Role of Artificial Intelligence and Vibrational Spectroscopy in Cancer Diagnostics

Ihtesham U Rehman[1*], Rabia Sanam Khan[1] and Shazza Rehman[2]

1. Bioengineering | Engineering Department, Faculty of Science and Technology, Lancaster University, Lancaster (UK)

2. Department of Medical Oncology, Airedale NHS Foundation Trust, Airedale General Hospital, Steeton, West Yorkshire, (UK)

*Corresponding author: Email: i.u.rehman@lancaster.ac.uk

Role of Artificial Intelligence and Vibrational Spectroscopy in Cancer Diagnostics

Abstract:

Raman and Infrared spectroscopic techniques are being used for the analysis of different types of cancers and other biological molecules. It is possible to identify cancers from normal tissues both in fresh and fixed tissues. These techniques can be used not only for the early diagnosis of cancer but also for monitoring the progression of the disease. Furthermore, chemical pathways to the progression of the disease process can be understood and followed. More recently, Artificial Intelligence (AI), Neural Network (NN) and Machine Learning are being combined with spectroscopy, which is making it easier to understand the chemical structural details of cancers and biological molecules more precisely and accurately. In this report, these aspects are being outlined by using breast cancer as a specific example.

Article highlights:

1- Analysis of breast cancer using Raman spectroscopy
2- Principal Component Analysis of spectral data of breast cancer
3- Cluster analysis of normal and cancerous normal breast tissue
4- Role of artificial intelligence and machine learning.
5- Early diagnosis and monitoring of the progression of cancer.

Introduction:

Tumorigenesis is a complex dynamic process accompanied by changes in the chemistry and biochemistry of cells, tissue and the microenvironment [1]. If these changes can be monitored and quantified over time before morphological or systemic expression of the disease, diagnosis can be done at a much earlier stage, improving prognosis. The information required for this spans over a number of biological levels from molecular structure to tissue composition. Raman and infrared spectroscopic techniques collect highly specific information making these suitable for cell and tissue analysis. This approach may help not only in early diagnosis but also in characterizing the disease state at various points of its progression. This is critical in planning the therapeutic approach of individual patients. To understand chemical structural properties of cancer and associated changes with the progression of disease, it is important to understand the biology and chemistry of the cancer. [2]

Most cancers (approximately 90%) are carcinomas. This can be explained by two factors: cell proliferation mainly occurs in epithelia, and they are more frequently exposed to various forms of physical and chemical damage that favour the development of cancer [3]. Dedifferentiation or anaplasia denotes the loss of normal characteristics, which may be attributed to chemical structure and these chemical structures can be very well studied with

the use of spectroscopic techniques. These techniques can help in identifying and establishing a chemical pathway by which the cancer progresses. The biological pathway of cancer progression is very well established, but the chemical pathway is still a matter of debate. It is strongly believed that vibrational spectroscopy offers excellent potential to study the chemical structural characteristics of the biological tissues. This is only possible if the spectral bands are precisely and accurately assigned to specific chemical bonds and functional groups.

In recent years, applications of IR and Raman spectroscopy have increased a great deal, the application of these spectroscopic methods on biological studies, particularly on clinical studies related to malignancy, and cancer detection has attracted significant attention among both clinical and non-clinical researchers. These methods have been reported on several biological tissues including bone[4-6], skin[7-9], colon[10], lung[11,12], breast[13-18], heart[19,20], cornea[21,22], liver[23,24], prostate[25,26], gastric[27,28], larynx[29], oral[30,31], bladder[32], endometrium[33], multiple sclerosis [34] and brain [35].

Spectroscopic techniques have immense potential in the diagnosis of human diseases including infections, inflammatory conditions and various cancers. Currently, although imaging can aid in diagnosis of many disease processes, histopathological tissue diagnosis remains the gold standard for diagnosis of most clinical conditions. In this respect, spectroscopic techniques can be of immense value to aid histopathological diagnosis.

The spectroscopic measurements can often be carried out with minimal sample preparation or need for consumables. Unlike IR spectroscopy, the material may absorb power from the laser or be induced to release part of its own power. Both cases are associated with a displacement between energy levels resulting in a shift in frequency between the incident and scattered radiation.

The spectra represent molecule specific peaks that are sensitive to composition, stress state, orientation and concentration. Namely, the spectra are a fingerprint of the material based on its chemical composition and structure. Its features are interpreted as a function of control variables; either light (amplitude, frequency, etc.) or of the sample (change of physical properties). The extraction of quantitative data from the spectra therefore requires a reference. Having an appropriate control is essential for the reliable discrimination of spectra.

In literature, significant information is available to clinicians and spectroscopists, including clinical, biological and biospectral chemical data and outputs of related spectral imaging. A combination of this data provides information that requires precise and accurate evaluation of pathology, chemical changes that may occur due to the progression of the disease process either for early diagnosis or monitoring of the disease process. To bring the entire process from "lab to patients" there is a need to understand the entire process in a rapid and accurate way that can be added to the armoury of pathologists and clinicians that will help in avoiding misdiagnosis. To achieve this, employment of artificial intelligence, computer aided diagnosis;

machine learning and artificial neural networks will play a pivotal role in the future. A combination of these methodologies will help in developing learning algorithms capable of analysis and interpretation of clinical data and in integrating them into categorized, defined outputs.

