engineering proceedings

MDPI

Proceeding Paper

A Novel Ensemble of FTIR Spectroscopic Biosensing and Deep Learning Postprocessing for Diagnosis of Endometrial Cancer †

Ejay Nsugbe [1,*], Dephney Mathebula [2] and Dawn Adams [3]

1 Nsugbe Research Labs, Swindon SN1 3LG, UK
2 University of South Africa; mathed2@unisa.ac.za
3 Ulster University; dawn.adams@southerntrust.hscni.net
* Correspondence: ennsugbe@yahoo.com
† Presented at the 10th International Electronic Conference on Sensors and Applications (ECSA-10), 15–30 November 2023; Available online: https://ecsa-10.sciforum.net/.

Abstract: Cancers are prevalent worldwide, affecting a substantial amount of the global population, while the early and proactive diagnosis of the disease continues to be a global medical challenge. Endometrial cancer represents a gynecological variant which is not only difficult to diagnose but also produces symptoms that are not distinct or exclusive to just the cancer itself. Blood spectroscopy has recently prevailed as a means towards a high-throughput and largely inexpensive method of diagnosing endometrial cancer. Using this method, and with the postprocessing of the accompanying spectra alongside the use of multivariate statistics, an inference can be formed which gives an indication of the presence and extent of the cancer. Previous work in this area has shown that the prediction results for this cancer could be improved with the use of signal decomposition models alongside machine learning prediction machines, thus demonstrating the potential appeal of decomposition models in the processing pipeline of the spectroscopy data. As part of this exploratory study, we employ for the first time the use of deep learning in the form of deep wavelet scattering, for the processing of acquired Fourier transform infrared (FTIR) spectra, which allows for a fully unsupervised decomposition and feature extraction of the resulting spectra, coupled with prediction machines capable of predicting the presence of the cancer. The obtained results show that the use of deep learning allows for enhanced predictions of endometrial cancer, whilst allowing for a clinical decision support platform which carries a greater degree of autonomy and therein diagnosis throughput.

Keywords: cancer; AI; machine learning; obstetrics; gynecology; FTIR; deep learning

Citation: Nsugbe, E.; Methebula, D.; Adams, D. A Novel Ensemble of FTIR Spectroscopic Biosensing and Deep Learning Postprocessing for Diagnosis of Endometrial Cancer. *Eng. Proc.* **2023**, *56*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Name

Published: 15 November 2023

1. Introduction

Endometrial cancer directly affects the lining of the uterus; it is one of the most diagnosed forms of cancer, and is also more prevalent in developing regions [1–5]. The formation of the cancer first involves structural changes within the endometrium due to hormonal variations, where prolonged exposure to certain hormones within the endometrium results in different initial variants of the cancer [1–5]. Risk factors include age, hormonal imbalances, genetic markers, and obesity, to name a few [1–5].

A symptom and direct manifestation of endometrial cancer is unusual uterine bleeding, some of the more frequently used diagnostic methods include endometrial biopsy processes, alongside transvaginal ultrasound methods [1–5]. Common treatment methods include: hysterectomies, vaginal brachytherapy, as well as medications depending on the overall stage of the cancer, followed by close monitoring of the behavior of the cancerous cells themselves [1–5].

Current means towards diagnosis of the cancer have been shown to carry undesired shortcomings, which has spurred the need for the exploration of other diagnosis mechanisms [1–5]. More effective means of diagnosis carry both cost saving implications as well as reducing the need for severe interventions such as significant clinical, pharmacological and surgical treatments, including hysterectomies [1–5]. Recent work has shown the promise of the use of blood biomarkers, alongside spectroscopic measurements, as a high-throughput means of triage and initial diagnosis, to be followed by invasive observations in the patients [1–5]. An illustration of the endometrial cancer disease can be seen in Figure 1.

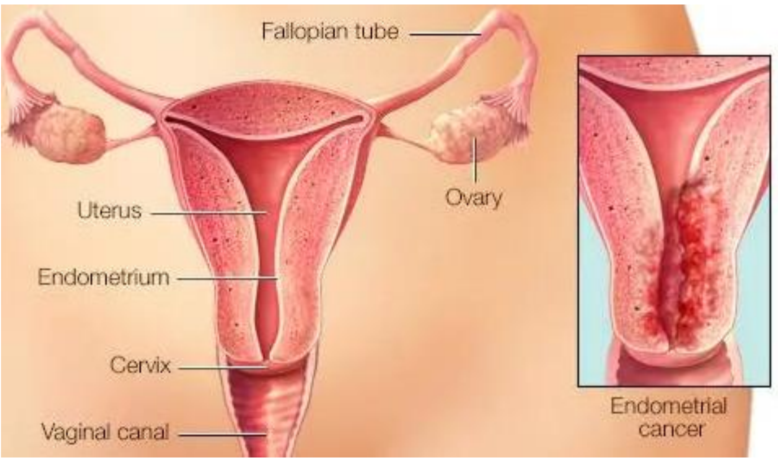

Figure 1. Illustration of endometrial cancer [6].

