إبراهيم[3]، حنان محمد عبدالسلام علي[4]
[1،2،3]قسم الفيزياء وشعبة الفيزياء الطبية، كلية الآداب والعلوم المرج، جامعة بنغازي، المرج، ليبيا
[4]المعهد العالي للعلوم والتقنية، المرج، ليبيا

*Corresponding author: nadir.driza@uob.edu.ly

Received: February 11, 2026	Accepted: March 26, 2026	Published: April 07, 2026

Abstract:
Getting an accurate and early diagnosis of cancer is still a big problem because traditional diagnostic methods are often invasive and don't target specific molecules. This study introduces a comprehensive methodology utilizing Raman spectroscopy and artificial intelligence (AI) to enhance cancer detection and diagnosis. Raman spectroscopy is a non-invasive, label-free way to look at the biochemical makeup of biological tissues by using molecular fingerprinting. This lets researchers find changes in nucleic acids, proteins, and lipids that are linked to cancer or live tissues affected by cancer. The study delineates the essential principles of Raman scattering and the arrangement of clinically relevant systems, encompassing fiber-optic probes for in vivo measurements. An AI-based analysis framework is used to improve diagnostic performance by using spectral preprocessing, feature extraction, and machine learning classification. The results exhibit elevated sensitivity and specificity across various cancer types, signifying the dependability of the suggested methodology. It's possible uses in real-time diagnosis and treatment monitoring also show how important it is for improving non-invasive cancer diagnostics.

Keywords: Artificial Intelligence (AI), Biochemical Profiling, Cancer Diagnostics, Molecular Fingerprinting, Raman Spectroscopy, Spectral Preprocessing.

الملخص:
لا يزال التشخيص الدقيق والمبكر للسرطان يمثل تحديًا كبيرًا، نظرًا لأن طرق التشخيص التقليدية غالبًا ما تكون جراحية ولا تستهدف جزيئات محددة. تقدم هذه الدراسة منهجية شاملة تستخدم مطيافية رامان والذكاء الاصطناعي لتحسين الكشف عن السرطان وتشخيصه. تُعد مطيافية رامان طريقة غير جراحية ولا تتطلب استخدام علامات لدراسة التركيب الكيميائي الحيوي للأنسجة البيولوجية من خلال البصمة الجزيئية. يُمكّن هذا الباحثين من تحديد التغيرات في الأحماض النووية والبروتينات والدهون المرتبطة بالسرطان أو الأنسجة الحية المصابة به. توضح الدراسة المبادئ الأساسية لتشتت رامان وترتيب الأنظمة ذات الصلة سريريًا، بما في ذلك مجسات الألياف البصرية للقياسات الحيوية. يُستخدم إطار تحليل قائم

على الذكاء الاصطناعي لتحسين الأداء التشخيصي من خلال المعالجة المسبقة للطيف، واستخلاص الميزات، وتصنيف التعلم الآلي. تُظهر النتائج حساسية ونوعية عاليتين عبر أنواع مختلفة من السرطان، مما يدل على موثوقية المنهجية المقترحة. كما تُظهر استخداماتها المحتملة في التشخيص الفوري ومراقبة العلاج مدى أهميتها في تحسين التشخيص غير الجراحي للسرطان.

الكلمات المفتاحية: الذكاء الاصطناعي، التحليل البيوكيميائي، تشخيص السرطان، البصمة الجزيئية، مطيافية رامان، المعالجة المسبقة للطيف.

Introduction:

Cancer is marked by atypical cellular proliferation and molecular changes within tissues. Conventional diagnostic methods, including histopathology and imaging techniques like MRI or CT, often involve invasiveness, prolonged processing times, or limited molecular specificity. Raman spectroscopy, a technique based on light scattering, provides biochemical profiles of tissues, and its combination with artificial intelligence has shown potential to improve detection accuracy [1].

Due to its non-invasive, label-free properties and capacity for real-time analysis, Raman spectroscopy can be incorporated into clinical settings, such as operating rooms or outpatient clinics, to support diagnosis and inform therapeutic decisions [2,3].

Principles of Raman Scattering:

In this section, we focus our work on the physical mechanism of Raman technology used in medical diagnosis and the biochemical basis of Raman-based cancer diagnosis, which is one of the most accurate and successful technologies and whose effect on living tissues is considered almost non-existent. Figure (1) outlines schematic diagram of Raman experimental setup.

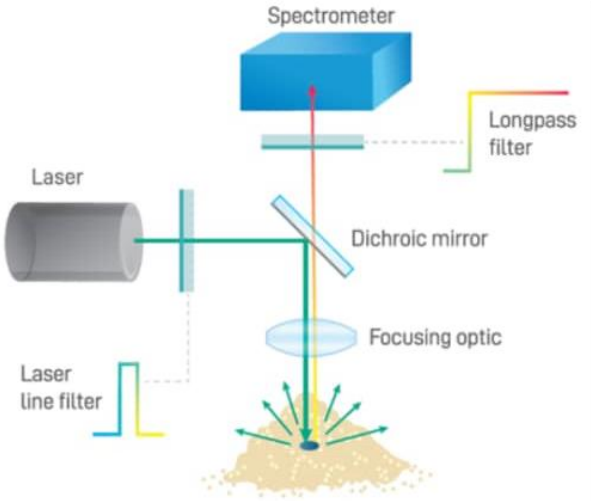

Figure (1): Schematic diagram of Raman experimental setup. This diagram shows the optical path in "micro" setup for Raman light scattering. It presents the basic principles of Raman light scattering and interaction with tissue.

Physical Mechanism:

When tissue is illuminated by light, typically from a laser source [4], a small fraction of photons undergo inelastic scattering due to interactions with molecular vibrations. This process generates a Raman spectrum characterized by peaks that correspond to specific chemical bonds or molecular components such as proteins, lipids, and nucleic acids [5].

The Biochemical Basis for Raman-based Cancer Diagnostics:

The application of Raman spectroscopy in oncology relies on its capacity to detect molecular fingerprints, which are unique inelastic scattering patterns associated with the vibrational modes of chemical bonds [6-8]. Since cancer arises from metabolic and genetic dysregulation, it produces significant alterations in the cellular microenvironment. Raman spectroscopy captures these pathological changes by identifying distinct biochemical deviations, as described below [6,9].

Increased Nucleic Acid Content:

A characteristic feature of malignant transformation is uncontrolled cellular proliferation, wherein cancer cells undergo rapid mitotic cycles supported by extensive genetic replication. This results in increased concentrations of DNA and RNA within the tissue [10]. Such changes are detectable in Raman spectra as elevated peak intensities related to phosphodiester bonds near 1090 cm^{-1} and

nitrogenous bases, especially the ring breathing modes of guanine and adenine around 1575 cm^{-1} [11,12]. These nucleic acid signals tend to correlate with higher malignancy grades and a larger nuclear-to-cytoplasmic ratio [13], parameters commonly assessed in histopathology.