Spectroscopic evaluation of cancer progression:

It is widely believed that cancer progression can be modelled using 'somatic evolution theory' [36]. The theory describes the tissue and substrate constraints applicable to the landscape in which evolution takes place. The tissue constraints represent both intracellular and extracellular changes and control initial tumour growth. Once the tumour has grown to a certain critical size, progression is limited by substrate constraints. It is important to distinguish the two different development states as;

- the adaptive phase, during which tissue constraints may alone present identifiable markers in the spectra (early diagnosis)
- the active growth phase, where substrate constraints may provide a better measure of the 'stage' of cancer.

It is envisaged that mutations (intracellular changes) play a more permissive than a constructive role in progression, their mutual link to extracellular changes and substrate utilization mechanics establishes a positive feedback loop. This feedback loop has to be taken into account when limiting the number and type of constraints being measured and used in the spectra [36].

Actively targeting either constraint can be a problem as the evolutionary landscape of cancer is highly complex, dynamic and heterogeneous. The effects of cancer progression extend beyond the primary tumour boundary even before it reaches a metastatic state. However, due to this heterogeneity, the products of the cancer and cancer-associated cells (both signalling pathways and metabolites) and their substrate utilization rates may remain regional and occur in temporary bursts. It is clinically unfeasible to collect Raman spectra from a large number of focal planes or over an extended time period during diagnosis. This raises two important considerations for reliable diagnosis [37];

i.   How to select the location(s), size and number of focal spots for spectra collection that are truly representative of the cancer population
ii. Out of the variables of location and time, is it possible to fix one, i.e. are there any constraints that are time-independent

There are some key points in defining the corresponding functional groups of every spectral peak. These points can play an outstanding role in the process of characteristic peak analysis, and they are of an outstanding importance in the process of having a clear understanding from the spectroscopic technique. There is a considerable amount of studies on spectroscopy of cancer tissues and the use of this technique is growing rapidly. There is a growing need to identify key spectral peaks and to correctly assign them to their respective chemical

structures. Therefore, it is of utmost importance to have a trustable spectral database widely available to researchers working on vibrational spectroscopy of biological molecules. In addition, it is equally important to have an approach which provides quantitative analysis as rapidly and precisely as possible.

Analysing cancer with a combination of spectroscopy, machine learning and artificial intelligence:

Just to highlight, substantial changes in chemical structure that are caused by cancer and their different grades and subtypes using vibrational spectroscopy can be more precisely picked up with a combination of spectroscopy, machine learning and artificial intelligence. The spectra data on breast cancer given in figures 1 and 2 highlights the importance of FTIR and Raman spectroscopy, chemometrics analysis and their outstanding clinical potential. It is clear that more understanding is required to improve the technology and facilitate its clinical use and data processing. As stated earlier, the role of spectroscopy in analysis of biological molecules has increased significantly over the last decade and more and more clinical studies are being reported. To accomplish more clinical multidisciplinary collaboration between researchers, companies, and clinicians is crucial.

Analysing breast cancer with Spectroscopy and AI:

Both infrared and Raman spectroscopic techniques in combination with artificial intelligence is the future with early and precise diagnosis and monitor progression of disease processes and analysing biological molecules.

While analysing normal tissue and breast cancer tissue with Raman spectroscopy in combination with Principal component analysis applied to the general spectral region, a clear separation and grouping was achieved with the PC-1 and PC-2 accounting for 57% of the variance, as it can be seen in Figure 1. The loadings associated with these principal components identified the following peaks as responsible for the grouping and separation: 2884, 1583, 1450, 1336, 1245, 1159, 1122, and 1004 cm^{-1}.

The peak located at $\approx$2884 cm^{-1} has been assigned to the CH_2 and CH stretching of lipids and proteins and was identified to contribute more in the cancer area in comparison with normal breast. This difference in content might be caused by the accelerated and active cell component production, which is characteristic of cancer development and progression [38].

The peak at $\approx$1583 cm^{-1} has been associated with tryptophan and the C=C bending mode of phenylalanine[185] and presented greater intensities on the cancer samples. This component forms a part of proteins and its contribution might be associated with the structural components in highly dense samples. This behaviour was confirmed by the peak at $\approx$1450 cm^{-1} also presenting higher intensities than the normal counterpart and which is representative of the CH_2 bending characteristic of malignant tissues [38].

Furthermore, the peak at 1000-1004 cm^{-1} presented greater contributions in cancerous samples and was assigned to phenylalanine [38]. Different authors focus on this peak as a

collagen specific assignment. The literature has reported collagen fibre deposition, rearrangement of fibres and overproduction occurring when tumours develop and grow [39]. Consequently, an increase of intensity in peaks associated with stromal overproduction were expected as classifiers of cancerous samples. Nonetheless, stromal components were an important and dominant part of the normal breast sections and they could have been included in the Raman acquisition. Besides, a phenylalanine assignment can be correlated to proteins and structural components of the epithelial cells.

Similar trends were found when analysing normal and breast cancer samples with FTIR. When compared, the spectral profile of both tissue types presented characteristic bands associated with the presence of carbohydrates, lipids, nucleic acids and proteins as shown in Figure 2. Overall, the bands found in the cancerous area spectrum are shifted towards lower wavenumbers in comparison with the normal breast bands. This downshift has been reported in cancerous samples by different research groups as well [40,41].

Furthermore, differences in the intensity of peaks were identified between normal and cancerous tissues as seen in Figure 2. The cancerous area spectra had a higher overall intensity which suggested biochemical changes associated with the disease. The increased intensity in C-H bands can be considered as an indicator of an increase of cellular components such as lipids, nucleic acids and proteins. Likewise, the shifting and increased intensity in the 1700-1500 cm^{-1} area associated with nucleic acids and amide-stretching peaks might suggest changes in the conformational structure of proteins in the cancerous tissue [14].

When PCA was applied to the spectral general region, good separation was obtained with principal component 3 and 5, accounting for 15% of the variance as is seen in the score plot presented in Figure 2. In this figure, a clear grouping can be seen for the cancerous area and the normal breast. CA and NB samples are presented at opposite sides of both principal components axes. The peaks responsible for the separation of cancerous and normal tissue on the score plots are 3275, 2963, 2916, 2849, 1669,1633, and 1530, cm^{-1}.

While applying the PCA model, created by using the whole spectral range and validated using Linear discriminant analysis (LDA), a sensitivity of 92% and specificity of 86% were each confirmed. In addition, cluster analysis was performed for the cancerous area and normal breast samples. The analysis was set up to recognize 10 clusters to facilitate separation as seen in Figure 3. The fourth cluster presented in plot D is formed by most of the normal breast samples. However, some of the normal samples are located on B, E, and H. The infiltration of normal samples into a cluster mainly formed by cancer samples confirmed the heterogeneity of our cohort. The cancerous samples analysed with this method represented different grades and subtypes, and as a result, present different characteristics that might resemble the normal tissue. Therefore, the clustering of both type of samples.

Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra. In this study, Deep learning methods for the prediction of molecular excitation spectra are presented. For the example of the electronic density of states of 132k organic molecules, three different neural network architectures: multilayer perceptron (MLP), convolutional neural network (CNN), and deep tensor neural network (DTNN) are trained and assessed. (48)

Support vector machines (SVM) in near infrared (NIR) spectroscopy have been employed to study model complex non-linear boundaries through the use of adapted kernel functions. They have been introduced in chemometrics and have proven to be powerful in NIR spectra classification. In this study, they highlighted one of the major drawbacks of SVM, which is that training the model requires optimization of the regularization and kernel meta-parameters in order to control the risk of overfitting and the complexity of the boundary. They proposed a methodological approach to guide the choice of the SVM parameters based on a grid search for minimizing the classification error rate but also relying on the visualization of the number of support vectors (SVs). Their optimized SVM models were quite parsimonious, relying on 8 and 35 support vectors respectively, and good classification performances were obtained (classification rate of 98.9% and 91% on the test sets, respectively). (49)

Applications of THz laser spectroscopy and machine learning for medical diagnostics have been reported which allows analysing molecular rotations associated with hydrogen bond breaking. By using the optimization algorithm, they carried out the classification of THz spectra of the most informative areas obtained in-vivo from the lymphedema-affected leg tissue (result of surgery) and obtained from healthy leg tissue. The results adequately show the separation of lymphedema tissues from healthy tissues in the space of the principal components (50)

The use of a self-learning classifier would be an ideal next step for this research. In general, terms, in the training stage, a model is created containing several rules. These rules encompass relevant information that is characteristic of specific classes, for example, the relevant database of the normal and malignant breast tissue biochemical behaviour identified in this work. When new data is included in this model, it should evolve and produce more accurate classification rates, adapting and improving the robustness of the final model.

Additionally, once the new and improved algorithm has been improved, the new classification algorithm could be tested in different fixated samples. These systems will then provide the opportunity of immediate feedback to clinicians and pathologists, significantly improving patient care. This will help in increasing confidence among the clinicians and rapidly bringing these techniques to the armoury of surgeons, oncologists, and pathologists.

Conclusion:

To summarise, vibrational spectroscopy combined with data mining and machine learning has the potential to offer a real-time analysis in biological samples, including cancer, with excellent accuracy and is capable of maintaining this accuracy for early diagnosis and monitoring of the progression of disease. Moreover, FTIR and Raman spectroscopic techniques can provide practitioners and researchers with chemical information reflecting changes at tissue and cellular levels. In clinical practice, these changes influence the treatment strategies. In practical research, these insights help to understand carcinogenesis and invasive progressions, accelerating the development of new treatments.

Expert opinion

Employing different methodologies for precise and rapid analysis through big data management techniques and working in partnership with clinicians is a way forward for all new researchers exploring the use of vibrational spectroscopy to study cancer and other disease processes. These will help in understanding the chemical pathway to the progression of disease opening possibilities of early diagnosis and continuous monitoring. It would be interesting to carry these findings further and focus some research on improving the sensitivity and specificity of the prediction model.

This could be achieved using machine learning and taking advantage of the big cohort data. The use of artificial intelligence and machine learning can represent an important tool in the spectroscopic analysis of vibrational data. The combination of vibrational spectroscopy and these strategies and algorithms can offer an effective and low-cost solution to healthcare and diagnostics.

Machine learning approaches consist of a combination of several classification systems. The use of several classifiers allows an improvement in the accuracy of the classification. These algorithms are designed to maximise the number of accurate classifications. Ideally, artificial intelligence (AI) and machine learning algorithms should be able to adapt to new data without retraining. AI and machine learning have been widely used in different fields using vibrational spectroscopy data, for example, the analysis of ovarian and breast cancer tissue, cervical cytology, food analysis, drugs concentrations, and soil analysis [42 – 47] and this trend is increasing rapidly.

A pathway depicting the pathway to combine vibrational spectroscopy with AI and ML has been suggested in Figure 4, which has immense potential in predicting various stages of different disease processes, in particular, with cancer diagnosis, staging and designing treatment. This will result in improved patient care pathways. A wealth of data now exists, describing chemical structural changes associated with various human disease conditions, including cancers. There still is the unmet need for creation of a single comprehensive and standardised database of these chemical change spectra for future applications and reference from where both the AI and ML can play a pivotal role in clinical applications. In addition, machine learning and data mining approaches have been advancing rapidly in recent years, with exciting new applications and it is anticipated that in the near future these combined approaches will provide a number of solutions not only to tackle cancer but also to facilitate a clearer understanding of other disease processes and biomaterials in a way that allows tailoring drugs and materials for a number of specific clinical applications.
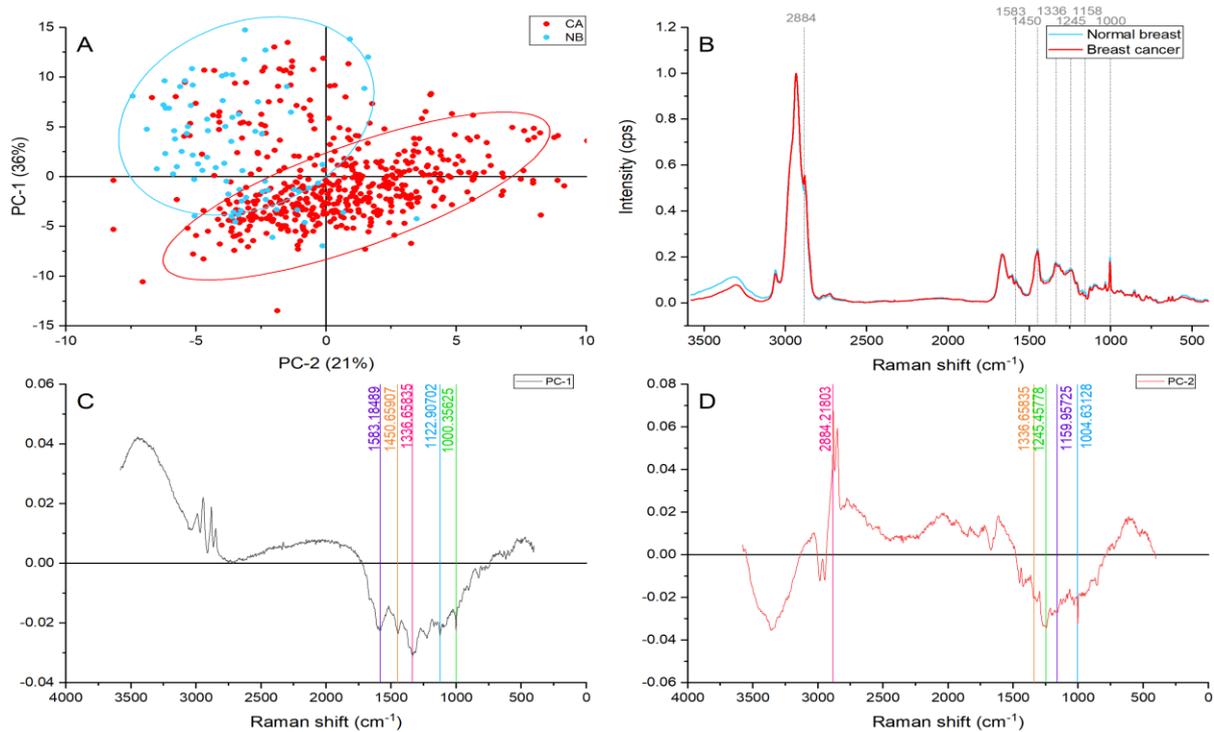
Acknowledgement:

Figure 1; Principal component analysis on the general region (4,000-400 cm^{-1}). (A) Score plot using PC-1 and PC-2 accounting for 57% of the variance, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 Loading, (D) PC-2 loading.
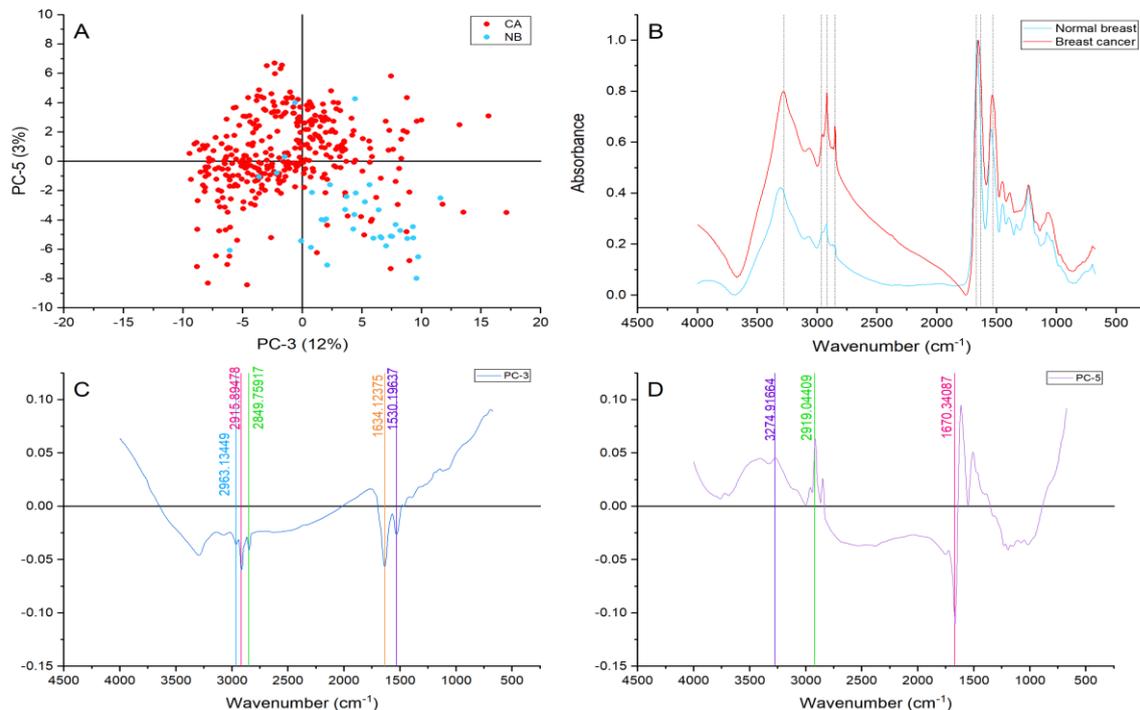

Figure 2; Principal component analysis of the general region (4,000-675 cm^{-1}). (A) Score plot using PC-3 and PC-5 accounting for 15% of the variance, (B) Average spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-5 loading.
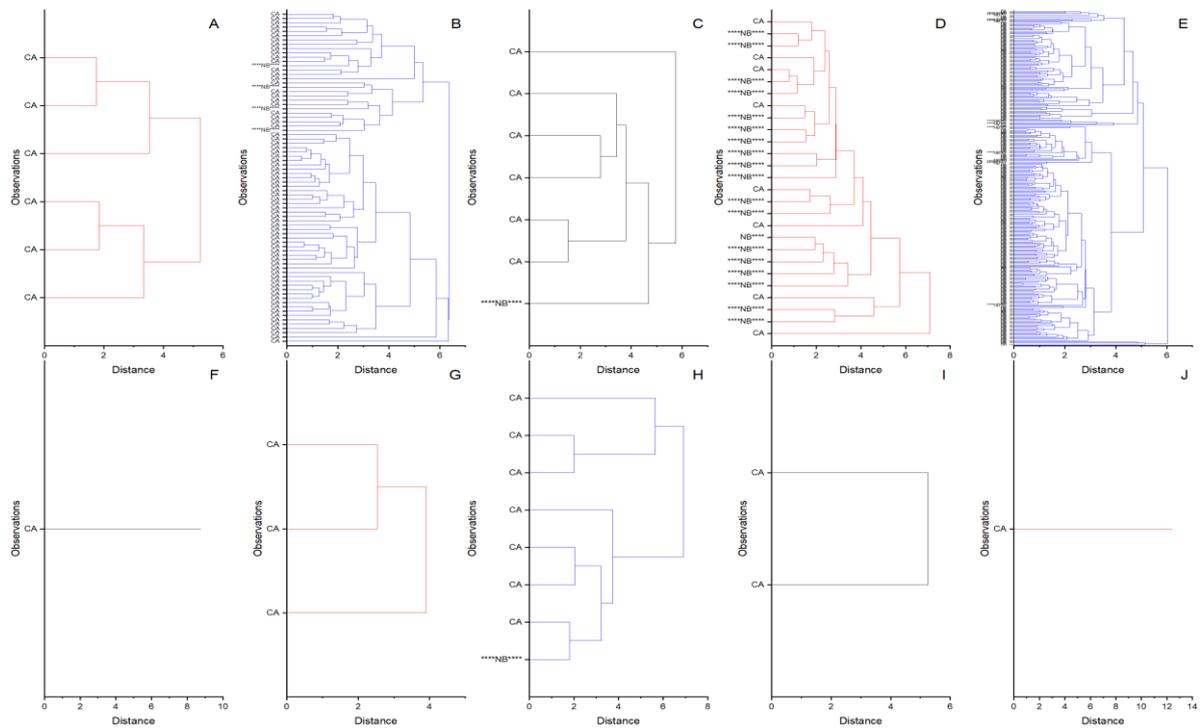
Figure 3; Cluster analysis of Cancerous area (CA) and Normal breast (****NB****). The analysis was set up to identify ten groups presented from A to J.
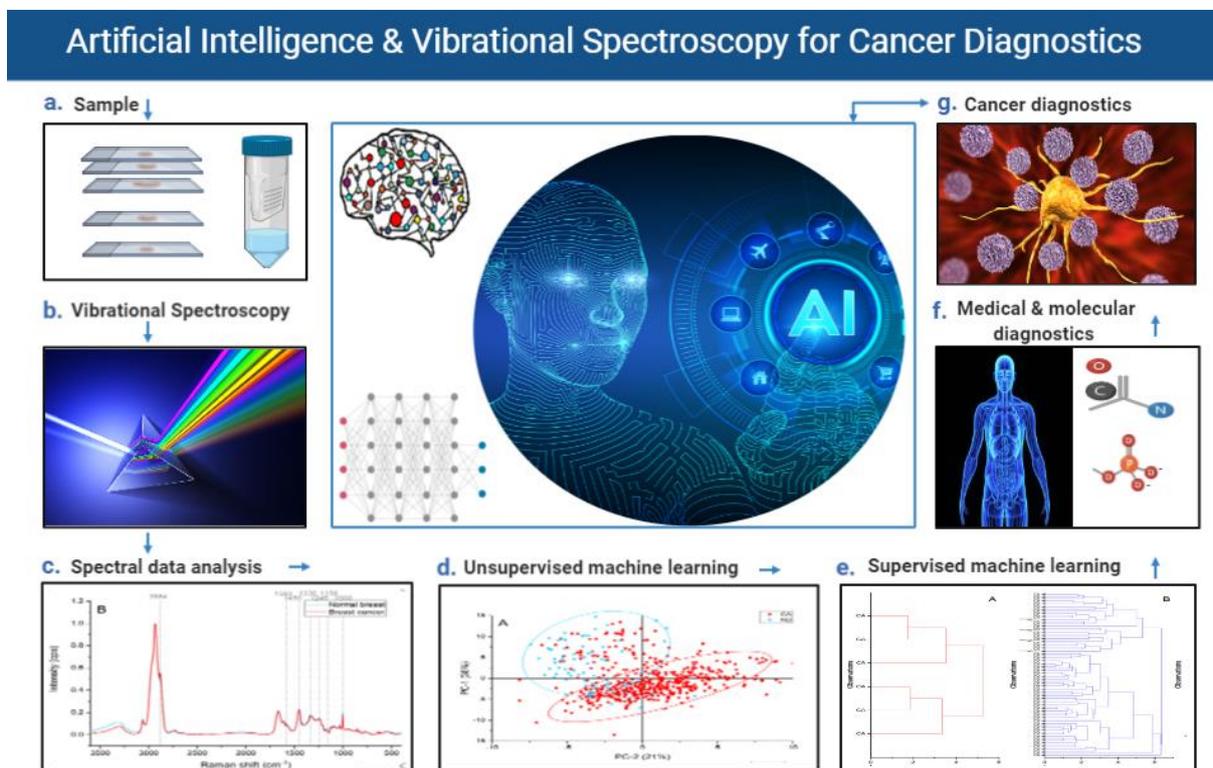

Figure 4; A pathway to cancer diagnosis by using vibrational spectroscopy and artificial intelligence.

References:

1. M. Wang, J. Zhao, L. Zhang, et al., Role of tumor microenvironment in tumorigenesis; J Cancer. 2017; 8(5): 761–773. doi: 10.7150/jca.17648
2. IU. Rehman, Z. Movasaghi, S. Rehman, Title: Vibrational Spectroscopy for Tissue Analysis, Series in Medical Physics and Biomedical Engineering, CRC Press; A Taylor and Francis Book, ISBN 978-1- 439-3608-8, October 2012
3. GM. Cooper and MA. Sunderland, The Cell: A Molecular Approach. 2nd edition, Sinauer Associates (2000), ISBN-10: 9780878931064
4. IU Rehman, R. Smith, L. L. Hench and W. Bonfield; Structural evaluation of human and sheep bone and comparison with synthetic hydroxyapatite by FT-Raman Spectroscopy, Journal of Biomedical Materials Research, Volume. 29(10), Pages 1287-1294. DOI: doi.org/10.1002/jbm.820291016 (1995)
5. R. Smith and IU. Rehman, Fourier Transform Raman Spectroscopic studies of Human Bone, Journal of Material. Science; Materials in Medicine, Volume. 5, Issue no. (9-10), Pages. 775-778, DOI: doi.org/10.1007/BF00120375   (1994)
6. AF Khan, MA Younas, AS Khan, S Tabassum, AA Chaudhry, IU Rehman, Raman Spectroscopy of Natural Bone and Synthetic Apatites; Applied Spectroscopy Reviews; Volume 48(4), Pages 329-355, DOI: doi.org/10.1080/05704928.2012.721107   (2013)
7. M. Larraona-Puy, A. Ghita, A. Zoladek, W. Perkins, S. Varma, I.H. Leach, A.A. Koloydenko, H. Williams, and I. Notingher, "Discrimination between Basal Cell Carcinoma and Hair Follicles in Skin Tissue Sections by Raman Micro-Spectroscopy," Journal of Molecular Structure. vol. 993, no. 1-3, pp. 57-61, (2011).
8. S. Sigurdsson, P.A. Philipsen, L.K. Hansen, J. Larsen, M. Gniadecka, and H.C. Wulf, "Detection of Skin Cancer by Classification of Raman Spectra," Ieee Transactions on Biomedical Engineering. vol. 51, no. 10, (2004).
9. WT. Cheng, MT. Liu, HN. Liu, and SY. Lin, "Micro-Raman Spectroscopy Used to Identify and Grade Human Skin Pilomatrixoma," Microscopy Research and Technique. vol. 68, no. 2, (2005).
10. W. Liu, Z. Sun, J. Chen and C. Jing, Raman Spectroscopy in Colorectal Cancer Diagnostics: Comparison of PCA-LDA and PLS-DA Models, Journal of Spectroscopy, Volume 2016, Article ID 1603609, 6 pages ,doi.org/10.1155/2016/1603609 (2016).
11. Y. Oshima, H. Shinzawa, T. Takenaka, C. Furihata, and H. Sato, "Discrimination Analysis of Human Lung Cancer Cells Associated with Histological Type and Malignancy Using Raman Spectroscopy," Journal of Biomedical Optics. vol. 15, no. 1, (2010).
12. X.Z. Li, T.Y. Yang, and J.H. Ding, Surface Enhanced Raman Spectroscopy (Sers) of Saliva for the Diagnosis of Lung Cancer, Spectroscopy and Spectral Analysis. vol. 32, no. 2, pp. 391-393. (2012).
13. ACS. Talari, CA. Evans, I. Holen, RE. Coleman and IU Rehman, Raman spectroscopic analysis differentiates between breast cancer cell lines. Journal of Raman Spectroscopy, 46(5), 421-427. doi:10.1002/jrs.4676 (2015).
14. S. Rehman, Z. Movasaghi, JA. Darr and IU Rehman, Fourier Transform Infrared Spectroscopic Analysis of Breast Cancer Tissues; Identifying Differences between Normal Breast, Invasive Ductal Carcinoma, and Ductal Carcinoma In Situ of the Breast; Applied Spectroscopy Reviews   Volume: 45   Issue: 5   Pages: 355-368; doi: 10.1080/05704928.2010.483674 (2010).
15. A.S. Haka, K.E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R.R. Dasari, and M.S. Feld, "Diagnosing Breast Cancer by Using Raman Spectroscopy," Proceedings of the National Academy of Sciences of the United States of America. vol. 102, no. 35, pp. 12371-12376 (2005).
16. A. Saha, I. Barman, N.C. Dingari, S. McGee, Z. Volynskaya, L.H. Galindo, W. Liu, D. Plecha, N. Klein, R.R. Dasari, and M. Fitzmaurice, "Raman Spectroscopy: A Real-Time Tool for Identifying Microcalcifications During Stereotactic Breast Core Needle Biopsies," Biomedical Optics Express. vol. 2, no. 10, pp. 2792-2803(2011).

17. A.S. Haka, K.E. Shafer, M. Fitzmaurice, R.R. Dasari, and M.S. Feld, "Distinguishing Type Ii Microcalcifications in Benign and Malignant Breast Lesions Using Raman Spectroscopy," Laboratory Investigation. vol. 82, no. 1, 36A-36A (2002).

18. G. Shetty, C. Kendall, N. Shepherd, N. Stone, and H. Barr, "Raman Spectroscopy: Elucidation of Biochemical Changes in Carcinogenesis of Oesophagus," British Journal of Cancer. vol. 94, no. 10, (2006).

19. M. Ogawa, Y. Harada, Y. Yamaoka, K. Fujita, H. Yaku, T. Takamatsu, Label-free biochemical imaging of heart tissue with high-speed spontaneous Raman microscopy, Biochem Biophys Res Commun.; 382(2):370-4. doi: 10.1016/j.bbrc.2009.03.028. Epub (2009).

20. A. Chaichi, A. Prasad, MR Gartia, Raman spectroscopy and microscopy applications in cardiovascular diseases: from molecules to organs, Biosensors (Basel) 8(4): 107, doi: 10.3390/bios8040107 (2018).

21. NJ. Bauer, F. Hendrikse, WF. March, In vivo confocal Raman spectroscopy of the human cornea, Cornea, 18(4):483-8. (1999)

22. R. Erckens, F. Jongsma, J. Wicksted, F. Hendrikse, W. March, M. Motamedi, Raman Spectroscopy in Ophthalmology: From Experimental Tool to Applications In Vivo. Lasers in medical science. 16. 236-52. Doi: 10.1007/PL00011360 (2001)

23. D. Sheng *et al.*, "A study of structural differences between liver cancer cells and normal liver cells using FTIR spectroscopy," *J. Mol. Struct.*, vol. 1099, pp. 18–23, (2015).

24. P. R. Palaniappan and V. Vijayasundaram, FTIR study of arsenic induced biochemical changes on the liver tissues of fresh water fingerlings Labeo rohita, *Rom J Biophys*, vol. 18, pp. 135-144, (2008).

25. Z. Cheng, C. Namhyun, R. Wang, S. Lee, K. Moon, SY. Yoon, L. Chen, J. Choo, Simultaneous Detection of Dual Prostate Specific Antigens Using Surface-Enhanced Raman Scattering-Based Immunoassay for Accurate Diagnosis of Prostate Cancer, ACS Nano, 11 (5), 4926-4933 DOI: 10.1021/acsnano.7b01536 (2017).

26. RE. Kast, SC. Tucker, K. Killian, M, Trexler, KV. Honn, GW. Auner, Emerging technology: applications of Raman spectroscopy for prostate cancer, Cancer and Metastasis Reviews, Volume 33, Issue 2–3, pp 673–693, doi.org/10.1007/s10555-013-9489-6 (2014).

27. S. Feng, J. Pan, Y. Wu, D. Lin, Y. Chen, G. Xi, J. Lin, and R. Chen, "Study on Gastric Cancer Blood Plasma Based on Surface-Enhanced Raman Spectroscopy Combined with Multivariate Analysis," Science China-Life Sciences. vol. 54, no. 9, pp. 828-834 (2011).

28. 16. T. Kawabata, H. Kikuchi, S. Okazaki, M. Yamamoto, Y. Hiramatsu, J. Yang, M. Baba, M. Ohta, K. Kamiya, T. Tanaka, and H. Konno, "Near-Infrared Multichannel Raman Spectroscopy with a 1064 Nm Excitation Wavelength for Ex Vivo Diagnosis of Gastric Cancer," Journal of Surgical Research. vol. 169, no. 2, pp. E137-E143 (2011).

29. MS. Bergholt; K. Lin; W. Zheng; Z. Huang; D. Lau, In vivo, real-time, transnasal, image-guided Raman endoscopy: defining spectral properties in the nasopharynx and larynx, J. of Biomedical Optics, 17(7), doi.org/10.1117/1.JBO.17.7.077002 (2012).

