SCIENTIFIC Publishing Limited

Digital Technologies Research and Applications

http://ojs.ukscip.com/index.php/dtra

Article

On the Use of Raman Blood Spectroscopy and Prediction Machines for Enhanced Care of Endometriosis Patients

Ejay Nsugb *

Nsugbe Research Labs, UK

* Correspondence: ennsugbe@yahoo.com

Received: 29 September 2022; **Accepted**: 30 May 2023; **Published**: 6 June 2023

Abstract: Endometriosis is a prevalent disease of the female endometrium which affects women of all ethnicities and has been seen to be most common in the 25–35 years age group. The disease does not have a definitive cure, hence care and management are the essential components towards dealing with the disease. At present, the predominant means towards the diagnosis of the presence of the disease involves different imaging modalities alongside laparoscopy, where the instrumentation is expensive to acquire and requires clinical expertise. Recently, work has been done by an author who leveraged Raman blood spectroscopy alongside machine learning towards an affordable high throughput means towards the prediction of endometriosis. This work utilises the Raman blood spectroscopy dataset alongside advanced signal processing, machine learning and clinical cybernetics, towards the design of a prediction machine which sits within a clinical framework to facilitate Human-Machine interaction for an enhanced care strategy for patients with endometriosis. The prediction machine is designed to initially predict whether a patient has the disease, and is then followed by the use of unsupervised learning to form an inference means towards predicting the extent of the disease. The results showed that a combination of the adopted methods could allow for a high prediction of the endometriosis disease. Subsequent work in this area would now include further optimisation of the prediction machine in order to potentially maximise the prediction accuracy.

Keywords: machine learning; public health; endometriosis obstetrics and artificial intelligence

1. Introduction

As far back as the 1600s, Daniel Shroen, a German physician, was amongst the premier clinicians to describe endometriosis in his seminal work "Disputatio Inauguralis Medica de Ulceribus Ulceri", alongside its chronic inflammatory symptoms [1].Endometriosis is a condition where there exists the presence of endometrial glands and soma outside of the uterine, and has been viewed as an oestrogen-dependent disease which is characterised by chronic pelvic pain and infertility, where it is seen to be prevalent in females in the age range of 25–35 years[2–5]. There exist multiple theories around the pathogenesis of the en- dometriosis phenomenon, one of which is based around the retrograde menstruation theory where endometrial fragments arepulsed out of the fallopian tubes from a force gradient emanating from dyssynergic uterine contractions [6]. Once these endometrial fragments arrive at the peritoneal cavity, they ultimately begin to grow and eventually commence with an invasion of the pelvis and its surroundings, where this phenomenon is heightened by factors that amplify the regurgitation of the endometrium such as early age at commencement and/or prolonged menstrual flows, as well as cellular manifestations that encourage cell implantation and growthat anomalous locations [1,6].

With endometriosis being an oestrogen-dependent disease, progesterone is viewed as a antagonistic hormone in dealing with endometriosis since its actions trigger an inhibition of oestrogen-based proliferation of various cells [1]. A common characteristic associated with endometriosis is inflammation due to the presence of ectopic tissue within the peritoneal cavity, resulting from a heightened production of proteins and lipids such as prostaglandins and cytokines [7].

An illustration showing the varied locations of endometriosis across the pelvis and the overall abdomen can be seen in Figure 1, while an image of left ovarian endometriosis can be seen in Figure 2.

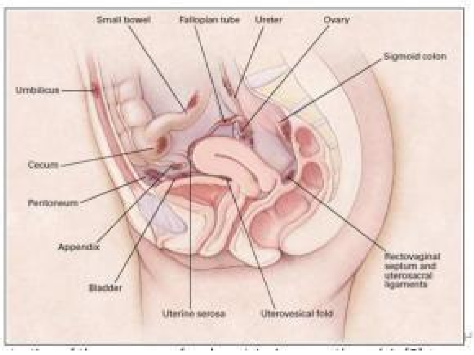

Figure 1: An illustration of the presence of endometriosis across the pelvis.[8]

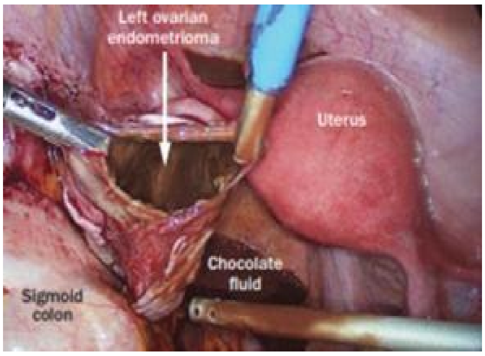

Figure 2: Image of ovarian endometriosis. [9]

In addition to early and prolonged menstrual flows, some other prominent risk factors for endometriosis include: a low body mass index (BMI), Caucasian ethnicity, consumption of 10 grams or more of alcohol per day, and a combination of a low BMI alongside infertility [1].

Food and diet have been said to play some role in the prevention of endometriosis, where the consumption of vegetables and fruits have been known to support the immune system due to their removal of free radicals, while on the other hand red meats have been seen to harness a negative effect due to their contents being responsible for driving up oestrogen contents in the body [1] The endometriosis disease has also been seen to carry considerable economic implications, where in European countries, the cost of care for each patient has been estimated to be close to £10,000, while in the USA medical costs have been seen to be in the region of $700 per month, and finally in Canada extrapolated figures have yielded an annual estimate of $1.8 billion for combined treatment and care costs [1,10], in addition to social and healthcare implications.

One of the pathological means towards the distinguishing of endometriosis is the use of four classes/stages, namely, minimal, mild, moderate and deep infiltrating endometriosis. Table 1 shows the various symptoms associated with the classes [11].

Table 1: Various classes of endometriosis and their associated primary symptoms.

Class of Endometriosis	Primary Symptoms
Minimal and mild endometriosis	Painful menstrual periods Irregular menstrual periods Dysmenorrhea Heavy menstrual flow
Moderate and deep infiltrating endometriosis	Tenderness Dysmenorrhea Restricted mobility and pelvic pain Heavy menstrual flows

Although the main mode of the diagnosis of endometriosis has been seen to be via histopathological methods, recently alternative methods spanning gynaecological examinations, imaging, laparoscopy, and biochemical characterisations have been investigated [1].The imaging aspects typically involve an ultrasound scan which by example can pick up endometrial cysts as well as other congenital conditions which prompt the flow of menstrual blood into the peritoneal cavity [12–14] Occasionally, magnetic resonance imaging (MRI) is employed as per the patient need, while the (minimally) invasive laparoscopic surgery has been widely viewed as the most frequently used method for the diagnosis of endometriosis [12–14].