Related work has shown further results in the investigation of this theory, for example, Paraskevaidi et al. [7] and Nsugbe et al. [5] in the use of a combination of blood biomarkers, and FTIR spectra towards the classification and recognition of different variants of endometrial cancer infections. Paraskevaidi et al.'s work utilized primarily multivariate statistics towards creating various discriminatory based models, while Nsugbe et al. utilized a novel approach based around spectra decompositions and machine learning to assemble prediction models [5,7–9]. Nsugbe et al.'s work brought to light the potential clinical value of the application of multiresolution and signal decomposition algorithms within the area of spectroscopy postprocessing [5,8,9]. Deep wavelet scattering (DWS) represents a multiresolution-based approach which also allows for unsupervised feature extraction, and is structurally an ensemble of both the classical wavelet transform and the deep learning-based convolutional neural network (CNN) [10]. Recent work has seen the application of DWS in various capacities within clinical medicine, which has shown to be beneficial in not requiring any expert knowledge regarding the feature extraction aspect of the process, whilst also being able to perform a decomposition act [11–13]. The majority of this has been carried out on primarily on time-series data. In this work we investigate the use of DWS for the first time on the use of spectroscopic data for the prediction of various kinds of cancers using Paraskevaidi et al.'s FTIR spectroscopic data [7,11–13].

From this, it is hypothesized that a combination of the blood spectroscopy, FTIR and DWS, alongside pattern recognition models, can help form a rapid high-throughput means for an initial triage and diagnosis of endometrial cancer, which requires minimal expert intervention due to its unsupervised nature.

2. Materials and Methods

2.1. Dataset

The FTIR data utilized in this study comprised of 242 noncancerous patients, 258 with type 1 endometrial cancer, and 64 with type 2 endometrial cancer; further insights on the patient cohort can be found in the publication by Paraskevaidi et al. [7]. The recruitment of the participants was done by the Manchester University NHS Foundation Trust, the Salford Royal Foundation Trust, and the Lancashire Teaching Hospital, with ethical approval given and patient consent provided prior to the start of the study. All the biopsy samples were labelled by certified gynecological pathologists as either normal or a variant of the endometrial cancerous disease [7]. The spectra were obtained from the blood samples using the Tensor 27 FTIR spectrometer with a Helios ATR attachment containing a diamond ATR crystal by Bruker Optics Ltd.

2.2. DWS

The DWS is based around the multiscale extraction of features in an unsupervised fashion, in a way which they are robust and continuous, and its architecture comprises of a merger between the wavelet transform and the CNN [10]. In an attempt to minimize the overall computational complexity of the method, preset values of the filters are set, which null out the need for iterative estimations, and make the method adept at working with a small set of samples due to this multiscale properties [10]. In the DWS, the deep CNN is used for the iterative applications, whilst the convolution is done via wavelets and nonlinear modules, as well as an averaging function. The implementation of the DWS in this paper involved a Gabor mother wavelet, a scale invariance of 1 s, the filter banks of 8 wavelets per octave in the first filter bank, as well as 1 wavelet per octave in the second filter bank.

2.3. Machine Learning Models

The discriminant analysis model, i.e., linear and quadratic (LDA and QDA), was employed, while the K-nearest neighbor was also utilized as part of this work with K selected as 1 [14]. Where these models have been specifically chosen largely due to their computational efficiency. All models were validated using the K-fold cross validation approach with K chosen 10, while the SMOTE algorithm was utilized for the purpose of class balancing.

3. Results

The results for the various machine learning exercises can be seen in Table 1, from which it can be seen that the DWS appears to be producing a better prediction accuracy, with the best performance of 71.6%, when compared with the prior method utilized in a previous publication [15]. This has thus provided a degree of statistical evidence that the DWS can indeed be utilized towards a combination of both spectra decomposition, whilst also performing unsupervised feature extraction, therein negating the need for a set of expert knowledge dependent feature extraction process.

Table 1. Accuracy of the machine learning exercises.

Model	Postprocessing Method from Nsugbe and Sanusi [15] (without LSDL) (%)	DWS (%)
LDA	57	59.7
QDA	n/a	64.1
KNN	71.3	71.6

Subsequent work to be done in this area would involve further optimization exercises in order to determine if the performance of the DWS can be improved, while also training the data on various other machine learning models with nonlinear decision boundaries. The results in Table 1 appear to suggest that these kinds of models are optimal for the case study being investigated.

4. Conclusions

Endometrial cancer is an increasingly common cancer variant which ranks as one of the more frequently diagnosed forms of the disease, with symptoms that typically feature uterine bleeding of various degrees. The use of blood spectroscopy has begun to be investigated in the literature, where FTIR spectroscopy has been used as a means towards the postprocessing of the blood samples acquired from the patients.

This work has investigated the use of DWS for the first time, which is a multiresolution unsupervised feature extraction method for the design of prediction models from FTIR spectra. The interim results from the case study carried out in this paper show potential and applicability of the approach towards the prediction of the presence of the endometrial cancer in patients with the disease. Further work would involve the use of a broader sample size of patients with the cancer, and optimization exercises to tune and improve the performance of the DWS, whilst also training other available machine learning models on the dataset to find a best fit model for the desired application. Furthermore, the exploration of unsupervised learning pattern recognition/machine learning models would be implemented.

Author Contributions: All authors contributed equally to the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data is available from a cited repository within the manuscript.

Acknowledgments: The authors would like to thank Brian Kerr from Kerr Editing for proofreading the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Amant, F.; Moerman, P.; Neven, P.; Timmerman, D.; Van Limbergen, E.; Vergote, I. Endometrial Cancer. *Lancet* **2005**, *366*, 491–505. https://doi.org/10.1016/S0140-6736(05)67063-8.
2. Parkin, D.M.; Pisani, P.; Ferlay, J. Global Cancer Statistics. *CA Cancer J. Clin.* **1999**, *49*, 33–64. https://doi.org/10.3322/canjclin.49.1.33.
3. Madison, T.; Schottenfeld, D.; James, S.A.; Schwartz, A.G.; Gruber, S.B. Endometrial Cancer: Socioeconomic Status and Racial/Ethnic Differences in Stage at Diagnosis, Treatment, and Survival. *Am. J. Public Health* **2004**, *94*, 2104–2111.
4. Clement, P.B.; Young, R.H. Endometrioid Carcinoma of the Uterine Corpus: A Review of Its Pathology with Emphasis on Recent Advances and Problematic Aspects. *Adv. Anat. Pathol.* **2002**, *9*, 145–184. https://doi.org/10.1097/00125480-200205000-00001.
5. Nsugbe, E. On the Use of Spectroscopy, Prediction Machines and Cybernetics for an Affordable and Proactive Care Approach for Endometrial Cancer. *Biomed. Eng. Adv.* **2022**, *4*, 100057. https://doi.org/10.1016/j.bea.2022.100057.
6. O'Hara, J. Mayo Clinic Q&A Podcast: Why Are More People Dying of Uterine Cancer? Available online: https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-qa-podcast-why-are-more-people-dying-of-uterine-cancer/ (accessed on 28 July 2023).
7. Paraskevaidi, M.; Morais, C.L.M.; Ashton, K.M.; Stringfellow, H.F.; McVey, R.J.; Ryan, N.A.J.; O'Flynn, H.; Sivalingam, V.N.; Kitson, S.J.; MacKintosh, M.L.; et al. Detecting Endometrial Cancer by Blood Spectroscopy: A Diagnostic Cross-Sectional Study. *Cancers* **2020**, *12*, 1256. https://doi.org/10.3390/cancers12051256.
8. Nsugbe, E. Particle Size Distribution Estimation of a Powder Agglomeration Process Using Acoustic Emissions. Ph.D. Thesis, Cranfield University, Cranfield, UK, 2017.
9. Nsugbe, E.; Williams Samuel, O.; Asogbon, M.G.; Li, G. Contrast of Multi-Resolution Analysis Approach to Transhumeral Phantom Motion Decoding. *CAAI Trans. Intell. Technol.* **2021**, *6*, 360–375. https://doi.org/10.1049/cit2.12039.
10. Andén, J.; Mallat, S. Deep Scattering Spectrum. *IEEE Trans. Signal Process.* **2014**, *62*, 4114–4128. https://doi.org/10.1109/TSP.2014.2326991.
11. Nsugbe, E. On the Application of Metaheuristics and Deep Wavelet Scattering Decompositions for the Prediction of Adolescent Psychosis Using EEG Brain Wave Signals. *Digit. Technol. Res. Appl.* **2022**, *1*, 9–24. https://doi.org/10.54963/dtra.v1i2.40.
12. Nsugbe, E.; Connelly, S. Multiscale Depth of Anaesthesia Prediction for Surgery Using Frontal Cortex Electroencephalography. *Healthc. Technol. Lett.* **2022**, *9*, 43–53. https://doi.org/10.1049/htl2.12025.

13. Nsugbe, E.; Olorunlambe, K.; Dearn, K. On the Early and Affordable Diagnosis of Joint Pathologies Using Acoustic Emissions, Deep Learning Decompositions and Prediction Machines. *Sensors* **2023**, *23*, 4449. https://doi.org/10.3390/s23094449.
14. Nsugbe, E.; Phillips, C.; Fraser, M.; McIntosh, J. Gesture Recognition for Transhumeral Prosthesis Control Using EMG and NIR. *IET Cyber-Syst. Robot.* **2020**, *2*, 122–131. https://doi.org/10.1049/iet-csr.2020.0008.
15. Nsugbe, E.; Sanusi, I. Towards an Affordable Magnetomyography Instrumentation and Low Model Complexity Approach for Labour Imminency Prediction Using a Novel Multiresolution Analysis. *Appl. AI Lett.* **2021**, *2*, e34. https://doi.org/10.1002/ail2.34.