78531_Auto_Edited.docx

WORD COUNT 4474
TIME SUBMITTED 12-OCT-2022 07:46PM
PAPER ID 91486596

Name of Journal: *World Journal of Gastroenterology*

Manuscript NO: 78531

Manuscript Type: MINIREVIEWS

Vibrational spectroscopy – are we close to finding a solution for early pancreatic cancer diagnosis?

Solution for early pancreatic cancer diagnosis

Krzysztof Szymoński, Łukasz Chmura, Ewelina Lipiec, Dariusz Adamek

Abstract

Pancreatic cancer (PC) is an aggressive and lethal neoplasm, ranking seventh in the world for cancer deaths, with overall 5-year survival rates still below 10%. The knowledge about PC pathogenesis is rapidly expanding. New aspects of tumor biology, including its molecular and morphological heterogeneity, are being reported daily and explain the complicated "cross-talk" that occurs between the cancer cells and the tumor stroma or the nature of PC-associated neural remodeling (PANR). Nevertheless, currently, there are no specific and sensitive diagnosis options for PC. Vibrational spectroscopy (VS) shows promise in becoming a major player in the development of early diagnosis technology. In this review, we summarize recent reports about improvements in spectroscopic methodologies, briefly explain and highlight the drawbacks of each of them, and discuss available solutions. The important aspects of spectroscopic data evaluation with multivariate data analysis and a convolutional neural network methodology are depicted. We conclude by presenting a study design for systemic verification of the VS-based methods in the diagnosis of PC.

Key Words: spectroscopic cancer diagnosis; Raman spectroscopy; pancreatic cancer diagnosis; DNA methylation; liquid biopsy biomarkers; convolutional neural networks

Szymoński K, Chmura Ł, Lipiec E, Adamek D. Vibrational spectroscopy – are we close to finding a solution for early pancreatic cancer diagnosis? *World J Gastroenterol* 2022; In press

Core Tip: Vibrational spectroscopy (VS) may become a major player in the development of early diagnosis technology for pancreatic cancer. As with every technique, VS has promising attributes as well as drawbacks that can be hard to capture. We summarize recent reports about improvements in spectroscopic methodologies, briefly explain and highlight the drawbacks of each of them, and discuss available solutions. Additionally, the important aspects of spectroscopic data

evaluation with multivariate data analysis and a convolutional neural network methodology are depicted.

INTRODUCTION

Pancreatic cancer (PC) is a very aggressive and lethal neoplasm, ranking seventh in the world for cancer deaths [1]. In 2020 there were an estimated 466 003 new PC-related deaths worldwide [2]. Despite a greater understanding of the nature of pancreatic tumors, 5-year survival rates have not improved and remain below 10%. Late-stage disease at diagnosis is a significant issue that contributes to poor overall survival rates. Another prognostically relevant, yet complex fact is related to the morphological and molecular heterogeneity of the tumor cells and the surrounding stroma. On one hand, the need for proper pathological evaluation with a detailed prognostic assessment is required, on the other, more importantly, the variability of PC leads to observed chemoresistance [3]. Moreover, differentiating PC from large tumors of the ampulla of Vater remains challenging [4]. Our previous article highlighted recent trends in PC pathology and research [3]. Vibrational spectroscopy (VS)-based methods will play some prominent roles in the early diagnosis of PC (Figure 1). Detailed studies regarding the molecular nature of PC are required to reveal novel early and precise diagnostic technologies, thus improving survival rates.

EARLY DIAGNOSIS OF PANCREATIC CANCER

The lack of specific and sensitive early diagnosis options for PC screening results in late-stage disease at the time of diagnosis and is one of the reasons for the overall poor PC survival rates. Out of the available options, serum antigen levels, such as carbohydrate antigen 19-9 (Ca19-9), are insufficient because of poor specificity and sensitivity for malignancy detection [5,6]. Some studies highlighted the usefulness of measuring interleukin-6 (IL-6) serum levels to differentiate pancreatic ductal adenocarcinoma (PDAC) patients from chronic or acute pancreatitis [7–9]. Recently, leukemia inhibitory factor (LIF) was reported to be a promising serum biomarker of

pancreatic malignancy [10], a therapy response monitoring indicator [11–14], and a metastatic disease predictor for PDAC patients [15].

