Article

On the Application of Metaheuristics and Deep Wavelet Scattering Decompositions for the Prediction of Adolescent Psychosis Using EEG Brain Wave Signals

Ejay Nsugbe *

Nsugbe Research Labs, UK

* Correspondence: ennsugbe@yahoo.com

Abstract: Schizophrenia is a common psychotic disorder which affects a substantial amount of the population, where the paranoid variant is viewed as the most common form of the disorder. This form of psychosis has been seen to affect both adults and adolescents; where in the case of adolescents, it is increasingly challenging to diagnose with traditional means involving clinical interviews. The use of electroencephalography (EEG) signals has proven to be an effective means of non-invasively diagnosing brain disorders, alongside having the ability to mitigate any form of subjective bias from the diagnosis process. This paper explores the use of acquired EEG signals, metaheuristics and deep wavelet scattering decomposition, and a combination of supervised and unsupervised learning, for the automated prediction of adolescent schizophrenia. The results showed the best accuracy for the metaheuristic decomposition alongside the candidate learning methods, in the region of 95%+ across the various classification metrics, which showcases an enhanced means of prediction of adolescent schizophrenia. Further work would now explore the use of Long Short-Term Memory and Convolution Neural Networks to investigate the classification performances.

Keywords: signal processing; evolutionary computing; pattern recognition; decision support; eeg; schizophrenia

1. Introduction

Psychosis generally involves the inability to perceive things for what they are and are accompanied by bouts of hallucinations and delusions [1,2]. Schizophrenia is a prominent psychotic disorder characterised by symptoms such as delusions, depressions and cognitive impairment to name a few; and these symptoms prevent the individual from forming functional relationships, in addition to fitting effectively in a professional setting [3]. These factors make apparent that the impacts of the schizophrenia disorder span beyond the individual alone, as the inability to work comes with economic consequences to society [3]. The roots, sources and origins of schizophrenia continue to be a subject for debate, but some of the established causes include genetics and environmental factors [4]. There are also various categories of schizophrenia, with the paranoid variant being the most common and prevalent [5].

Recent medical research has helped uncover the notion that there exists varying extents and severity to the schizophrenia disorder a summary of the four stages alongside their characteristics, means of diagnosis, accompanying disability and care strategy can be seen in Table 1 [6].

As there are yet to be established biomarkers for the diagnosis of schizophrenia, the means towards diagnosis of the illness are via questionnaire style interviews, series of observations and a review of the patient's medical health records [7]. These means of diagnosis have received wide criticism for their subjectivity and lack of replicability, and are made tedious by the fact that schizophrenia also carries overlap with other forms of mental psychotic disorders [8].

Unlike other psychotic disorders, schizophrenia is one which has been known to occur in the early stage of life, as adolescents (i.e., 10–17 years) have also been seen to have the disorder [9]. Due to the established methods for clinical diagnosis of schizophrenia being formed with data from adult patients, the means towards the diagnosis of adolescents have applied these biased diagnosis practices, which have been reported to yield false positives on occasions [9]. Other challenges faced

in the application of classical interview style clinical diagnosis towards adolescent schizophrenic patients include the ability to differentiate between genuine explanation of delusions against adolescent fantasies - a trait which is common amongst that age group due to their naïve psychology [9].

However, it is worth mentioning that neuroimaging results have shown consistencies in the neuro circuitry configuration between both adolescent and adult schizophrenics, therein showing on a neurobiological scale that similarities exist between the two groups [9]. Depending on the stage of the diagnosed schizophrenic disorder, antipsychotic medications are usually the default resort for these patient, but specifically in the case of the adolescent patients, in addition to antipsychotics there exists a high emphasis of the

combination of medication with psychosocial interventions such as psychiatric awareness and counselling, in addition to a thorough assessment of the social needs of the adolescent patient [9]. The research presented as part of this paper involves the application of signal processing models on a dataset from a brain-machine interface (BMI) towards the diagnosis of adolescent schizophrenia.