Altered Protein Expression:

Cancer induces significant modifications in the cellular proteome, including overexpression of certain enzymes, presence of mutated proteins, and remodeling of extracellular matrix components such as collagen [14-17]. In Raman spectra, these alterations appear as changes in the Amide I (1600–1690 cm^{-1}) and Amide III (1230–1300 cm^{-1}) bands, where shifts in peak positions reflect modifications in protein secondary structures, for instance, transitions from *α*-helices to *β*-sheets [18,19]. Because proteins are central to cellular function, these spectral changes offer important insights into the tissue's physiological and metabolic state, allowing detection of early biochemical events linked to tumor development [20].

Changes in Lipid Profiles:

Reprogramming of lipid metabolism is a hallmark of cancer, characterized by increased de novo fatty acid synthesis supporting membrane formation during rapid proliferation, along with enhanced lipid breakdown in some tumor types to satisfy elevated energy demands [21,22]. Raman spectral analysis reveals these metabolic shifts through variations in the C–H stretching region (2800–3000 cm^{-1}) and CH_2 bending modes near 1440 cm^{-1} [23-25]. A notable decrease in the lipid-to-protein ratio is often observed in epithelial malignancies, reflecting changes in cellular composition and tissue architecture [26-28]. Examining these lipid-associated spectral features provides valuable information on tumor metabolism and assists in distinguishing lipid-rich normal tissue from protein-dense malignant tissue [29]. Together, these biochemical and spectral characteristics produce distinct Raman signatures that support sensitive and specific identification of cancerous transformations.

Methods:

Clinical Raman System Configuration:

Adapting Raman spectroscopy from controlled laboratory conditions to clinical settings requires a tailored optoelectrochemical system design. The clinical Raman apparatus must balance high sensitivity with portability and functionality within environments constrained by medical procedures, such as surgical suites or endoscopy rooms (see Figure 2) [30,31]. The system architecture comprises four principal subsystems [32]:

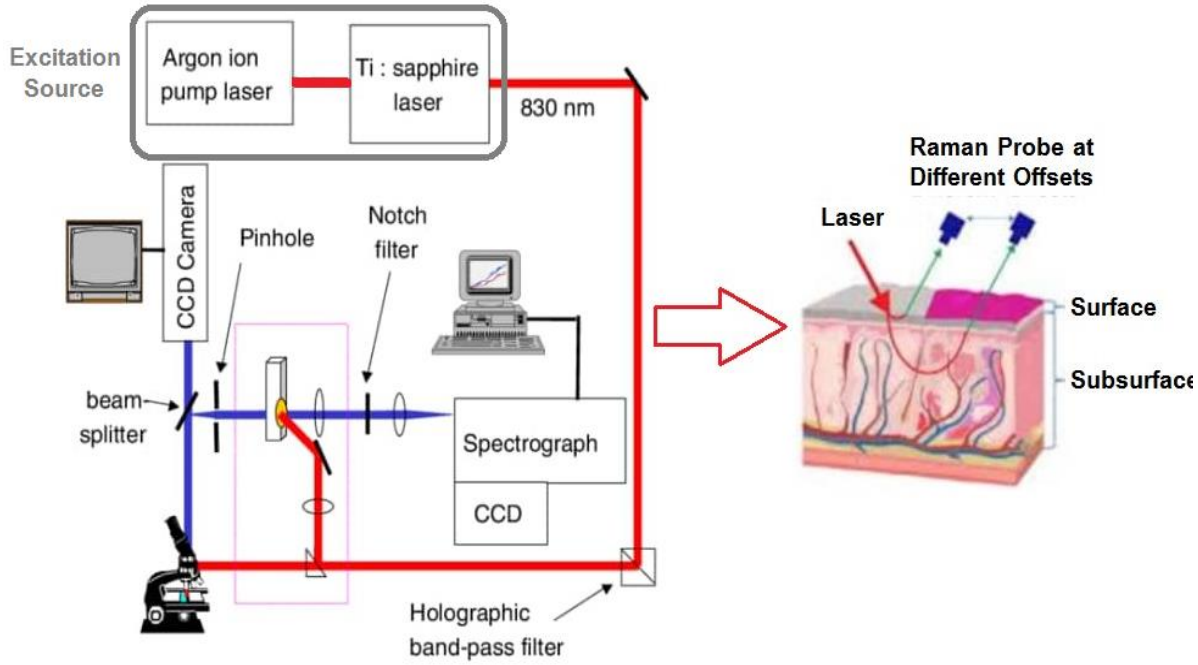

Figure (2): Examples of Raman systems used in clinical settings.

The Excitation Source (Light Delivery):

The excitation source commonly employed in clinical Raman spectroscopy is a diode-stabilized laser operating in the near-infrared (NIR) range, typically at wavelengths around 785 nm or 830 nm. This spectral region is chosen to exploit the biological window, where tissue autofluorescence - which often interferes with spectral measurements - is reduced compared to visible wavelengths. Additionally, NIR

radiation allows for deeper penetration into biological tissues and reduces the likelihood of photothermal damage, thereby supporting safe and effective in vivo data acquisition.

Fiber-Optic Probe Architecture:

In contrast to conventional fixed microscope stages used in laboratories, in vivo clinical applications necessitate a flexible interface. This role is fulfilled by a fiber-optic probe that facilitates minimally invasive signal collection from biological tissues (refer to Figure 3).