LIQUID BIOPSY BIOMARKERS

Circulating tumor DNA

Recently, some new circulating biomarkers, particularly those associated with early-stage disease, have been intensely studied. These include a group of tumor nucleic acids circulating in the patient's blood serum, such as circulating tumor DNA (ctDNA), cell-free DNA (cfDNA), cell-free RNA (cfRNA), circulating tumor cells (CTCs), or extracellular vesicles (EVs) (Figure 2)[16]. Detection of aberrant methylation markers specific for certain malignancies, including PDAC, was previously shown using genetic methods, such as next-generation sequencing (NGS) and droplet digital PCR (ddPCR) [17–20], but the sensitivity of these methods is still insufficient [20].

Extracellular vesicles

Both tumor and healthy cells release lipid-membrane vesicles, termed "extracellular vesicles" (EVs), into the bloodstream. EVs include oncosomes, apoptotic bodies, microvesicles, and exosomes of different sizes and biogenesis [21,22]. They have an established role in cancer cell communication and metastasis [23,24]. EVs are released in a passive mode (apoptotic bodies) from dying cells (apoptosis, necrosis) or in an active mode (microvesicles, exosomes) from living cells [25]. They make up a large part of the ctDNA (single-strand DNA and double-strand DNA) in a liquid biopsy, but also carry other biomarkers, such as tumor protein antigens [26,27], or microRNA (miRNA) [28,29]. The cargo within an EV creates a unique spectroscopic fingerprint specific for certain tumors, including PDAC. The molecular characteristics obtained using VS combined with the use of convolutional neural networks (CNN), which can specify the result and

pinpoint valuable information from the general all-in-one data, together create a tool for successful diagnosis.

DNA methylation

DNA methylation plays an important role in the regulation of gene expression, and cellular differentiation. Aberrant methylation is associated with the development and progression of various cancers [16,30,31]. Moreover, these specific aberrations may act as cancer biomarkers, enabling early diagnosis. Standard methylation status evaluation techniques include bisulfite conversion assay, the sequencing or melting curve analysis, restricted enzymes-based assay, and affinity capture using methylated DNA binding proteins [31]. Some of these methods suffer from a laborious workflow and false positives. Others require specific reagents such as enzymes and binding proteins, which are costly and time-consuming. Therefore, we will discuss the usefulness of label-free spectroscopic methods, such as surface-enhanced Raman spectroscopy (SERS), in DNA methylation status detection, and we will address potential issues that arise when using these techniques.

PDAC HETEROGENEITY AND CHEMORESISTANCE

PC is well known to be very heterogeneous in molecular and morphological phenotype. It is one of the reasons, aside from the lack of adequate early diagnosis methods, for patients' poor prognosis, because current treatment options do not consider tumor heterogeneity and thus give insufficient results [32,33]. When designing studies based on PC diagnosis, one must differentiate the results concerning the histomorphological type of the tumor. This distinction is crucial for patients care, due to different molecular pathways governing the development and evolution of these tumors, as well as the prognosis assessment and potential therapy options [3]. In our previous study, we demonstrated that Raman spectroscopy (RS) is capable of detecting ampullary cancer in pancreatic tumor tissue slides [34]. A subsequent step could then differentiate between

