National Research University Higher School of Economics

as a manuscript

Yaniv Cohen

IR spectrography and tomography of Human Tissue and their analysis by methods of machine learning

PhD Dissertation Summary

for the purpose of obtaining academic degree

Doctor of Philosophy in Engineering

Academic supervisor:

Doctor of Technical Sciences, Professor

Evgenii A. Krouk

Moscow – 2022

The Relevance of the Research

Cancer diagnostics is a particularly active area with publications including cervical, lung, prostate, colon, esophageal, gastric, brain and skin cancers, etc. During the recent decades, it was stated by the scientific society some common intrinsic cancers related to openings (orifices) and the current art diagnostic methods.

Nowadays the gold standard for diagnosis involves histopathology based on sectioning and staining of tissues. However, the main problem with this method is that it takes rather long time (several days to several weeks) to obtain pathology results, due to lengthy procedures for preparation of tissues and thorough procedure for pathology evaluation. Furthermore, during the biopsy session, elaborate removal of tissue is required, which can be dolorous and expensive; however, a limited number of samples can be taken. No less important are the problems associated with storage, transportation and further expert analysis of biopsy samples in a laboratory; this increases the expenditures, increases the likelihood that samples will be mismanaged, contaminated or wasted, and also introduces a significant deceleration in the process of receiving results. Finally, to interpret the biopsy results, microscopic analysis is usually used, leading to qualitative judgmental results that cannot be interpreted consistently.

Therefore, the medical society seeks to elaborate safe diagnosis technologies with more efficient sensitivity, for non-operative detection of internal cancers in their primary phases and differentiation between cancer, neoplasms benign and unspecified malformations of internal tissue (including cysts and polyps).

An important area of investigation in this area is currently the search for opportunities for the earliest possible and non-invasive detection of damaged tissue areas for their prompt removal. Different methods involving optical, acoustical, magnetic and X-ray devices and dedicated techniques have been

used in recent studies. In the works by Minsky, confocal Scanning Microscopy was developed, that is an optical imaging technique for increasing optical resolution and contrast. Godfrey Hounsfield presented X-ray computed tomography, CT Scan, which is a medical imaging technique to be used in radiology to get detailed images of the body noninvasively for diagnostic purposes. In the works by Mansfield and Lauterbur, magnetic resonance imaging (MRI) was developed, that is a medical imaging technique to be used in radiology to form pictures of the anatomy and the physiological processes of the body.

For example, in Barrett's and ulcerative colitis, the hallmark for carcinogenesis is flat dysplasia. The current state of the art in this case is taking random biopsy specimens in multiple spots going on a proverbial wild-goose chase. However, obtaining the great amount of biopsy specimens taken in instinctive spots is overlong, high cost, adds potential complexity, and results in possible false negatives (because typically <5% of mucosa is sampled).

As another example, endeavors to refine detection of adenoma were applied towards endoscopic blind spots (to the rear of folds, flexures) or visualized subtle lesions in the regions of vision (eg, flat and depressed lesions). Active efforts at present have been directed toward improving the endoscope (high-definition, narrow-band imaging, autofluorescence, etc.) or contrast agents (molecular imaging, chromoendoscopy). The mentioned attempts improve the detecting ability but do not destroy the problem completely.

In several studies, methods involving Fourier-transform infrared spectroscopy and imaging in combination with attenuated total reflection (ATR-FTIR) were used for the analysis of salivary exosomes from oral cancer patient or for the detection of stem cells in cancer research and to the characterization of cell-cycle variations. IR spectra can be used as a sensitive marker of structural changes of cells and of reorganization occurring in cells and most

biomolecules give rise to IR absorption bands between 1800 and 700 cm^{-1}, that are known as the “fingerprint region” or primary absorption region.

