A Review on Artificial Intelligence Applied On Raman Spectroscopy for Multitudinous Cancers

Pallavi Badarala [1], Dr.Bhaskararaju Vatchavai [2], Tejaswi Geddada [3*], Sri Vyshnavi Sudarsanam [4*], Lakshmi Sowjanya Rapolu [5*], Krishna Goutami Gudala [6*]

[1]*Assistant Professor, Sri Vasavi Institute of Pharmaceutical Sciences*

[2] *Principal, Sri Vasavi Institute of Pharmaceutical Sciences*

[3*, 4*, 5*,6*] *Students, Sri Vasavi Institute of Pharmaceutical Sciences*

Abstract- **Raman spectroscopy is a label-free optical method that has been used extensively for tumor diagnosis. Using conventional diagnostic techniques, tumors can be categorized as benign, malignant, or subtypes based on the various Raman technologies. First, the cells were evaluated structurally and spectrally. Following that, the cells' Raman data was efficiently examined utilizing a variety of machine learning methods, such as neural networks and multivariate models, which were all assessed simultaneously. The reports unequivocally demonstrate the effectiveness of AI assisted Raman spectroscopy for neoplasm cell classification and prediction, with an accuracy of correct predictions on a single spectrum. A molecular probe previously reported the ability of Raman spectroscopy to discriminate between tumor and healthy tissue. Present a quick, quantitative, probabilistic tumor assessment that incorporates real-time error analysis.**

Key words: Artificial Intelligence, Diagnosis, Machine Learning, Raman spectroscopy, Tumor

1.HISTORY

In 1923, Smekal established the theoretical prediction of Raman scattering, which was empirically confirmed in 1928 by two separate sets of researchers: Landsberg and Mandelstam in Russia and Raman and Krishnan in India. Regretfully, this phenomenon was only recognized as having been discovered by the first group. This aspect bears Raman's name, who won the 1930 Nobel Prize in Physics for his achievements. However, the first practical Raman spectrometer using mercury as the radiation source at 435.8 nanometers was not built until 1953, after the monochromatic was discovered. Thus, more than 40,000 compounds could be identified with this design prior to the 1960s. However, the application was challenging due to the source employed, particularly for colorful and photosensitive materials that absorbed this radiation aggressively. With the invention of the laser in 1960 and its application as a light source in Raman spectrometers, created by Brazilian scientists Sergio Porto and Wood in 1962, this resulted in yet another important breakthrough in this type of spectroscopy. This made the application of this kind of spectroscopy simpler and quicker.

As mentioned before, it was first suggested in 1966 to link a microscope to a Raman spectrometer. Nevertheless, Delahey and Dhamelin court did not develop this technology until 1974. Such a novel technique allowed surface mapping with Raman spectroscopy. Using charge-coupled device (CCD) detectors was another important development in 1987.

Using CCD detectors has improved the sensitivity of Raman and increased its possible applications by addressing the signal's inherent weakness from Raman scattering. The 1950s saw the development of a number of Raman spectroscopy variations, including coherent anti-Stokes Raman spectroscopy (CARS), surface enhanced Raman spectroscopy (SERS), resonance Raman spectroscopy (RRS), tip enhanced Raman spectroscopy (TERS), and stimulated Raman spectroscopy (SRS), which significantly improved speed and/or resolution when compared to spontaneous Raman (Mitsutake H, 2019)

II. INTRODUCTION

RAMAN SPECTROSCOPY -

The frequency shift of inelastic scattered light from a material is measured using a method known as Raman spectroscopy. This occurs when a photon of light strikes a molecule and produces a scattered shot. Either a photon with a lower frequency than the original photon known as Stokes Raman scattering or one with a higher frequency known as anti-Stokes Raman scattering will result in scattered light. In the latter case, when the photon is initially in an excited vibration state, it will receive energy from the molecule's connection. Raman's fundamental concept is to measure the energy shift of the outgoing photon. How the wavelength of the scattered light changes is determined by the chemical.

The displacement of the constituent atoms from their equilibrium positions due to molecular vibrations will cause the change in molecular polarizability, according to the Raman selection rule (Rostron P, 2016).

DISEASE:

"Disease" is a term used to describe a condition that disrupts an organism's normal physiological functions. Disease can be caused by a wide range of factors, including genetic variants, environmental factors, lifestyle decisions, and pathogens (such as bacteria, viruses, and parasites). Not only can illnesses have a major impact on an individual's physical health, but they can also have substantial social, psychological, and economical consequences (Brown JS, 2023).

CANCER:

Numerous tempo-spatial alterations in cell physiology are a part of the complicated disease known as cancer, which eventually results in malignant tumors. The disease's biological endpoint is abnormal cell proliferation, or neoplasia. Tumor cell invasion into adjacent tissues and distant organs is the primary cause of illness and death for most cancer patients. The metabolic process that transforms healthy cells into malignant ones has been the subject of extensive research in the biomedical sciences for many years. Finding a cure or long-term management strategies for metastatic cancer is still as challenging as it was when President Richard Nixon declared a war on cancer forty years ago, notwithstanding this research (Seyfried TN, 2010).

III. TYPES OF CANCER

BRAIN CANCER - Intracranial cancer is still challenging to identify and cure in its initial stages. The fact that several brain tumors are located next to or inside anatomical areas essential for fundamental motor, cognitive, reflexive, and other activities complicate this condition. However, because edema, or excess extracellular water, can accumulate around the tumor site, making precise tumor margin discriminating challenging, these imaging studies of malignant human brain tumors do not readily allow quantification of the actual tumor volume. The blood-brain barrier (BBB) also makes it inefficient to administer contrast chemicals (Koo YE, 2006).

BREAST CANCER - Breast cancer is now considered a serious illness in the modern world. The most common form of this type of breast cancer, ductal carcinoma, begins in the lining of the milk duct. Thin tubes carry milk from the lobules in the breast to the little nipple. Another type of breast cancer that begins in the lobules of the breast is called lobular carcinoma. Invasive breast cancer occurs when breast cancer spreads to the surrounding normal tissue from its original site in the breast ducts or lobules. Though it is uncommon in men, breast cancer can strike both sexes. The discovery of a new lump or alteration in the breast skin may be obscured by early burning indicators of breast cancer. These are the symptoms and indicators that help to detect breast cancer early. The patient will find it easier to spot any changes in her breasts if she does monthly breast self-examinations (Joshi J, 2014).

MYELOMA CANCER - When clonal plasma cells proliferate in the bone marrow (BM), monoclonal immune globulins are released, which is a hallmark of multiple myeloma (MM), a hematologic cancer .With an incidence of 4–6 per 100,000 year, MM is the second most common hematologic malignancy in the US .The response and survival rates of MM patients have significantly increased due to advancements in supportive care, autologous stem cell transplantation, and the introduction of novel drugs such as proteasome inhibitors and immune modulators. Complex and mutually reinforcing interactions between the MMSC and the surrounding bone marrow (BM) milieu sustain MMSC survival and self-renewal (Gao M, 2016).

LUNG CANCER - In the United States of America (USA), lung cancer is the top cause of mortality for both men and women. It is a common malignancy that is highly aggressive and spreads quickly. The regulatory circuits that regulate healthy cell division and homeostasis are flawed in lung cancer cells. A sequence of genetic and epigenetic changes is believed to initiate the transition from a normal to malignant lung cancer phenotype, which then progresses to invasive cancer through clonal growth .Compression of the mediastinal structures is always linked to progressive involvement of lymph nodes, which can result in tracheal compression, venous congestion and congestion related to collateral circulation, or esophageal compression and difficulties swallowing. Before a primary lung lesion is known to exist, there are indications of metastatic illness affecting distant locations such the liver, brain, or bone (Lemjabbar-Alaoui H, 2015).

LIVER CANCER -A leading cause of deaths from cancer The most common primary liver cancer worldwide and one that has a high death rate is Hepatocellular carcinoma (HCC).This implies that the premalignant environment of the liver depends on hepatic fibrosis. It is essential to appropriately diagnose and stage hepatocellular carcinoma in order to improve treatment outcomes and patient survival rates. It is challenging to diagnose HCC early, though, especially in people with chronic liver disease. An important aspect of hepatic fibrosis and liver cancer is their strong association; over 80% of hepatocellular carcinomas (HCCs) start in cirrhotic or fibrotic livers (Wang L, 2024).

The only one of the top five most fatal cancers to have an annual percentage rise in occurrence is liver cancer, which ranks fifth in the US and is the primary cause of cancer-related deaths globally (El-Serag HB, 2008).