three groups of pancreatic tumors, specifically conventional pancreatic ductal adenocarcinoma (cPDAC), pancreatic ductal adenocarcinoma derived from intraductal papillary mucinous neoplasm (IPMC), and ampulla of Vater adenocarcinoma (AVAC). cPDAC is the most common form of PC developing *via* pancreatic intraepithelial neoplasia (PanIN). It arises in the ductal epithelium localized in "normal" pancreatic tissue, sometimes with signs of chronic pancreatitis, or is combined with so-called "acinar-to-ductal metaplasia" (ADM) regions. The KRAS mutation is the initiating event in this pathway of carcinogenesis [3]. IPMCs are carcinomas that arise from intraductal papillary mucinous neoplasms (IPMNs), cystic tumors, that develop in the main or peripheral branches of pancreatic ducts. Guanine nucleotide-binding protein, alpha stimulating activity polypeptide (GNAS) proto-oncogene mutation, which is not found anywhere else in pancreatic tumors, plays a significant role in IPMN development [3]. The third relevant group, AVAC, is a cancer of the duodenal ampulla of Vater. Oftentimes AVAC tumors grow too large diameters and deeply infiltrate the pancreatic tissue. The histomorphological similarities of AVAC and cPDAC tumors, often make them hard to distinguish from each other. Moreover, these tumors are often treated clinically and diagnostically in the same way, but the latest reports suggest that they differ regarding prognostic factors' incidence, such as tumor differentiation, perineural invasion, venous invasion, or lymph node involvement [35]. The early occurrence of bile duct obstruction symptoms in IPMC or AVAC enables earlier diagnosis and thus may lead to a better prognosis. However, the reports supporting this are ambiguous [3]. Further subgrouping of PC tumors into morphologically distinct entities, such as large duct (cystic papillary), foamy glands, clear cell, adenosquamous, vacuolated-cell, or colloid, which are described in detail elsewhere [3], may benefit patients and clinicians, because of the different prognosis of some of these groups.

Recently, Mukhopadhyay *et al* [36] showed that the nuclear factor-erythroid 2–related factor 2 (NRF2) was responsible for gemcitabine chemoresistance, and the NRF2 expression level in PDAC tissues correlated with poor patients outcome. Another study by Patzak *et al* [37] described cytosolic 5′-nucleotidase 1A (NT5C1A) as a mediator of this

resistance by reducing intracellular gemcitabine metabolites and limiting its efficacy. Gemcitabine is a standard chemotherapeutic for PDAC. The tumor's resistance to the therapy is among the main reasons for PC's drastically low 5-year survival rates. Thus, recognizing this chemoresistance is of great importance. The utilization of VS is a step closer in this direction. VS may prove to be beneficial for identifying chemoresistant pancreatic tumors. More studies are required to evaluate NRF2 or/and NT5C1A expression levels in PDAC tissues, and compare them with spectroscopic data, to identify spectral markers that correlate with gemcitabine chemoresistance. Performing these studies will help select patients that might benefit from gemcitabine therapy.

Molecular spectroscopy methods obtain all of the information about the studied sample with a single measurement. There is no need for special labeling or selecting areas of interest, that other genetic or biochemical methods may require. Analyzing spectral data plays a pivotal role. All data is ready for interpretation and provides information about the tumor, such as its subtype, differentiation level, specific chemoresistance, and other tumor-specific poor prognostic factors (hepatocyte nuclear factor-1B - HNF1B expression, or cancer stem cells). Whether or not we can decipher this information will determine the efficacy of this methodology in diagnostics.

MOLECULAR SPECTROSCOPY

Molecular VS was confirmed to play an important role in the characterization of the chemical structure and composition of malignant tissues [38–41] and the analysis of human blood serum [42]. Due to high chemical selectivity, RS and infrared spectroscopy (IR) can become efficient tools supporting screening for pancreatic malignancy [43]. These methodologies provide information about the various biologically significant molecules and functional groups in a tumor including phospholipids and triglycerides, proteins, nucleic acids, phosphates, and carbohydrates. Electromagnetic radiation (photons) can be absorbed and/or scattered by a sample because the energy of phonon excitations and vibrations, as well as oscillations of functional groups in molecules, correspond to the energy in the infrared region of the electromagnetic radiation

spectrum. Therefore, the presence of functional groups in the analyte can be detected *via* interaction with light, providing information about the molecular structure and composition of the investigated sample. IR and RS take advantage of absorption and inelastic scattering, respectively. The results obtained using these methods indicate differences in metabolic pathways typical for various neoplasms. The main advantage of the molecular spectroscopic approach is achieving information about samples in a label-free and noninvasive manner. The research potential of spectroscopic methods has not yet been fully explored in the investigation of PC. However, preliminary results are promising [34,44].