The two techniques: Fourier-transform infrared spectroscopy and imaging spectroscopy of frozen sections (cryosections) have gained wide research interest aiming for malignancy diagnostics. Achieved results are to a great extent similar to the routine frozen section (FS) pathology examination. These methods received the recognition among the professional society, but still they have not entered the clinical practice due to their shortcomings. It is only relatively recently that major improvements in technology have afforded the sensitivity required to study biological molecules and it is only within the last decade or so that FTIR imaging has become available. Some of the most important technological developments include the interferometer, highly sensitive detectors and array detectors, powerful light sources and attenuated total reflection (ATR) technology. Advances in computing power have also enabled rapid processing of large datasets. Thus, developments should be implemented in the Medical Infra-Red Optical System (IROS).

Though the field has advanced greatly in the latest years, the principal task remains to develop, translate and implement in the clinical practice a rapid and reliable method that can be used for diagnosis of cancer during the operation. These methods are generally complex and time-consuming, and so are difficult to translate to a clinical setting.

The approach to diagnostic process involving ATR-FTIR seems very promising. However, the following issues remain unresolved: a) methods of identification of spectral patterns that distinguish cancerous from healthy tissue; b) methods of datasets analysis; c) simple and stable instrumentations.

The Objectives and Goals of Dissertation

Thus, **the objective** of this work is to provide a non-invasive, fast, compact, remote, portable and highly sensitive intraoperative diagnostic methodology and the corresponding tool. To resolve this issue, we plan to employ measurements of the contrast of the temperature difference against the background involving ATR-FTIR.

To achieve this goal, it is necessary to solve the following **tasks**:

1. To develop reproducible IR imaging and spectroscopy techniques that might be developed into clinical tools for earlier non-invasive diagnosis of human inner tissue anomalies – tumors and diseases.
2. To select and find more attractive from existing current diagnostic analytical and statistical methods to obey and mitigate some of lacking, and propose more optimal techniques for raw data analysis observed experimentally.
3. To use a thermal IR-camera and FTIR spectrometer in order to make such selection of existing techniques, which can be simple and stable instrumentations enough for future hospital and clinical installation.
4. To develop the attractive and optimal machine learning methods that are sufficiently simple and inexpensive so that translation into clinical usage could become feasible.
5. To explore and find the best machine-learning tool for use with spectral analysis of human tissue earlier non-invasive diagnosis and differentiation of cancerous and non-cancerous skin anomalous.

The Scientific Novelty of the Study

The present thesis contributes to solving problems of earlier non-invasive diagnosis of human tissue anomalies – tumors and diseases, based on the theoretical framework of IR tomography and spectrography, on more attractive instrumentation usage for this purposes, on the analytical and statistical methods

of raw data analysis, and on elements of Artificial Neural Networks (ANN) and Machine Learning algorithms. Thus:

1. we employed a cloud data base with machine learning methods in the area of cancerous skin identification and classification, where this has not previously been applied, which allows to increase the ability of the system to improve the identification and classification rate of different biochemical composition of tissues and to help the physician as a real time on-site decision support system;
2. we experimentally established the heat signatures of different human body tissues corresponding to normal and abnormal states;
3. we proposed a new method of detecting and identification of gastric, colorectal and cervical cancers, that is different with respect to other methods in that the conventional methods are based on the subjective inspection in the visible spectral range of the physician, whereas in our method we are based on the overheated characteristic of cancerous tissues and detecting these anomalies with a NIR or MIR camera;
4. we proposed a new method of preliminary classification of anomalies, that is different with respect to other methods in that it is done using IR optical band in the range of 3–12 μm;
5. we proposed a new method for early detection or Cancer removal by thermal IR approach based on passive imaging of heat signature changes in tumors due to minute changes of environmental temperatures around these tumors by applying heating and cooling methods on the tissue, this method allows for screening and for characterizing of these tumors more accurately at the earlier stages of their spatial-temporal evolution which cannot be detected by human or conventional machine vision systems. These new methods of heating and cooling are important as they show that cancer removal can be monitor non-invasively to ensure complete removal of cancer inside/outside of the human body.