The treatment of endometriosis is an area of uncertainty due to the nature of the chronic disease, where current treatment therapies are primarily based around pharmacological, surgical and physiotherapy care measures [1]. The choice of what treatment pathway is taken typically depends on a mixture of patient preference alongside symptoms, and a perceived estimation of the severity of the disease [1,15]. The administering of medication to patients has been typically seen to be for patients who are in their reproductive years, where the aim of the medication-based treatment is to alleviate pain and inhibit the further spread of the disease towards neighbouring regions [1]. A brief example of medications administered to patients who are nursing endometriosis includes: hormonal medications, non-steroidal anti-inflammatory medication for pain relief and reduction of swellings, progesterone receptor modulators, and aromatase inhibitors[1].

Surgical treatment in endometriosis is usually prompted by patients with chronic pelvic pain, requiring fertility treatments or those who have ovarian cysts where, as mentioned, the minimally invasive laparoscopic surgery is the preferred means of surgery [16].

Physiotherapy is also used as part of the rehab of endometrial patients who begin to struggle with mobility due to the disease causing chronic inflammations in and around the pelvis [17–19]. The physiotherapy courses are based around supporting mobility with an ultimate restoration of the tissues around the pelvis area, while exercise has also been seen to be beneficial in combatting endometriosis as it has been said to combat inflammations, ovarian stimulations, muscle relaxations and also provide psychological reprieves from anxiety and depression in women who may be undergoing hormone replacement therapies [17–19].

Prior work in the area of machine learning and endometriosis has involved the leveraging of various data sources, alongside the relevant computational models, towards the prediction of a patient's genetic susceptibility to endometriosis, and the use of various kinds of clinical imaging alongside deep learning towards interpreting patient imaging data [20–26]. The prediction models from the mentioned case studies have proposed their models as independent systems without a concise account for the Human-Machine interaction that is necessary for the potential clinical deployment of the model.

Parlatan et al. [26] utilised a unique approach comprising of Raman blood spectroscopy alongside machine learning to serve as a means towards an affordable screening with a high throughput for the prediction of the presence of endometriosis in patients. Their modelling approach utilised the raw spectroscopy data as input without employing feature extraction and therein leading to high dimensionality which could ultimately cause for model overfit [26]. Furthermore, their model validation approach appears to have been done on a small subset of the overall data [26].

As part of the work done and the contributions made as part of this paper, we aim to utilise signal processing techniques, alongside cybernetics, in the design of a prediction machine that is capable of predicting endometriosis and is hosted within a clinical cybernetic loop which is poised towards enhancing care strategies for endometrial patients[27] .

This work adopts the approach of using the principles of cybernetic steering towards the creation of a system that hosts both a prediction ma- chine alongside clinical expertise for an enhanced decision support clinical platform [27,28]. This framework was showcased in the area of pregnancy medicine for the prediction of preterm that allowed for collaboration, shared decision support and an information flow approach between the midwife and obstetrician, alongside the prediction machine which produces indications of potential preterm pregnancies and a subsequent recommendation based on patient information [28]. This approach has also recently been adopted as part of the growing reprieves from anxiety and depression in women who may be undergoing hormone replacement therapies [17–19] .area of intelligent gynaecology and oncology, which is based around the utilisation of tools such as artificial intelligence and cybernetics towards achieving things such as early diagnosis, affordable screening for low income settings and, of course, greater patient care [29]. The particular case study in which the authors utilised the cybernetic principle within the area of oncology was used for the early prediction of cervical cancer from patient medical health records, which is a prominent gynaecological disease [29]. That work hosted the prediction machine in a loop that also included clinical experts as per the trend with these systems, where the goal was to not only leverage patient in- formation as a means towards early detection of the disease, but also as a means towards an affordable and high throughput patient care, which can also be appealing and applicable to low income settings who may lack resources for cutting edge diagnosis instrumentation [29].

Likewise, this work sees a contribution within this growing area where we utilise a data source from Raman blood spectroscopy towards the assembly of a cybernetic system for an affordable means towards predicting the presence of endometriosis in female patients, in a framework which also includes oncology clinical experts, and is ultimately poised towards greater and enhanced care for patients with the disease. Specifically speaking, the contributions of this manuscript are as follows:

- The proposition of a clinical cybernetic system which hosts a prediction machine and clinical experts

- The use of signal processing and machine learning methods for the prediction of the presence of endometriosis from Raman blood spectroscopy

- The utilisation of unsupervised learning for the development of a distance and proximity framework for the estimation of the extent of the endometriosis disease inpatients.

2. Dataset

The dataset utilised as part of this study was acquired by Parlatan et al. [26] and received ethical approval from the Faculty of Medicine, Acibadem University where each participant produced a written consent prior to the commencement of the data collection process. The sample collection involved the acquisition of blood samples which were centrifuged to isolate the serum, where the subsequent serum samples were stored at 4 degrees Celsius for a maximum of two days following its collection, and for the measurement exercise, 0.5 ml of the serum was prepared in a quartz cuvette [26].

The sample analysis involved the use of a water immersion microscope 60X, NA, Olympus, a single mode diode laser with a wavelength of 785 nm and power of 100 mW for the Raman excitation [26]. A Faraday isolator was used for optical filtering to remove unwanted back reflected beams, while a laser line was utilised to cleanse the laser profile around the 785 nm waveband, and samples were illuminated through a focused lens of sorts and then the subsequently back scattered light at 180 degrees was acquired [26]. Further filtering was done with Rayleigh scattered photons using sequentially scattered Raman edge filters, while the focusing of the Raman scattered beams was done using a 100 micrometre slit spectrometer with an achromatic lens with a focal length of 50 mm [26]. The spectrometer that was used comprised of 600 lines/mm grating alongside a thermoelectric cooled CCD camera.

As part of the pre-processing of the data, calibration, background subtraction, baseline correction and normalisation were performed on the data. Further details on the experimental acquisition of the dataset can be seen in Parlatan et al. [26].

For the signal processing exercise detailed in the subsequent sections, a decomposition algorithm was also applied to the dataset as comparison with the analysis with the raw data. Thus, due to the data analytical limitations of the algorithm, a total of 144 Raman blood spectroscopy samples were used for the work done in this paper, amounting to a total of 72 patients' (36 endometriosis and 36 normal) data that was used for the analysis, considering measurements were repeated twice for each patient within 24 hours, where the data files comprised of Raman data in the waveband of 350–1800 nm.

3. Raman Spectroscopy

The Raman spectroscopic technology is an example of light-matter interaction. The Raman scattering of light by molecules is based on and was initially predicted by Smekal in 1923 using quantum theory before being subsequently validated by Raman in 1928, where to date, there now exist 25 distinct kinds of Raman spectroscopic technologies [30,31]. Raman technologies are now of prominent use for optical measurements which involve liquids, due to their absorbance properties within the infrared region. Within the area of medicine and biology, the Raman technology is one which is of keen interest for substance characterisation as it does not require fluorescent marker molecules, is non-invasive and can be used to gain molecular and concentration information from an array of biological substances [30,32].

The fundamental operating principle behind the Raman spectroscopy is based around the interaction of light with matter, where the oscillatory electromagnetic field causes a perturbation of the charge distribution within matter that causes a bi-directional exchange of energy and the associated momentum while leaving the source matter in a modified state, examples of this being the rotational vibrations in liquids and gases [30,33].

In the generic form of the Raman spectroscopy, the technology takes the form and abides to the operating principles of spontaneous emissions spontaneous Raman, which involves the interaction of an incident photon with a molecule which is subsequently scattered [30]. The resulting scattering in the case of light particles is mostly elastic and is referred to as Rayleigh scattering, whereas inelastic scattering—where the incident energy does not equal the energy of the scattered photon— is referred to as the Raman effect [31,34]. For an inelastic process in the case of a lattice, the energy transfer process creates a vibration quanta, also referred to as quasi-particle/phonon [30]. The Raman scattering effect can also give rise to spin waves whose shift in angular frequency (w) of scattered light can be described as Equation (1):

$$w_{scat} = w_p \pm w_{osc} \quad (1)$$

Where scat is the scattered light, p is the incident/pump photon, and osc denotes the molecular and lattice vibration. The binary operator ± is dictated by the energy conservation.

On the occasion where the energy of the scattered photon is lower than that of the incident photon, this phenomenon is referred to as Stokes Raman scattering, and likewise when the energy of the scattered photon is higher than the incident photon, this is termed as anti-Stokes Raman scattering [30]. The Raman process obeys the law of momentum conservation and is expressed as equation 2 in wavevector form:

$$\vec{\ }_{d} = \vec{\ } \pm \vec{\ } \quad (2)$$

Where $\vec{\ }_{d}$, $\vec{\ }$, $\vec{\ }$ represent the wave vectors for the scattered light, incident light and the molecular vibration respectively. In molecules and crystals, charge distributions have a preferred equilibrium which they tend, where an external field has the ability to modify the charge distribution based on its ability to form anisotropic dipoles, referred to as polarizability and dielectric susceptibility [30].

The classical Raman effect has said to be associated with a modulation of the polarizability, which is due to the oscillatory dynamics of the inter atomic displacements [30] Mathematically, given a polarisation vector of a certain material, $\vec{\ }$, with *j* and *k* representing a set of vector components in the 3-dimensional *x*, *y* and *z* directions, where the first order approximation of the *j*th component of $\vec{\ }$ is linked to the oscillatory electric field vector $\vec{\ }$, which is associated with light in the form:

$$p_j^{\,1} = \varepsilon_0 X_{jk}^{1} E_k \quad (3)$$

Where ε_0 represents the permittivity of free space, $x_{j\,k}^{1}$ is a rank 2 tensor for dielectric susceptibility, while superscript 1 indicates a first order polarisation contribution, where the polarisation tensor represents a function of the nuclear coordinates, which is also dependent on the vibrational frequency [35]. For a small and borderline negligible modulation, this dependency can be expressed with the Taylor series in relation to the axis of vibration:

$$X^1_{jk}(\vec{k}_p, W_p) \approx X^1_{jk}(\vec{k}_p, W_p)\Big|_{u=0} + u_l\left(\frac{\partial X^1_{jk}\left(\vec{k}_p, W_p\right)}{\partial u_l}\right)_{u=0} + u_l u_m\left(\frac{\partial^2 X^1_{jk}(\vec{k}_p, W_p)}{\partial u_l \partial u_m}\right)_{u=0} + \cdots$$

where $\vec{u}$ is the nuclear displacement vector, indices *j*, *k*, *l* and *m* embody various spatial coordinates, expressing the accompanying electric field which accompanies the light as Equation (5):

$$\vec{E}(\vec{r}, t)=\vec{E}(\vec{k}_p,)\cos(\vec{k}_p \cdot \vec{r} - t) \quad (5)$$

While the nuclear displacement is expressed as Equation (6):

$$\vec{u}(\vec{r}, t)=\vec{u}(\vec{q}, {}_c)\cos(\vec{q} \cdot \vec{r} - {}_c t) \quad (6)$$

Where a direct dependency on time can be found by a substitution exercise involving the variables for monochromatic light and displacement. The particular term which relates to first order Raman scattering coefficients is derived from the middle term in Equation (7), and ultimately yields.

$$P_j(, t, \vec{u})=$$

$$\frac{1}{2}\varepsilon_0\left(\frac{\partial X^1_{jk}\left(\vec{k}_p, W_p\right)}{\partial u_l}\right)_{u=0} u_l\left(\vec{q}\right)\left(\vec{k}_p, W_p\right)$$

$$X\{\cos[(\vec{k}_p + \vec{q}) \cdot \vec{r} - (W_p + {}_c)t] + \cos[(\vec{k}_p - \vec{q}) \cdot \vec{r} - (W_p - {}_c)t]\} \quad (7)$$

These terms comprise both the Stokes and anti-Stokes frequencies, which also indicate the conservation of momentum. Figure 3 shows an energy level figure for a Raman process with Stokes and anti-Stokes emissions (W_S and W_{AS}) [30].

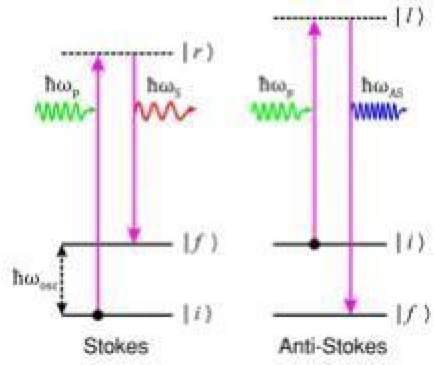

Figure 3: Energy level diagram for a Raman process.