Raman hyperspectral mapping

RS is a VS method that can deliver a detailed molecular fingerprint of a studied sample in almost real-time, without the need for special labeling. A technique called Raman hyperspectral mapping (RHM) provides high-resolution imaging at a relatively low cost compared to other well-established medical imaging techniques, such as magnetic resonance imaging (MRI) [43]. RHM relies on multiple measurements of adjacent "pixels" of tissue and combining the resulting spectra into a single map image. Such imaging allows the selection of points of interest in the studied sample and precise distinction between particular tissue elements, such as cancer cells (with nuclei and cytoplasm) or the stroma compartment (Figure 3).

The RHM methodology of tissue samples usually involves slicing 2.5 µm thick tissue sections with a microtome from a standard formalin-fixed paraffin-embedded tissue block. Because glass gives substantial interference in the Raman readings, for the slide mounting, special calcium fluoride (CaF_2) windows are used for the slide mounting instead. Another important aspect of RHM is the selection of the region of interest which is usually a part of the cancerous glands or stroma compartment. This should be done by an experienced pathologist. Subsequently, a complete paraffin removal

procedure ought to be conducted involving a 12-h xylene bath and graded ethanol rehydration. On such preprocessed tissue slides, the Raman measurements can be done. This methodology was already described in our previous paper on ampullary cancer detection with RHM [34].

RS is accurate, and the information obtained is characterized by good resolution. Nevertheless, RS requires a sophisticated methodology and equipment because the Raman effect is very weak in nature [43,45]. Another drawback of RS is the substantial sample pre-processing, manual selection of points of interest, and further data analysis.

Surface-enhanced Raman spectroscopy

To augment the Raman signal strength, another method, called surface-enhanced Raman spectroscopy (SERS) is used [43,45]. SERS is a label-free, ultrasensitive tool, capable of DNA methylation analysis. SERS utilizes the same physical phenomenon as RS (Raman effect), but the effect is significantly enhanced using specially synthesized plasmonic nanoparticles [46] that the molecules (such as DNA) are absorbed into [31]. Furthermore, SERS is efficient for liquid biopsy measurements, and it does not require special labeling. However, the production of nanoparticles requires an experienced team and a proper methodology.

The unique ability for such sensitivity of SERS is achieved *via* the use of plasmonic nanostructures (SERS substrates), between which so-called "hot spots" are localized. The traditional methodology of SERS measurements involved mixing the sample with gold (Au) nanoparticles (20-50 nm diameter). One method of SERS substrate production prepares Au nanoparticles *via* chemical reduction of tetrachloroauric (III) acid using trisodium citrate under specific reaction conditions according to the procedure described by Frens *et al* [46]. However, this synthesis method is characterized by random and nonuniform hot spot distribution, which leads to the poor reproducibility of SERS [31,47,48]. Another issue, especially regarding DNA methylation studies, is the difficulty

distinguishing between DNA methylation signals and the adjacent nucleotide signals due to their similarity [49]. Additionally, surfactants and/or capping agents can affect signal purity [50]. Overcoming these issues is crucial to utilizing SERS for diagnostics largely based on DNA methylation analysis. To achieve this, a methodology of proper SERS substrate development is required. Such a substrate is characterized by a large Raman enhancement, regularly arranged hot spots, and an open and easily accessible surface topology [31]. Luo *et al* [31] proposed the use of a plasmonic gold nanohole array (PGNA) as a SERS substrate. Originally, the authors described the use of electron beam lithography (EBL) for the PGNA substrates production, but a focused ion beam (FIB) might be even better, due to its higher resolution [51]. Both EBL and FIB can be used to obtain a periodic matrix of holes (plasmonic nano-holes array) in a Au layer evaporated onto an atomically flat non-plasmonic substrate [31] (Figure 4).

