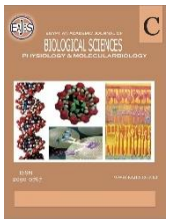
.

Citation: *Egypt.Acad.J.Biolog.Sci. (C.Physiology and Molecular biology) Vol. 16(1) pp417-430 (2024)*
DOI: 10.21608/EAJBSC.2024.360286

Egypt. Acad. J. Biolog. Sci., 16(1):417-430 (2024)

Egyptian Academic Journal of Biological Sciences
C. Physiology & Molecular Biology
ISSN 2090-0767
www.eajbsc.journals.ekb.eg



Impact of Machine Learning on Raman and Raman Optical Activity (ROA) Spectroscopic Analyses of ribonucleic acid structure

Omer Azher
Department of Laboratory Medicine, Faculty of Applied Medical Sciences, Al-Baha University, Al-Baha, Saudi Arabia
***E-mail:** oazher@bu.edu.sa

REVIEW INFO
Review History
Received:2/4/2024
Accepted:12/6/2024
Available:16/6/2024

Keywords:
Artificial Intelligence, RNA, spectroscopy, analysis, structural organization.

ABSTRACT

The potential of machine learning (ML) to revolutionize analytical sciences, especially, Raman and ROI spectra collection for RNA nucleotides is paramount. ML provides exceptional opportunities for the speedy extraction of vast information from the complex dataset generated by various analytical techniques including spectroscopy which could expedite the determination of the behaviour of complex molecules with the utmost accuracy. Ribonucleic acid (RNA) molecules exist in all living cells. These polymers play significant roles in various biochemical processes, such as translation and protein synthesis. The function of RNA as a catalyst for several cellular reactions in addition to its significant role in gene expression shapes the biological system. The functional versatility of RNAs depends on their ability to fold in various structural conformations, which necessitates delineating the motifs and elements' structures in RNA to gain a comprehensive insight into the functional versatilities of these biopolymers. Moreover, the pivotal role of these polymers in diagnosis and therapy could be comprehended by functional activity analysis of RNAs using Raman and ROA spectroscopy in conjunction with ML and artificial intelligence. The current review aimed to shed light on the impact of ML algorithms on Raman and ROA spectroscopic RNA structural data analysis. Additionally, this review summarizes the RNA structural organization and methodological approaches of ML-assisted Raman and ROA spectroscopies for RNA in tandem with traditional algorithms. The future directions of the ML-assisted Raman and ROA for RNA structural analysis have also been highlighted to boost biomolecular research efficiency and accuracy.

INTRODUCTION

Raman (R) and Raman Optical Activity (ROI) spectroscopies are highly sensitive measurement techniques based on the principle of light-chemical bond vibrational interaction in molecules of the material (Hobro *et al.*, 2008; Madey & Yates Jr 2013), that have been employed extensively in analytical sciences (Ayres *et al.*, 2021; Fan *et al.*, 2011). Light interaction with chemical-bond electron density results in molecular vibration-excitation and light frequency shift explains the Raman effect (Ahmed & Jackson 2014). Moreover, the effect is also observed when the elastic scattering of light and energy exchange with material excitation, for instance, lattice vibration in solid material occurs. Therefore, the vibrational fingerprint inherently associated with a specific molecular structure could be inquisitively analyzed to gain insight into molecular identification and characterization (Das & Agrawal 2011; Garcia-Rico *et al.* 2018).

The application of Raman analysis has been recognized in various scientific areas including pharmaceutical science for characterizing unknown biomolecules of medicinal interest (Craig *et al.*, 2013; Li-Chan 1996; Movasaghi *et al.*, 2007). Despite the high potential of the Raman spectroscopy technique, it is complicated to process and extract valuable information from the spectral data of complexity with random noise that necessitates the robust processing technique (Gautam *et al.*, 2015; Pelletier 2003). The key components of the Raman spectroscopy and Raman scattering phenomenon are illustrated in Figures 1a & 1b (Orlando *et al.*, 2021; Rostron *et al.*, 2016).

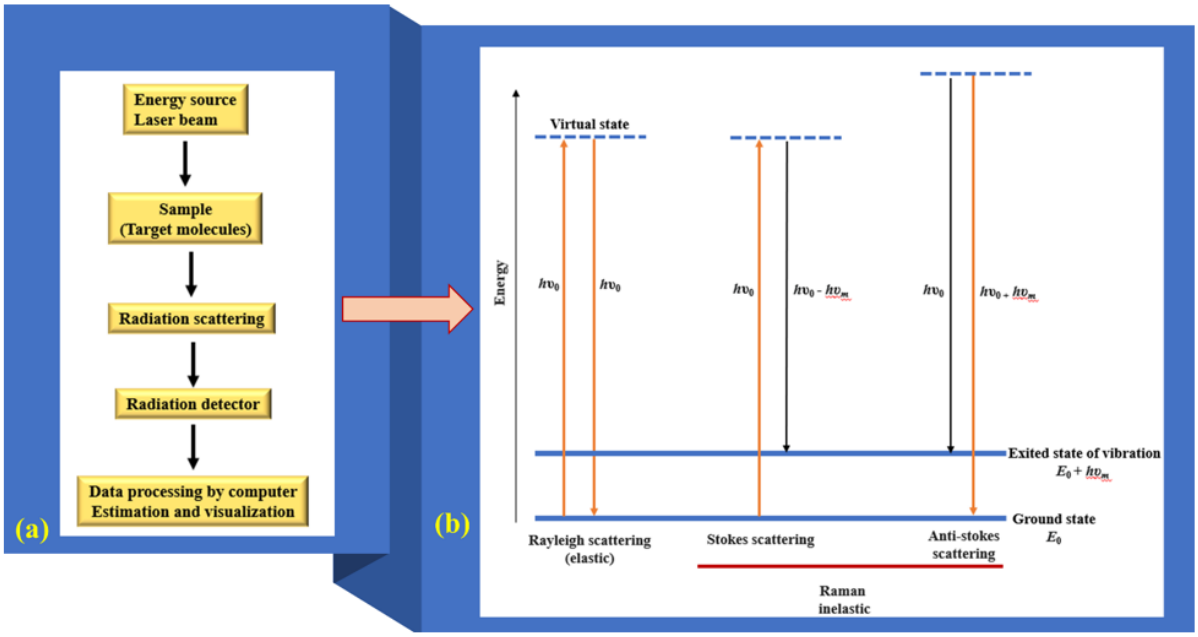


Fig. 1. Illustration of the (a); key components of the Raman spectroscopy and (b); Raman scattering phenomena. Plank's constant (h), frequency of the incident light (υ_0), Vm: frequency of molecular vibration (υ_m), and energy at the ground level (E_0).