The Practical Significance

1. The practical applications of the proposed approach, based on combined platform of multidisciplinary techniques, theoretical and experimental, resulted in creation of a Medical IROS that is a tabletop device for real-time tissue diagnosis that utilizes FTIR spectroscopy and the ATR principle to accurately diagnose the tissue.
2. Within the scope of our findings, we developed a new endoscope based on an array of infrared and visual optic micro detectors operating in the waveband range of 3–12 μm working both in integral and spectral regimes based on the sign and amplitude of the contrast of cancerous anomalies with respect to those for normal living tissues.
3. A trained Artificial Neural Network was created serving to predict cancer and other pathologies based on measurements by FTIR-ATR device. The reliability of the proposed neural network method was examined on the data collected through Medical IROS (FTIR-ATR) device and obtained by a biopsy.
4. By using a combination of IR thermal imaging and machine learning spectral analysis, a decision support system for real time, on site, early detection and identification of cancer was built.
5. The medical IROS along with the decision support system serve as a basis to perform a real-time optical biopsy or spectral histopathology (SHP) of tissues on-site in the operation room. This is a powerful tool for real-time, on-site, cost-effective, simple-to-use, non-destructive, non-operator, early non-invasive detection and identification of different kinds of human tissue inner cancers.

The Key Findings of the Thesis to Be Defended

1) A new approach to IR tomography and spectrography with the corresponding theoretical and conceptual background.

2) An improved method of early detection by thermal IR approach based on passive imaging of heat signature changes in tumors due to minute changes of environmental temperatures around these tumors after applying heating and cooling methods on the tissue, which allows for screening and for characterizing of these tumors more accurately at the earlier stages of their spatial-temporal evolution.
3) New experimental techniques based on usage of ATR-FTIR spectroscopy as a diagnostic tool for quick identification of cancerous diseases during operation.
4) Adaptation of Artificial Neural Network structure algorithms to investigation of tissue anomalies – tumors, cancers, and metastases, for express analysis of such kinds of anomalies.
5) A unified platform for combination of elements of theoretical framework, ATR-FTIR spectroscopy, neuron structure algorithms and elements of machine learning, which allows a specialist in medicine to obtain a simple but powerful inexpensive tool for early non-contact detection and identification of tumors and the ability to monitor tumor removal during the operation by using a sensitive thermal camera and by heating or cooling technique.

Methodology used in Dissertation

Using a cloud database with Machine learning methods increases the ability of the system to improve the identification / classification rate of different biochemical composition of tissues and to help to physician as a real time on-site decision support system. In this research, we investigated the possibility of using untreated, freshly excised GI tissue samples in conjunction with ATR-FTIR spectroscopy as a diagnostic tool for quick identification of cancerous diseases during operation.

Endoscopy is the most important imaging technique in the field of gastroenterology. It assists the clinician in making better decisions, diagnose more accurately and of course is a powerful therapeutic tool. The prediction and final survival rate of gastrointestinal cancers are dependent on the stage of disease. The ideal would be to detect those gastrointestinal lesions at early stage or even premalignant forms that are difficult to detect by conventional endoscopy with white light optical imaging as they show minimum or no changes in morphological characteristics and are thus left untreated. Since the introduction of video endoscopes in the 1990, there has been a big advance in image quality. High definition endoscopes and screens are now becoming standard of care. In parallel to the development of endoscopes, we continued to seek other methods for optical enhancement. Some of the technologies that are available are: Chromoendscopy, NBI, FICE Autofluorescence and Confocal Laser Endoscopy. In our research we account for some disadvantages met during these investigations, such as time consuming, price and need of operator training. All this limits their widespread use. Accounting for the experimentally obtained fact that human body radiates naturally in the IR range and each tissue either normal or not normal, has its heat signature, we in our dissertation have proposed improving of early detection by Thermal IR approach based on passive imaging of heat signature changes in tumors due to minute changes of environmental temperatures around these tumors after applying heating and cooling methods on the tissue, that may allow for screening and for characterizing these tumors more accurately at the earlier stages of their spatial-temporal evolution.