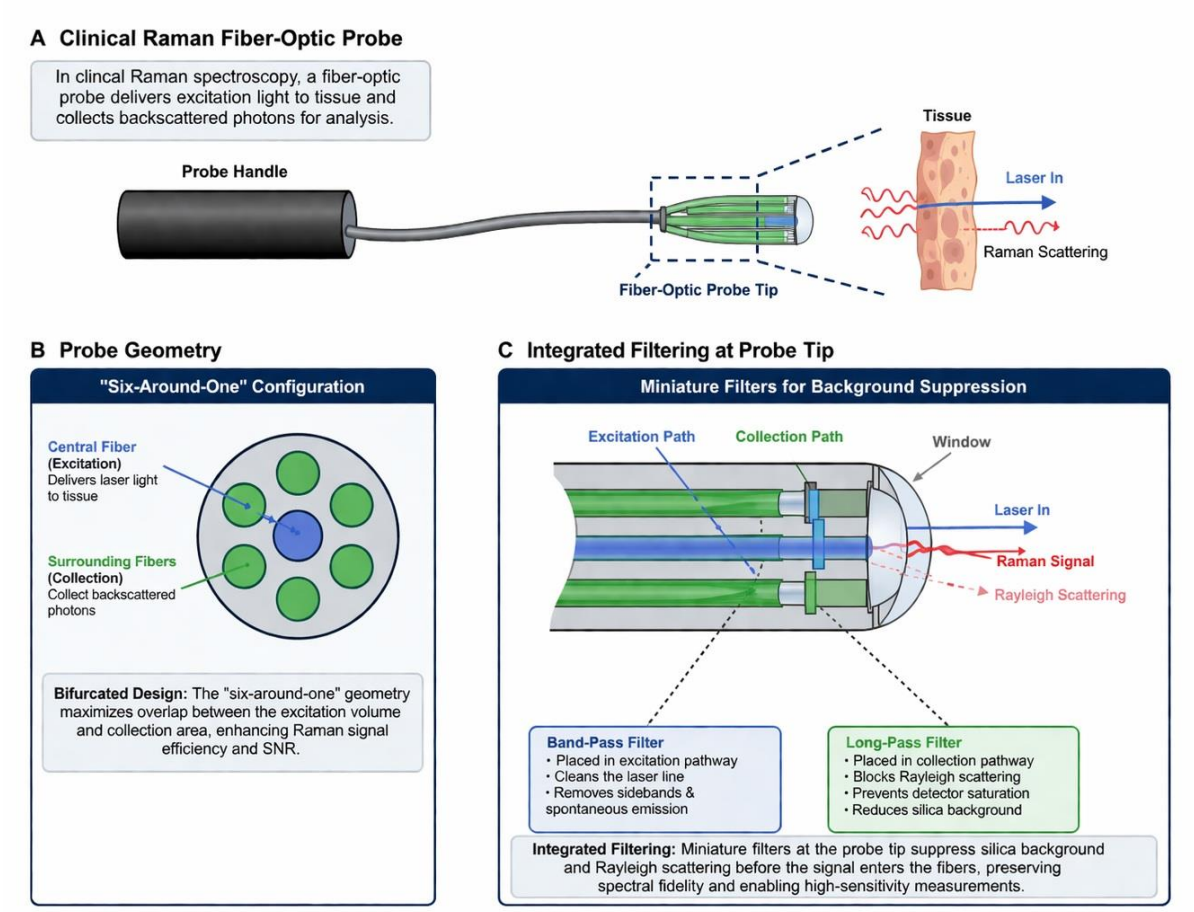

Figure (3): Clinical Raman Fiber-Optic Probe Configuration. Schematic illustration of a fiber-optic Raman probe used for *in vivo* clinical measurements. The probe serves as an interface between the Raman spectroscopy system and biological tissue, delivering excitation laser light to the sample and collecting backscattered photons. The distal probe tip is positioned in contact with the tissue, enabling minimally invasive, real-time spectroscopic analysis.

The probe typically employs a bifurcated six-around-one arrangement, featuring one central fiber for delivering excitation light and six surrounding fibers for collecting backscattered photons. This configuration enhances the spatial overlap between illumination and collection zones, leading to improved Raman signal collection efficiency and enhanced signal-to-noise ratio. To address spectral contamination from the fibers themselves - particularly silica-induced background signals - optical filters are integrated at the probe tip. Band-pass filters are placed in the excitation path to maintain laser spectral purity, while long-pass filters are positioned in the collection path to reduce intense Rayleigh scattering. Positioning these filters distally serves to minimize background contributions before signal transmission, preserving spectral fidelity (see Figure 4).

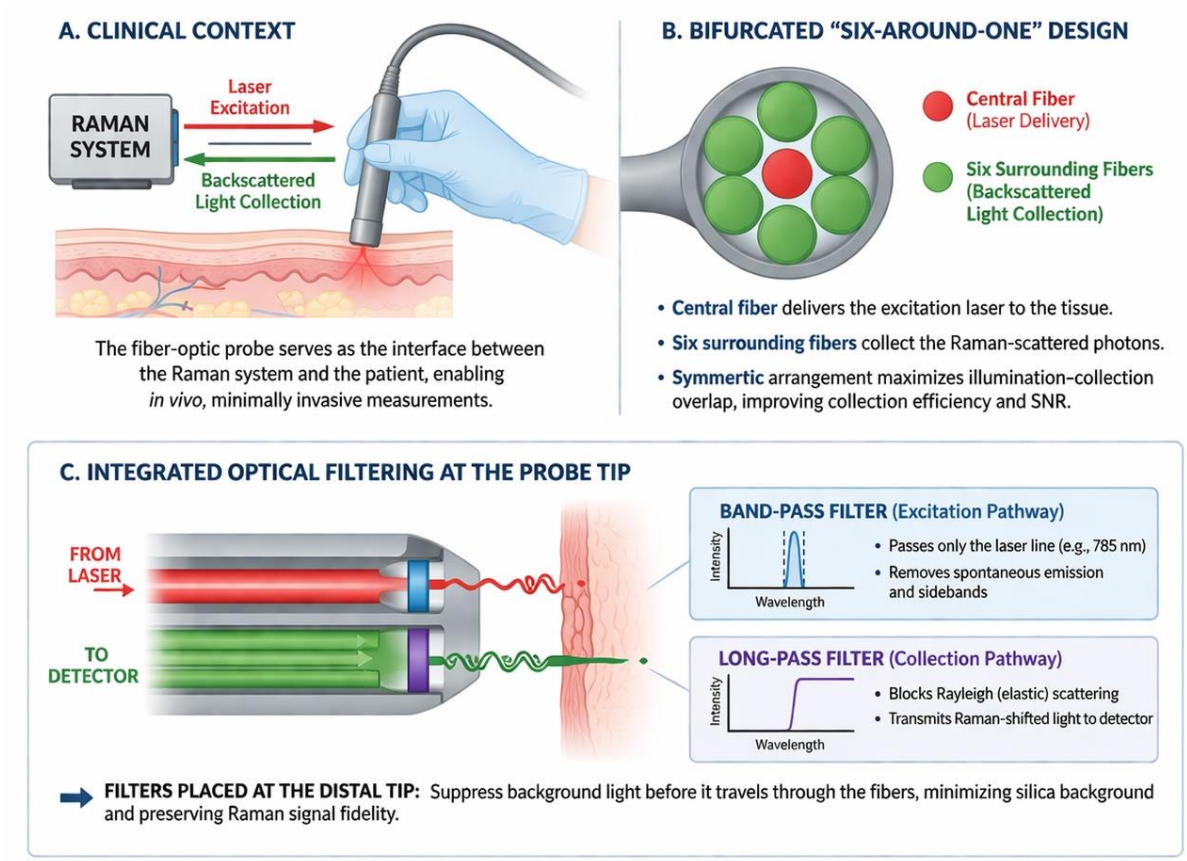

Figure (4): Bifurcated "Six-Around-One" Geometry and Integrated Optical Filtering. Detailed schematic of the fiber-optic probe architecture. (A) Cross-sectional view of the "six-around-one" configuration, where a central fiber delivers the excitation laser and six surrounding fibers collect Raman-scattered light, maximizing illumination–collection overlap and signal-to-noise ratio. (B) Longitudinal view of the probe tip showing integrated optical filters: a band-pass filter in the excitation pathway ensures spectral purity of the incident laser, while a long-pass filter in the collection pathway suppresses Rayleigh scattering and silica background, thereby preserving Raman signal fidelity. (C) By integrating optical filters at the distal end of the probe, fiber-generated silica Raman background is suppressed at the source. This ensures that the collection fibers transmit a high-purity signal, maintaining the spectral fidelity required for accurate biochemical analysis.