SERS is characterized by a large amount of work required for substrate production, but the sample pre-processing is minimal. Raw blood serum samples can be analyzed, without requiring preselection by a specialist. This is a valuable asset in the search for a diagnostic option.

Attenuated Total Reflection Fourier Transform Infrared Spectroscopy (ATR-FTIR)

ATR-FTIR is a complementary method to SERS and RS, which measures absorption as an infrared spectrum of a biochemical fingerprint. It is widely accessible, inexpensive, and easy to use and implement, while still providing substantial information about the sample being analyzed. ATR-FTIR is characterized by high sensitivity in biomedical diagnostics [52] and can detect spectral markers of many pathologies in physiological liquids such as saliva or blood serum. ATR-FTIR was used to detect lung [53], bile duct [54], ovary [55,56], breast [57], and brain [58,59] tumors in the blood serum samples. In a study by Butler *et al* [59], the team described the methodology of using ATR-FTIR in an early screening of brain tumor patients with sensitivity and specificity values of 93.2% and

92.8%, respectively. The high sensitivity of ATR-FTIR can be achieved due to the design of the instrumental set-up. In the ATR-FTIR, a sample droplet is placed on an internal reflection element (IRE), also called the ATR crystal. IREs with high refractive indexes, such as diamond, zinc selenide, germanium, or silicon are used. The incident IR beam is directed through the IRE and the resultant evanescent wave extends beyond the ATR crystal surface and penetrates the sample. As with SERS, there is no need for sample pre-processing or special labeling. No additional substrate production is required as well. However, achieving sensitive information about the samples, and their interpretation remains challenging. Employing CNNs to analyze ATR-FTIR data is a viable solution that has already shown success [59].

Molecular spectroscopy in diagnostics

Multiple studies have shown the usefulness of VS in the detection of DNA methylation status of cells, as well as other diagnostically important factors. For example, DNA methylation aberrations were successfully detected using SERS [31,60–63], and FTIR [64–70] methods. Recently, in an interesting study by Ho *et al* [71], the authors successfully utilized RS and deep neural networking that enabled culture-free serum bacteria identification and antibiotic susceptibility testing with approx. 97% accuracy. In other studies [72,73], RS and SERS were used to detect previously untraceable concentrations of biomarkers (matrix metalloproteinase 7 – MMP7, and mucin 4 – MUC4) in the serum of PDAC patients.

To date, no report has combined detecting aberrant DNA methylation markers of PDAC, obtained by analyzing patients' serum with the use of VS methods, specifically the ATR-FTIR combined with CNN classification. Although partial results are available and show promise, additional investigations are needed to support the combination of VS methods with CNN methodology for PDAC detection.

Residual disease monitoring

Currently, there are no efficient methods of patient monitoring for minimal residual disease (MRD) in PC. Generally, postoperative surveillance methods, including monitoring of clinical symptoms, blood tumor markers, and CT or MRI imaging are used, but these methods lack sensitivity and specificity for MRD [74]. Recently, circulating tumor DNA (ctDNA) detection using genetic methods in a liquid biopsy was highly advocated for in terms of MRD monitoring [74–77]. VS could be a better option for MRD monitoring compared to genetic mutation detection, because of VS's ability to identify DNA methylation aberrations [20]. However, there is currently no data supporting this. Notably, as highlighted by Henriksen *et al* [78], surgical trauma elevates the serum ctDNA levels up to 4 wk after the surgery. This should be taken into consideration when designing a MRD monitoring study using liquid biopsy analysis.

MULTIVARIATE DATA ANALYSIS

In spectroscopic data evaluation, it is very important to draw proper conclusions. Therefore, various methods of multivariate data analysis are used to help with data interpretation [79]. Hyperspectral mapping with K-means clustering (KMC), and principal component analysis (PCA) is commonly performed to explore spectral variation [80–82]. Pre-processing of the spectroscopic data involves cosmic rays removal, baseline correction, and smoothing (adaptive multi-round smoothing based on the Savitzky-Golay filter). Minimal necessary operations are applied to explore marker bands of significant aspects, such as DNA methylation while preventing the loss of important spectral information. Multivariate data analysis is applied to reduce data dimensionality and extract the most important parameters from the acquired information. Some of these methods are briefly described below.