The combination of IR thermal imaging and machine learning spectral analysis, may lead to a new platform for real time, on site, early detection and identification of cancer and finally, to establish a worldwide cost effective, simple to use, non-destructive, non-operator depended, decision support

system based on a machine-learning network to perform a real-time optical biopsy or spectral histopathology of tissues on-site in the operation room.

Authenticity of the results

The authenticity of the results obtained in the dissertation is confirmed by:

- the correspondence of experimental data obtained by the proposed methods and by standard methods;
- compliance of the obtained results with fundamental physical and biological principles;
- numerical modeling on modern calculation packages.

The author's personal contribution

- Development and implementation of mathematical models and algorithms;
- Proofs of the theorems, developing algorithms and their testing;
- Collecting datasets, performing computational, pre-clinical and clinical experiments.

Approbation of Research Results

The work underwent approbation at the following pier International Conferences:

1) "Method of Infrared Thermography for Earlier Diagnostics of Gastric Colorectal and Cervical Cancer." Innovation in Medicine and Healthcare 2016. InMed 2016. Smart Innovation, Systems and Technologies, Tenerife, Spain,15-17 June 2016.
2) "Real-Time, On-Site, Machine Learning Identification Methodology of Intrinsic Human Cancers Based on Infra- Red Spectral Analysis – Clinical Results". 11th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2020) Orlando, USA, March, 2020.

(Chosen as “best 25%-30% of the papers, which authors registered for IMCIC 2020”)

3) “Artificial Neural Network in Predicting Cancer Based on Infrared Spectroscopy”’. Intelligent Decision Technologies. IDT 2020. Smart Innovation, Systems and Technologies, Split City, CROATIA, 17-19 June 2020. KES Virtual Conference center.

The list of the published articles where the main scientific results of the thesis are reflected

- *The list of articles published in the journals included in international citation system Scopus:*

1) B.-Z. Dekel, A. Zilberman, N. Blaunstein, Y. Cohen, M.B. Sergeev,L.L. Varlamova and G.S. Polishchuk (2016), Method of Infrared Thermography for Earlier Diagnostics of Gastric Colorectal and Cervical Cancer. In: Chen YW., Tanaka S., Howlett R., Jain L. (eds) Innovation in Medicine and Healthcare 2016. InMed 2016. Smart Innovation, Systems and Technologies, vol 60. Springer, Cham. Pages 83-92 (DOI: https://doi.org/10.1007/978-3-319-39687-3_8)
2) Y. Cohen, ,A. Zilberman, B.-Z. Dekel, E. Krouk, "Real-Time, On-Site, Machine Learning Identification Methodology of Intrinsic Human Cancers Based on Infra- Red Spectral Analysis – Clinical Results. Proceedings of The 11th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2020) Orlando, USA, March, 2020, pages 72 – 77;.
3) Cohen Y., Zilberman A., Dekel B.Z., Krouk E. (2020) Artificial Neural Network in Predicting Cancer Based on Infrared Spectroscopy. In: Czarnowski I., Howlett R., Jain L. (eds) Intelligent Decision Technologies. IDT 2020. Smart Innovation, Systems and Technologies,

vol 193, pp 141-153. Springer, Singapore (DOI: https://doi.org/10.1007/978-981-15-5925-9_12)

4) Y. Cohen, ,A. Zilberman, B.-Z. Dekel, E. Krouk, "Real-Time, On-Site, Machine Learning Identification Methodology of Intrinsic Human Cancers Based on Infra- Red Spectral Analysis – Clinical Results. Journal of Systemics, Cybernetics and Informatics (JSCI) , vol. 18, No. 2 , 2020.

- *The list of articles published in Medical journals:*

5) Y. Cohen, B.-Z. Dekel, E. Krouk, and N. Blaunstein "Method of Infrared Thermography for Earlier Diagnostics of Gastric Colorectal and Cervical Cancer". EC Gastroenterology and Digestive System, Vol. 6.10 , pp. 916-933., 2019.

The Structure of Dissertation

In the Introduction chapter, there are presented: the relevance of the work, literature review, the goal and aims, scientific novelty, practical significance, methodology and the proposed techniques.