Long data processing time and error-prone analytical results with traditional computational approaches are not sufficient to meet the multidimensional Raman spectroscopy-based research about biological molecules including structural insight of RNA (Antonio & Schultz 2014; Butler *et al.*, 2016; Guo *et al.*, 2021). Therefore, with the advent of artificial intelligence (AI), recently, machine-learning (ML) emerged as a potential analytical tool/technique to address such issues with its capabilities to make automated predictions and mine deep complex data including spectral data (Kusters *et al.*, 2020; Xu *et al.*, 2021). AI is accomplished based on training with pre-labeled data to provide predictions on fresh data input which plays a crucial role in expediting the experimental and computational analysis (Duarte & Ståhl 2019; Lewis & Denning 2018). Consequently, ML could be used to analyze the spectral datasets obtained from Raman spectroscopy to unfold the structural complexity of the RNA structural organization (Qi *et al.*, 2023). ML is advantageous over traditional chemometrics and qualitative and quantitative statistical methods because it can analyze high-dimensional datasets efficiently and find significant connections and patterns beyond the functional groups level in a molecule (Adhikari *et al.*, 2023; Leardi 2002; Rocha *et al.*, 2020). The applications of decision trees, support vector machines, random forests, and artificial neural networks ML algorithms have been described and

reported in recently published scientific reports (Bhatti *et al.*, 2023; Charbuty & Abdulazeez 2021; Ding *et al.*, 2011), in the recent past which could play a vital role in Raman spectral comprehensive data analysis more effectively compared to the traditional data processing technique (Carey *et al.*, 2015).

ML algorithms in tandem with traditional processing techniques such as principal component analysis (PCA), partial least-square regression (PLS), linear-regression (LR), linear-discriminant analysis (LDA), least-square (LS), and quadrant-discriminant analysis (QDA) in conjunction with spectral preprocessing techniques have been reported to be employed to automated classification of the spectral data of biomolecules (RNA), therefore, these algorithms have been identified as the remarkable research subject in the last few years (Fan *et al.*, 2023; Han *et al.*, 2022; Luo *et al.*, 2022; Zhang *et al.*, 2020). The application of artificial intelligence potentially expedites the determination of molecular patterns and connections based on analyzing a given data set and predicting valuable results. Though there are review articles on the application of ML in various scientific areas are, however, information on the impact of ML on Raman spectroscopic analysis of RNA has not sufficiently published. Therefore, this review aimed to summarize the structural organization of ribonucleic acid and the impacts of ML on Raman and ROA spectral analysis of motifs and elements in RNA structure along with the future direction of spectral research with the application of ML. The methodological approaches for the ML-assisted Raman and ROA spectroscopies are depicted in Figure 2.

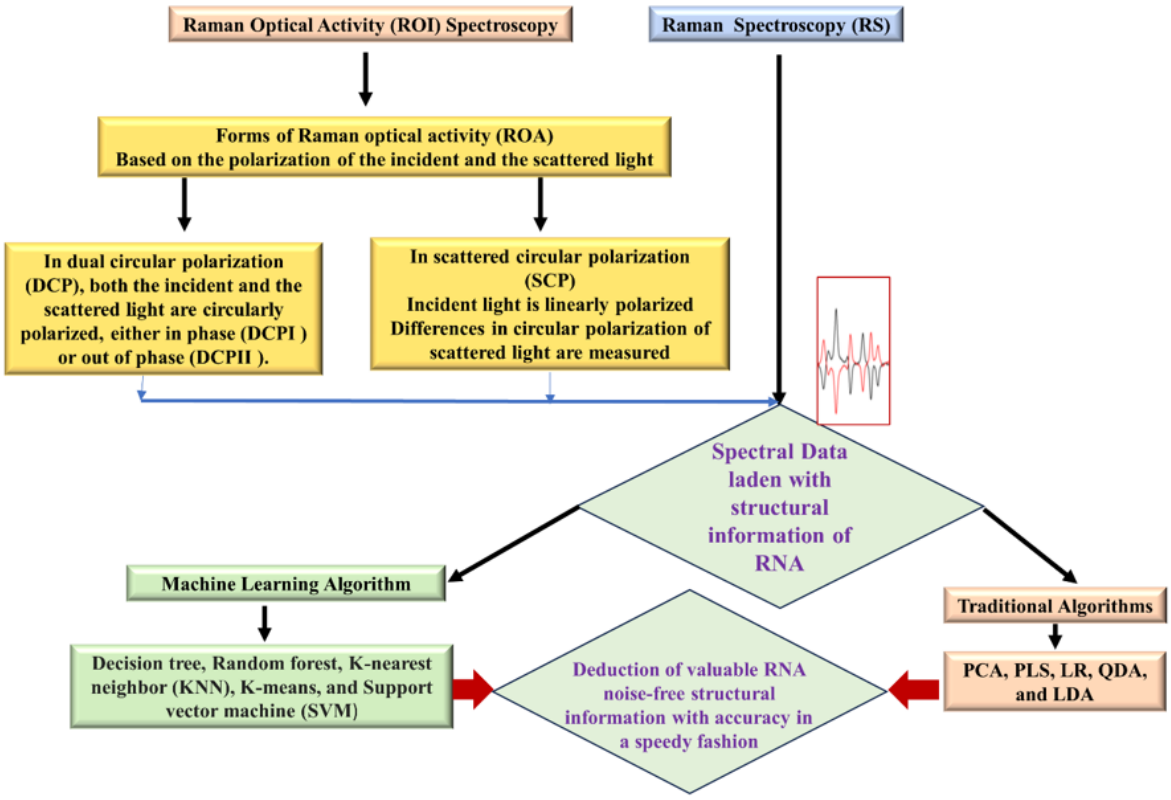


Fig. 2. Depiction of the Machine Learning Approaches assisted the Raman and Raman Optical Activity (ROI) spectroscopies for RNA structures. PCA; principal component analysis, PLS; partial least square regression, LR; linear regression, LDA; linear discriminant analysis, LS; least square, and QDA; quadrant discriminant analysis**.**

Structural Organization of RNAs:

NA negatively charged biopolymer with 2'-hydroxyl group imparting which differentiates it from DNA thermodynamically and conformationally that reflects in the structure and the function of RNA (Fohrer *et al.*, 2006). A variety of RNA exists each with specific functional features including catalysis, protein synthesis, and gene expression regulation (Minchin & Lodge 2019). RNA conformation is explained under 3 different structural organization levels: (A) primary structure, (B) secondary structure, and (C) tertiary structure. The sequence of nucleotides in RNA defines the primary organizational level while the secondary structure involves different base pairing modes (canonical and noncanonical) that describe the two-dimensional (2D)-folding of the biopolymer. Moreover, the tertiary structure involves the interaction between various secondary structural motifs that leads to overall three-dimensional arrangements of RNA (Abraham *et al.*, 2008; Dirheimer *et al.*, 1994; Eric Westhof & Pascal Auffinger 2000).

The primary structural organization encompasses an arrangement with a sequence of four different nucleotides of ribonucleic acid (Madison 1968). Although RNA (a single-stranded molecule) has the potential for folding itself to form diverse structural motifs. The pairing of the adjacent nucleotides is what defines the secondary structure of the ribonucleic acid. Unlike proteins, the secondary structural stability of RNA is unrelated to its tertiary conformation. As a result of it, RNA folding involves first the formation followed by the consolidation of its secondary structure (Dima *et al.*, 2005).

RNA base pairing follows a canonical (Watson–Crick) base-pairing (interaction of Guanine-cytosine while adenine-uracil with H-bonds). Moreover, the Watson-Crick interaction is prevalent among RNA molecules, and it has a remarkable attribute in the formation of RNA helices. In addition, it also follows non-canonical (Non-Watson-Crick) base pairing which constitutes about 40% of the total base pairing in RNA molecules (Lemieux & Major 2002). This mode of pairing provides distinctive sites for interactions with proteins, ligands, and metals. Furthermore, noncanonical base pairing is critical for the existence of the A-form structure of RNA (Lemieux & Major 2002; Sharma *et al.*, 2010). Common noncanonical pairings include the G-U-Wobble pair, A+:C, and G-A-pairs. The G-U wobble-pair is the most frequently detected base pair in RNA. Moreover, wobble interaction produces distinctive structural, chemical, and ligand binding capabilities. In addition, G-U base-pairs are thermodynamically more stable. This stability allows the wobble base to be involved in various biological activities (Halder & Bhattacharyya 2013). On the other hand, A+: C base pairs are observed in ribozymes and some RNA loops. In this type of interaction, the addition of a proton exposes a hydrogen (H)-bond to the cytosine-carbonyl group. This feature provides further chemical diversity to RNA (Chen *et al.*, 2012; Halder & Bhattacharyya 2013). Moreover, G: U-base-pairs are spotted commonly in internal loops of RNA tertiary structure, therefore, they assist the folding of RNA and also enhance the ligand binding capability of RNA molecules (Chen *et al.*, 2005).

Suborganization of Secondary RNA Structures:

RNA secondary structures include a. stems, b. loops and c. pseudoknots. A stem develops when two or more adjacent complementary nucleotides are paired. On the other hand, the unpaired nucleotides in the stems are called loops (Holbrook 2005; Svoboda & Cara 2006). There are different types of RNA loops present in various locations within the biopolymer, examples include: (i) Hairpin, (ii) Internal loop, (iii) Bulge, and (iv) Multibranch loop or junction. Hairpin loops are one of the fundamental RNA secondary structures (Jia *et al.* 2004; Svoboda & Cara 2006). They are formed in various parts of different types of RNA. Each stem-loop has distinctive criteria, such as

nucleotide sequence, size of the loop as well as stem length (Holbrook 2005). The prevalence and versatility of hairpin loops reflect their important role in different biological functions, for example, regulation of gene expression, stimulation of RNA folding, and recognition of RNA binding proteins (Svoboda & Cara 2006).

Types of RNA Secondary Structure:

The presence of four nucleotides in an RNA loop forms a structure called tetraloops. According to the sequence of the residues, there are three main types of tetraloops: GNRA, UNCG as well as CUUG (A, U, G/C stand for N, and R stand for A/G). Although each family has distinct nucleotide sequences, they are structurally very similar. Functionally, various roles of different types of tetraloops have been reported. For example, the GNRA tetraloop acts as a site for protein interaction (Thapar *et al.*, 2014). Moreover, the UUCG tetraloop serves as a site for RNA folding and prohibits clustering of large molecules whereas; GAAA plays a critical role in interactions stabilizing tertiary structure (Nicolas Leulliot *et al.*, 1999; Thapar *et al.*, 2014).

Internal loops are another type of RNA secondary motif. These loops are formed because of the unpaired nucleotides present between two stems (Schroeder & Turner 2000). In addition, internal loops have two subdivisions: symmetric internal loops, which include an equal number of the residues or strands, and asymmetric internal loops, which involve unequal numbers of nucleotides. Internal loops are crucial for many biologically significant functions, with one of these being to provide free energy for RNA folding (Hammond *et al.*, 2010). Bulges can be defined as unpaired regions of nucleotides that arise only from one RNA strand. The size of the bulge varies from single to numerous residues.

Furthermore, bulges influence the assembly of RNA architecture (Danaee *et al.*, 2018). The fourth kind of RNA loop is the Multibranch loop. The M-loop is a complex structure from which several loops exit. Numerous Multibranch loops are present in rRNA since they are critical for configuring RNA secondary structure (Diamond *et al.*, 2001). Moreover, pseudoknots are considered one of the most prominent structures of RNA (Staple & Butcher 2005). They have evolved because of the pairing of a hairpin-loop having a single-stranded complementary sequence. Sometimes, base-pairing phenomena occur between 2 or more than 2 hairpin loops. The formation of pseudoknots in catalytic RNAs is more obvious than in other RNA types (Hajdin *et al.*, 2013). In addition, RNA pseudoknots are required for many biological functions of human RNA, such as telomerase activity (Theimer *et al.*, 2005), therefore, in addition, the presence of pseudoknots in viral RNA is essential for replication and gene expression (Brierley *et al.*, 2007).

RNA Structural Motifs and Structural Elements:

Structural motifs and structural elements are two terms used for further understanding of various structures of molecules (Butcher & Pyle 2011; Hendrix *et al.*, 2005). RNA motifs are specific areas within the molecule with defined lengths and sequences of nucleotides. They usually behave as one unit and perform specific structural or biological functions (Kinjo & Nakamura 2012). Motifs in RNAs are primarily identified by a unique sequence of nucleotides in some areas of functional RNAs, such as tRNA and rRNA (Hendrix *et al.* 2005), examples of some of these motifs include different tetraloops, the kink-turn, the sarcin–ricin loop, and the T-loop are tabulated in Table 1.

Table 1. Tabulation of structural details of RNA motifs.

RNA Structural motifs and elements	Description
GNRA tetraloop	This tetraloop is composed of four unpaired nucleotides following the GNRA sequence where N stands for the A, C, G, or U and R stands for A/G.
Lonepair triloop	This motif includes a single base-paired nucleotide coated by three three-nucleotide hairpin loops. (Lee *et al.* 2003)
T-loop	One of the structural components of tRNA. This loop contains five unpaired nucleotides that form a U-turn structure which is surrounded by a noncanonical base-pair. (Krasilnikov 2003)
Sarcin- ricin loop	Composed of two secondary structures: GAGA tetraloop and bulged G motif. (Korennykh *et al.* 2007)
Kink turn	Made from two helices in which a three-nucleotide-bulge is positioned righltly on its 3'-side by A-G as well as G-A base pairs and canonical base pairing on its 5' side. Forming kink in the backbone of the helix. (Schroeder *et al.* 2010)
D-loop	One of the motifs in tRNA is composed of dihydrouracil in addition to 7-11 base pairs. (Nicolas *et al.* 2002)
Hook turn	Occurs in RNA double strands where two asymmetric internal loops are present. One of the two helices is short (S) while the other one is intended to be longer (L). An A-turn helix and the (S) strand bend forming a hook turn. (Zhong & Zhang 2012)
C-loop	This motif includes two asymmetric helices with two or more base triples which are produced by 2 stacked canonical base-pairs having interacted with loop-bases. (Afonin & Leontis 2006)
Kissing loop	Occurs when two hairpins interact with each other via canonical base pairing. (Salim *et al.* 2012)
U-turn	Sharp bends in the RNA backbone as a result of UNR sequence coated by pyrimidine (Y) in the form of Y-Y, Y-A, or G-A base-pair (Gutell *et al.* 2000).
S-turn	Two continuous bends in RNA backbone resemble 'S' shape (Hendrix *et al.* 2005).
Cross-strand stack	Base pairing between one helix and an opposite helix (Lee *et al.* 2006).

RNA Tertiary Structure:

The diverse types of secondary structural elements and motifs interact with one another developing a more sophisticated structural organization that determines the overall comprehensive architecture of the molecule (Abraham *et al.*, 2008). Despite its role in formulating RNA overall conformation, a tertiary configuration is also directly related to many biological functions performed by ribonucleic acid (E. Westhof & P. Auffinger 2000). Moreover, the interconnection between different secondary structural motifs involved in RNA tertiary structure can be divided into three main interactions: (a) between two double strands, (b) between a helical strand and an unpaired region, and (c) between two unpaired regions (Abraham *et al.*, 2008). Interaction between two double strands can be subdivided into: a) Coaxial stacking and b) adenosine platform. Coaxial stacking occurs when two double strands are next to each other this causes stacking of their terminal base pairs (Tyagi &

Mathews 2007; Zhang *et al.*, 2011). Interaction between helical strand and unpaired region in which the binding between a double strand and a single region involves four types of organizations: a) triplex, b) tetra loop, c) metal core, and d) ribose zipper. The triplex includes the binding of a double helical strand with one single strand.

The triplex -forming oligonucleotides (TFOs) present in the single helix bind with the double strand via non-Watson –Crick base pairing. (Buske *et al.*, 2011). Tetraloops are important structural motifs with numerous biological functions. Furthermore, tetraloops can produce another critical type of binding known as tetra loop-tetraloop receptor interaction (Moore 1999). This type of interaction is an important RNA motif that can bring different structures to proximity.

The binding of tetraloop with a receptor is mainly established between GNRA tetraloop and a target receptor containing a GAAA motif in the minor groove of RNA. Moreover, hydrogen bonds are formed between the OH groups present in the receptor as well as the GNRA. In addition to the H-bonding, the adenosine platform performs a crucial role in binding with A^2 from the tetraloop adding more stability to the motif (Westhof & Fritsch 2000).

Raman and ROA Spectroscopies for Structural Analysis of Ribonucleic Acid:

The role of vibrational-spectroscopies in structural biology is paramount because of their sensitivity to unfold vast structural information and their applicability to diverse biomolecules under various conditions (Hobro *et al.*, 2008). Raman spectroscopy and ROA are the two principal vibration spectroscopies (Ashton *et al.*, 2007)**.** ROA operates based on the Raman-scattering phenomenon of light and measures chirality allied with the Raman transition as illustrated in Figure 3 (Batista Jr *et al.*, 2015). Measuring Raman and ROA spectra simultaneously from the same specimen are achieved they exhibit high sensitivity to diverse attributes of macro-molecular-structure, therefore, the data retrieved shows complementarity (Batista Jr *et al.* 2015). The combined potential of the Raman and ROA have been exploited to investigate RNA structure to a greater extent (Hobro *et al.*, 2007). Despite being complementary techniques, the Raman and ROA could also be used independently for RNA structural analysis (Barron *et al.*, 2003). Moreover, these techniques have been used, recently, to identify structural conformations, for instance, the GNRA tetraloop(Hernández *et al.* 2003; N Leulliot *et al.* 1999). Novel RNA structural information could be obtained by ROA (Blanch *et al.*, 2002). Additionally, Raman and ROA spectroscopic spectra may be inquisitively evaluated to delineate the alteration in RNA sequences and structure by analyzing the specific spectral changes (Blanch *et al.*, 2002).

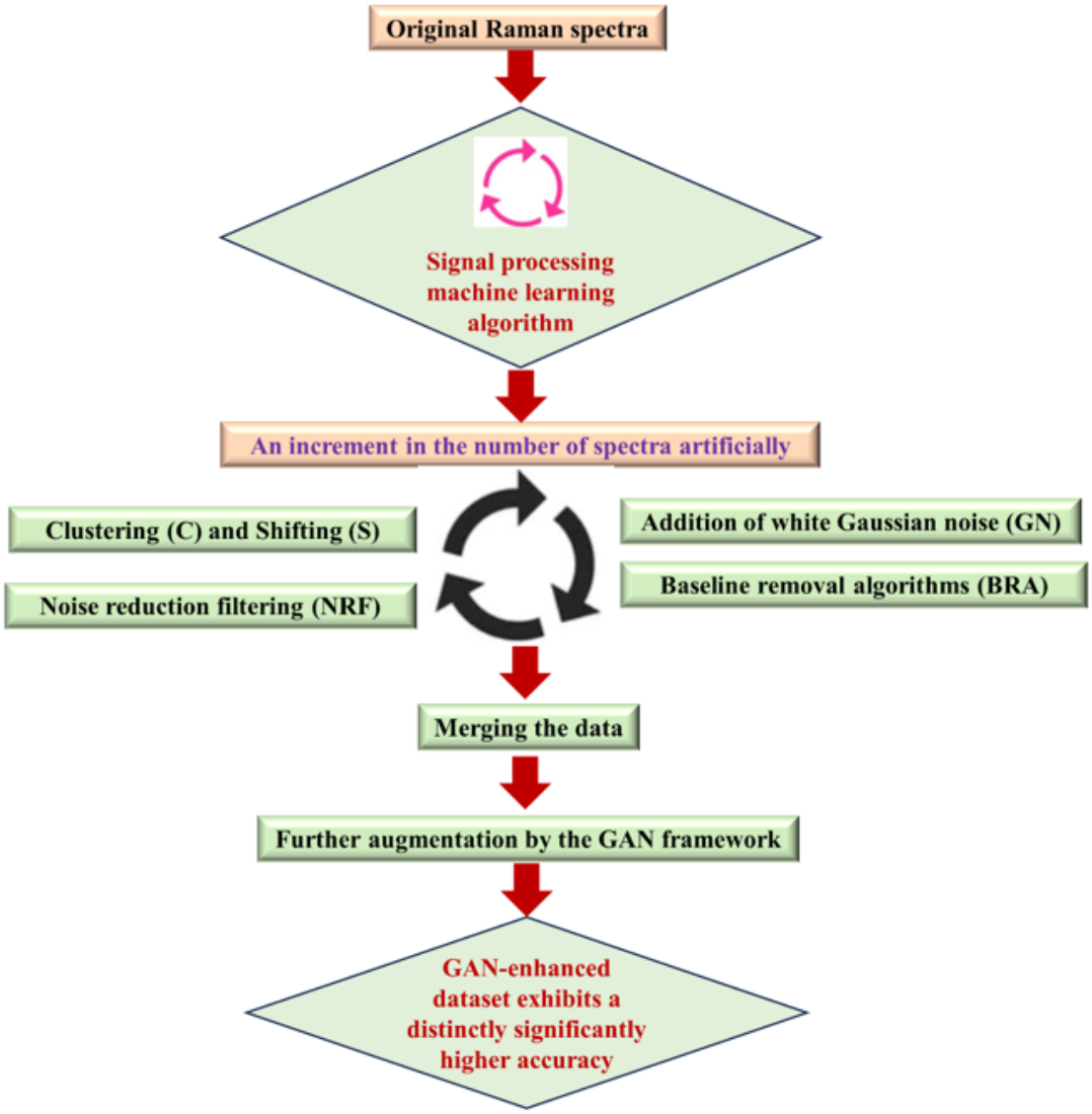


Fig. 3. Illustration of Rama data enhancement strategies. GAN; generative adversarial network

Data Enrichment for Raman and ROA Spectra:

One of the crucial factors for machine learning is the training dataset size which impacts the performance of ML and deep learning significantly (Dargan *et al.*, 2020). For multi-classification studies of ML-assisted Raman and ROA spectra, the amount of Raman data to the considerable level is an additional challenge because its availability with researchers usually remains minimal (Shorten & Khoshgoftaar 2019). Additionally, small data-based deep learning throws the issue of overfitting (Shorten & Khoshgoftaar 2019). Therefore, to address this issue, an ordered/hierarchical data augmentation/enhancement strategy is applied using a generative adversarial network (*GAN*) method (Kim & Lee 2022). The data enrichment hierarchy starts with a Raman spectrum and advances to encompass various signal processing ML algorithms to enhance the spectra counts artificially in a dataset (original) which includes the addition of white-Gaussian-noise (GN) to the originally generated signal data, employing baseline-removal-algorithms (BRA), noise-reduction-filtering (NRF), clustering process (C), shifting (S), and finally merging the data followed by further enhancement by employing the GAN framework (Kim & Lee 2022). Figure 3 illustrates the data enhancement strategy (Kim & Lee 2022).

Future Perspectives and Conclusions:

ML-assisted Raman and ROA spectroscopies reflect great potential in expediting ribonucleic acid research with high precision and considerable accuracy. Machine learning is based on the analysis of huge data divided into training and test datasets, which helps in automatizing the whole process to provide accurate, valuable prediction results. These vibration spectroscopic techniques are one of the most appropriate analytical tools for studying RNA sequences, motifs, and elements and their conformation alterations under various conditions. However, the limited data size of Raman spectra remains a potential challenge for the ML-assisted Raman and ROA spectroscopies which needs to be addressed with priority. In addition, establishing a public database with standard normalization and data processing methods for obtaining Raman spectra worldwide would be the future direction to address the Raman spectral data size. Furthermore, minimizing the time taken to retrieve the required number of Raman spectroscopic imageries with the enhanced spectrophotometric device efficiency would for extracting reliable RNA structural information. Moreover, the miniaturization of a spectrophotometer and the advancement of ML techniques, in the future, would be a powerful combinatorial tool for future in-depth analysis of RNA structures along with other biomolecular structures. ML-assisted data Raman spectroscopy would also guide more effective analysis of the huge and complex biomolecular data and therefore, it would be able to revolutionize the nucleic acid and protein research's speed, accuracy, and reliability along with minimizing manpower and analysis cost.

Declarations:

Ethical Approval: It is not applicable.

Conflict of interests: Theres no conflict of interest

Authors Contributions: I hereby verify that the author mentioned on the title page have made substantial contributions to the conception and design of the study, have thoroughly reviewed the manuscript, confirm the accuracy and authenticity of the data and its interpretation, and consent to its submission.

Funding: No funding was received.

Availability of Data and Materials: All datasets analysed and described during the present study are available from the corresponding author upon reasonable request.

Acknowledgements: Not applicable

REFERENCES

Abraham M, Dror O, Nussinov R, & Wolfson H J (2008). Analysis and classification of RNA tertiary structures. *RNA*, 14(11): 2274-2289.

Adhikari M, Houhou R, Hniopek J, & Bocklitz T (2023). Review of Fluorescence Lifetime Imaging Microscopy (FLIM) Data Analysis Using Machine Learning. *Journal of Experimental and Theoretical Analyses*, 1(1): 44-63.

Afonin K A, & Leontis N B (2006). Generating new specific RNA interaction interfaces using C-loops. *Journal of the American Chemical Society*, 128(50): 16131-16137. https://doi.org/10.1021/ja064289h

Ahmed W, & Jackson M J. (2014). *Emerging nanotechnologies for manufacturing* .William Andrew.

Antonio K A, & Schultz Z D (2014). Advances in biomedical Raman microscopy. *Analytical chemistry*, 86(1): 30-46.

Ashton L, Barron L D, Hecht L, Hyde J, & Blanch E W (2007). Two-dimensional Raman and Raman optical activity correlation analysis of the α-helix-to-disordered transition in poly (L-glutamic acid). *Analyst*, 132(5): 468-479.

Ayres L B, Gomez F J V, Linton J R, Silva M F, & Garcia C D (2021). Taking the leap between analytical chemistry and artificial intelligence: A tutorial review. *Analytica Chimica Acta*, 1161: 338403. https://doi.org/10.1016/j.aca.2021.338403

Barron L D, Blanch E W, McColl I H, Syme C D, Hecht L, & Nielsen K (2003). Structure and behaviour of proteins, nucleic acids and viruses from vibrational Raman optical activity. *Spectroscopy*, 17(2-3): 101-126.

Batista Jr J M, Blanch E W, & da Silva Bolzani V (2015). Recent advances in the use of vibrational chiroptical spectroscopic methods for stereochemical characterization of natural products. *Natural Product Reports*, 32(9): 1280-1302.

Bhatti S, Ahmad S R, Asif M, & Farooqi I u H (2023). Estimation of aboveground carbon stock using Sentinel-2A data and Random Forest algorithm in scrub forests of the Salt Range, Pakistan. *Forestry*, 96(1): 104-120.

Blanch E W, Hecht L, Syme C D, Volpetti V, Lomonossoff G P, Nielsen K, & Barron L D (2002). Molecular structures of viruses from Raman optical activity. *Journal of general virology*, 83(10): 2593-2600.

Brierley I, Pennell S, & Gilbert R J (2007). Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nature Reviews Microbiology*, 5(8): 598-610.

Buske F A, Mattick J S, & Bailey T L (2011). Potential in vivo roles of nucleic acid triple-helices. *RNA Biology*, 8(3): 427-439. https://doi.org/10.4161/rna.8.3.14999

Butcher S E, & Pyle A M (2011). The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of chemical research*, 44(12): 1302-1311.

Butler H J, Ashton L, Bird B, Cinque G, Curtis K, Dorney J, Esmonde-White K, Fullwood N J, Gardner B, & Martin-Hirsch P L (2016). Using Raman spectroscopy to characterize biological materials. *Nature protocols*, 11(4): 664-687.

Carey C, Boucher T, Mahadevan S, Bartholomew P, & Dyar M (2015). Machine learning tools formineral recognition and classification from Raman spectroscopy. *Journal of Raman Spectroscopy*, 46(10): 894-903.

Charbuty B, & Abdulazeez A (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01): 20-28.

Chen G, Znosko B M, Kennedy S D, Krugh T R, & Turner D H (2005). Solution structure of an RNA internal loop with three consecutive sheared GA pairs. *Biochemistry*, 44(8): 2845-2856.

Chen J L, Dishler A L, Kennedy S D, Yildirim I, Liu B, Turner D H, & Serra M J (2012). Testing the nearest neighbor model for canonical RNA base pairs: revision of GU parameters. *Biochemistry*, 51(16): 3508-3522.

Craig A P, Franca A S, & Irudayaraj J (2013). Surface-enhanced Raman spectroscopy applied to food safety. *Annual review of food science and technology*, 4: 369-380.

Danaee P, Rouches M, Wiley M, Deng D, Huang L, & Hendrix D (2018). bpRNA: large-scale automated annotation and analysis of RNA secondary structure. *Nucleic acids research*, 46(11): 5381-5394.

Dargan S, Kumar M, Ayyagari M R, & Kumar G (2020). A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 27: 1071-1092.

Das R S, & Agrawal Y (2011). Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational spectroscopy*, 57(2): 163-176.

Diamond J M, Turner D H, & Mathews D H (2001). Thermodynamics of three-way multibranch loops in RNA. *Biochemistry*, 40(23): 6971-6981.

Dima R I, Hyeon C, & Thirumalai D (2005). Extracting stacking interaction parameters for RNA from the data set of native structures. *Journal of molecular biology*, 347(1): 53-69.

Ding S, Su C, & Yu J (2011). An optimizing BP neural network algorithm based on genetic algorithm. *Artificial intelligence review*, 36: 153-162.

Dirheimer G, Keith G, Dumas P, & Westhof E (1994). Primary, secondary, and tertiary structures of tRNAs. tRNA: Structure, biosynthesis, and function: 93-126.

Duarte D, & Ståhl N (2019). Machine learning: a concise overview. *Data Science in Practice:* 27-58.

Fan M, Andrade G F, & Brolo A G (2011). A review on the fabrication of substrates for surface enhanced Raman spectroscopy and their applications in analytical chemistry. *Analytica chimica acta*, 693(1-2): 7-25.

Fan X, Wang Y, Yu C, Lv Y, Zhang H, Yang Q, Wen M, Lu H, & Zhang Z (2023). A Universal and Accurate Method for Easily Identifying Components in Raman Spectroscopy Based on Deep Learning. *Analytical Chemistry*, 95(11): 4863-4870.

Fohrer J, Hennig M, & Carlomagno T (2006). Influence of the 2'-hydroxyl group conformation on the stability of A-form helices in RNA. *Journal of Molecular Biology*, 356(2): 280-287. https://doi.org/10.1016/j.jmb.2005.11.043

Garcia-Rico E, Alvarez-Puebla R A, & Guerrini L (2018). Direct surface-enhanced Raman scattering (SERS) spectroscopy of nucleic acids: From fundamental studies to real-life applications. *Chemical Society Reviews*, 47(13): 4909-4923.

Gautam R, Vanga S, Ariese F, & Umapathy S (2015). Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation*, 2: 1-38.

Guo S, Popp J, & Bocklitz T (2021). Chemometric analysis in Raman spectroscopy from experimental design to machine learning–based modeling. *Nature protocols*, 16(12): 5426-5459.

Gutell R R, Cannone J J, Konings D, & Gautheret D (2000). Predicting U-turns in ribosomal RNA with comparative sequence analysis. *Journal of Molecular Biology*, 300(4): 791-803.

Hajdin C E, Bellaousov S, Huggins W, Leonard C W, Mathews D H, & Weeks K M (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proceedings of the National Academy of Sciences*, 110(14): 5498-5503.

Halder S, & Bhattacharyya D (2013). RNA structure and dynamics: a base pairing perspective. *Progress in Biophysics and Molecular Biology*, 113(2): 264-283.

Hammond N B, Tolbert B S, Kierzek R, Turner D H, & Kennedy S D (2010). RNA Internal loops with tandem AG pairs: the structure of the 5′ G AG U/3′ U GA G loop can be dramatically different from others, including 5′ A AG U/3′ U GA A. *Biochemistry*, 49(27): 5817-5827.

Han R, Ketkaew R, & Luber S (2022). A concise review on recent developments of machine learning for the prediction of vibrational spectra. *The Journal of Physical Chemistry A*, 126(6): 801-812.

Hendrix D K, Brenner S E, & Holbrook S R (2005). RNA structural motifs: building blocks of a modular biomolecule. *Quarterly reviews of biophysics*, 38(3): 221-243.

Hernández B, Baumruk V, Leulliot N, Gouyette C, Huynh-Dinh T, & Ghomi M (2003). Thermodynamic

and structural features of ultrastable DNA and RNA hairpins. J*ournal of molecular structure*, 651: 67-74.

Hobro A J, Rouhi M, Blanch E W, & Conn G L (2007). Raman and Raman optical activity (ROA) analysis of RNA structural motifs in Domain I of the EMCV IRES. *Nucleic acids research*, 35(4): 1169-1177.

Hobro A J, Rouhi M, Conn G L, & Blanch E W (2008). Raman and Raman optical activity (ROA) analysis of RNA structural motifs. *Vibrational Spectroscopy*, 48(1): 37-43. https://doi.org/10.1016/j.vibspec.2007.11.007

Holbrook S R (2005). RNA structure: the long and the short of it. *Current opinion in structural biology*, 15(3): 302-308.

Jia M, Luo L, & Liu C (2004). Statistical correlation between protein secondary structure and messenger RNA stem-loop structure. *Biopolymers: Original Research on Biomolecules*, 73(1): 16-26.

Kim Y, & Lee W (2022). Distributed Raman spectrum data augmentation system using federated learning with deep generative models. *Sensors*, 22(24): 9900.

Kinjo A R, & Nakamura H (2012). Composite structural motifs of binding sites for delineating biological functions of proteins. *PloS one*, 7(2): e31437.

Korennykh A V, Plantinga M J, Correll C C, & Piccirilli J A (2007). <Linkage between Substrate Recognition and Catalysis during Cleavage of Sarcin: Ricin Loop RNA by Restrictocin. *Biochemistry*, 46: 12744-12756.

Krasilnikov A S (2003). On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*, 9(6): 640-643. https://doi.org/10.1261/rna.2202703

Kusters R, Misevic D, Berry H, Cully A, Le Cunff Y, Dandoy L, Díaz-Rodríguez N, Ficher M, Grizou J, & Othmani A (2020). Interdisciplinary research in artificial intelligence: challenges and opportunities. *Frontiers in big data*, 3: 577974.

Leardi R (2002). Chemometrics: From classical to genetic algorithms. *Grasas y Aceites*, 53(1): 115-127.

Lee J C, Cannone J J, & Gutell R R (2003). <The Lonepair Triloop- A New Motif in RNA Structure. *Journal of Molecular Biology*, 325. https://doi.org/10.1016/S0022-2836 (02) 01106-3

Lee J C, Gutell R R, & Russell R (2006). The UAA/GAN internal loop motif: a new RNA structural element that forms a cross-strand AAA stack and long-range tertiary interactions. *Journal of molecular biology*, 360(5): 978-988.

Lemieux S, & Major F (2002). RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic acids research*, 30(19): 4250-4263.

Leulliot N, Abdelkafi M, Turpin P-Y, Ghomi M, Baumruk V, Namane A, Gouyette C, & Huynh-Dinh T (1999). Unusual nucleotide conformations in GNRA and UNCG type tetraloop hairpins: evidence from Raman markers assignments. *Nucleic acids research*, 27(5): 1398-1404.

Leulliot N, Baumruk V, Gouyette C, Huynh-Dinh T, Turpin P-Y, & Ghomi M (1999). Aqueous phase structural features of GNRA tetraloops formed in short hairpins as evidenced by UV absorption and Raman spectroscopy. *Vibrational spectroscopy*, 19(2): 335-340.

Lewis T G, & Denning P J (2018). Learning machine learning. *Communications of the ACM*, 61(12): 24-27.

Li-Chan E C (1996). The applications of Raman spectroscopy in food science. *Trends in Food Science & Technology*, 11(7): 361-370.

Luo R, Popp J, & Bocklitz T (2022). Deep learning for Raman spectroscopy: A review. *Analytica*, 3(3): 287-301.

Madey T E, & Yates Jr J T. (2013). *Vibrational spectroscopy of molecules on surfaces* (Vol. 1). Springer Science & Business Media.

Madison J (1968). Primary structure of RNA. *Annual review of biochemistry*, 37(1): 131-148.

Minchin S, & Lodge J (2019). Understanding biochemistry: structure and function of nucleic acids. Essays Biochemestry, 63(4): 433-456. https://doi.org/10.1042/ebc20180038

Moore P B (1999). Structural motifs in RNA. Annual. Review of Biochemistry 68: 287-300.

Movasaghi Z, Rehman S, & Rehman I U (2007). Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 42(5): 493-541.

Nicolas J C, Levine A J, & Moreau J (2002). <tRNA Structure Goes from L to λ. *Oncogene*, 14: 1427-1433.

Orlando A, Franceschini F, Muscas C, Pidkova S, Bartoli M, Rovere M, & Tagliaferro A (2021). A comprehensive review on Raman spectroscopy applications. *Chemosensors*, 9(9): 262.

Pelletier M J (2003). Quantitative analysis using Raman spectrometry. *Applied spectroscopy*, 57(1): 20A-42A.

Qi Y, Hu D, Jiang Y, Wu Z, Zheng M, Chen E X, Liang Y, Sadi M A, Zhang K, & Chen Y P (2023). Recent progresses in machine learning assisted Raman spectroscopy. *Advanced Optical Materials*, 11(14): 2203104.

Rocha W F d C, Prado C B d, & Blonder N (2020). Comparison of chemometric problems in food analysis using non-linear methods. *Molecules*, 25(13): 3025.

Rostron P, Gaber S, & Gaber D (2016). Raman spectroscopy, review. laser 21: 24.

Salim N, Lamichhane R, Zhao R, Banerjee T, Philip J, Rueda D, & Feig A L (2012). Thermodynamic and kinetic analysis of an RNA kissing interaction and its resolution into an extended duplex. *Biophysical journal*, 102(5): 1097-1107. https://doi.org/10.1016/j.bpj.2011.12.052

Schroeder K T, McPhee S A, Ouellet J, & Lilley D M (2010). A structural database for k-turn motifs in RNA. *RNA*, 16(8): 1463-1468. https://doi.org/10.1261/rna.2207910

Schroeder S J, & Turner D H (2000). Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry*, 39(31): 9257-9274.

Sharma P, Chawla M, Sharma S, & Mitra A (2010). On the role of Hoogsteen: Hoogsteen interactions in RNA: Ab initio investigations of structures and energies. *RNA*, 16(5): 942-957.

Shorten C, & Khoshgoftaar T M (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1): 1-48.

Staple D W, & Butcher S E (2005). Pseudoknots: RNA structures with diverse functions. *PLoS biology*, 3(6): e213.

Svoboda P, & Cara A D (2006). Hairpin RNA: a secondary structure of primary importance. *Cellular and Molecular Life Sciences CMLS*, 63: 901-908.

Thapar R, Denmon A P, & Nikonowicz E P (2014). Recognition modes of RNA tetraloops and tetraloop-like motifs by RNA-binding proteins. Wiley Interdisciplinary Reviews: *RNA*, 5(1): 49-67.

Theimer C A, Blois C A, & Feigon J (2005). Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Molecular cell*, 17(5): 671-682.

Tyagi R, & Mathews D H (2007). Predicting helical coaxial stacking in RNA

multibranch loops. RNA 13(7): 939-951. https://doi.org/10.1261/ rna. 305307

Westhof E, & Auffinger P (2000). RNA tertiary structure. *Encyclopedia of analytical chemistry:* 5222-5232.

Westhof E, & Auffinger P (2000). <RNA Tertiary Structure. *Encyclopedia of Analytical Chemistry:* 10.

Westhof E, & Fritsch V (2000). <RNA folding- beyond Watson–Crick pairs. *Structure*, 8: 55-65.

Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, Liu X, Wu Y, Dong F, & Qiu C-W (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation* 2(4).

Zhang M, Perelson A S, & Tung C-S (2011). RNA Structural Motifs. https://doi.org/10.1002/9780470015902.a0003132.pub2

Zhang R, Xie H, Cai S, Hu Y, Liu G k, Hong W, & Tian Z q (2020). Transfer-learning-based Raman spectra identification. *Journal of Raman Spectroscopy*, 51(1): 176-186.

Zhong C, & Zhang S (2012). Clustering RNA structural motifs in ribosomal RNAs using secondary structural alignment .*Nucleic Acids Research*, 40(3): 1307-1317. https://doi.org/10.1093/nar/gkr804