Unsupervised hierarchical cluster analysis (HCA) is a clustering algorithm designed to group the obtained spectra or to produce false-colored maps based on spectral similarity and variability.

PCA is based on a linear transformation of the data to a new space described by orthogonal axes, the so-called principal components. The most significant results are the "score" values, which represent the data in multidimensional space corresponding to the principal components, and the loading values, which identify the variables causing the data separation according to their influence on the scores. Additionally, the results of the PCA are complemented by the explained variance. The explained variance gives the percentage of the total variance of the original data set, which is explained by a certain principal component.

Partial least squares regression (PLSR) involves a linear transition of numerous original descriptors to a new variable space based on a small number of orthogonal factors (latent variables). In other words, PLSR allows the construction of predictive models when the factors are highly collinear. This analysis estimates unobservable variables as exact linear combinations of their empirical indicators. The estimated proxies are treated as substitutes for the latent variables. The selected case values capture most of the independent variables' variance. This variance is used for predicting the dependent variable.

Non-negative matrix factorization (NMF) is a useful tool for the analysis of high-dimensional data. Besides detecting a compressed representation, NMF delivers insights into the structure and features of the given data by extracting easily interpretable factors. With the use of NMF, basic spectral components for proteins, lipids, phospholipids, or nucleic acids can be compared.

All of these methods require human decision-making, although minimal. As a result, these methods might disturb seemingly irrelevant data, and lose it. To overcome this and enhance the sensitivity and specificity of spectral data interpretation, deep neural networks are used.

CONVOLUTIONAL NEURAL NETWORKS

RS is characterized by a very low strength of the measured effect (Raman effect) and is thus very sensitive to distortion factors, such as fluorescence, thermal noise, the quality

of the measuring equipment, and research team experience. Removing noise artifacts requires various pre-processing methods on acquired spectra (i.e., cosmic rays removal, baseline correction, and smoothing). All of these depend on human input and definition. On one hand, this prevents automation, but on the other, some seemingly irrelevant data might be lost during the pre-processing. The successful use of CNNs in spectroscopic data evaluation and classification was shown in multiple studies [83,84] including those using RS [85], SERS [71,86], and ATR-FTIR [87]. As a source, the CNN is fed with raw spectral data, without human interaction-related pre-processing. This approach gives better results in classification than using conventional methods [83] and makes the methodology more universal. CNN objectives should be clear and simple. Similarly, proper selection of training data is of great importance. For each objective, different training sets should be created. All spectral results from each group are split (2:1) to form the training (two-thirds of data) and testing (one-third of data) datasets. Training datasets should include positive and negative cases, preferably with many variations. For example, when designing a training dataset for the CNN that will decide whether the results are from a malignant pancreatic tumor or not from a malignant pancreatic tumor, one might include cases from PC, but as a negative control also include malignancies of other sites (i.e., colorectal or breast carcinomas), and benign pancreatic entities (i.e., IPMN, mucinous cystic neoplasm, or groove pancreatitis).

There are some issues related to utilizing neural networks (NNs) with multiple layers (deep neural networks), specifically very deep NNs. One obstacle is the vanishing/exploding gradient problem. A training process of a NN in simplest words usually involves facing some issue and updating the "weights" of the algorithm to better cope with the problem. A great benefit of a CNN is that it extracts features of the task on its own. Features extraction is done using an optimization algorithm, such as "gradient descent", which simply finds values of a function's parameters to minimize the cost function. If gradients that update the weights shrink, the weights are no longer updated, and the learning stops. This is called a vanishing gradient problem. Similarly, if gradients grow, weights do not update reasonably, and the learning becomes

unstable, resulting in the exploding gradient problem. The so-called "skip connection" technique is utilized to overcome the vanishing/exploding gradient problem [84,88]. This is a basis for a Residual Network CNN architecture (ResNet) [89]. Another issue of CNNs is data overfitting. When a model used to train the CNN is very complex, and there is a limited amount of learning data available, the CNN learns to know the training dataset well, but performs poorly against any new data (i.e., validating/testing dataset or the implementation data). One way to improve this shortcoming is by using proper data augmentation techniques. Usually, in spectral data analysis, augmentation requires making additional artificial spectra by small spectral shifting, expanding spectral range, adding Gaussian noise, or superimposing the spectra for each real result [84].

