Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 270 (2025) 2666–2675

Procedia Computer Science

www.elsevier.com/locate/procedia

29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)

Machine Learning meets Raman spectroscopy: a systematic review of literature in cancer diagnostics

Bogdan Oancea[a*], Marian Necula[b], Eduard-Costin Milea[b], Alexandru Amărioarei[a,b], Ion Petre[b,c], Mihaela-Marinela Păun[a,b]

[a]*University of Bucharest, Bd. Regina Elisabeta no. 4–12, Sector 3, 030108 Bucharest, Romania*
[b]*Bioinformatics Department, National Institute for R&D for Biological Sciences, Splaiul Independentei 296, 060031 Bucharest, Romania*
[c]*Department of Mathematics and Statistics, University of Turku, Finland*

Abstract

The integration of machine learning (ML) techniques with Raman spectroscopy has emerged as a promising strategy for advancing cancer diagnostics through label-free, high-resolution molecular analysis. This review aims to map and synthesize current research directions in this rapidly evolving field by conducting a structured review of existing review articles. Using a curated dataset of 70 reviews retrieved from Scopus and Web of Science, we applied Latent Dirichlet Allocation (LDA) topic modeling to uncover dominant thematic clusters across the literature. Our findings reveal five key research axes: (1) instrumentation and signal acquisition, (2) data preprocessing and spectral denoising, (3) classification models and algorithmic pipelines, (4) biomedical applications in oncology, and (5) emerging trends including deep learning and hybrid methods. This thematic structure highlights both the maturity and fragmentation of the current knowledge landscape. We also discuss the limitations of our approach, including database and article-type restrictions, and the use of LDA as a single modeling method. By identifying underexplored areas and recurring methodological challenges, this review contributes to a clearer understanding of the research gaps and future opportunities at the intersection of ML and Raman spectroscopy for cancer research.

Keywords: Raman spectroscopy, machine learning, cancer diagnostics, spectral data processing, computational oncology

1. Introduction

Cancer continues to represent one of the greatest public health challenges worldwide, with its incidence steadily increasing due to aging populations and lifestyle factors. Early detection and precise characterization of cancer are

* Corresponding author.
E-mail address: bogdan.oancea@faa.unibuc.ro

10.1016/j.procs.2025.09.388

critical to improving patient outcomes, survival rates, and reducing healthcare burdens [1,2]. Traditional cancer diagnostics and treatments, including conventional biopsy, imaging techniques, and histopathology, face significant limitations such as invasiveness, inter-observer variability, and late-stage detection [3,4]. To address these challenges, research has increasingly turned towards developing novel diagnostic technologies that are minimally invasive, highly specific, sensitive, and capable of real-time and multiplex detection [5–8].

Raman spectroscopy (RS) is a non-invasive optical technique that analyzes the vibrational modes of molecules to generate a spectral fingerprint of a sample. In cancer research, this method offers significant advantages because it provides detailed information about the biochemical composition of tissues without the need for labeling or staining. The unique spectral profiles obtained can help distinguish between normal and cancerous cells based on their molecular structure and composition [9,10], helping to identify the solid tumor contours[11]. It can also help discriminate tumor sub-types that are very difficult to separate in the clinical setup using imaging or pathology [12–14].

Technologies like (spontaneous) Raman Spectroscopy (RS), Resonance Raman Spectroscopy (RRS), Surface-Enhanced Raman Spectroscopy (SERS), or Stimulated Raman Spectroscopy (SRS) have emerged as promising diagnostic tools due to their high sensitivity, specificity, and capability for multiplex biomarker detection [5,11,15,16]. In recent years, researchers have increasingly applied RS to identify and classify various types of cancer, such as brain cancer [12], breast cancer [17,18], laryngeal cancer [19]. By analyzing the differences in vibrational spectra, scientists have been able to detect subtle changes in tissue biochemistry that are associated with specific malignancy. For example, differences in lipid, protein, and nucleic acid content have been observed between healthy tissues and tumors, offering promising diagnostic markers for early detection and treatment monitoring [20,12,14]. However, Raman spectra are typically high-dimensional and can be affected by noise and background interference, which makes manual interpretation challenging. This is where machine learning techniques come into play. Algorithms such as support vector machines, neural networks, and random forests have been successfully employed to process these complex spectral datasets, enabling automated classification and prediction of cancerous tissues with high accuracy [21]. The integration of artificial intelligence (AI) into the diagnostic process is revolutionizing medical image analysis and spectroscopic data interpretation, substantially improving diagnostic accuracy and consistency [3,8,22]. This multidisciplinary approach, combining spectroscopy, nanotechnology, and AI, is poised to significantly advance cancer diagnostics and management strategies, offering more personalized and effective patient care.

Combining RS with machine learning (ML) has led to significant improvements in diagnostic performance. For instance, models have been developed to extract features from raw spectral data, effectively reducing dimensionality while preserving important discriminatory information [12,19,23]. These advanced methods have demonstrated robust performance in differentiating between various cancer types, often outperforming traditional statistical techniques [24].

This paper presents a comprehensive overview of the integration of RS and ML in cancer research, emphasizing how their combination enhances early detection, classification, and treatment monitoring. We review recent advances, discuss methodological challenges, and highlight emerging trends to inform both researchers and clinicians. The Methodology section outlines the literature search process and the use of Latent Dirichlet Allocation (LDA) for topic modeling. The Discussion synthesizes thematic clusters identified through LDA, focusing on innovations and clinical applications. The Conclusions summarize key insights, practical implications, and directions for future research.