4. Methods

4.1. Cybernetic System

The concept of cybernetic steering emanates from Greek, which implies a methodical steering and guidance towards a desired outcome. This term was formalised and popularised by the seminal work of Norbert Wiener and is now a baseline from which multi-component systems can be synthesised via its hierarchical framework, which is comprised of both the 1st and 2nd order cybernetic framework [27,28,36]. The 1st order cybernetics analyses the technical components of the system and its feedback elements which allow for the systematic steering ofthe system itself, while the 2nd order cybernetic framework provides a broad scope lens through which the end-to-end operation of the system is philosophically analysed and is therein referred to as the "cybernetics of cybernetics" [37,38]. We would be adopting this hierarchical framework towards the characterisation, synthesis and analysis, of the proposed cyber-human cybernetic system shown in Figure 4, which is poised towards a clinical expert/prediction machine collaboration for enhanced care for patients with endometriosis [39,40].

(where |i> = initial state, |f> = final state and |r> = incident pump photon, and his Plank's constant) [30].

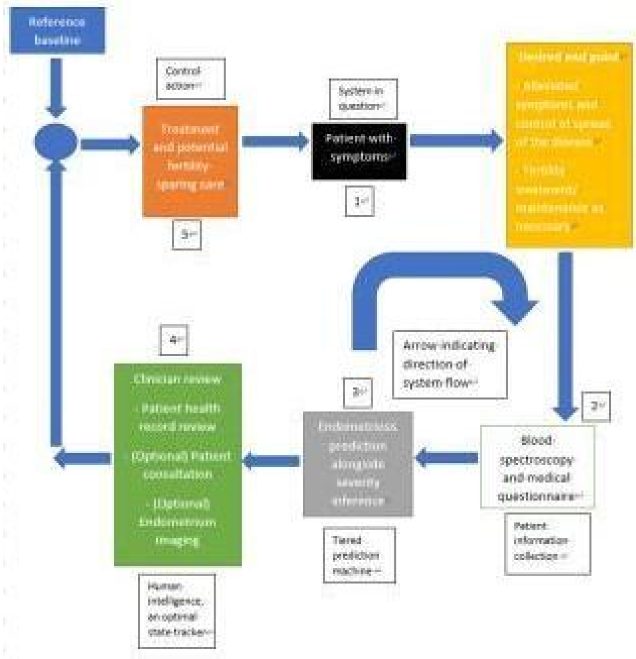

Figure 4: A system diagram showing the proposed cybernetic system.

4.2. 1st Order Cybernetics

In this section, the system components, their functions and associated interfacing and interactions for the system presented in Figure 4 are described as follows:

4.2.1. System/Patient with Endometriosis: the system which is the focal emphasis of the cybernetic system is a female patient who is experiencing symptoms which are related to the various stages of endometriosis detailed in Table 1.

4.2.2. Data Acquisition: the proposed means of information and data collection is pitched to be a non-invasive approach which involves the acquisition of blood samples from a patient, which is analysed using a Raman spectroscopic method and serves as the primary information source. This means of data acquisition has the merit of being minimally invasive and affordable, alongside possessing a high throughput turnaround.

Although not covered in this paper, an auxiliary source of data collection can also be in the form of questionnaires, as shown in the work done by Bendifallah et al. [20] where they leveraged information from a questionnaire which contained information about the patients' age, BMI, menstrual cycle, intercourse, and the presence of blood in the urine, to build predictive models.

4.2.3. Prediction Machine/State Estimation: the prediction machine is a machine learning underpinned model which is capable of, in this case, predicting whether a patient has endometriosis. The prediction machine for this particular case study is of a tiered format, where the rest set of prediction is of a binary nature to indicate whether a patient has endometriosis. If yes, an inference method is used to predict the potential stage of the disease. The models and algorithmic methods utilised as part of the design of the prediction machine is detailed subsequently.

4.2.4. Clinician Review/Optimal State Tracker: the purpose of this stage is to embody the role of an optimum state tracker, where a clinical expert reviews the prediction made by the machine and analyses the information with respect to the patient medical health record to detect potential false positives. In settings where resources permit, trans- vaginal ultrasound could also be done to verify the presence of endometriosis as part of this stage.

4.2.5. Control Action/Treatment: the control action as part of the proposed cybernetic loop is the application of a course of treatment therapy to the patient. As endometriosis is an oestrogen-dependent disease, the general course of treatment involves the administering of progesterone-based medication which contributes towards hormonal balancing and a potential prevention of the spread of the endometriosis disease. In an attempt to standardise treatment practices for endometriosis, Olive et al. proposed two sets of treatment flows for patients with the disease, as shown in Figure 5, where the first pathway is based around the initial administering of medication, followed by a culmination of laparoscopic surgery if the medications do not suffice [8]. On the other hand, the second pathway comprises of a combination of disease and fertility treatment where laparoscopic treatment is rendered, followed by assisted reproductive measures.

4.3. 2nd Order Cybernetics

As described, the 2nd order cybernetics effectively describes the "cybernetics of cybernetics", where the end-to-end operability of the system is synthesised and analysed [41]. The proposed system is of a cyberhuman nature which hosts both a clinical expert (human intelligence) and a prediction machine (machine intelligence), where the input from both of these parties are fused together to provide an enhanced care strategy and platform for patients with suspected endometriosis. The presence of the prediction machine serves as a supporting source of information and decision support to the clinician for more informed decision making. With this framework it is anticipated that overall better care and prioritisation can be provided with patients who are suffering from the disease. Some of the achievables that would be in place due to the proposed collaborative dynamic in the cybernetic framework include a potential earlier and affordable means of detecting the presence of endometriosis with a high throughput, with the ability to prioritise care based on the severity of the diagnosis. In addition to this, patients can also be instantaneously placed on fertility-sparing treatment based on the perceived extent and severity of the disease.

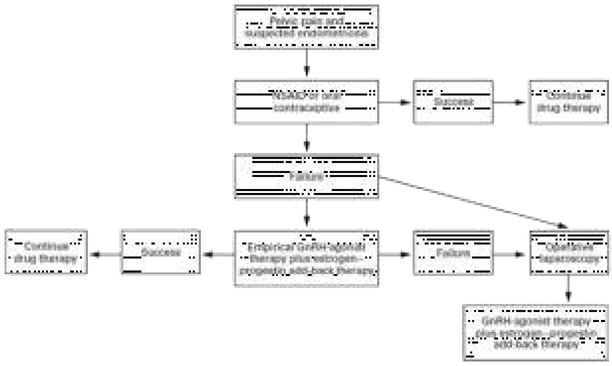

Figure 5: Potential treatment pathways for endometriosis patients, as proposed by Olive et al. [8]

4.4. Signal Processing and Machine Learning

4.4.1. Spectroscopy Decomposition

As part of the signal processing work done in this paper, comparison work was done utilising the raw spectroscopy data, which was subsequently processed and the decomposed spectroscopy waveform put through the Linear Series Decomposition Learner (LSDL), which provided positive results and showed movement in prediction exercises involving infrared spectroscopic data in previous studies[42–47]. The LSDL has been seen to be a meta-heuristically driven signal decomposition method which, although was incepted and primarily poised for the decomposition time-series signals, the application of the algorithm has resulted in the minimisation of the overall uncertainty within a sample time-series waveform[43–47]. As mentioned, its application has now been extended towards spectroscopic data where prediction accuracies were noted to be of improvement, thus showing signs to also be able to decompose spectroscopy waveforms; therefore, we seek to observe the applicability of the LSDL towards Raman spectroscopic data for the prediction of endometriosis. For a given waveform expressed as | | ,the comprehensive lists of heuristics and parameters used in the tuning of the LSDL can be seen in Nsugbe et al. [43] and Nsugbe and Sanusi [46].

The result of the selection of the optimal threshold region for the LSDL showed that the 2nd iteration of the lower threshold produced the optimal waveform decomposition. The result of the obtained values for the performance decomposition performance index can be seen in Table 2.

4.4.2. Feature Extraction

As for the characterisation of the waveform spectra, feature extraction was employed as opposed to the feeding the machine learning model the full raw spectroscopic data, which can make the classifier suffer from the curse of high dimensionality and increasing the potential of a model overfit [48]. The following features were extracted from the candidate spectra: mean absolute value (MAV), waveform length (WL), slope sign change (SSC), root mean square (RMS), fourth order autoregressive coefficient (AR), cepstrum (Ceps), maximum fractal length (MFL), Higuchi fractal dimension (HFD), median frequency (MF) and number of peaks (NP) [28,49,50]. The value of 1 µv was utilized for features which required thresholds, while k was selected as 10 for the HFD. The group of features comprise an ensemble of multiple kinds of features which could make for a thorough characterisation of the Raman spectra.It can be said that the list of features assembled and utilised in this study stems from and is inspired by previous studies, of which these group of features have been deemed to be capable of reliably modelling physiological signals from various sources[43,49].

Table 2: Signal decomposition results for the LSDL.

	1st Iteration	**2nd Iteration**	**3rd Iteration**	**4th Iteration**
Upper Threshold Region	n/a	n/a	n/a	n/a
Lower Threshold Region	2.0001	2.0008	2.0001	2.0001

4.4.3. Machine Learning

Supervised Learning

The following Supervised Learning methods were utilised as part of this study as they represent a range and variety of different model architectures of varying complexities, which can allow for the observation of the optimal model given the nature of the dataset in this study.

- Decision Tree (DT): is a machine learning model which works with a Boolean-like logic flow towards sorting data into various distinct classes in a tree-like manner, and is also a non-parametric approach [51].

- Discriminant Analysis: is a statistical method primed towards first the projection of a feature vector into a lower dimensional space, followed by the instillation of class boundaries between classes which can either be linear (LDA) or non-linear (QDA) [49,50].

- K-Nearest Neighbour (KNN): is an approach which uses a voting strategy based on a nearest neighbour criteria towards assigning data samples to a category, in this work the Euclidean distance was used as the distance metric of choice, and the k value was chosen to be 1 [52].

- Support Vector Machine (SVM): is an iterative classification model which utilises a subset of the data —referred to as a set of support vectors—towards the instillation of class boundaries to separate classes of data from each other, where the implementation of the class boundaries is typically done in a higher dimensional space using a technique termed as a 'kernel trick' [53]. Different variants of the SVM algorithm have been adopted for use in this paper, namely, the linear (LSVM), quadratic (QSVM), cubic (CSVM), and fine Gaussian (FGSVM) kernels.

- Multi Layer Perceptron Neural Network(MLPNN): is a data driven function approximation model that is capable of mapping the relationship between input data and output label in a non-parametric manner, where data classes can be separated using a non-linear decision boundary when a non-linear activation function is selected [54]. The implemented version of the MLPNN comprised a single hidden unit with 30 neurons, while the backpropagation algorithm was utilised for the training of the network and the softmax was used in the output layer.

- Logistic Regression (LR): is a statistically driven means of binary classification where a sample is designated a class based on its value with respect to a threshold and asigmoid curve [55].

All of the supervised learning models were validated using the K-Fold cross-validation method where K was selected as 10, with the data split being in the format of 80:20, with 80% of the dataset used towards the training of the models and the remainder 20% utilised for cross-validation purposes. The MATLAB classification learner application was utilised for these exercises.

Unsupervised Learning

The K-Means algorithm was selected for use in this study due to a combination of computational efficiency and capability in carrying out unsupervised learning exercises.

- K Means: is an iterative means towards the sorting of data into various clusters based on the Euclidean distance criteria, where the algorithm seeks to maximise the distance between various clusters, while minimising inter-cluster variability with the placement of the cluster centroid [56]. The clustering assignment method works with the expectation-maximisation (E-M) method, where the E step involves the initial assignment of the clusters, while the M is the stage where results are iteratively updated.

5. Results

For the characterisation of the model performance, four select model performance metrics were selected for use as adopted in a previous study, namely: accuracy (ACC), sensitivity (SEN), specificity (SPEC), and area under the curve (AUC)[46].

5.1. Endometriosis Prediction

The results for the endometriosis prediction for the raw spectroscopic data can be seen in Table 3, where it can be seen that the LDA, MLPNN and LR produced the best model performances across the various classifiers which were trialled as part of the study. From Table 3 it can be seen that the utilisation of high order kernels with the case of the SVM did not yield a higher classification accuracy. Although it is not precisely apparent why the mentioned models produced the best classification performance, it can be seen that these model architectures produced compatibility with the dataset being used as part of this study.

Table 3: Endometriosis prediction using raw spectroscopy data.

Machine Learning Model	**ACC (%)**	**SEN (%)**	**SPEC (%)**	**AUC (%)**
DT	61	61	61	61
LDA	71	69	73	71
QDA	65	64	67	66
KNN	60	60	59	60
LSVM	69	68	70	69
QSVM	67	66	68	67
CSVM	60	60	61	61
FGSVM	61	60	63	62
MLPNN	72	74	71	73
LR	70	69	71	70

The results in Table 4 show the model performance for the LSDL pre-processed data, where it can be seen that the best results belong to the DT, although the results can be seen to be down from Table 3, which therein shows that that the LSDL decomposition does not have additional benefits for the prediction performance for the Raman spectroscopy data. Although in a previous study, the LSDL provided improved performance for infrared spectroscopy data, it can be seen that this does not generalise across the spectroscopic data from the Raman.

Table 4: Endometriosis prediction using LSDL spectroscopy data.

Machine Learning Model	**ACC (%)**	**SEN (%)**	**SPEC (%)**	**AUC (%)**
DT	**62**	**63**	**62**	**63**
LDA	**61**	**62**	**61**	**62**
QDA	**53**	**54**	**53**	**54**
KNN	**51**	**51**	**51**	**51**
LSVM	**58**	**60**	**57**	**59**
QSVM	**52**	**52**	**52**	**52**
CSVM	**51**	**51**	**51**	**51**
FGSVM	**50**	**50**	**50**	**50**
MLPNN	**62**	**63**	**62**	**63**
LR	**60**	**61**	**60**	**61**

5.2. Endometriosis Stage Prediction

The K-Means algorithm was adopted for an unlabelled means towards the partitioning of the data, where the results of the exercise can be seen in Table 5, and where it can be noted again that the raw data provided the best results for the clustering exercise. It is worth mentioning that, due to the K-Mean's random initialisation, the algorithm was run five times with the best clustering performance chosen as the optimal, as shown in Table 5.

5.2.1. Endometriosis Stage Prediction Algorithm

Leveraging the centroid and Euclidean distance information, we are able to devise a means of inferring the extent of the endometriosis dis- ease from the K-Means assignment. Given the centroid of the cluster, all samples have a given spread and distance from the centroid, where the closer a sample.

Table 5. Endometriosis stage prediction using the K-Means algorithm.

	ACC (%)	Class 1	Class 2
LSDL	60	47/72	39/72
Raw	64	41/72	51/72

Point is to the centroid the greater the confidence in the cluster estimation and thus the potential later stage a disease is. The further it is from the cluster centroid, the lower the confidence in its cluster assignment, and therein can be inferred an early stage of the disease. Thus, using distance as a proxy for the extent of the endometriosis disease, the following was created for the endometriosis cluster/Class 1 as shown in Table 1, where it should be noted that only the correctly clustered samples were used while the SMOTE synthetic sample generator was utilised for class balancing purposes, which resulted in a sum total of 50 samples used for the subsequent prediction exercises.

Driven largely by the distribution of the data, two classes of endometriosis extent were created, namely, Minimally-Mild Endometriosis and Mod- erate-Deep Endometriosis.

Table 6. Data class and Euclidean distance range.

Data Class		Euclidean Distance Range
Minimal-Mild dometriosis/Class 1	En-	1 : (0.5* sample with maximum Euclidean distance from centroid)
Moderate-Deep dometriosis/Class 2	En-	(0.5* sample with maximum Euclidean distance from centroid): (sample with maximum Euclidean distance from centroid)

The results of the endometriosis stage prediction for the two classes created in Table 1, can be seen in Table 7, from which a generally high prediction accuracy can be seen across the various different models, with the high order kernel/SVM producing the best classification performance, albeit by a slender margin compared to the other models. This goes to suggest that a non-linear decision boundary is highly effective in the separation of the samples in the various classes, but further work would need to investigate if this is consistent with a broader sample set.

A flow diagram of the various stages involved in the disease prediction can be seen as follows in Figure 6.

Table 7: Endometriosis stage prediction results.

Machine Learning Model	ACC (%)	SEN (%)	SPEC (%)	AUC (%)
DT	94	100	89	94
LDA	88	100	81	90
QDA	n/a	n/a	n/a	n/a
KNN	94	100	89	94
LSVM	88	100	81	90
QSVM	96	100	93	96
CSVM	96	100	93	96
FGSVM	96	93	100	96
MLPNN	92	100	86	93
LR	88	100	81	90

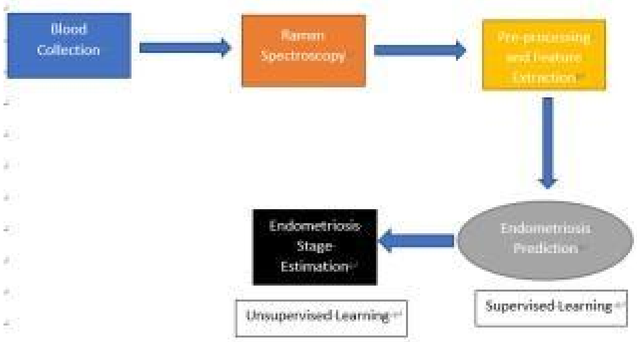

Figure 6: A stage flow diagram of the proposed disease prediction.

6. Conclusions

Endometriosis is a disease that affects a large number of women primarily in the age band of 25-35 years, and is characterised by painful periods, chronic pelvic pain and causes long term infertility issues, where health economics statistics appear to indicate that the disease is one with a considerable financial implication to society. The diagnosis of the diseases is typically done with a mixture of imaging modality, with laparoscopic means forming the bedrock for a formal diagnosis of the disease in areas where there is expertise, and the technology can be afforded. The main treatment options for the disease involve the administration of progesterone medications due to endometriosis being an oestrogen-dependent disease, and in certain cases surgical procedure for the removal of infected tissues and any possible organ adhesions that may have occurred.

As with most other areas of clinical medicine, the use of machine learning has steadily begun to gain traction in the diagnosis of endometriosis, where researchers have initially adopted the use of image recognition towards interpreting images from the endometrium as part of a means of detecting the presence of the disease. Parlatan et al. [26] proposed an alternate means of diagnosis involving the use of blood spectroscopy and Raman spectroscopy towards a high throughput means of diagnosis.

In this study, we applied advanced signal processing and machine learning alongside clinical cybernetics towards the design of a prediction machine that is capable of first predicting the presence of endometriosis from blood spectroscopy, followed by a subsequent inference of the extent of the disease. This prediction machine is scheduled to sit within the proposed clinical cybernetic loop that can make for a platform that allows for enhanced patient care. In the first instance, the results showed a reasonable prediction accuracy with the MLPNN, LDA and LR producing the best predictions made with the raw undecomposed spectroscopy data. The follow up on the prediction of the extent of the endometriosis disease which was done using the K-Means unsupervised learning, where the centroid and Euclidean distance metrics were leveraged towards an inference method for both early and lat- er stage endometriosis, which can help in the prioritisation of treatment and care in patients. The results showed a high prediction performance in and around the 90 % region across the various machine learning models when the clustered data were used to train various models as part of the pilot.

Subsequent work in this area is likely to involve the extraction and ranking of further features in order to understand the features driving the model performances, while also looking to seek out further features which could contribute towards boosting the overall prediction performance, in addition to the testing of the proposed cybernetic system.

Acknowledgements

The author would like to thank Michael Provost for reviewing the manuscript and Brian Kerr from Kerr Editing for proofreading this manuscript. The author would also like to thank Laura Williams for providing inspiration for the research conducted as part of this paper.

References

1. Smolarz B, Szyłło K, Romanowicz H. Endometriosis: Epidemiology, Classification, Pathogenesis, Treatment and Genetics (Review of Literature). Int J Mol Sci 2021;22:10554. https://doi.org/10.3390/ ijms221910554.
2. Giudice LC. Clinical practice. Endometriosis. N Engl J Med 2010;362:2389–98. https://doi.org/10.1056/NEJMcp1000274.
3. Missmer SA, Cramer DW. The epidemiology of endometriosis. Obstet Gynecol Clin North Am 2003;30:1–19, vii. https://doi.org/10.1016/s0889-8545(02)00050-5.
4. Parazzini F, Vercellini P, Pelucchi C. Endometriosis: Epidemiology, and Etiological Factors. In: Giudice LC, Evers JLH, Healy DL, editors. Endometriosis: Science and Practice, Wiley-Blackwell; 2012.
5. Viganò P, Parazzini F, Somigliana E, Vercellini P. Endometriosis: epidemiology and aetiological fac- tors. Best Pract Res Clin Obstet Gynaecol 2004;18:177–200. https://doi.org/10.1016/j.bpobgyn.2004.01.007.
6. Burney RO, Giudice LC. Pathogenesis and patho-physiology of endometriosis. Fertil Steril 2012;98:511–9. https://doi.org/10.1016/j.fertnstert.2012.06.029.
7. Bulun SE. Endometriosis. N Engl J Med 2009;360:268–79. https://doi.org/10.1056/ NEJMra 0804690.
8. Olive D, Pritts E. Treatment of Endometriosis. N Engl J Med 2000;345:266–75. https://doi.org/10.1056/NEJM200107263450407.
9. Vercellini P, Viganò P, Somigliana E, Fedele L. Endometriosis: pathogenesis and treatment. Nat Rev Endocrinol 2014;10:261–75. https://doi.org/10.1038/nrendo.2013.255.
10. Simoens S, Dunselman G, Dirksen C, Hummelshoj L, Bokor A, Brandes I, et al. The burden of endo-metriosis: costs and quality of life of women with endometriosis and treated in referral centres. Human Reproduction 2012;27:1292–9. https://doi.org/10.1093/humrep/des073.
11. SankaravadivelV, Thalavaipillai S. Symptoms based endometriosis prediction using machine learning. Bulletin of Electrical Engineering and Informatics 2021;10:3102–9. https://doi.org/10.11591/eei.v10i6.3254.
12. Raimondo D, Youssef A, Mabrouk M, Del Forno S, Martelli V, Pilu G, et al. Pelvic floor muscle dys- function on 3D/4D transperineal ultrasound in pa- tients with deep infiltrating endometriosis: a pilot study. Ultrasound Obstet Gynecol 2017;50:527–32. https://doi.org/10.1002/uog.17323.
13. Dietz HP. Pelvic floor ultrasound: a review. Am J Obstet Gynecol 2010;202:321–34. https://doi.org/10.1016/j.ajog.2009.08.018.
14. Youssef A, Montaguti E, Sanlorenzo O, Cariello L, Salsi G, Morganelli G, et al. Reliability of new three-dimensional ultrasound technique for pelvic hiatal area measurement. Ultrasound Obstet Gynecol 2016;47:629–35. https://doi.org/10.1002/uog.14933.
15. Kilpatrick RD, Chiuve SE, Leslie WD, Wegrzyn LR, Gao W, Yang H, et al. Estimating the Effect of Elagolix Treatment for Endometriosis on Postmenopausal Bone Outcomes: A Model Bridging Phase III Trials to an Older Real-World Population. JBMR Plus 2020;4:e10401. https://doi.org/10.1002/jbm4.10401.
16. Schipper E, Nezhat C. Video-assisted laparoscopy for the detection and diagnosis of endometriosis: safety, reliability, and invasiveness. Int J Womens Health 2012;4:383–93. https://doi.org/10.2147/ IJ WH.S24948.
17. Vural M. Pelvic pain rehabilitation. Turk JPhys Med Rehabil2018;64:291–9. https://doi.org/10.5606/tftrd.2018.3616.
18. Febbraio MA. Exercise and inflammation. J Appl Physiol (1985) 2007;103:376–7. https://doi.org/10.1152/japplphysiol.00414.2007.

19. Koppan A, Hamori J, Vranics I, Garai J, Kriszbacher I, Bodis J, et al. Pelvic pain in endometriosis: pain- killers or sport to alleviate symptoms? Acta Physiol Hung 2010;97:234–9. https://doi.org/10.1556/APhysiol.97.2010.2.10.
20. Bendifallah S, Puchar A, Suisse S, Delbos L, Poil- blanc M, Descamps P, et al. Machine learning algorithms as new screening approach for patients with endometriosis. Sci Rep 2022;12:639. https://doi.org/10.1038/s41598-021-04637-2.
21. PouliakisA, Margari N, Karakitsou E, Valasoulis G, Koufopoulos N, Koureas N, et al. Artificial Intelli- gence via Competitive Learning and Image Analysis for Endometrial Malignancies: Discriminating En- dometrial Cells and Lesions. International Journal of Reliable and Quality E-Healthcare (IJRQEH) 2019;8. https://doi.org/10.4018/IJRQEH.2019100102.
22. Kleczyk E, Peri A, Yadav T, Komera R, Peri M, Guduru V, et al. Predicting Endometriosis Onset Using Machine Learning Algorithms. 2020. https://doi.org/10.21203/rs.3.rs-135736/v1.
23. Akter S, Xu D, Nagel SC, Bromfield JJ, Pelch K, Wilshire GB, et al. Machine Learning Classifiers for Endometriosis Using Transcriptomics and Methylomics Data. Frontiers in Genetics 2019;10.
24. Visalaxi S, Muthu TS. Automated prediction of endometriosis using deep learning. International Journal of Nonlinear Analysis and Applications 2021;12:2403–16. https://doi.org/10.22075/ijnaa.2021.5383.
25. Lee JH, Kwon S-Y, Chang J, Yuk J-S. Machine Learning Approach to find the relation between Endometriosis, benign breast disease, cystitis and non-toxic goiter. Sci Rep 2019;9:5410. https://doi.org/10.1038/s41598-019-41973-w.
26. Parlatan U, Inanc M, Ozgor BY, Oral E, Baştu E, Unlu MB, et al. Raman spectroscopy as a non-invasive diagnostic technique for endometriosis. Scientific Reports 2019. https://doi.org/10.1038/s41598-019-56308-y.
27. Wiener N. Cybernetics or control and communication in the animal and the machine. 2. ed., reprint. Cambridge, MA, USA: MIT Press; 2007.
28. Nsugbe E, Obajemu O, Samuel OW, Sanusi I. Enhancing care strategies for preterm pregnancies by using a prediction machine to aid clinical care decisions. Machine Learning with Applications 2021;6:100110.
29. Nsugbe E. Towards the use of cybernetics for an enhanced cervical cancer care strategy. Intelligent Medicine 2022. https://doi.org/10.1016/j.imed.2022.02.001.
30. Jones RR, Hooper DC, Zhang L, Wolverson D, Valev VK. Raman Techniques: Fundamentals and Frontiers. Nanoscale Res Lett 2019;14:231. https://doi.org/10.1186/s11671-019-3039-2.
31. Das RS, Agrawal YK. Raman spectroscopy: Recent advancements, techniques and applications. Vibra- tional Spectroscopy 2011;2:163–76. https://doi.org/10.1016/j.vibspec.2011.08.003.
32. Butler HJ, Ashton L, Bird B, Cinque G, Curtis K, Dorney J, et al. Using Raman spectroscopy to char- acterize biological materials. Nat Protoc 2016;11:664–87. https://doi.org/10.1038/nprot.2016.036.
33. Long DA. The Raman effect: a unified treatment of the theory of Raman scattering by molecules. Chichester ; New York: Wiley; 2002.
34. Everall N, Hahn T, Matousek P, Parker AW, Towrie M. Photon Migration in Raman Spectroscopy. Appl Spectrosc 2004;58:591–7. https://doi.org/10.1366/000370204774103426.
35. Gardiner DJ, Graves PR, Bowley HJ. Practical Ra- man spectroscopy. Berlin; New York: Spring- er-Verlag; 1989.
36. Nsugbe E. A cybernetic framework for predicting preterm and enhancing care strategies: A review. Biomedical Engineering Advances 2021;2:100024. https://doi.org/10.1016/j.bea.2021.100024.
37. Attainable Utopias : First Order Cybernetics. At- tainable Utopias n.d. http://attainable-utopias.org /tiki/FirstOrderCybernetics#:~:text=First%20Order %20Cybernetics%20describes%20aspects,its%20whole%20situation%20or%20environment (accessed March 20, 2022).
38. Heylighen F, Joslyn C. Second-Order Cybernetics. VubAcBe 2001. http://pespmc1.vub.ac.be/SECORCYB.html (accessed April 21, 2022).
39. Malone TW. Superminds: the surprising power of people and computers thinking together. First edition. New York: Little, Brown and Company; 2018.
40. Provost MJ. Everything works wonderfully: an overview of servitization and asset man- agement. 2014.
41. Glanville R. SECOND ORDER CYBERNETICS 2008.https://www.pangaro.com/glanville/Glanville-SECOND_ORDER_CYBERNETICS.pdf.
42. Nsugbe E. On the use of Spectroscopy, Prediction Machines and Cybernetics for an Affordable and Proactive Care Approach for Endometrial Cancer Unpublished results.
43. Nsugbe E, William Samuel O, Asogbon MG, Li G. Contrast of multi-resolution analysis approach to transhumeral phantom motion decoding. CAAI Transactions on Intelligence Technology 2021;6:360–75. https://doi.org/10.1049/cit2.12039.
44. Nsugbe E. Particle size distribution estimation of a powder agglomeration process using acoustic emissions. Thesis. 2017.
45. Nsugbe E, Starr A, Foote P, Ruiz-Carcel C, Jennions I. Size differentiation of a continuous stream of par- ticles using acoustic emissions. IOP Conf Ser: Mater Sci Eng 2016;161:012090. https://doi.org/10.1088/1757-899X/161/1/012090.
46. Nsugbe E, Sanusi I. Towards an affordable magne- tomyography instrumentation and low model com- plexity approach for labour imminency prediction using a novel multiresolution analysis. Applied AI Letters 2021;2. https://doi.org/10. 1002/ail2.34.
47. Nsugbe E, Ruiz-Carcel C, Starr A, Jennions I. Es- timation of fine and oversize particle ratio in a het- erogeneous compound with acoustic emissions. Sensors 2018;18:851. https://doi.org/10.3390/s18030851.
48. Yiu T. The Curse of Dimensionality. Medium 2021. https://towardsdatascience.com/the-curse-of-Reliable and Quality E-Healthcare (IJRQEH) 2019;8. https://doi.org/10.4018/IJRQEH.2019100102.
49. Nsugbe E, Phillips C, Fraser M, McIntosh J. Gesture recognition for transhumeral prosthesis control using EMG and NIR. IET Cyber-Systems and Robotics 2020;2:122-31. https://doi.org/10.1049/iet-csr.2020.0008.
50. Nsugbe E, Samuel OW, Asogbon MG, Li G. Phan- tom motion intent decoding for transhumeral prosthesis control with fused neuromuscular and brain wave signals. IET Cyber-Systems and Robotics 2021;3:77–88. https://doi.org/10.1049/csy2.12009.
51. Sharma H, Kumar S. A Survey on Decision Tree Algorithms of Classification in Data Mining. International Journal of Science and Research (IJSR) 2016;5.
52. Kozma L. k Nearest Neighbors algorithm (kNN). Helsinki University of Technology; 2008.

53. Suthaharan S. Support Vector Machine. Machine Learning Models and Algorithms for Big Data Classification, vol. 36, Boston, MA: Springer US; 2016, p. 207-35. https://doi.org/10.1007/978-1-4899-7641-3_9.
54. Yegnanarayana B. Artificial neural networks. 11.print. New Delhi: Prentice Hall of India; 2005.
55. LaValley MP. Logistic regression. Circulation 2008;117:2395-9. https://doi.org/10.1161/CIRCULATIONAHA.106.682658.
56. Nsugbe E, Samuel OW, Asogbon MG, Li G. A Self-Learning and Adaptive Control Scheme for Phantom Prosthesis Control Using Combined Neuromuscular and Brain-Wave Bio-Signals. Engineering Proceedings 2020;2:59. https://doi.org/10.3390/ecsa-7-08169.