THYROID CANCER – Thyroid cancer is the most common malignant neoplasia of the endocrine system, characterized by the malignant growth of thyroid gland cells. Over the past few decades, incidence rates have increased, mostly because to improvements in diagnostic procedures, and they frequently rank among the top 10 most common cancers, but they vary by country. Thyroid follicular epithelial cell-derived cancer commonly manifests in three main forms: follicular carcinoma (FC), follicular variant of papillary thyroid carcinoma (FV-PTC), and papillary thyroid carcinoma (PTC).Tumors with a follicular pattern may present diagnostic challenges during the histological evaluation of surgically removed thyroid glands because the presence of malignant features, such as capsular or vascular invasion, may not be sufficient. These cases may be categorized as follicular tumors of uncertain malignant potential (FT-UMP)8, which could cause a dubious assessment of the patient's risk (Gerecke C, 2016)

UTERINE CANCER - The development of the female reproductive system can be linked to the paramesonephric duct. The fallopian tubes are formed by bilateral paramesonephric ducts, which subsequently unite to produce the uterus and the top most part of the vaginal canal. A number of ligaments support the uterus, which gives it its anteverted and anteflexed posture in the pelvis. The embryo becomes a fetus in the uterus after embedding. Mother and fetal blood exchange nutrients and oxygen through the placenta, which is connected to the endometrium. When the endometrium is not pregnant, it sheds every month, resulting in cyclic menstrual bleeding.

The innermost layer is called the endometrium, followed by the intermediate layer called the myometrium, and the outermost layer is called the serosa . The myometrium and the endometrium can both give rise to uterine cancer. Although uncommon, uterine sarcomas, which originate from the middle layer of muscles, are frequently aggressive and require immediate detection and treatment. Endometrial carcinoma is the most prevalent gynecological cancer in affluent nations (Faizan U, 2023).

OVERIAN CANCER -In women worldwide, ovarian cancer ranks as the second most frequent cause of gynecologic cancer-related deaths. Over 180 thousand people died from ovarian cancer in 2018, out of the approximately 300,000 cases that were identified worldwide. The primary cause of this regrettable circumstance is the absence of certain symptoms in the early stages, which is evidently a contributing factor to the low survival rate. There is currently no reliable and sensitive biomarker for early detection, despite massive global efforts (Chen F, 2022).

IV. TYPES OF RAMAN SPECTROSCOPY:

Raman spectroscopy has significantly expanded its analytical capabilities in the last decade. Modern technology has made it feasible to buy affordable, portable Raman equipment for on-site analysis.

1.Surface Enhanced Raman Spectroscopy:

Highly sensitive structural detection of low concentration analytes is made possible by surface-enhanced Raman spectroscopy (SERS), one of the most sensitive instruments, which amplifies electromagnetic fields created by the excitation of localised surface Plasmon's (LSP) of adsorbate molecules on the roughened metal surface.

2.Surface enhanced hyper Raman scattering (HR):

Two photons disperse an inelastic sum of frequencies to generate surface enhanced hyper Raman scattering, whereas a single photon causes regular Raman scattering. The overall surface enhancement factors of SERS may be evaluated to be around 1014. Chemical electronic resonance might be the source of the amplification, where the chemical transition matches the cumulative frequency of the incoming photons.

3. Tip enhanced Raman spectroscopy (TERS):

The term "tip enhanced Raman spectroscopy" (TERS) refers to the combination of Raman spectroscopy and scanning probe microscopy. At the same time, TERS can provide spectral/chemical and topographical information via SPM and Raman, respectively.

4.Coherent anti – Stokes Raman Spectroscopy:

Coherence anti -stoke Raman spectroscopy, The Raman signal can be enhanced using a nonlinear Raman technique called Raman scattering spectroscopy, or CARS. It is known as an anti-stoke frequency technique because, as the name suggests, it generates a signal with a frequency higher than the excitation frequency using coherent laser beams.

5.Resonance Raman Spectroscopy (RRS):

The resonance When the energy of the incident light's photons is almost equivalent to the energy needed for an electronic transition, a device known as Raman spectroscopy measures the change in photon frequency. The resonant excitation may increase the electrons' oscillation charge displacement. As a result, the induced dipole moment will increase, directly increasing the efficiency of Raman scattering.

6.Confocal Raman microscopy:

It was in 1955 that Marvin Minsky invented confocal Raman microscopy. The probe head in confocal Raman microscopy directs laser light onto the material through the microscope objective. As a spatial filter, the pinhole will also refocus the backscattered Raman signal. The signal will then be collected by a charge-coupled device camera (detector) to produce a spectrum.

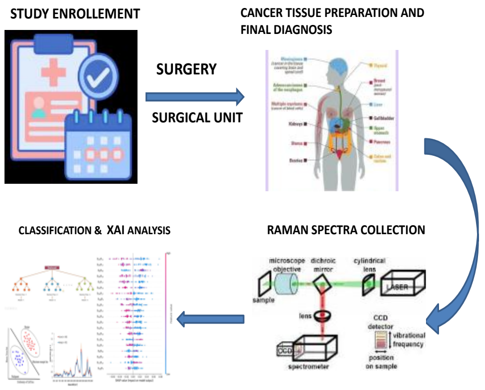
7.Raman imaging microscopy:

It is possible to produce a comprehensive chemical representation of the Raman spectrum using the Raman imaging approach, which was first developed by Delhaye and Dhamelincourt in 1975.Raman imaging generally combines digital imaging techniques with Raman spectroscopy. While the chemical composition, phase, structure, and crystal structure of the material determine the color pictures, the direct Raman imaging depends on the spatial coordinates and Raman intensity (Rostron P, 2016).

V. REVOLUTIONIZING RAMAN SPECTROSCOPY WITH AI :

Recently, Raman spectroscopy (RS) has gained a lot of attention due to several beneficial characteristics. Without the use of dyes or probes, this method allows for the examination of molecular and chemical changes in living tissues. Several components of the sample can be quantified using the same data set by using information-rich Raman spectra and chemical fingerprints. Inelastic light scattering, also known as Raman scattering, is the basis of the mechanism, which depends on the change in charge distribution that results in molecular polarizability. To find its composite wavelengths, Raman scattered light including stokes and anti-stokes that has passed through the optical filter is transferred to a diffraction grating. The position and intensity of each dispersed, composite wavelength are transformed into the Raman spectrum. The quality, quantity, symmetry, and orientation of the molecule are all revealed by these characteristic Raman peaks, which also reveal information on the biological components of the cell. Brain, ovarian, thyroid, myeloma, lung, breast, and uterine cancers are among the many tumors that have lately been successfully identified using Raman micro spectroscopy and machine learning techniques.

When combined with improved AI and ML, Raman spectroscopy has emerged as a potent tool for accurate disease prediction. In earlier research, we distinguished between normal and cancerous cell lines, hypoxia and proliferative regions, and even individual cells in 3D scaffolds using AI and ML-based systematic intelligent control-based techniques. These techniques, which use fuzzy control, genetic algorithms, PCA, LDA, and cluster analysis, have aided in the display of several aspects of single cell and tissue biopsies.



Moreover, Raman Spectroscopy revealed high-resolution spectrum features and information at the cellular level. Without the use of an external marker, the different cell lines were measured using RS in a potentially non-invasive manner. The spectra of the cell lines show that the amounts of biological materials such as proteins, lipids, and nucleic acids have altered. Because PCA could detect variations in the strength of the Raman peaks, the three cell lines were successfully and precisely differentiated. When predicting the LDA model of three cell lines, a 100% sensitivity and 91% specificity were attained. This method's specificity has been shown to be sensitive enough to detect biological components within cells.

Two important areas of artificial intelligence, machine learning (ML) and deep learning (DL), have seen a surge in attention in recent years due to the growing requirement to analyze and understand complex patterns within varied datasets (Conforti PM, 2024).

Raman spectral measurements:

Biopsies were used to produce high-quality Raman spectra using a 532 nm laser and a non-invasive dispersive micro-Raman instrument (Thermo Nicolet). The equipment was carefully calibrated using polystyrene as a reference before collecting spectral data from biopsies. By calculating its signal to noise ratio, the laser's sensitivity was 1000:1. Previous communications have demonstrated that biological materials are insensitive to the 10-mw laser light. A 50× long working distance 48 objective and a 50-micron pinhole spectrograph aperture were used. The exposure period for spectral collecting was set at 50 s with five exposures. 20 spectra were obtained for each cell line in the 400–3200 cm1 spectral region. Data collection was done by Themo Nicolet OMNICTM software (Talari, 2019).