Electroencephalography (EEG) represents a specialised form of BMI that has been used in varying degrees for different functions, and which also involves the identification of neurodegenerative and psychotic diseases [10]. EEG signals effectively represent neural oscillations associated with the flow of bioelectrical signals across the

Table 1. A summary of the four stages of schizophrenia. [6]

	Stage 1	Stage 2	Stage 3	Stage 4*
Characteristic	- Genetic-based susceptibility	- Cognitive impairment - Social and behavioural deficits	- Psychomotor abnormalities	- The disorder morphs from being mainly psychological to also medical
Diagnosis Method	- Genome sequencing	- Structured interview for prodromal syndrome (SIPS)	- Hallucinations - Disorganised thought - Questionnaire-based interview by psychiatrist	- Questionnaire-based interview by psychiatrist
Disability	- Family history review - Sparse, sometimes a light cognitive deficit	- Cognitive assessments - Notable change in social function	- Steady loss of function	- Chronic manifestation of symptoms
Care Intervention	- N/A	- Counselling support - Polyunsaturated fatty acid	- Antipsychotic medication - Psychosocial care-based interventions	- Societal seclusion due to esoteric behaviours - Antipsychotic medication - Steady rehabilitation programme

*It should be noted that the stage 4 level of schizophrenia is not a universal outcome for all patients.

brain. They hold merits over other brain-based analysis counterparts such as magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI) due to greater affordability, and requiring a much smaller segment of data in order to infer a neural state, i.e., in the or- der of milliseconds, relative to the likes of fMRI which re- quires minute-long data segments at the very minimum [11]. Due to the dynamic nature of the brain, acquired EEG signals are typically stochastic non-stationary signals which require state of the art signal processing and machine learning models to decode the underlying information embedded within a time varying signal [12]. The feature extraction aspect of the signal processing involves the extraction of signals from multiple perspectives which span linear, frequency and complexity/nonlinear-based features, as have been

applied in other studies involving stochastic physiological signals [13,14]. The machine learning aspect has seen the application of models with soft and hard interpretability for the classification and pattern recognition element [14]. It should be noted that in the majority of cases, the recognition capability of the machine learning model hinges upon the quality and effectiveness of the signal processing mechanism applied towards the acquired signal which, in addition to the extraction of appropriate features, also includes the decomposition of a candidate signal to minimize redundancy.

In signal processing, the decomposition of a signal al- lows for the expression of that same signal as a superposition of its constituent parts, thus a given signal $f(t)$ can be decomposed into $a_1.g_1(t)+a_2.g_2(t)......a_n.g_n(t)$,

where is a constant and gn is a component part [15]. This approach is useful for a variety of reasons in a broad number of examples which includes: mixture signals where a source separation is desired; analysis of seismic waves in seismology; the detection of an underlying anomaly embedded deep within a signal, i.e., an intermittent heart problem or a blip in the stock market; denoising of signals whose information is convolved within noise and uncertainties; and the analysing of physiological data such as that which results from a set of brain waves and comprise of a convolution of multiple frequency scales and neural oscillations [16–19]. For a decomposed signal, the optimal decomposed constituent typically reflects the part which contains the rich information within the signal that is relevant towards the signal processing task at hand, and is usually determined with the use of a select cost function/performance index [15].

The canonical decomposition method is the Fourier transform, which applies geometric functions as its basis for decomposing a candidate signal into component parts with respect to frequency. The Fourier approach towards signal decomposition formed the inspiration for the wave let decomposition, which uses a select basis function from a library of bases (i.e., Har, Daubechies and Moorlet) to obtain a time-frequency representation of a signal [15,20] . Research work on source separation of mixture signals conducted by Nsugbe et al. yielded a metaheuristic and therein an artificial intelligence-based means towards the decomposition of a signal to estimate its constituent parts [16,21–23]. The approach uses a set of heuristics and linear thresholds as a basis function towards systematically decomposing a mixture signal, while also learning for the quality of information within each constituent decomposition produced with respect to a performance index. The approach, now named the linear series decomposition learner (LSDL), although originally conceived for source separation of mixtures, has been applied towards classification tasks given an acquired physiological signal in the areas of pregnancy and rehabilitation medicine. The LSDL was applied towards achieving enhanced labour prediction from acquired magnetomyography signals from womb contractions, and for a "mind-based" control of a bionic prosthesis limb for a transhumeral amputee using acquired EEG signals [24,25]. In the case of the mixture separation and bionic prosthesis limb, the LSDL was benchmarked against the wavelet decomposition (with Daubechies mother wavelet), where it was seen that it possessed a superior performance when compared [24].

In the machine learning literature, convolutional neural networks (CNN) represent a form of deep learning architecture that is capable of an unsupervised feature learning, where deep multiscale characteristics of the training samples are learned, and which have produced impressive results in the area of image recognition, but have been criticised for a heavy computational demand alongside the need for a broad and augmented amount of training samples in order to learn effectively [26]. The wavelet scattering (WS) represents a form of automated feature extraction method inspired by the learning process carried out by the CNNs, and allows the user to use the extracted multiscale features to train machine learning models of their choice. Chiefly speaking, the WS involves three main stages enroute towards arriving at its final feature set: convolution using wavelets; nonlinearity by taking the modulus; followed by filtering and averaging using wavelet low pass filters, and is also analogous to pooling [27].

There exists a substantial amount of literature in the area of the application of EEG and machine learning, for the prediction of the schizophrenia psychosis. However, these works utilise a varied source of EEG data with further variations in the acquisition electronics, which ultimately makes it challenging to form a like-for-like comparison of the literature. That being said, the range of methods and techniques used span the application of statistical methods, non-linear complexity theory, nature inspired computational methods, and feature selection alongside the occasional use of decomposition methods, all of which have been applied in a varying capacity towards forming an intelligence system that is capable of pattern recognising the schizophrenia psychosis from an acquired EEG signal [8,28–36].

The application of the LSDL to EEG signals for the prediction of adolescent schizophrenia presents a novel opportunity for the enhancement of the prediction capability of schizophrenia from EEG signals due to the capability of the LSDL to reduce redundancy from the signal, which allows for a boost in accuracy, as observed in previous studies. Thus, in this study, the performance of the LSDL prediction of schizophrenia will be compared with modelling done purely with the raw signal, which represents the state of the art in the majority of the cases presented in the literature, albeit via a varied source of EEG signals. The work will be complemented with the inclusion of analysis from the WS which, as mentioned, represents an unsupervised feature extraction and signal decomposition option and, as per the literature, is yet to be applied to the prediction of schizophrenia.

In this paper, the respective performance of the LSDL and WS will be contrasted in their ability to produce an enhanced means of predicting adolescent schizophrenia from a set of EEG signals. This will be combined with unsupervised learning models as part of strides towards an automated platform for clinical decision support. Specifically speaking, the contribution of this manuscript is as follows:

● A contrast of the abilities of the LSDL and WS decomposition methods in producing an enhanced means of predicting the presence of schizophrenia in adolescents, while being benchmarked with predictions made with the raw EEG signals, which represent the state of the art.

● A novel application of the K-Means unsupervised learning for the automated labelling of the data, followed by the application of supervised learning methods as a means towards intelligent self-learning and classification between the Schizophrenic and Normal adolescents, which could lead to a clinical decision support platform that provides a greater autonomy and minimal external intervention in the prediction of the psychosis.

2. Materials

The human brain consists of over a billion neurons whose interaction and signalling produces a bioelectrical signal that abides to Maxwell's equations and the laws of electrodynamics, as shown in previous studies [37]. For a continuous stream of EEG signals acquired from an array of channels, they can be represented as a multidimensional vector where $x(t)=(x_1(t),\ldots\ldots x_d(t))$, where d represents the number of electrode channels for a fixed time interval.

$t \in [0, T]$. Based on the discrete sampling sequence of EEG acquisition systems, an acquired set of EEG signals can be expressed as a series in the following form $(x(0), x(\frac{t}{n}), x(\frac{2t}{n})\ldots\ldots x(T)$. A sample image of an EEG acquisition process can be seen in Figure 1, while Figure 2 provides an illustration of small electrical fields generated by the synaptic currents in the pyramidal cells within the brain. In order for candidate EEG electrodes to be able acquire signals through the skull and thick host of tissues, cells in the order of thousands need to simultaneously fire for a cumulant of a bioelectrical signal to be able to propagate through and be acquired by the EEG electrodes [38].
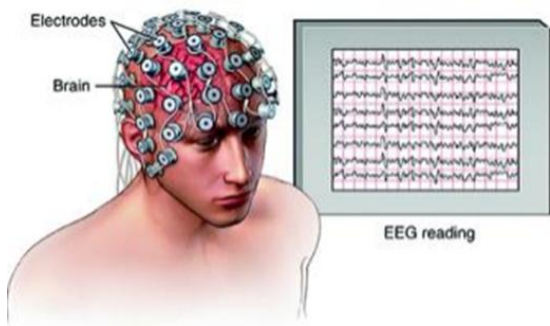

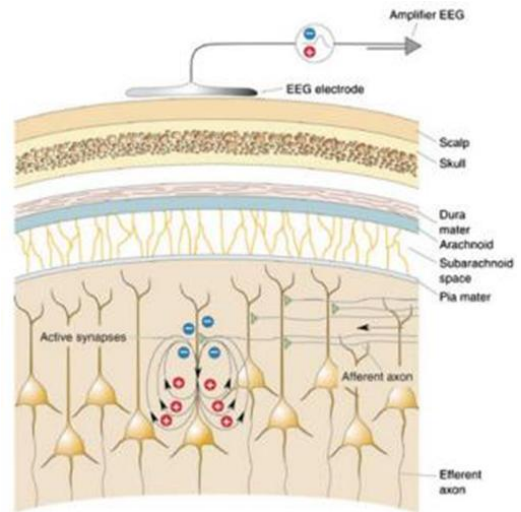**Figure 1.** Image of an EEG acquisition process. [38]

Figure 2. A cross-sectional view of the human skull and the various layers while little electrical fields are being generated by the pyramidal cells. [38]

3. Experimental Method

The EEG data used for the analysis in this study was taken from an opensource database with data from Normal and Schizophrenic adolescents aged 10-14 years who were clinically diagnosed with schizophrenia at the Research Center for Psychological disorders of the Russian Academy of Medical Sciences [39]. The EEG electrode placement was done following the 10/20 electrode scheme with a reference electrode placed on the earlobe, as seen in Figure 3 [39]. It was also confirmed that the patients were not on any antipsychotic medications prior to the acquisition of the EEG signals. Thus the influence of psychiatric medication on the acquired EEG signals can be negated [39].
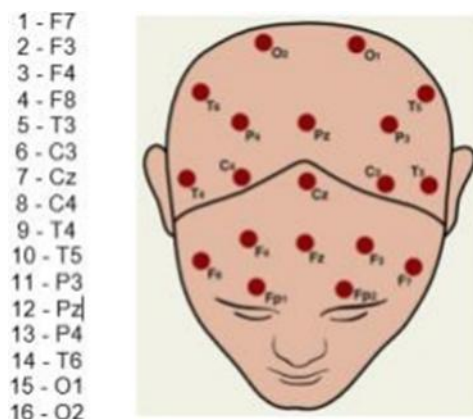

Figure 3. Image showing electrode placement location on patients. [39]

The acquisition parameters for the experiment include a sampling of 128 Hz and spanned 60 seconds, thus producing a signal vector of 7680 Hz in length. As part of the proof-of-concept work done in this paper, a

sum total of 20 patients ' EEG data was utilised and comprised of 10 Normal and 10 Schizophrenic patients.