CONCLUSION

When designing a study investigating VS methods in PC diagnosis it is important to bear in mind a couple of aspects. First, the study is required to address all the drawbacks of RS, SERS, and ATR-FTIR methodology, described briefly above. Another crucial issue is the proper cooperation of multidisciplinary teams, including medical specialists, such as pathologists and clinicians, and spectroscopic specialists, such as physicists and chemists. For example, comparing PC cases with pancreatic neuroendocrine neoplasms is not reasonable because the malignancies represent different entities with very distinct pathways governing their initiation and progression, as much as patient survival. When dealing with PC, detailed knowledge of the tumor's molecular and pathological nature is required. In particular, one should interpret the results of measurements in the context of PC subtypes and some other prognostic factors [3]. Of great importance is the proper analysis of measured spectroscopic data. For example, when conducting a RS on complex tissue samples, such as PC sections, the random spots of measurements will bring ambiguous results. The spectral data might be disturbed by numerous interfering phenomena, such as inflammation, tumor necrosis, or fibrosis. Additionally, the cytoplasmic and nuclear regions of cancer cells significantly differ as well. Interpretation of the PC stroma

compartment, including the complicated cancer-stroma "cross-talk", is another aspect to address [3]. Careful selection of regions of interest is important too. This should be done by a specialized pancreatic pathologist familiar with the spectroscopic methodology. Finally, in the search for a PC diagnostic tool, we look for universality and automation. Thus, the interpretation of spectral data obtained from liquid biopsy, which relies on human-dependent pre-processing is not a good path to follow. CNNs are invaluable here, but proper design and training them is the key to success. Following these aims, we designed a study that will comprehensively evaluate the diagnostic potential of using VS methods in diagnosing PC, by systemically evaluating liquid biopsy samples (Figure 5). In conclusion, VS seems to be leading the way in the race, with most of the methodology drawbacks resolved, at least partially (Table 1).

ACKNOWLEDGEMENTS

This work is supported by the National Science Centre, Poland under the "OPUS 19" project (Reg. No. UMO-2020/37/B/ST4/02990).

78531_Auto_Edited.docx

ORIGINALITY REPORT

9%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.mdpi.com Internet	186 words — 4%
2	Xiaojun Luo, Yingfang Xing, Daniel David Galvan, Erjin Zheng, Ping Wu, Chenxin Cai, Qiuming Yu. "Plasmonic Gold Nanohole Array for Surface-Enhanced Raman Scattering Detection of DNA Methylation", ACS Sensors, 2019 Crossref	44 words — 1%
3	www.ifj.edu.pl Internet	28 words — 1%
4	deepai.org Internet	24 words — 1%
5	www.spiedigitallibrary.org Internet	20 words — < 1%
6	ruj.uj.edu.pl Internet	19 words — < 1%
7	Kenny Kong, Catherine Kendall, Nicholas Stone, Ioan Notingher. "Raman spectroscopy for medical diagnostics — From in-vitro biofluid assays to in-vivo cancer detection", Advanced Drug Delivery Reviews, 2015 Crossref	18 words — < 1%
8	ijcem.com Internet	

17 words — < 1%

9	docplayer.net Internet	16 words — < 1%
10	www.explorationpub.com Internet	15 words — < 1%
11	Konstantin Fackeldey, Jonas Röhm, Amir Niknejad, Surahit Chewle, Marcus Weber. "Analyzing Raman spectral data without separabiliy assumption", Journal of Mathematical Chemistry, 2021 Crossref	14 words — < 1%
12	www.pjps.pk Internet	13 words — < 1%
13	www.ncbi.nlm.nih.gov Internet	12 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS