A Novel Ensemble of Fourier Transform Infrared Spectroscopic Biosensing and Deep Learning Postprocessing for Diagnosis of Endometrial Cancer †

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

Abstract: Cancers are prevalent worldwide, affecting a substantial amount of the global population, while 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 models, 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 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

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 and 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 hysterecromies [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](image-url)

Related work has shown further results in the investigation of this theory; for example, Paraskevaidi et al. [7] and Nsugbe et al. [5] used a combination of blood biomarkers and FTIR spectra for the classification and recognition of different variants of endometrial cancer infections. Paraskevaidi et al.’s work utilized primarily multivariate statistics to create various discriminatory-based models, while Nsugbe et al. utilized a novel approach based on 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 primarily on time-series data. In this work, we investigate the use of DWS for the first time on 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 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 242 noncancerous patients: 258 with type 1 endometrial cancer and 64 with type 2 endometrial cancer; further insights into the patient cohort can be found in the publication by Paraskevaidi et al. [7]. The recruitment
of the participants was performed 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 of the biopsy samples were labelled by certified gynecological pathologists as either normal or a variant of endometrial cancer [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. (Ettlingen, Germany).

2.2. DWS

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 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 the need for iterative estimations and make the method adept at working with a small set of samples due to these multiscale properties [10]. In DWS, the deep CNN is used for iterative applications, whilst the convolution is performed via wavelets and nonlinear modules, as well as an averaging function. The implementation of DWS in this paper involved a Gabor mother wavelet, a scale invariance of 1 s, and 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]. 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 to be 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 showing that DWS can indeed be utilized towards spectra decomposition whilst also performing unsupervised feature extraction, therein negating the need for an expert knowledge-dependent feature extraction process.

Table 1. Accuracy of the machine learning exercises.

<table>
<thead>
<tr>
<th>Model</th>
<th>Postprocessing Method from Nsugbe and Sanusi [15] (without LSDL) (%)</th>
<th>DWS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>57</td>
<td>59.7</td>
</tr>
<tr>
<td>QDA</td>
<td>n/a</td>
<td>64.1</td>
</tr>
<tr>
<td>KNN</td>
<td>71.3</td>
<td>71.6</td>
</tr>
</tbody>
</table>

Subsequent work to be conducted 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 the potential and applicability of the approach towards the prediction of endometrial cancer in patients. Further work should 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 should be implemented.

**Author Contributions:** Conceptualization, E.N.; methodology, E.N.; software, E.N.; validation, E.N.; formal analysis, E.N.; investigation, E.N.; resources, E.N.; data curation, E.N.; writing—original draft preparation, E.N.; writing—review and editing, E.N., D.M. and D.A.; visualization, E.N.; supervision, E.N.; project administration, E.N.; funding acquisition, E.N.; D.M. and D.A. 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.

**Institutional Review Board Statement:** The study was conducted within ethical standards, details of which have been included within the manuscript.

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

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

**Acknowledgments:** The authors would like to thank Brian for proofreading the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


15. Nsugbe, E.; Sanusi, I. Towards an Affordable Magnetomyography Instrumentation and Low Model Complexity Approach for Labour Imminence Prediction Using a Novel Multiresolution Analysis. *Appl. AI Lett.* **2021**, *2*, e34. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.