3.1. Signal Processing and Machine Learning

3.1.1. LSDL

The LSDL was originally conceived for the source separation of mixture signals acquired from acoustic emission signals of powders in the micron scale striking a medium. It works with the hypothesis that source activities manifest themselves in the time domain in the form of an impulse signal of a fixed magnitude that can also be characterised by its magnitude [16, 21-23]. As these impulse signals are dynamic, an accompanying decay typically accompanies them and thus a unit impulse signal can be mathematically expressed as s(t)=e -wu()+ for u(t)=$\begin{cases}0, a < 0 \\ 1, a \geq 0\end{cases}$ where, u(t) is a step function, indicates that the function is 0 until $\geq$ 0, and is a noise source associated with the decay of the activity of the impulse signal. From a discretisation perspective, the decay of the impulse signal manifests itself in the form of "false" impulse signals in the time domain which dynamically carries a lower magnitude than the original impulse signal. In the case of events which occur simultaneously, the resulting time domain signal has been seen to contain overlapping impulse, with the magnitude (also referred to as amplitude in this case) being a key characteristic towards an effective characterisation of the source events, as described by Nsugbe et al. [16 ,21-23]. This gives rise to the hypothesis that an optimal amplitude region within a candidate signal is one whose information quality is maximised as adjudged by a designated performance index, and therein contains minimal uncertainties within it.

In order to separate a candidate signal into various "amplitude bands", tuned linear thresholds of varying amplitudes are employed in the process, which work with a set of heuristics. These thresholds are used to create subsignal components and therein a decomposition of the original signal. The thresholding, and therein the signal decomposition process , works first by the application of an initialisation threshold which splits the signal into two constituent parts; X1 1 and X12, belonging to the set x={xuxa, which hosts both signal decompositions, where xx represents the upper-amplitude region and the 1st threshold iteration/1st decomposition of the signal, and X2 represents the lower amplitude region and the 1st threshold iteration/2nd decomposition. The initialisation threshold is typically selected to be 0 5 of the maximum of the signal itself in order to effectively separate the signal into two equal parts. To assess the quality of the decompositions within the set, select features are extracted from each candidate decomposition from which their quality is evaluated using a performance index J.

The decomposition is then repeated using a set of defined threshold scaling heuristics to further decompose xn.andxa, where for each subsequent decompositions J is computed to form an array where the argmax(J) is chosen. At the end, the iterations are terminated once a conditional minimum is found, with the assumption of a convex optimisation problem. The performance index used as part of the LSDL algorithm is the normalised Euclidean distance and is expressed as follows, for the calculation of the distance in Euclidean space between two candidate decompositions from two different source signals which require classification as shown in Equations (1)–(3) [40]:

$$ED(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \qquad (1)$$

$$\sigma = \sqrt{\frac{\sum_{N=1}^{N_W} (R_W - \mu)^2}{N_W}} \qquad (2)$$

$$J = (p,q) = \frac{ED(p,q)}{\sigma_m} \qquad (3)$$

where ED is the Euclidean distance given coordinates p and q, which in this case correspond to the features within the feature vectors from a specific electrode channel, w is the wth feature within the feature vector, Ne, r_w are specific feature within the feature vector, μ is the mean of the features, while σ_m represents the mean of the standard deviations of the of the two candidate features in question from two signal decompositions. Note that by means of standardisation, the decomposition levels need to be equivalent for the two signals used for the computing of1, i.e., for two candidate signals with their accompanying decompositions, S1={1x2} and S2={ 1x2}, once the feature vectors are formed for the respective decomposi tions, s(xa(fctre) and s2=xn(features) and not S1=xn(features) and s2={x2(features)3.

The threshold scaling works with a set of heuristics which can be modified by the user as per the requirements and resources available. For a sample signal=…Table 2 shows the defined threshold parameters adopted for the scaling used in this study, where it can be noted that the parameters for the Upper threshold region have been updated from what was proposed previously in studies by Nsugbe and Sanusi [24], and Nsugbe et al. [25].

Table 2. Threshold parameters for implementation of LSDL.

Iteration	1	2	3	n
Upper threshold region parameter (Upper)	$T_{1\_upper\_1}$ $= 50\%$ of max $\lvert s_n \rvert$	$T_{1\_upper\_2}$ $= \dfrac{\max \lvert s_n \rvert + T_{1\_upper\_1}}{2}$	$T_{1\_upper\_3}$ $= \dfrac{\max \lvert s_n \rvert + T_{1\_upper\_2}}{2}$	$T_{1\_upper\_n}$ $= \dfrac{\max \lvert s_n \rvert + T_{1\_upper\_3}}{2}$
Lower threshold region parameter (Lower)	$T_{1\_lower\_1}$ $= 50\%$ of max $\lvert s_n \rvert$	$T_{1\_lower\_2} = \dfrac{T_{1\_lower\_1}}{2}$	$T_{1\_lower\_3} = \dfrac{T_{1\_lower\_2}}{2}$	$T_{1\_lower\_n} = \dfrac{T_{1\_lower\_n-1}}{2}$

where T and T:ta are the thresholds corresponding to the Upper and Lower amplitude regions of the signal respectively. Expressing the respective threshold regions as a series as shown in Equations (4)-(6):

$$\text{Upper}(x) = T_{1\_upper\_1}(x) + T_{1\_upper\_2}(x) + T) \ldots \ldots \ldots T \quad (4)$$

$$\text{Lower}(x) = T_{1\_upper\_1}(x) + T_{1\_upper\_2}(x) + T) \ldots \ldots \ldots T \quad (5)$$

$$T(x) = \text{upper}(x) + \text{Lower} \quad (6)$$

Using the law of superposition, T(x) refers to the recovered and reconstructed version of the original signal.

A comprehensive list of the steps and heuristics used as part of the LSDL process for classification purposes can be seen in Nsugbe and Sanusi [24], and Nsugbe et al. [25], and the optimisation objective can be formulated as can be seen in Equation (7):

$$\arg\max J x_i E R \quad (7)$$

where R is a set of real numbers and x_i is a value within the set.

3.1.2. Optimal LSDL Signal Region

Using candidate signals from each set of classes (i.e., Schizophrenic and Normal) and for three iterations and decomposition levels, the result for the LSDL can be seen in Table 3.

Table 3. LSDL decomposition results.

Iteration	1	2	3
Upper	2.761	2.773	2.547
Lower	2.826	2.439	2.827

It can be seen from Table 3 that both sets of signals can be said to be of a high quality, as reflected by the concise range of performance indexes in the table. This could also be said to represent a sense of optimalityin the

acquisition process of the experiment,i.e., sample rate etc. Expressing the optimisation objective from the perspective of evolutionary metaheuristic algorithms, the selection process is one of "elitism" where the "fittest" item is selected from the batch, as has also been the approach in prior work [24,25,41]. As part of the work done in this paper, the "crossover" and "mutation" evolutionary rules were applied as part of the selection process, where the top two items that produced the maximum J were selected as part of the crossover process and the signals from theirrespective threshold regions were superimposed, i.e. $T_{1\ lower\ 1}(x) + T_{1\ lower\ 3}(x)$ due to these two regions having the highest indices for J. The resulting J value for the crossover and mutation exercise was 2.826, which shows a minor degradation in performance when compared with the elitism value of 2.827, therein implying that superimposing LSDL decomposition regions does not yield an optimal and enhanced quality as part of the LSDL method. Thus, the optimal LSDL decomposition region can be seen to be from iteration 3 of the Lower threshold region. This threshold region was generalised across subsequent analysis of further signals in this work where the LSDL was applied. In terms of computational complexity, it has been seen that the LSDL's complexity is of the order O(n) [42,43].