2. Methodology

We initiated our review study by defining a comprehensive set of queries, which we then executed on two of the most prestigious bibliographic databases—Scopus and Web of Science. These queries were designed to capture a broad spectrum of literature relevant to our research focus. By leveraging the extensive indexing and robust search functionalities of both databases, we systematically identified, collected, and screened a large corpus of publications. This rigorous and methodical literature search provided a solid foundation for our subsequent analysis and synthesis.

The steps used to select review papers are presented in Figure 1. The queries (combination between Raman spectroscopy and artificial intelligence/machine learning methods) returned a number of 11764 papers, published between 1991 and 2025, out of which we ended up with 70 review papers based on the article type, namely reviews only, deduplicated and filtered using the following keywords from cancer/oncology lexical field: {"neoplasia", "tumor", "carcinoma", "sarcoma", "melanoma", "leukemia", "lymphoma", "metastasis", "oncology", "malignant", "cancer"}.

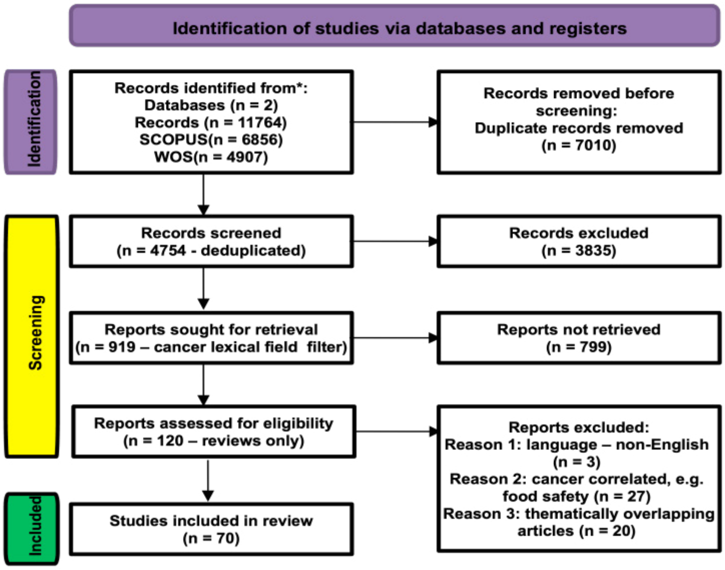

Figure 1. Flow diagram of the article selection process. This diagram summarizes the screening workflow—from 11,764 initially retrieved records, through duplicate removal, title/abstract screening, and filtering by cancer-related keywords—to the final inclusion of 70 review articles.

To identify the most relevant topics from the corpus, we used Latent Dirichlet Allocation LDA [25], an unsupervised topic modeling method that represents each document as a mixture of latent topics. LDA was selected for its ability to handle high-dimensional text data without extensive feature engineering and to uncover meaningful semantic structures. Unlike traditional clustering methods, it offers a probabilistic, interpretable framework well-suited for large bibliographic datasets such as Scopus and Web of Science. The analysis was performed using the R packages lda[26] and ldatuning[27], and the results are presented in Table 1. The optimal number of topics identified by ldatuning, which was determined to be 8, was subsequently used as an input parameter for the lda function.

Table 1. The representative keywords summarizing the primary topics extracted via LDA from 70 review articles on Raman spectroscopy and machine learning applications in cancer diagnostics.

Topic words	
sers, liquid, surface enhanced, biomarkers, biopsy, detection, label free, evs, challenges, early detection	tissue, vibrational, breast, esophageal, bone, included, gastric, egc,m etastases, networks
Surgical,patient, imaging, outcomes, intraoperative, modalities, tomography, surgery, disorders, lspr	imaging, tissue, scattering, microscopy, malignant, glioma, ramanbased, srs, hyperspectral, normal
raman, spectroscopy, cancer, diagnosis, clinical, analysis, methods, techniques, review, learning	article, error, published, lrs, stochastic, searched, base, among, biotherapy
skin, accuracy, noninvasive, diagnostic, evaluation, fibrosis, review, melanoma, detection, liver	network, data, neural, deep, learning, cnn, feature, convolutional, models, model

Our analysis highlights five key research axes driving recent advances in the field. First, instrumentation and signal acquisition ensure reliable capture of high-quality spectral data. Second, effective preprocessing and denoising enhance signal clarity. Third, advanced classification models, including advanced ML methods, enable accurate analysis of complex datasets. Fourth, these techniques have direct biomedical applications, especially in cancer diagnostics. Finally, emerging approaches such as deep learning and hybrid methods integrate modern computational tools with traditional analytics. Together, these axes reflect the interdisciplinary progress shaping the future of diagnostic technologies.

Table 2 provides an integrated summary of the diagnostic strategies reviewed. It lists the cancer types investigated, the primary Raman spectroscopy techniques employed (CRS, SERS, CARS, SRS/Laser RS) to obtain molecular fingerprints of tissues, and the machine learning and AI methods (SVMs, ANNs, CNNs/Deep Learning, chemometric, and ensemble methods) used to analyze spectral data and support diagnostic accuracy.

Table 2. Overview of Cancer Types, Raman Spectroscopy Techniques, and ML/AI Methods

Cancer Types		
Breast Cancer	Prostate Cancer	Brain Cancer (including gliomas)
Bladder Cancer	Oral Cancer	Colon and Cervical Cancer
Pancreatic and Gastric Cancer	Skin Cancer (including melanoma)	Liver Cancer
	Laryngeal Cancer	Hematological Cancers

Raman Spectroscopy Techniques	ML/AI Methods
CRS: Captures molecular fingerprints.	SVMs: Robust for high-dimensional data.
SERS: Enhances detection of low-concentration biomarkers.	ANNs: (incl. stochastic variants) yield outputs and confidence scores.
CARS: Improves imaging contrast & speed.	CNNs/Deep Learning: Automates feature extraction and provides high accuracy classifications.
SRS/Laser RS: Boosts signal strength/speed in advanced imaging.	Chemometric Methods: (e.g., PCA, LDA, PLSDA) for preprocessing and pattern recognition in spectral data.
	Ensemble Methods: (e.g., Random Forests) combine models for enhanced performance.

We also acknowledge methodological limitations that may introduce bias. Our literature search was limited to Scopus and Web of Science, potentially omitting relevant studies from other databases. Keyword selection focused on high-frequency terms related to RS, ML, and cancer, possibly overlooking alternative terminology. Additionally, the review was restricted to secondary literature due to time constraints, excluding recent primary research. Finally, while LDA was effective for topic modeling, it may lack the sensitivity to capture nuanced language patterns, affecting topic granularity. Alternatives to LDA are for example BERTopic (and its science-focused variant SciBERT), Top2Vec, Contextualized Topic Models CTM, KeyBERT, pyLDAvis, that may offer different insights into this data. These limitations should be considered when interpreting our findings.

3. Discussion

Among the methodologies employed across studies utilizing RS, various innovative diagnostic approaches have emerged, collectively aimed at transforming cancer detection. Spectroscopic techniques, particularly SERS, stand out prominently due to their sensitivity and molecular-level accuracy, frequently enhanced by integrating nanoparticle substrates [6,7,15]. Concurrently, advancements in nanotechnology have provided novel diagnostic platforms facilitating precise tumor targeting, high-resolution imaging, and integrated therapeutic interventions known as theranostics [5,8]. Additionally, liquid biopsy approaches have become increasingly significant as minimally invasive techniques for detecting cancer biomarkers, such as circulating tumor DNA, microRNAs, and extracellular vesicles, employing microfluidic and spectroscopic analyses [1–3]. Artificial intelligence, especially ML algorithms, has enhanced diagnostic precision by efficiently interpreting complex biomedical data from both spectroscopic analyses and medical imaging [3,22,28]. Complementing these approaches, advanced optical imaging techniques, including fluorescence-guided surgery, intraoperative confocal microscopy, and optical coherence tomography (OCT), offer critical real-time visualization during surgical procedures, thereby significantly improving surgical precision, patient safety, and clinical outcomes [29,30].

Beyond conventional RS, SERS has emerged as a transformative variant that addresses the low signal intensity issue by employing nanostructured metallic substrates to amplify Raman signals—sometimes by up to eight orders of magnitude [31]. Advances in nanosol SERS quantitative analysis further improve reproducibility and sensitivity by integrating bioenzyme amplification and aptamer-based selectivity. A comprehensive review on how nanomaterials—ranging from plasmonic nanostructures to 2D and 3D-ordered substrates—enhance SERS performance and enable highly sensitive detection of cancer-related biomarkers is presented in [32].

Early clinical work focused on portable RS systems for in vivo cancer diagnostics. For example, [33] demonstrated a portable Raman device distinguishing malignant from benign skin lesions. Reviews [34] detail how RS and SERS were adapted for rapid, on-site biomarker detection. SERS, leveraging nanostructured substrates, enhances weak spectral signals and detects low-concentration biomarkers [35]. Integration with aptamers or enzymatic amplification yielded diagnostic accuracies up to 98% [36]. Due to SERS's rich data, deep learning and ensemble methods often outperform classical chemometrics in real-time applications. Studies consistently combine enhanced Raman signals with ML to improve accuracy, reduce sample prep, and support clinical scalability. SERS-based biosensors with ML enable fast, label-free diagnostics [37], while NIR and Raman setups evolve despite early signal and matrix challenges [38]. Deep learning approaches decode complex optical data in real time [39], and multivariate classifiers achieve over 90% accuracy in metabolic fingerprinting [40]. Nanomaterials and ML remain key. Nano-patterned substrates improve SERS signal quality, enhancing ML interpretation for low-level biomarker detection [41]. For intraoperative use, fast data acquisition, standardized prep, and ML classification are essential [39]. Beyond cancer, AI-driven SERS and IR methods support rapid viral detection (e.g., COVID-19) [42] and complex classification in microbiology and food safety [43], with broad relevance to clinical diagnostics. Raman and SERS detect key cellular molecules like carotenoids and lipids [44], while SERS-enhanced exosome detection—often ML-assisted—reaches single-exosome sensitivity for cancer diagnosis [45]. Fiber-based in vivo RS, with classifiers like PLS or RF, shows 90–95% accuracy for colon and cervical cancer [46]. Deep learning improves single-cell SERS accuracy to >95% by resolving overlapping signals [47]. Blood cancer detection via SERS and ML reaches ~90% accuracy [48], and integration into endoscopy enables 90–95% accuracy in gastric lesion classification [49]. Ensemble models and CNNs boost SERS biosensor performance by 5–10% over simpler models [50]. Intraoperative glioma margin detection exceeds 90% accuracy using label-free SERS and advanced analytics [14]. SERS used as an “omics” tool shows >90% accuracy in tissue and fluid phenotyping with deep models [51], and SERS-based pathogen/cancer detection achieves ~95% with RF and CNNs [52]. For skin cancer, SERS plus ML counters melanin/water confounders with >90% diagnostic accuracy [53]. Exosome-based SERS hits 95% accuracy in liver/pancreatic tumours [2], while single-cell SERS with NIR substrates and neural networks achieves 95–97% [54]. Studies on HCC and aging highlight the potential of SERS fused with next-gen ML [55]. Hematology-focused SERS studies report >90% accuracy even in small cohorts [56]. SERS paired with advanced analytics (from PLS to deep learning) reliably achieves >85% accuracy in near-clinical settings, though broader use requires standardization and large-scale validation.

Although RS with its ability to capture a unique “molecular fingerprint” of tissues and biofluids has long been recognized as a promising tool for non-invasive cancer detection, its inherently weak and complex signals (often obscured by fluorescence and biological noise) present significant challenges for direct clinical interpretation. In recent years, researchers have increasingly turned to ML to extract diagnostically relevant information from Raman spectra [12]. By coupling RS with ML, studies have improved sensitivity and specificity for detecting malignancies and enabled real-time, intraoperative decision-making [57]. As the technology advanced, researchers began pairing RS with more powerful ML schemes (e.g., SVMs, ensemble methods, and deep learning), increasing both the robustness and the sensitivity of diagnostic models.

In [58] the authors report that combining RS with advanced chemometric methods can achieve high diagnostic accuracy across a wide range of cancers, suggesting its potential as a universal, noninvasive diagnostic tool. Their review reveals that spontaneous RS, when paired with machine learning techniques such as PCA, LDA, PLS-DA, and SVM, has consistently yielded sensitivity and specificity values often exceeding 85%, with some studies reporting accuracies approaching or even surpassing 95% in differentiating cancerous from healthy tissues. They highlight the ability of this integrated approach to analyze diverse biological samples—tissues, cells, and body fluids—capturing subtle biochemical variations that are critical for early diagnosis and effective treatment monitoring, thereby paving the way for large-scale clinical validation and eventual implementation in routine cancer screening.

A major evolution in the field has been the transition from traditional chemometric methods to advanced ML algorithms. Early studies using PCA-LDA achieved classification accuracies of 85–90% for differentiating cancerous from normal tissues [59]. With the advent of deep learning, convolutional neural networks (CNNs) have been employed to automatically learn salient spectral features; in several studies, CNN-based models have pushed classification accuracies beyond 90%, with some reports approaching 95% accuracy on controlled datasets [60,61]. Deep learning models—especially CNNs—were also used to automatically extract key spectral features without requiring extensive manual data transformation. CNN-based approaches often surpassed 90% classification accuracy in rigorously controlled trials, with certain experiments nearing 95% on well-curated spectral datasets [62].

For surgical applications, particularly in breast cancer, stochastic backpropagation neural networks have been developed not only to classify tissues but also to provide a probability score and associated error estimate for each prediction. This dual output is critical for intraoperative margin detection, where real-time decisions can reduce re-excision rates. In a notable review study [61], it is reported that stochastic networks yielded both a classification decision and a measure of confidence, meeting the need for transparent and quantifiable diagnostics. For margin detection in surgical contexts, [63] also reported that stochastic backpropagation networks not only distinguish healthy from malignant tissue but also generate confidence scores for real-time decision-making.

In [64] the authors reported that combining biosensor outputs with SVM classifiers achieved point-of-care diagnostic accuracies around 90%. In another domain, SERS combined with SVM has been applied for prostate cancer screening, achieving a diagnostic accuracy as high as 98.1% based on serum SERS spectra [36]. Such studies underscore the promise of label-free, noninvasive screening methods using blood samples.

The reviewed body of literature highlights significant advancements at the intersection of AI, ML, and RS for cancer diagnostics. In the context of neuro-oncology, in [65] the authors present a comprehensive review of AI/ML systems applied to optical spectral data, emphasizing the use of various Raman modalities—spontaneous, stimulated, resonance, SERS, and CARS—alongside algorithms such as PCA, LDA, SVM, and artificial neural networks. These methods achieved classification accuracies ranging from 70% to 95%, demonstrating strong potential for intraoperative tumor delineation in central nervous system cancers.

Focusing on SERS, [66] discusses how AI and advanced data analytics can help overcome the translational challenges of SERS in clinical applications. They argue that combining SERS with deep learning and multivariate analysis could enable high-accuracy biomarker detection and patient stratification, despite current limitations in reproducibility and standardization. In the domain of bladder cancer, RS, particularly when integrated with SERS and classification models like PCA-LDA and SVM, enables effective differentiation between benign, inflammatory, and malignant bladder tissue [67]. The findings presented in [67] indicate strong diagnostic potential in both tissue-section and endoscopic applications while [68] addresses oral cancer detection, reviewing low-cost and non-invasive screening technologies, including RS, with a focus on integration with AI-based decision support. The authors highlight that such tools can be deployed in primary care or low-resource environments, where early detection can dramatically reduce mortality. In contrast, studies on bone metastases [69] do not employ RS directly, but they illustrate how AI/ML techniques contribute to enhanced diagnostic workflows, emphasizing AI's role in radiological and molecular analysis for metastatic bone lesions, notably from prostate, lung, and breast cancers. Together, these studies underscore that the fusion of RS and AI holds substantial promise across multiple cancer types. Performance metrics are encouraging, with accuracies often exceeding conventional methods. However, widespread clinical adoption depends on overcoming technical challenges such as data standardization, spectral variability, and device interoperability. A systematic review of optical methods—including RS—for brain tumor detection. Is presented in [70]. The study reports accuracy ranging from 54% to 100%, depending on the method and study design, confirming that Raman-based methods offer clinically useful precision when combined with AI for differentiating tumor from normal tissue. Intraoperative neurosurgical histology and a description of the utility of Raman-based techniques like

Stimulated Raman Scattering (SRS), CARS, and SERS in rapid, label-free diagnosis is also presented in [71]. These methods are often complemented by deep neural networks to enhance classification accuracy, highlighting their impact on surgical outcomes in brain cancer. In the field of pancreatic cancer, [72] discusses vibrational spectroscopy—including RS—as a powerful diagnostic tool, particularly when enhanced with CNNs. The report underscores the urgent need for early diagnosis in pancreatic ductal adenocarcinoma and how multivariate analysis and deep learning could distinguish cancer stages more effectively than conventional approaches. Another comprehensive review [73] emphasizes the potential of RS—particularly spontaneous and SERS variants—coupled with machine learning (e.g., PCA, LDA, SVM, PLS-DA, ANN, Random Forests) for the universal diagnosis of cancers, infections, and degenerative diseases. They report classification accuracies of over 90% for several cancer types using cross-validation and external validation protocols, underlining the robustness of such models in clinical scenarios. Another key study [74] reports diagnostic accuracies exceeding 90% for breast, cervical, and colorectal cancers using methods like PCA, LDA, SVM, logistic regression, ANN, and MRDF. It highlights the potential of non-invasive, real-time detection, especially with fiber-optic or nanoparticle-enhanced SERS techniques. Additionally, in [75] it is shown an over 99% classification accuracy in distinguishing breast cancer cell lines using RS with hierarchical classifiers and multiple validation schemes. This underscores the technique's potential for cell-level cancer detection, validated through rigorous chemometric analysis.

In the context of brain cancer detection, Fourier-transform infrared (FT-IR) and Raman spectroscopy have shown significant improvements in sensitivity over time. In [76] the authors report that these techniques can achieve accuracies exceeding 90% in classifying brain cancer subtypes. The complementary nature of FT-IR and Raman spectroscopy allows them to serve as powerful tools in clinical theatres, detecting subtle biochemical changes in tissue composition and enhancing intraoperative margin assessment [59,77].

Moreover, in the surgical setting, reviews of Laser RS for breast cancer detection emphasize that state-of-the-art ML methods—such as stochastic backpropagation neural networks—can provide not only classification outcomes but also uncertainty estimates. This capability is crucial for guiding surgical decisions where complete tumor resection with negative margins directly impacts patient outcomes [61].

Despite significant progress, challenges such as small sample sizes, heterogeneous datasets, and the absence of standardized spectral acquisition and preprocessing protocols continue to hinder the clinical translation of RS-ML systems. Reviews such as [70,78] emphasize the need for large, standardized spectral databases to enable more robust model training. Looking forward, integrating RS with deep learning and uncertainty quantification techniques could improve diagnostic accuracy and provide confidence estimates for predictions, supporting more informed clinical decision-making. As both optical and computational methods advance, the convergence of these technologies may ultimately enable universal, noninvasive, and cost-effective cancer diagnostics.

In summary, the integration of RS with ML represents a multi-faceted advancement in cancer diagnostics. Early portable RS systems, chemometric techniques, and state-of-the-art ML algorithms—from CNNs and stochastic neural networks to SVMs—have collectively enhanced the sensitivity, specificity, and clinical utility of RS while emerging SERS technologies further amplify these benefits.

4. Conclusions

The convergence of Raman spectroscopy and machine learning represents a significant leap forward in cancer diagnostics. From portable, in vivo systems for skin cancer to intraoperative tools for breast cancer margin detection, ML techniques—from PCA-LDA and SVMs to CNNs and stochastic neural networks—have markedly improved the accuracy, speed, and reliability of RS-based diagnostics. With reported accuracies frequently exceeding 90% and some deep learning models reaching near 95%, these integrated approaches are poised to reduce diagnostic uncertainty and improve patient outcomes. As challenges such as data heterogeneity and limited sample sizes are addressed through larger benchmark studies, the clinical translation of RS-ML systems will become increasingly feasible, paving the way for noninvasive, real-time cancer detection and personalized treatment strategies. Challenges remain also in standardizing RS methods and integrating them into clinical workflows. Variability in instrumentation, sample preparation, and environmental factors can impact the reproducibility of spectral measurements. Researchers continue to work on developing calibration protocols and robust algorithms that can adapt to these variabilities, ensuring consistent performance across different settings [28].

Several limitations in our review design should be acknowledged. First, the literature search was restricted to only two databases—Scopus and Web of Science—which, although highly reputable, may have excluded relevant studies indexed elsewhere. Second, the inclusion criteria focused solely on review articles, potentially omitting recent primary studies that could offer valuable insights. Third, the topic modeling relied exclusively on Latent Dirichlet Allocation, which, while robust for high-dimensional text data, may not capture all thematic nuances as effectively as some newer NLP techniques.

Looking forward, the integration of RS and ML holds great promise for real-time, non-invasive cancer diagnostics. As computational power increases and more comprehensive spectral databases become available, these techniques are expected to become even more accurate and accessible. The ongoing interdisciplinary collaboration among engineers, chemists, and clinicians is critical to translating these advanced methodologies into practical clinical tools that can improve patient outcomes [51].

Acknowledgements

This work was supported in part by the MRID, project PNRR-I8 no 842027778, contract no 760096 and by the Core Program within the National Research, Development and Innovation Plan 2022-2027, carried out with the support of MRID, project no. 23020101(SIA-PRO), contract no 7N/2022.

References

[1] Attia AM, Rezaee-Zavareh MS, Hwang SY, et al. Novel Biomarkers for Early Detection of Hepatocellular Carcinoma. Diagnostics 2024;14.
[2] Liu X, Jia Y, Zheng C. Recent progress in Surface-Enhanced Raman Spectroscopy detection of biomarkers in liquid biopsy for breast cancer. Frontiers in Oncology 2024;14.
[3] Grignaffini F, Barbuto F, Troiano M, et al. The Use of Artificial Intelligence in the Liver Histopathology Field: A Systematic Review. Diagnostics 2024;14.
[4] Rahaman A, Anantharaju A, Jeyachandran K, et al. Optical imaging for early detection of cervical cancer: state of the art and perspectives. Journal of Biomedical Optics 2023;28.
[5] Duffield C, Rey Gomez LM, Tsao SC-H, et al. Recent advances in SERS assays for detection of multiple extracellular vesicles biomarkers for cancer diagnosis. Nanoscale 2024.
[6] Yang G, Zhang K, Xu W, et al. A review of clinical use of surface-enhanced Raman scattering-based biosensing for glioma. Frontiers in Neurology 2024;15.
[7] Vázquez-Iglesias L, Stanfoca Casagrande GM, García-Lojo D, et al. SERS sensing for cancer biomarker: Approaches and directions. Bioactive Materials 2024;34:248–68.
[8] Ahmed Taha B, Kadhim AC, Addie AJ, et al. Advancing cancer diagnostics through multifaceted optical biosensors supported by nanomaterials and artificial intelligence: A panoramic outlook. Microchemical Journal 2024;205.
[9] Orlando A, Franceschini F, Muscas C, et al. A Comprehensive Review on Raman Spectroscopy Applications. Chemosensors 2021;9:262. https://doi.org/10.3390/chemosensors9090262.
[10] Amharref N, Beljebbar A, Dukic S, et al. Discriminating healthy from tumor and necrosis tissue in rat brain tissue samples by Raman spectral imaging. Biochim Biophys Acta 2007;1768:2605–15.
[11] Kuzmin AN, Pliss A, Rzhevskii A, et al. BCAbox algorithm expands capabilities of Raman microscope for single organelles assessment. Biosensors (Basel) 2018;8:106.
[12] Lita A, Sjöberg J, Păcioianu D, et al. Raman-based machine-learning platform reveals unique metabolic differences between IDHmut and IDHwt glioma. Neuro-Oncology 2024;26:1994–2009.
[13] Jermyn M, Desroches J, Mercier J, et al. Raman spectroscopy detects distant invasive brain cancer cells centimeters beyond MRI capability in humans. Biomed Opt Express 2016;7:5129–37.
[14] Zhang Y, Yu H, Li Y, et al. Raman spectroscopy: A prospective intraoperative visualization technique for gliomas. Frontiers in Oncology 2023;12.
[15] Fan M, Weng Y, Liu Y, et al. Surface-Enhanced Raman Spectroscopy-Based Optical Biosensor for Liquid Biopsy: Toward Precision Medicine. Laser and Photonics Reviews 2024;18.
[16] Hollon T, Jiang C, Chowdury A, et al. Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging. Nature Medicine 2023;29:828–32.
[17] Zhang L, Li C, Peng D, et al. Raman spectroscopy and machine learning for the classification of breast cancers. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy 2022;264.

[18] Kothari R, Jones V, Mena D, et al. Raman spectroscopy and artificial intelligence to predict the Bayesian probability of breast cancer. Scientific Reports 2021;11. https://doi.org/10.1038/s41598-021-85758-6.
[19] Li Z, Li Z, Chen Q, et al. Machine-learning-assisted spontaneous Raman spectroscopy classification and feature extraction for the diagnosis of human laryngeal cancer. Computers in Biology and Medicine 2022;146. https://doi.org/10.1016/j.compbiomed.2022.105617.
[20] Zhang S, Qi Y, Tan SPH, et al. Molecular Fingerprint Detection Using Raman and Infrared Spectroscopy Technologies for Cancer Detection: A Progress Review. Biosensors (Basel) 2023;13:557. https://doi.org/10.3390/bios13050557.
[21] Luo R, Popp J, Bocklitz T. Deep Learning for Raman Spectroscopy: A Review. Analytica 2022;3:287–301. https://doi.org/10.3390/analytica3030020.
[22] Horta-Velázquez A, Arce F, Rodríguez-Sevilla E, et al. Toward smart diagnostics via artificial intelligence-assisted surface-enhanced Raman spectroscopy. TrAC - Trends in Analytical Chemistry 2023;169.
[23] He C, Zhu S, Wu X, et al. Accurate Tumor Subtype Detection with Raman Spectroscopy via Variational Autoencoder and Machine Learning. ACS Omega 2022;7:10458–68. https://doi.org/10.1021/acsomega.1c07263.
[24] Conforti PM, Lazzini G, Russo P, et al. Raman Spectroscopy and AI Applications in Cancer Grading: An Overview. IEEE Access 2024;12:54816–52. https://doi.org/10.1109/ACCESS.2024.3388841.
[25] Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3:993–1022.
[26] Jonathan Chang. lda: Collapsed Gibbs Sampling Methods for Topic Models 2009:1.5.2. https://doi.org/10.32614/CRAN.package.lda.
[27] Nikita M. ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters 2015:1.0.2. https://doi.org/10.32614/CRAN.package.ldatuning.
[28] Qi Y, Liu Y, Luo J. Recent application of Raman spectroscopy in tumor diagnosis: from conventional methods to artificial intelligence fusion. PhotoniX 2023;4.
[29] Bin-Alamer O, Abou-Al-Shaar H, Gersey ZC, et al. Intraoperative Imaging and Optical Visualization Techniques for Brain Tumor Resection: A Narrative Review. Cancers 2023;15.
[30] Sekhar LN, Juric-Sekhar G, Qazi Z, et al. The Future of Skull Base Surgery: A View Through Tinted Glasses. World Neurosurgery 2020;142:29–42.
[31] Fernández-Lodeiro A, Constantinou M, Panteli C, et al. Breath Analysis via Surface Enhanced Raman Spectroscopy. ACS Sensors 2025.
[32] Yuan K, Jurado-Sánchez B, Escarpa A. Nanomaterials meet surface-enhanced Raman scattering towards enhanced clinical diagnosis: a review. Journal of Nanobiotechnology 2022;20.
[33] Kourkoumelis N, Balatsoukas I, Moulia V, et al. Advances in the in vivo Raman spectroscopy of malignant skin tumors using portable instrumentation. International Journal of Molecular Sciences 2015;16:14554–70.
[34] Khristoforova Y, Bratchenko L, Bratchenko I. Raman-Based Techniques in Medical Applications for Diagnostic Tasks: A Review. International Journal of Molecular Sciences 2023;24.
[35] Li D, Yao D, Li C, et al. Nanosol SERS quantitative analytical method: A review. TrAC - Trends in Analytical Chemistry 2020;127.
[36] Li S, Guo Z, Liu Z. Surface-enhanced Raman spectroscopy + support vector machine: A new noninvasive method for prostate cancer screening? Expert Review of Anticancer Therapy 2014;15:5–7.
[37] Ding Y, Sun Y, Liu C, et al. SERS-Based Biosensors Combined with Machine Learning for Medical Application. ChemistryOpen 2023;12.
[38] Dumont E, De Bleye C, Sacré P-Y, et al. From near-infrared and Raman to surface-enhanced Raman spectroscopy: Progress, limitations and perspectives in bioanalysis. Bioanalysis 2016;8:1077–103.
[39] Fang J, Swain A, Unni R, et al. Decoding Optical Data with Machine Learning. Laser and Photonics Reviews 2021;15.
[40] Ellis DI, Dunn WB, Griffin JL, et al. Metabolic fingerprinting as a diagnostic tool. Pharmacogenomics 2007;8:1243–66.
[41] Manoharan Nair Sudha Kumari S, Thankappan Suryabai X. Sensing the Future-Frontiers in Biosensors: Exploring Classifications, Principles, and Recent Advances. ACS Omega 2024;9:48918–87.
[42] Garip Ustaoğlu Ş, Kaygusuz H, Bilgin MD, et al. Novel approaches for covid-19 diagnosis and treatment: A nonsystematic review. Turkish Journal of Biology 2021;45:358–71.
[43] Feng Y-Z, Sun D-W. Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review. Critical Reviews in Food Science and Nutrition 2012;52:1039–58.
[44] Allakhverdiev ES, Khabatova VV, Kossalbayev BD, et al. Raman Spectroscopy and Its Modifications Applied to Biological and Medical Research. Cells 2022;11.
[45] Yu D, Li Y, Wang M, et al. Exosomes as a new frontier of cancer liquid biopsy. Molecular Cancer 2022;21.
[46] Barik AK, Pavithran MS, Lukose J, et al. In vivo spectroscopy: optical fiber probes for clinical applications. Expert Review of Medical Devices 2022;19:657–75.
[47] Watanabe TM, Sasaki K, Fujita H. Recent Advances in Raman Spectral Imaging in Cell Diagnosis and Gene Expression Prediction. Genes 2022;13.
[48] Savinova I, Fournier A, Grieve S, et al. Plain language summary of a fingerprinting technique to monitor blood cancer. Nanomedicine 2022;17:1751–6.
[49] Liu K, Zhao Q, Li B, et al. Raman Spectroscopy: A Novel Technology for Gastric Cancer Diagnosis. Frontiers in Bioengineering and Biotechnology 2022;10.
[50] Lin C, Li Y, Peng Y, et al. Recent development of surface-enhanced Raman scattering for biosensing. Journal of Nanobiotechnology 2023;21.