In Chapter 1, we show the ability to detect small tumors and metastasis on mice that are not visible by naked eye or by visual camera, using a thermal camera with heating or cooling technique that can also be used for tumor removal monitoring. In addition, in-vitro thermal imaging of Laparoscopic procedure for removing tumors filmed to show the thermal difference between normal tissue and tumors. Furthermore, a theoretical background including a mathematical model is presented.

Chapter 2 presents the combined device and method for the in-vitro analysis of tissue and biological cells, by using the device Medical IROS system.

Discussing the basic principles of the FTIR-ATR detection, means of operation, basic elements of the developed device, data flow of information and

presenting the full coupled system of data transfer from each patient of different hospitals into the Center of collection information and its decision made by medical personnel after analyzing results of machinery learning.

In Chapter 3 we choose and build a classifier that can distinguish between cancer, normal and other tissue pathologies from the measured FTIR spectroscopy data. Discussing the problem description, Data preparation and pre-processing, Machine Learning approach for classification. The classification methods and results presented are based on: Partial least square regression (PLSR); Principal component regression (PCR); Linear Discriminate Analysis (LDA) classifier; Naive Bayesian classifier (NBC), including a review of the Spectral biomarkers for discrimination between Normal and Malignant cells, shows the Malignancy indicators in the Mid IR spectral region.

The main aim of **Chapter 4** is to present results of evaluation of the Artificial Neural Network in predicting cancer and other pathologies based on measurements by FTIR-ATR device. The feed-forward backpropagation neural network with supervised learning (the real outputs are known in advance) was used to classify the disease: cancer/non-cancer or cancer-polyp-normal. The preliminary results are based on 76 wet tissue samples from a clinical trial. Discussing the ANN concept, Biological and Artificial Neural networks, the Feed-forward ANN classifier design and the Network training Algorithms.

Chapter 5 present the main summary of the research presented in previous chapters. Thus, excellent results were achieved by applying machine learning (ML) based classification methods to 76 wet / Fresh samples: "Cancer", ”Polyp” or "non-Cancer:". Overall, 98.7% (75 / 76 biopsies) of the predictions are correctly classified and 1.3% (1 / 76 biopsies) was wrong classified

In **Chapter 6**, we show the possible practical applications of algorithms, techniques, and methods, proposed for imaging and identification of tissue anomalies, including various tumors and intrinsic cancers, based on our experience and results obtained during 4-year research.

Finally, in **Annex**, we give overview of the matter of investigation, based on following mathematical and bio-physical aspects, such as: The Electromagnetic wave spectrum; Fourier Transform Infrared Spectroscopy (FTIR Spectroscopy); FTIR spectroscopy applied to medical diagnostics; Vibrational Spectroscopy and IR-spectroscopic sampling modes including FTIR spectroscopy modes commonly used for the interrogation of cellular materials

Conclusion

The human body sometimes develops tumors and other internal lesions, which destroys the functionality of affected organs and ultimately lead to death. Their early detection and identification is critical for the health and survival of the patients.

1) A new approach to IR tomography and spectrography with the corresponding theoretical and conceptual background.
2) An improved method of early detection by thermal IR approach based on passive imaging of heat signature changes in tumors due to minute changes of environmental temperatures around these tumors after applying heating and cooling methods on the tissue, which allows for screening and for characterizing of these tumors more accurately at the earlier stages of their spatial-temporal evolution.
3) New experimental techniques based on usage of ATR-FTIR spectroscopy as a diagnostic tool for quick identification of cancerous diseases during operation.
4) Adaptation of Artificial Neural Network structure algorithms to investigation of tissue anomalies – tumors, cancers, and metastases, for express analysis of such kinds of anomalies.
5) A unified platform for combination of elements of theoretical framework, ATR-FTIR spectroscopy, neuron structure algorithms and elements of machine learning, which allows a specialist in medicine to obtain a

simple but powerful inexpensive tool for early non-contact detection and identification of tumors and the ability to monitor tumor removal during the operation by using a sensitive thermal camera and by heating or cooling technique.