30. SA. Mian, C. Yorucu, MS. Ullah, IU Rehman, Raman spectroscopy can discriminate between normal, dysplastic and cancerous oral mucosa: a tissue-engineering approach. Journal of Tissue Engineering and Regenerative Medicine. doi:10.1002/term.2234 (2016).

31. SA. Mian, H. Colley, MH. Thornhill, IU Rehman, Development of a dewaxing protocol for tissue-engineered models of the oral mucosa used for Raman spectroscopic analysis. Applied Spectroscopy Reviews, 49(8), 614-617. doi:10.1080/05704928.2014.882348 (2014).

32. A. Shapiro, O.N. Gofrit, G. Pizov, J.K. Cohen, and J. Maier, "Raman Molecular Imaging: A Novel Spectroscopic Technique for Diagnosis of Bladder Cancer in Urine Specimens," European Urology. vol. 59, no. 1, pp. 106-112. (2011)

33. M.F.K. Fung, M.K. Senterman, N.Z. Mikhael, S. Lacelle, and P.T.T. Wong, "Pressure-Tuning Fourier Transform Infrared Spectroscopic Study of Carcinogenesis in Human Endometrium," Biospectroscopy. vol. 2, no. 3, (1996).

34. Ramos, I. R., Lyng, F. M., **IU Rehman**, Sharrack, B., & Woodroofe, M. N. The use of vibrational spectroscopy to study the pathogenesis multiple sclerosis and other neurological conditions. Applied Spectroscopy Reviews, 52(10), 868-882. doi: 10.1080/05704928.2017.1336450 (2017).

35. M. Kirsch, G. Schackert, R. Salzer, and C. Krafft, "Raman Spectroscopic Imaging for in Vivo Detection of Cerebral Brain Metastases," Analytical and Bioanalytical Chemistry. vol. 398, no. 4, pp. 1707-1713 [2010].

36. R. Gatenby and TL. Vincent, An Evolutionary Model of Carcinogenesis, Cancer Research (63) (19) 6212-6220; (2003)

37. N. McGranahan, C. Swanton, Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future, Cell, Volume 168, Issue 4(9), Pages 613-628 (2017).

38. ACS. Talari, Z. Movasaghi, S. Rehman, IU Rehman, Raman Spectroscopy of Biological Tissues. Applied Spectroscopy Reviews, 50(1), 46-111. doi:10.1080/05704928.2014.923902 (2015).

39. M. W. Conklin *et al.*, "Aligned collagen is a prognostic signature for survival in human breast carcinoma," *Am. J. Pathol.*, vol. 178, no. 3, pp. 1221–1232, (2011).

40. R. Eckel, H. Huo, H.-W. Guan, X. Hu, X. Che, and W.-D. Huang, "Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue," *Vib. Spectrosc.*, vol. 27, no. 2, pp. 165–173, 2001.

41. T. Gao, J. Feng, and Y. Ci, "Human breast carcinomal tissues display distinctive FTIR spectra: implication for the histological characterization of carcinomas," (1999).

42. M. Sattlecker, R. Baker, N. Stone, and C. Bessant, "Support vector machine ensembles for breast cancer type prediction from mid-FTIR micro-calcification spectra," *Chemom. Intell. Lab. Syst.*, vol. 107, no. 2, pp. 363–370, (2011).

43. K. Gajjar et al., "Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer," Analyst, vol. 138, no. 14, p. 3917, (2013).

44. DI. Ellis, D. Broadhurst, S. J. Clarke, and R. Goodacre, "Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning," Analyst, vol. 130, no. 12, p. 1648, (2005).

45. A. Morellos et al., "Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy," Biosyst. Eng., vol. 152, pp. 104–116, (2016).

46. C. Cheng, J. Liu, C. Zhang, M. Cai, H. Wang, and W. Xiong, "An overview of infrared spectroscopy based on continuous wavelet transform combined with machine learning algorithms: application to Chinese medicines, plant classification, and cancer diagnosis," Applied. Spectroscopy Review, vol. 45, no. 2, pp. 148–164, (2010).

47. F. Amato, A. López, E. MaríaPeña-Méndez, P. Vaňhara, A. Hampl, J. Havel, Artificial neural networks in medical diagnosis, Journal of Applied Biomedicine, Volume 11, Issue 2, 2013, Pages 47-58, https://doi.org/10.2478/v10136-012-0031-x.

48. Kunal Ghosh Annika Stuke Milica Todorović Peter Bjørn Jørgensen Mikkel N. Schmidt Aki Vehtari Patrick Rinke, Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra, Advanced Science, Volume6, Issue9, May 3, 2019, 1801367, https://doi.org/10.1002/advs.201801367.

49. Olivier Devosa, Cyril Ruckebuscha, Alexandra Durandab, Ludovic Duponchela, JeanPierre Huvennea, Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation, Chemometrics and Intelligent Laboratory Systems, Volume 96, Issue 1, 15 March 2009, Pages 27-33, https://doi.org/10.1016/j.chemolab.2008.11.005

50. Yu. V. Kistenev, A. V. Borisov, A.I.Knyazkova, E. A. Sandykova, V. V. Nikolaev, D. A. Vrazhnov, Applications of THz laser spectroscopy and machine learning for medical diagnostics, EPJ Web of Conferences 195, 10006 (2018), TERA-2018, https://doi.org/10.1051/epjconf/201819510006