Preparing the spectra and extracting features:
Every spectrum is interpolated to a grid of Raman shifts with uniform spacing. After normalizing each spectrum to have a sum of one (i.e., an area under the curve equal to one), cubic spline smoothing is applied. Each preprocessed spectrum is then subjected to peak detection, and the local maxima that emerge are gathered from each spectrum (Bellantuono L, 2023 Oct 3).

- SUPPORTING VECTOR MACHINE:

A supervised learning technique known as SVM. SVM's fundamental concept is the construction of a hyper plane that maximizes the margin, or distance, between it and each class's closest elements, also known as support vectors. The adjustment of a regularization term, represented by C, usually balances the trade-off between the tolerance for misclassification and the previously indicated margin size. In multiclass classification cases, the procedure can also be used by successively applying a one-on-one method to each class to find its respective category (Conforti PM, 2024).

- XG BOOST:

A popular and perfect machine learning technique is tree boosting. This paper Tree boosting is o the most well-liked and effective machine learning methods. The scalable from beginning to end tree boosting system XG Boost is described in this study. Data scientists utilize it widely to get cutting-edge results on a range of machine learning problems. For approximation tree learning, we offer a novel sparsity-aware method for sparse data with weighted quantile drawings. More importantly, we provide recommendations for sharding, cache access patterns, and data compression in order to build a scalable tree boosting system. Combining these discoveries allows XGBoost to scale to billions of samples while outperforming existing systems in terms of resource use (Sulistiani I, 2022).

- LINER DISCRIMINATION ANALYSIS:

To classify the many types of cancer, one artificial intelligence technique, LDA, was employed. In all, 120 spectra (30 spectra per subtype) were selected, and for LDA predictive classification, they were processed using baseline correction and normalization. The whole spectral range of 3200–600 cm inverse was covered by LDA models (Talari, 2019).

- RANDOM FOREST ALGORITHM:

Cancer medical case diagnoses are evaluated using this technique. The result of several decision trees and the properties of several eigen values may be combined by this technique to increase prediction accuracy. To obtain precise classification results, the random forest ensemble learning technique may be utilized to aggregate the output of several weak classifiers. This method offers a high prediction accuracy. In practical terms, it is crucial for supplemental medical diagnostics (Dai B, 2018).

- CONVOLUTIONAL NEURAL NETWORK:

For image identification applications, Convolutional Neural Networks (CNNs) are a popular machine learning classifier. Later CNN models are built on top of the old CNN models, which train supervised networks using the gradient-based back propagation technique. Standard back propagation is used to train convolutional filters, which capture significant features of the input data. One-dimensional convolutional neural networks, or 1D-CNNs, share a structure with conventional CNNs. The primary distinction is that 1D-CNN modifies the forward and back propagation equations during training by applying a one-dimensional filter to the convolutional layer and using one-dimensional data as input (Ma D, 2021).

VI. APPLICATIONS

Raman spectroscopy's excellent sensitivity, low detection dosage, and quick speed make it useful in many different fields.

1. Non polar bonds information received by Raman spectroscopy is obtained from symmetrical skeleton vibrations.

2. The ability to study xylene isomers and their mixtures both qualitatively and quantitatively has been shown now a days by Raman spectroscopy.

3. As different functional groups have different Raman spectra peaks, the researchers established a novel mixture analysis method based on a non-negative elastic net.

4. For medical evaluation and poisoning therapy, medication composition and quantitative analysis are crucial.

5. Raman spectroscopy has gained popularity as an analytical method in the pharmaceutical industry due to its ability to produce intricate chemical fingerprints. The findings demonstrate that ternary mixture analysis is able to recognize it

6. The findings show that, with timely medical intervention, ternary mixture analysis may identify the cause of death and detect quantitative cocaine toxicity.

7. The combination of machine learning and Raman spectroscopy offers some hope for the creation of a noninvasive, quick, affordable, and precise tool for general medical diagnosis.

8. As a proof-of-concept platform for medical diagnostics, SERS has been designed to identify proteins at the single molecular level and to check changes in the SERS spectrum using principal component analysis.

9. Microfluidic Raman biochip exosome detection has emerged as one of the most promising methods for prostate cancer diagnosis.