[51] Cutshaw G, Uthaman S, Hassan N, et al. The Emerging Role of Raman Spectroscopy as an Omics Approach for Metabolic Profiling and Biomarker Detection toward Precision Medicine. Chemical Reviews 2023;123:8297–346.
[52] Beeram R, Vepa KR, Soma VR. Recent Trends in SERS-Based Plasmonic Sensors for Disease Diagnostics, Biomolecules Detection, and Machine Learning Techniques. Biosensors 2023;13.
[53] Delrue C, Speeckaert R, Oyaert M, et al. From Vibrations to Visions: Raman Spectroscopy's Impact on Skin Cancer Diagnostics. Journal of Clinical Medicine 2023;12.
[54] Zhang Y, Chang K, Ogunlade B, et al. From Genotype to Phenotype: Raman Spectroscopy and Machine Learning for Label-Free Single-Cell Analysis. ACS Nano 2024;18:18101–17.
[55] Akagi K, Koizumi K, Kadowaki M, et al. New Possibilities for Evaluating the Development of Age-Related Pathologies Using the Dynamical Network Biomarkers Theory. Cells 2023;12.
[56] Laskowska P, Mrowka P, Glodkowska-Mrowka E. Raman Spectroscopy as a Research and Diagnostic Tool in Clinical Hematology and Hematooncology. International Journal of Molecular Sciences 2024;25.
[57] Goff NK, Ashby L, Ashour R. Meta-Analysis of the Efficacy of Raman Spectroscopy and Machine-Learning–Based Identification of Glioma Tissue. World Neurosurgery 2024;189:26–32. https://doi.org/10.1016/j.wneu.2024.05.112.
[58] Ralbovsky NM, Lednev IK. Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy 2019;219:463–87.
[59] Mamede AP, Santos IP, Batista de Carvalho ALM, et al. A New Look into Cancer—A Review on the Contribution of Vibrational Spectroscopy on Early Diagnosis and Surgery Guidance. Cancers 2021;13:5336. https://doi.org/10.3390/cancers13215336.
[60] Blake N, Gaifulina R, Griffin LD, et al. Machine Learning of Raman Spectroscopy Data for Classifying Cancers: A Review of the Recent Literature. Diagnostics 2022;12.
[61] Kothari R, Fong Y, Storrie-Lombardi MC. Review of laser Raman spectroscopy for surgical breast cancer detection: Stochastic backpropagation neural networks. Sensors (Switzerland) 2020;20:1–20.
[62] Ramirez CMA, Greenop M, Almoshawah YA, et al. Advancing cervical cancer diagnosis and screening with spectroscopy and machine learning. Expert Review of Molecular Diagnostics 2023;23:375–90.
[63] Boichenko E, Kirsanov D. Optical spectroscopy and chemometrics in intraoperative tumor margin assessment. TrAC - Trends in Analytical Chemistry 2023;160.
[64] Kokabi M, Tahir MN, Singh D, et al. Advancing Healthcare: Synergizing Biosensors and Machine Learning for Early Cancer Diagnosis. Biosensors 2023;13.
[65] Savelieva T, Romanishkin I, Ospanov A, et al. Machine Learning and Artificial Intelligence Systems Based on the Optical Spectral Analysis in Neuro-Oncology. Photonics 2025;12.
[66] Mahmoud AY, Teixeira A, Aranda M, et al. Will data analytics revolution finally bring SERS to the clinic? TrAC - Trends in Analytical Chemistry 2023;169.
[67] Liu Z, Zhang P, Wang H, et al. Raman Spectrum-Based Diagnosis Strategy for Bladder Tumor. Urologia Internationalis 2022;106:109–15.
[68] Haj-Hosseini N, Lindblad J, Hasseus B, et al. Early Detection of Oral Potentially Malignant Disorders: A Review on Prospective Screening Methods with Regard to Global Challenges. Journal of Maxillofacial & Oral Surgery 2024;23:23–32.
[69] Papalia GF, Brigato P, Sisca L, et al. Artificial Intelligence in Detection, Management, and Prognosis of Bone Metastasis: A Systematic Review. Cancers 2024;16.
[70] Burström G, Amini M, El-Hajj VG, et al. Optical Methods for Brain Tumor Detection: A Systematic Review. Journal of Clinical Medicine 2024;13.
[71] Mohamed AA, Sargent E, Williams C, et al. Advancements in Neurosurgical Intraoperative Histology. Tomography 2024;10:693–704.
[72] Szymoński K, Chmura Ł, Lipiec E, et al. Vibrational spectroscopy—are we close to finding a solution for early pancreatic cancer diagnosis? World Journal of Gastroenterology 2023;29:96–109.
[73] Ralbovsky NM, Lednev IK. Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning. Chemical Society Reviews 2020;49:7428–53.
[74] Tu Q, Chang C. Diagnostic applications of Raman spectroscopy. Nanomedicine: Nanotechnology, Biology, and Medicine 2012;8:545–58.
[75] Bocklitz T, Putsche M, Stüber C, et al. A comprehensive study of classification methods for medical diagnosis. Journal of Raman Spectroscopy 2009;40:1759–65.
[76] Lilo T, Morais CL, Shenton C, et al. Revising Fourier-transform infrared (FT-IR) and Raman spectroscopy towards brain cancer detection. Photodiagnosis and Photodynamic Therapy 2022;38.
[77] Auner GW, Brusatori MA. A "New Radiation"—shining light on cancer. Cancer and Metastasis Reviews 2020;39:1019–20.
[78] Krafft C, Steiner G, Beleites C, et al. Disease recognition by infrared and Raman spectroscopy. Journal of Biophotonics 2009;2:13–28.