The Spectrograph:

Backscattered light collected by the probe is directed to a high-throughput spectrograph, which is responsible for dispersing the light into its spectral components. This is achieved through a volume phase holographic (VPH) grating, an optical element noted for its high diffraction efficiency and spectral resolution. The grating spatially separates incident photons according to their wavelength, allowing precise measurement of Raman shifts that correspond to molecular vibrational modes. Given the inherently weak Raman scattering signals, clinical Raman spectrographs are optimized to maximize photon collection efficiency. This typically involves designing the spectrograph with a low f-number (e.g., f/1.8), which increases the collecting solid angle and thus the throughput of scattered photons. Such optimization substantially enhances signal intensity and spectral quality, facilitating reliable detection of subtle biochemical features within tissues.

Detection and Cooling Systems:

Following spectral dispersion, the light is detected by a charge-coupled device (CCD), which acts as the primary sensor for photon detection and data acquisition. Clinical-grade Raman systems generally use back-illuminated, deep-depletion CCDs to improve quantum efficiency, especially within the NIR spectral region. This enhanced sensitivity is critical for capturing the weak Raman signals emitted by biological tissues. Moreover, managing detector noise is essential for accurate spectral acquisition. To reduce dark current - thermal electronic noise that can obscure low-intensity signals - the CCD is thermoelectrically cooled to temperatures between approximately −60 °C and −90 °C. This cooling minimizes thermal noise, thereby boosting the signal-to-noise ratio and enabling the detection of fine biochemical variations in the tissue spectra (see Table I).

Table (1): Hardware Specifications for Clinical Deployment [34].

Component	Clinical Requirement	Scientific Function
Laser Source	785 nm (Stabilized)	Minimizes fluorescence and prevents tissue damage.
Optical Probe	Sterilizable / Miniaturized	Enables *in vivo* access via endoscopes or needles.
Filters	Narrow Band-pass / Long-pass	Isolates weak Raman signal from Rayleigh noise.
CCD Detector	Thermoelectrically Cooled	Reduces thermal background for low-light detection.

Data Acquisition in Living Tissue:

Raman spectroscopic data acquisition from biological tissues can be classified into three main categories [33], each designed to address distinct clinical or experimental objectives: in vivo measurements, ex vivo analysis, and liquid biopsy-based methods.

In vivo measurements entail direct assessment of tissue within a living subject using a fiber-optic Raman probe. The probe is positioned against the tissue surface during clinical examination or surgical intervention [34], permitting real-time and non-invasive biochemical characterization. This technique allows for prompt diagnostic insights while maintaining the physiological integrity of the tissue. Nevertheless, it faces challenges including motion-induced artifacts, tissue heterogeneity, and stringent limitations on laser exposure to ensure safety.

Ex vivo analysis is conducted on freshly excised tissue obtained via biopsy or surgical removal. Measurements occur immediately post-excision to preserve the native biochemical profile and limit degradation [34]. This method tends to yield higher quality spectra relative to in vivo approaches, owing to greater experimental control, reduced motion effects, and the ability to fine-tune acquisition parameters. Furthermore, it facilitates direct comparison with histopathological evaluations, thereby strengthening diagnostic correlations.

Liquid biopsy offers a minimally invasive alternative, focusing on biofluids such as blood, saliva, or urine. Raman spectroscopy is applied to detect molecular patterns and cancer-associated biomarkers present in these fluids. While this approach improves patient comfort and sampling convenience, it typically demands sophisticated signal processing because of the low abundance of target molecules and the complex composition of the biological milieu.

In all cases, the selection of laser parameters warrants careful attention to balance patient safety and measurement quality. Commonly, low-power near-infrared (NIR) excitation sources at wavelengths around 785 nm or 830 nm are utilized. These wavelengths provide an effective compromise by minimizing tissue autofluorescence and thermal effects [34], while producing Raman signals of sufficient intensity for reliable spectral interpretation.

AI Processing Workflow:

The integration of artificial intelligence in processing Raman spectroscopy data improves both spectral reliability and diagnostic precision. Initially, raw spectral data are subjected to preprocessing steps such as baseline correction, normalization, and noise reduction aimed at mitigating fluorescence effects and other artifacts. Subsequently, pertinent features are extracted from characteristic Raman peaks. These features serve as input to machine learning classifiers, facilitating accurate discrimination and categorization of tissue types or biochemical conditions. The workflow of this AI-driven Raman analysis approach is illustrated as we can see in the Figure 5, which shows the AI-powered Raman analysis pipeline.

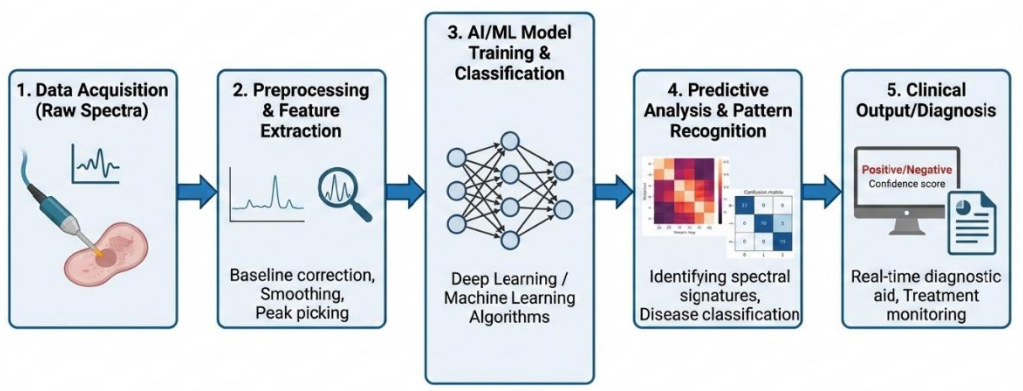

Figure (5): Schematic representation of the automated workflow for processing raw Raman spectral data through an artificial intelligence model to generate clinical diagnostic information.

Preprocessing:

Raw spectra are often contaminated by background noise, notably fluorescence signals. The preprocessing stage involves several critical steps, including baseline correction.

Baseline correction:

A well-recognized challenge in Raman spectroscopy arises from interference by strong fluorescence backgrounds. Such background signals stem from electronic excitations within the sample matrix or from impurities, typically appearing as broad, slowly varying continua overlaying the Raman spectrum. In many cases, fluorescence intensity can surpass that of Raman scattering by multiple orders of magnitude, obscuring weaker vibrational features and causing baseline drift or curvature. This produces an artificial slope within the spectral profile, complicating both qualitative peak identification and quantitative intensity measurements, as documented in earlier investigations [35,32].

To address this limitation, baseline correction is implemented as a fundamental preprocessing step in spectral analysis. The primary goal is to estimate and subtract the fluorescence contribution without compromising the integrity of the Raman peaks. Various computational strategies have been proposed, with Polynomial Fitting and Asymmetric Least Squares (ALS) smoothing among the most widely used. Polynomial fitting involves approximating the background by fitting a low-order polynomial that captures the overall baseline trend while disregarding high-frequency spectral details. In contrast, ALS smoothing applies an iterative optimization procedure with asymmetric penalties that suppress positive residuals associated with peaks, thus preserving the smooth baseline shape. This approach is especially effective for spectra exhibiting complex, varying background curvature [36].

Following correction, the estimated baseline is subtracted from the original spectrum to yield a leveled output. This adjustment establishes a consistent zero baseline across the spectral range, facilitating accurate determination of peak positions and intensities corresponding to molecular vibrational modes. Therefore, baseline correction plays a vital role in improving the robustness of subsequent analyses, including multivariate statistical techniques, compound identification, and concentration assessment. In fields such as biomedical diagnostics and material characterization, where detecting subtle spectral variations is crucial, precise baseline removal is essential for ensuring analytical validity and reproducibility.

Normalization:

In Raman spectroscopy, variations in experimental conditions contribute substantially to non-chemical differences observed in spectral data. Factors such as changes in laser power, detector sensitivity, integration time, and inconsistencies in sample thickness or optical alignment can cause significant fluctuations in overall spectral intensity. These fluctuations do not represent inherent molecular characteristics but introduce systematic bias, thus complicating direct spectral comparisons. It has been demonstrated in prior research that spectra obtained from chemically identical samples may differ solely due to such external measurement factors [35,32].

To mitigate this issue, normalization is employed as an essential preprocessing step that rescales spectral intensities to achieve uniformity and comparability. The key objective of normalization is to adjust signal magnitudes while preserving the relative distribution of spectral features. Commonly used methods include Min-Max normalization and Vector normalization. The former performs a linear scaling of intensity values to fit within a predetermined range, typically between 0 and 1, thereby maintaining spectral shape while standardizing amplitude. Vector normalization, alternatively called unit vector scaling, involves dividing each intensity by the Euclidean norm, calculated as the square root of the sum of squared intensities across the spectrum. This results in scaling the spectrum to unit length in multidimensional space, ensuring equal overall magnitude across spectra despite their initial intensity differences [36].

Normalization primarily aims to remove intensity variations arising from instrumental or physical inconsistencies rather than chemical disparities. By standardizing spectral scales, it facilitates downstream analyses - such as peak comparison, multivariate statistical modeling, or machine learning classification - to reflect authentic molecular differences. This approach thus supports more accurate and meaningful comparisons between spectra obtained from different samples or under various experimental settings. Particularly in advanced analytical pipelines dealing with large spectral datasets, normalization enhances data reproducibility, increases model stability, and underpins the validity of comparative interpretations.

Noise Filtering:

Raman spectral data are often affected by various noise sources that compromise signal quality and complicate accurate interpretation. Among these, high-frequency shot noise, stemming from the discrete nature of photon detection, and electronic noise related to the detector and instrumentation, are particularly prevalent. These noise components manifest as rapid, stochastic fluctuations superimposed on the spectra, which can obscure subtle Raman peaks and diminish the clarity of

spectral features. Previous research has demonstrated that such distortions may adversely impact both qualitative visualization and quantitative analysis of spectral data [35,32].

To address these challenges, noise filtering methods are routinely applied as a preprocessing step. The goal is to attenuate high-frequency noise while retaining essential spectral characteristics, such as the location, width, and intensity of Raman peaks. The Savitzky–Golay filter is commonly employed for this purpose. This method operates by sliding a window across the spectral data and fitting a low-degree polynomial to the points within the window, replacing the central point with the fitted value. Unlike simple averaging, this approach smooths the data while preserving the original peak shapes and resolution, avoiding the distortions often introduced by more aggressive smoothing techniques [36].

The fundamental aim of noise filtering is to enhance the signal-to-noise ratio (SNR), which facilitates the detection of meaningful spectral features. Improved SNR contributes to more reliable identification of weak Raman bands and more accurate estimation of peak parameters, both of which are crucial for subsequent analyses, including chemical identification, quantitative modeling, and machine learning classification. However, the filtering outcome depends critically on parameter choices such as window size and polynomial order in the Savitzky–Golay filter. Excessive smoothing risks loss of spectral detail and peak broadening, whereas insufficient smoothing may leave noise inadequately suppressed. Consequently, careful tuning and validation are necessary to strike an appropriate balance between noise reduction and spectral fidelity.

Feature Extraction:

Feature extraction constitutes a pivotal phase in applying artificial intelligence techniques to Raman spectroscopy in biomedical diagnostics. Following preprocessing, spectral data are converted into quantitative descriptors that encapsulate relevant biochemical information. Such features comprise peak positions, intensities, widths, and spectral patterns, which can differentiate molecular signatures of normal and cancerous tissues [37,38].

To manage the high dimensionality of spectral data, dimensionality reduction methods like Principal Component Analysis (PCA) are frequently employed to condense the dataset while maintaining principal variance components. The resulting features then serve as inputs for machine learning algorithms tasked with classification.

Overall, feature extraction enhances discrimination between pathological states by isolating spectrally meaningful characteristics, thereby contributing to improvements in the accuracy, efficiency, and objectivity of AI-supported diagnostic systems.

Machine Learning Classifiers:

Machine learning classifiers play a central role in analyzing Raman spectral data for medical diagnostics. Due to the complexity and high dimensionality of these data, supervised learning approaches are used to identify patterns corresponding to various tissue conditions. Traditional algorithms, such as Support Vector Machines (SVM) and Random Forests, are widely adopted for their robustness and efficacy in classifying spectral features, particularly in distinguishing normal from pathological tissue [40].

Simultaneously, deep learning techniques have demonstrated strong potential for automated feature extraction. Convolutional Neural Networks (CNNs), in particular, can learn hierarchical spatial and spectral representations directly from raw or minimally processed data, enabling accurate and rapid differentiation between benign and malignant tissues with substantial sensitivity and specificity [40].

These models are trained on labeled spectral datasets to recognize characteristic patterns associated with tissue classification, disease progression, and therapeutic outcomes. Consequently, the integration of machine learning frameworks with Raman spectroscopy holds promise for enhancing diagnostic accuracy and supporting objective, timely, and reliable clinical decision-making.

Clinical Performance and Diagnostic Efficacy:

The effectiveness of Raman spectroscopy in clinical settings is assessed by its sensitivity, reflecting its capacity to accurately detect cancerous tissue, and specificity, indicating its ability to correctly recognize healthy tissue. Recent clinical investigations and meta-analyses suggest that integrating Raman spectroscopy with sophisticated machine learning techniques, such as Principal Component Analysis (PCA) and Support Vector Machines (SVM), can enhance diagnostic accuracy in diverse oncological contexts [39].

Comparative Performance Metrics:

Table II presents a summary of principal clinical outcomes reported in recent studies, encompassing various tissue types and diagnostic settings.

Table (2): It presents the diagnostic performance of Raman-based screening methods in oncology, with sensitivity and specificity measures compared to those obtained through conventional histopathological analysis. Abbreviations used include SERS for Surface-Enhanced Raman Spectroscopy, BCC for Basal Cell Carcinoma, and SCC for Squamous Cell Carcinoma.

Cancer Type	Study Format	Sensitivity (%)	Specificity (%)	Key Biomarker Focus
Lung	Serum-based (SERS)	98.68%	91.81%	Nucleic acids & Proteins
Glioblastoma	Intraoperative (Margins)	90.0%	95.0%	Phenylalanine & Lipids
Esophageal	In vivo (Endoscopy)	94.4%	100.0%	Glycoproteins & DNA
Breast Cancer	Ex vivo (Biopsy)	80.0%	90.0%	Calcium Hydroxyapatite
Skin (BCC/SCC)	In vivo (Probe)	95.0%	87.0%	Keratin & Ceramides

The Role of Chemometrics in Clinical Accuracy:

The clinical applicability of Raman spectroscopy largely depends on the implementation of advanced chemometric techniques to manage the complex nature of high-dimensional spectral data. Raw spectral outputs often involve overlapping molecular vibrational modes and low signal-to-noise ratios, rendering manual interpretation inadequate for the stringent requirements of clinical practice. To address this challenge, multivariate methods, particularly the combination of Principal Component Analysis (PCA) with Linear Discriminant Analysis (LDA), are employed to reduce data dimensionality and highlight the most relevant spectral variations. More recently, the adoption of Deep Learning models, such as Convolutional Neural Networks (CNNs), has improved diagnostic accuracy in serum-based liquid biopsies, achieving performance metrics exceeding 94%. These computational approaches are adept at detecting subtle, non-linear spectral patterns and complex biochemical interactions that conventional statistical techniques may miss, thus providing the precision necessary for contemporary pathological classification [39].

Discussion of Results and Clinical Applications:

The strong diagnostic outcomes reported for a range of malignancies, including breast, lung, brain, and skin cancers, demonstrate the significant potential of Raman spectroscopy in clinical settings. Numerous studies have documented specificity rates above 90%, a level that holds considerable importance for reducing false-positive biopsy results. By employing AI-driven systems capable of recognizing nuanced biochemical signatures beyond the scope of traditional histopathology, this approach helps alleviate patient distress and lowers unnecessary medical costs. The technique's inherent benefits, non-invasiveness, elimination of labels, and no reliance on exogenous contrast agents, enable high molecular specificity alongside real-time analysis across various tissue types without disrupting the biological milieu [41,42].

Beyond its utility in initial diagnosis, Raman spectroscopy's real-time sensing capability proves valuable during surgical procedures, especially in complex tumor resections [43,44]. In cases such as glioblastoma and breast cancer, Raman probes aid surgeons in delineating tumor margins intraoperatively, facilitating the attainment of clear resection boundaries while sparing healthy tissue [43,45]. This precise guidance addresses the persistent issue of incomplete tumor excision, thereby reducing the risk of local tumor recurrence. Additionally, Raman technology offers a platform for continuous treatment monitoring; by observing spectral alterations longitudinally, clinicians can assess chemotherapy efficacy, detect emerging molecular resistance, and identify early recurrence signs below clinical thresholds, thereby enabling more individualized therapeutic strategies [44].

Despite these advances, several technological and systemic challenges must be resolved before Raman spectroscopy can be widely standardized in clinical practice. The inherently weak Raman scattering signal necessitates the use of highly sensitive, cooled detectors, while intrinsic tissue autofluorescence often interferes with target measurements. Moreover, the reliability of AI models hinges on access to extensive, rigorously annotated datasets, which remain under development internationally.

To mitigate these limitations, research is progressing toward enhanced methods such as Surface-Enhanced Raman Spectroscopy (SERS) [31], which leverages metallic nanoparticles to boost signal intensity by several orders of magnitude. Simultaneously, the transition from conventional machine learning techniques to deep learning frameworks is strengthening the robustness of spectral data classification. Finally, combining Raman spectroscopy with complementary imaging modalities like

Optical Coherence Tomography (OCT) or ultrasound is fostering a multimodal diagnostic approach that further elevates the clinical utility of optical biopsies [46,47].

Conclusion:

This study has demonstrated that integrating Raman light scattering with artificial intelligence offers a promising approach for cancer diagnosis and treatment. Raman spectroscopy delivers detailed, label-free biochemical insights by detecting molecular changes linked to malignancy, including modifications in nucleic acids, proteins, and lipid metabolism. Coupling these complex spectral data with sophisticated AI algorithms allows for efficient processing, facilitating accurate classification and supporting real-time clinical decision-making.

The findings indicate that AI-enhanced Raman technologies achieve notable sensitivity and specificity across various cancer types, suggesting their potential utility as dependable diagnostic modalities. Additionally, the non-invasive, rapid, and in vivo analytical capabilities of this approach make it a viable complement or alternative to traditional histopathological methods. Its use in intraoperative guidance and monitoring of therapeutic responses further highlights its relevance within precision medicine frameworks.

Although challenges remain, such as weak signal intensity, fluorescence interference, and the requirement for extensive annotated datasets, ongoing developments, particularly in surface-enhanced Raman scattering (SERS) and deep learning techniques, are anticipated to mitigate these limitations. In summary, the integration of Raman spectroscopy with artificial intelligence constitutes a meaningful advance toward enhancing the accuracy, efficiency, and personalization of cancer management.

References:

1. Jain, G. K., Verma, R., Chougule, A., & Singh, B., “Scientific Advances in Cancer Detection Using Raman Spectroscopy,” Asian Pacific Journal of Cancer Prevention: APJCP, 25(11), 3977, pp. 3977-3986, 2024. DOI:10.31557/APJCP.2024.25.11.3977.
2. Huang, L., Sun, H., Sun, L., Shi, K., Chen, Y., Ren, X., ... & Wang, Y., “Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning,” Nature Communications, 14(1), 48, pp. 1-14, 2023. https://doi.org/10.1038/s41467-022-35696-2.
3. Qi, Y., Liu, Y., & Luo, J., “Recent application of Raman spectroscopy in tumor diagnosis: from conventional methods to artificial intelligence fusion,” PhotoniX, 4(1), 22, pp. 1-42, 2023. https://doi.org/10.1186/s43074-023-00098-0.
4. Vaňková, L., Bufka, J., Šigutová, P., Holubová, M., & Křížková, V. (2026). “Raman Spectral Signatures of Human Neutrophils: A Single-Cell Approach,” Journal of Biophotonics, 19(3), e70251, pp. 1-8, 2026. https://doi: 10.1002/jbio.70251.
5. Shipp, D. W., Sinjab, F., & Notingher, I., “Raman spectroscopy: techniques and applications in the life sciences,” Advances in Optics and Photonics, 9(2), 315-428, pp. 315-428, 2017. https://doi.org/10.1364/AOP.9.000315.
6. Murugappan, S., Tofail, S. A., & Thorat, N. D., “Raman spectroscopy: a tool for molecular fingerprinting of brain cancer,” ACS omega, 8(31), pp. 27845-27861, 2023. https://doi.org/10.1021/acsomega.3c01848.
7. Zhang, S., Qi, Y., Tan, S. P. H., Bi, R., & Olivo, M., “Molecular fingerprint detection using Raman and infrared spectroscopy technologies for cancer detection: a progress review,” Biosensors, 13(5), 557, pp. 1, 2023. doi: 10.3390/bios13050557.
8. Pimenta, S., & Correia, J. H., “Biomedical Applications of Raman Spectroscopy: A Review,” Photochem, 5(4), 29, pp. 1. 2025. https://doi.org/10.3390/photochem5040029.
9. Driza Nadir, Ali Hanan, Ibrahim Ola, Hamad Rafa, Elgade Asma, “INTELLIGENT SPECTROSCOPY: MERGING RAMAN, AI, AND ELECTRONICS FOR ADVANCED MEDICAL DIAGNOSIS AND TREATMENT,” Electrical and Electronics Engineering: An International Journal (ELELIJ) Vol.14, No. ¾, pp 17-26, 2025.
10. Han, J., Dong, H., Zhu, T., Wei, Q., Wang, Y., Wang, Y., Lv Yu, Mu H., Huang S., Zeng Ke, Xu J., & Ding, J., “Biochemical hallmarks-targeting antineoplastic nanotherapeutics,” Bioactive Materials, 36, pp. 427-454, 2024. doi: 10.1016/j.bioactmat.2024.05.042.
11. Chen, Y., Dai, J., Zhou, X., Liu, Y., Zhang, W., & Peng, G., “Raman spectroscopy analysis of the biochemical characteristics of molecules associated with the malignant transformation of gastric mucosa,” PLoS One, Vol. 9, Issue 4, pp. 1-11, 2014. https://doi.org/10.1371/journal.pone.009306.
12. Ilin, Y., Choi, J. S., Harley, B. A., & Kraft, M. L., “Identifying states along the hematopoietic stem cell differentiation hierarchy with single cell specificity via Raman spectroscopy,” Analytical chemistry, 87(22): 11317-11324, pp. 1-17, 2015. DOI: 10.1021/acs.analchem.5b02537.

13. Kim, C., Hong, S., Ma, S. H., Lee, J., So, H., Kim, J. Y., Shin E., Lee K., Choi S., Park J., Park Y., Kim Y., Kim Ji, & Kim, J.," Replication stress–induced nuclear hypertrophy alters chromatin topology and impacts cancer cell fitness," Proceedings of the National Academy of Sciences, Vol. 122, No. 37, e2424709122, pp. 1-12, 2025. https://doi.org/10.1073/pnas.2424709122.
14. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., & Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nature communications, 11(1), 5120. https://doi.org/10.1038/s41467-020-18794-x.
15. Radisky, E. S., " Extracellular proteolysis in cancer: Proteases, substrates, and mechanisms in tumor progression and metastasis," Journal of Biological Chemistry, 300(6), 107347, pp. 1-36 , 2024. doi: 10.1016/j.jbc.2024.107347.
16. Ramberger, E., Sapozhnikova, V., Ng, Y. L. D., Dolnik, A., Ziehm, M., Popp, O., ... & Krönke, J., "The proteogenomic landscape of multiple myeloma reveals insights into disease biology and therapeutic opportunities," Nature cancer, Vol. 5, pp. 1267-1284. 2024. https://doi.org/10.1038/s43018-024-00784-3.
17. Popova, N. V., & Jücker, M., "The functional role of extracellular matrix proteins in cancer," Cancers, 14(1), 238, pp. 1-28, 2022. doi: 10.3390/cancers14010238.
18. Mesias, V. S. D., Zhang, J., Fu, W., Dai, X., & Huang, J., "Enhanced characterization of protein secondary structure transitions using Raman and SERS measurements combined with 2D correlation spectroscopy and principal component analysis," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Vol. 343, 126607, pp. 1-5, 2025. https://doi.org/10.1016/j.saa.2025.126607.
19. Ramos, S., & Lee, J. C. "Raman spectroscopy in the study of amyloid formation and phase separation," Biochemical Society Transactions, Vol. 52, pp. 1121-1130, 2024. doi: 10.1042/BST20230599.
20. Cutshaw, G., Uthaman, S., Hassan, N., Kothadiya, S., Wen, X., & Bardhan, R., "The emerging role of Raman spectroscopy as an omics approach for metabolic profiling and biomarker detection toward precision medicine," Chemical reviews, 123(13), 8297-8346, pp. 1-93, 2023. doi: 10.1021/acs.chemrev.2c00897.
21. Cheng, C., Geng, F., Cheng, X., & Guo, D., "Lipid metabolism reprogramming and its potential targets in cancer," Cancer communications, 38(1), 27, pp. 1-14, 2018.doi: 10.1186/s40880-018-0301-4.
22. Jin, H. R., Wang, J., Wang, Z. J., Xi, M. J., Xia, B. H., Deng, K., & Yang, J. L., "Lipid metabolic reprogramming in tumor microenvironment: from mechanisms to therapeutics," Journal of hematology & oncology, 16(1), 103, pp. 1-33, 2023.doi: 10.1186/s13045-023-01498-2.
23. Petersen, F. N. R., & Nielsen, C. H., "Raman spectroscopy as a tool for investigating lipid protein interactions," Spectroscopy (Santa Monica), Vol. 24, Issue 10, ISSN 0887-6703,pp. 26-35, 2009.
24. Sassi Paola, Caponi Silvia, Ricci Maria, Morresi Assunta, Oldenhof Harriëtte, Wolkers Willem F., Fioretto Daniele, "Infrared versus light scattering techniques to monitor the gel to liquid crystal phase transition in lipid membranes," Journal of Raman Spectroscopy, Vol. 46, Issue 7, pp. 644-651, 2015. https://doi.org/10.1002/jrs.4702Digital Object Identifier.
25. Lin, C., Li, Y., Peng, Y., Zhao, S., Xu, M., Zhang, L., Huang Z., Shi J., & Yang, Y. "Recent development of surface-enhanced Raman scattering for biosensing," Journal of Nanobiotechnology, Vol. 21(1): 149., pp. 1-37, 2023. doi: 10.1186/s12951-023-01890-7.
26. Ustaoglu, S. G., Ali, M. H. M., Rakib, F., Blezer, E. L. A., Van Heijningen, C. L., Dijkhuizen, R. M., & Severcan, F., "Biomolecular changes and subsequent time-dependent recovery in hippocampal tissue after experimental mild traumatic brain injury," Scientific reports, Vol. 11 (1), pp. 1–13. 2021.https://doi.org/10.1038/s41598-021-92015-3.
27. Sabtu, S. N., Sani, S. A., Looi, L. M., Chiew, S. F., Pathmanathan, D., Bradley, D. A., & Osman, Z. "Indication of high lipid content in epithelial-mesenchymal transitions of breast tissues," Scientific reports, 11(1), 3250, pp. 1-18, 2021. https://doi.org/10.1038/s41598-021-81426-x.
28. Janssens, M., van Smeden, J., Puppels, G. J., Lavrijsen, A. P. M., Caspers, P. J., & Bouwstra, J. A. "Lipid to protein ratio plays an important role in the skin barrier function in patients with atopic eczema," British Journal of Dermatology, Vol. 170, Issue 6, pp. 1248-1255, 2014. https://doi.org/10.1111/bjd.12908.
29. Szlasa, W., Zendran, I., Zalesińska, A., Tarek, M., & Kulbacka, J. "Lipid composition of the cancer cell membrane," Journal of bioenergetics and biomembranes, Vol. 52, pp. 321-342, 2020. doi: 10.1007/s10863-020-09846-4.
30. Pence, I., & Mahadevan-Jansen, A. "Clinical instrumentation and applications of Raman spectroscopy," Chemical Society Reviews, Vol. 45, 1958-1979, pp. 1-44, 2016. doi: 10.1039/c5cs00581g.

31. Wang, Y., Fang, L., Wang, Y., & Xiong, Z. “Current trends of Raman spectroscopy in clinic settings: opportunities and challenges,” Advanced Science, Vol. 11, 2300668, pp. 1-36, 2024. doi: 10.1002/advs.202300668.
32. Ramírez-Elías, M. G., & González, F. J., “Raman Spectroscopy for In Vivo Medical Diagnosis,” In Raman Spectroscopy. InTech, pp. 293-311, 2018. https://doi.org/10.5772/intechopen.72933.
33. Hanna, K., Krzoska, E., Shaaban, A. M., Muirhead, D., Abu-Eid, R., & Speirs, V. “Raman spectros copy: current applications in breast cancer diagnosis, challenges and future prospects,” British jou r--nal of cancer, Vol. 126, pp. 1125-1139, 2022. doi: 10.1038/s41416-021-01659-5.
34. Sheehy, G. “Bringing Real-Time Raman Spectroscopy Tissue Characterisation to In Vivo Applicati ons Including Cancer Diagnosis,” (Doctoral dissertation, Ecole Polytechnique, Montreal (Canada)) , pp. 1-167, 2025.
35. Vulchi, R. T., Morgunov, V., Junjuri, R., & Bocklitz, T., “Artifacts and anomalies in Raman spectros copy: a review on origins and correction procedures,” Molecules, Vol. 29, 4748, pp. 1-24, 2024. do i: 10.3390/molecules29194748.
36. Mostafapour, S., Dörfer, T., Heinke, R., Rösch, P., Popp, J., & Bocklitz, T., “Investigating the effec t of different pre-treatment methods on Raman spectra recorded with different excitation waveleng th,” Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Vol. 302, 123100, pp 1-8, 2023. https://doi.org/10.1016/j.saa.2023.123100.
37. Zhu, X., Zhao, Y., Zan, C., Ma, H., & Liu, J. “Recent advances in applications of artificial intelligen ce-assisted Raman spectroscopy in diagnosis of cancers,” Frontiers in Molecular Biosciences, Vol . 12, 1690063, pp. 1-15, 2025. doi: 10.3389/fmolb.2025.1690063.
38. Wang, M., Chang, W., & Zhang, Y., “Artificial Intelligence for the Diagnosis and Management of C ancers: Potentials and Challenges,” MedComm, Vol. 6, Issue 11, e70460, pp. 1-55, 2025. doi: 10. 1002/mco2.70460.
39. Sharma, N., Rao, S., Noothalapati, H., Mazumder, N., & Paul, B. “Raman spectroscopy in the dete ction and diagnosis of lung cancer: a meta-analysis;” Lasers in Medical Science, Vol. 40, Issue 1, No.164, pp. 1-12, 2025. doi: 10.1007/s10103-025-04421-y.
40. Wedin, M., & Bengtsson, I., “A Comparative Study on Machine Learning Models for Automatic Cla ssification of Cell Types from Digitally Reconstructed Neurons,” KTH ROYAL INSTITUTE OF TEC HNOLOGY SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE SWEDEN , pp. 1-50, 2021.
41. Short, M. A., Lam, S., McWilliams, A. M., Ionescu, D. N., & Zeng, H., “Using laser Raman spectroscopy to reduce false positives of autofluorescence bronchoscopies: a pilot study,” Journal of Thoracic Oncology, Vol. 6, pp. 1206-1214, 2011. doi: 10.1097/JTO.0b013e3182178ef7.
42. Chen, M., Zhao, M., Cai, Y., Zhang, Q., Peng, Z., Li, Q., & Wang, Z., “Surface-enhanced Raman spectroscopy for label-free cancer liquid biopsy: from fundamentals to clinical analysis of biofluid,” Frontiers in Chemistry, Vol. 13, 1696979, pp. 1-22, 2025. doi: 10.3389/fchem.2025.1696979.
43. Rivera, D., Young, T., Rao, A., Zhang, J. Y., Brown, C., Huo, L., ... & Schupper, A. J., “Current applications of Raman spectroscopy in intraoperative neurosurgery,” Biomedicines, Vol. 12, 2363, pp. 1-18, 2024. doi: 10.3390/biomedicines12102363.
44. Heng, H. P. S., Shu, C., Zheng, W., Lin, K., & Huang, Z., “Advances in real-time fiber-optic Raman spectroscopy for early cancer diagnosis: Pushing the frontier into clinical endoscopic applications,” Translational Biophotonics, Vol. 3, e202000018, pp. 1-31, 2021. DOI:10.1002/tbio.202000018.
45. Jermyn, M., Desroches, J., Mercier, J., St-Arnaud, K., Guiot, M. C., Leblond, F., & Petrecca, K., “Raman spectroscopy detects distant invasive brain cancer cells centimeters beyond MRI capability in humans,” Biomedical optics express, Vol. 7, No. 12, pp. 5129-5137, 2016. doi: 10.1364/BOE.7.005129.
46. Fitzgerald, S., Akhtar, J., Schartner, E., Ebendorff-Heidepriem, H., Mahadevan-Jansen, A., & Li, J. (2023). Multimodal Raman spectroscopy and optical coherence tomography for biomedical analysis. Journal of biophotonics, Vol. 16, e202200231, P1, 2023. https://doi.org/10.1002/jbio.202200231.
47. Ashok, P. C., Praveen, B. B., Bellini, N., Riches, A., Dholakia, K., & Herrington, C. S., “Multi-modal approach using Raman spectroscopy and optical coherence tomography for the discrimination of colonic adenocarcinoma from normal colon,” Biomedical optics express, Vol. 4, Issue 10, pp. 2179-2186, 2013. doi: 10.1364/BOE.4.002179.