10. Multiple imaging permeabilities and molecular phenotypes of bladder tissue can be obtained using SERS (Liangrui p, 2021).

VII. ACKNOWLEDGEMENT

I would like to thank all of the people who helped us complete this study on Artificial intelligence applied on Raman spectroscopy for multitudinous cancers."They have been invaluable in their help and knowledge. Thanks to the resources provided by the library and my instructors, as well as the databases and research articles that improved my understanding, I am appreciative. I would like to thank my friends for their support, to sum up.

VIII. CONCLUSION

Raman spectroscopy combined with artificial intelligence (AI) is a game-changing method for multifaceted cancer diagnosis and treatment. The extremely specific and non-invasive method of

Raman spectroscopy offers comprehensive molecular fingerprints of living tissues. The technique's sensitivity, specificity, and predictive accuracy for cancer detection are improved when it is coupled with AI, especially ML and DL algorithms.

REFERENCES

[1] Bellantuono L, T. R. (2023 Oct 3, Oct 3). An eXplainable Artificial Intelligence analysis of Raman spectra for thyroid cancer diagnosis. Scientific Reports.

[2] Brown JS, A. S. (2023 , Nov 1). Updating the definition of cancer . Molecular Cancer Research, 21(11), 1142-47.

[3] Chen F, S. C. (2022, January 15). Screening ovarian cancers with Raman spectroscopy of blood plasma coupled with machine learning data processing. Spectrochimica Acta Part A: Molecular and Biomolecular spectroscopy, 265.

[4] Conforti PM, L. G. (2024, Apr). Raman spectroscopy and AI applications in cancer grading:An overview. IEEE Access, 12, 54816-54852.

[5] Dai B, C. R. (2018 , Dec 6). Using random forest algorithm for breast cancer diagnosis. IEEE, 449-452.

[6] El-Serag HB, M. J. (2008 , May 1). Diagnosis and treatment of hepatocellular carcinoma. Gastroenterology, 134(6), 1752-1763.

[7] Faizan U, M. V. (2023 , Feb 12). Uterine cance. InStatPearls [Internet].

[8] Gao M, K. Y. (2016, Jun 6). Multiple myeloma cancer stem cells. Oncotarget, 7(23), 35466-35477.

[9] Gerecke C, F. S.-H. (2016, Jul). The diagnosis and treatment of multiple myeloma. Deutsches Ärzteblatt International , 113(470), 27-28.

[10] Joshi J, D. R. (2014, Oct). Diagnosis and prognosis breast cancer using classification rules. International Journal of Engineering Research and General Science, 2(6), 315-323.

[11] Koo YE, R. G. (2006 , Dec 1). Brain cancer diagnosis and therapy with nanoplatforms. Advanced drug delivery reviews, 58(14), 1556-77.

[12] Lemjabbar-Alaoui H, H. O. (2015 , Dec 1). Lung cancer: Biology and treatment options. Biochimica et Biophysica Acta , 1856(2), 189-210.

[13] Liangrui p, P. Z. (2021, Jul 15). A review of artificial intelligence methods combined with Raman spectroscopy to identify the composition of substances. Journal of Raman spectroscopy, 53, 6-19.

[14] Ma D, S. L. (2021 , Jul 15). Classifying breast cancer tissue by Raman spectroscopy with one-dimensional convolutional neural network. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 256.

[15] Mitsutake H, P. R. (2019, Oct 24). Raman imaging spectroscopy: history, fundamentals and current scenario of the technique . Journal of the Brazilian Chemical Society, 30(11), 2243-2258.

[16] Rostron P, G. S. (2016, September). Raman spectroscopy, review. laser. International Journal of Engineering and Technical Research, 6(1).

[17] Seyfried TN, S. L. (2010 , Dec 7). Cancer as a metabolic disease. Nutrition & metabolism, 1-22.

[18] Sulistiani I, W. W. (2022 , Jul 27). Breast cancer prediction using random forest and Gaussian naïve Bayes algorithms. . IEEE, 170-175.

[19] Talari, A. C. (2019, sep 08). Advancing cancer diagnostics with artificial intelligence and spectroscopy: identifying chemical changes associated with breast cancer. Expert Review of Molecular Diagnostics, 19(10), 929–940.

[20] Wang L, F. M. (2024, september 3). Artificial intelligence techniques in liver cancer. Frontiers in Oncology, 14.