3.1.3. Deep Wavelet Scattering

The deep wavelet scattering allows for a form of unsupervised feature extraction of low-variance features for a given time domain signal, where the obtained features are robust to translations and are also continuous [44]. Unlike a standard CNN, the wavelet scattering uses predefined wavelet and scaling filters as opposed to learning them from the data [44]. The incremental steps that have culminated in the deep wavelet scattering has seen combined work from Mallat, Bruna and Andén, who made strides towards establishing a mathematical formalism of CNNs, followed by work done by Andén and Lostanlen who provided the computational framework for the wavelet scattering [27,45–47]. Mallat

[45] has described the key properties of CNNs carried over to the deep wavelet scattering that allow for the extraction of useful features, including: multiscale contractions, the linearisation of hierarchical symmetries, and sparse representations. By having an a-priori set value for the filters, the deep wavelet scattering is said to be able to work effectively with small training data samples [44]. A diagram showing the various steps associated with the deep wavelet scattering can be seen in Figure 4.
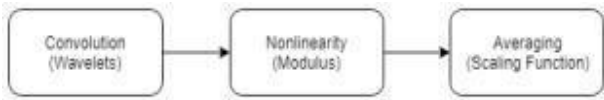
Figure 4. The key associated steps of the deep wavelet scattering. [44]

Working with the mathematical formalism proposed by Liu et al., assuming a sample signal f(t) being analysed with Ø being a low pass filter and a wavelet function of Ψ for filtering purposes, which spans the range offre-quencies of the signal. Assume $Ø_1(t)$ to be a low pass filter which provides localised translation invariance of f at a defined scale T [48,49] . The family of wavelet indices possessing an octave frequency resolution Q_k is denoted as $\wedge_k$, while the multiscale high pass filter banks $\{\Psi_{jk}\}_{jk\in\wedge_k}$ are constructed via a dilation of the wavelet Ψ[48,49].

The wavelet scattering network is implemented via a deep CNN which iterativelyconvolves through classical wavelets, nonlinear modulus, and an averaging scaling function as seen in Figure 4 [48]. The convolution part sof() =f*g (e) , where sa is the zero-order scattering coefficients, generate locally translation invariant features off which, although yielding a loss of high frequency information, can be recovered via the wavelet modulus transform lwl, expressed as shown in Equation (8) [48]:

$$|W_1|f = \left\{ S_0 f(t), |f * \Psi_{j1}(t)| \right\}_{j1\in\wedge_1} \quad (8)$$

In a cascading fashion, the first order scattering coefficients can be obtained by the averaging of the wavelet modulus coefficient with as shown in Equation (9):

$$S_1 f(t) = \left\{ |f * \Psi_{j1}(t)| * Ø_j(t) \right\}_{j1\in\wedge_1} \quad (9)$$

Once again, to recover the information lost from the averaging process, while bearing in mind that saf(t can be assumed to be the low frequency component of |f*Ψj1|, by applying the wavelet modulus the high frequency com - ponents can again be expressed as shown in Equation (10) [48]:

$$|W_2||f * \Psi_{j1}| = \left\{ S_1 f(t), ||f * \Psi_{j1}| * \Psi_{j2}(t)| \right\}_{j2\in\wedge_2} \quad (10)$$

And therein further defining the second order coefficients shown in Equation (11):

$$= = \quad 1\,2 \quad (11)$$

Sequentially iterating the defined process deduces the relevant wavelet modulus convolutions shown in Equation (12):

$$U_m f(t) = \left\{ \left|| |f * \Psi_{j1}| * ..... | * \Psi_{jm} |\right| \right\}_{j1\in\wedge_1}, i = 1,2, ...m. \quad (12)$$

where um is an mth order modulus.

How to obtain the mth order scattering coefficient, umf(t) with $Ø_j$, can be seen as shown in Equation (13):

$$S_m f(t) = \left\{ \left|| |f * \Psi_{j1}| * ..... | * \Psi_{jm} | * Ø_j(t) \right| \right\}_{j1\in\wedge_1}, i = 1,2, ...m \quad (13)$$

The defined approach is used to obtain a final scatter matrix sf(t)={snf()m, which concatenates the scattering coefficients from all orders as a means of characterising an input signal, where 1 represents the maximum decomposition level [48]. A tree-based visualisation of the scattering decomposition network ca be seen in Figure 5.
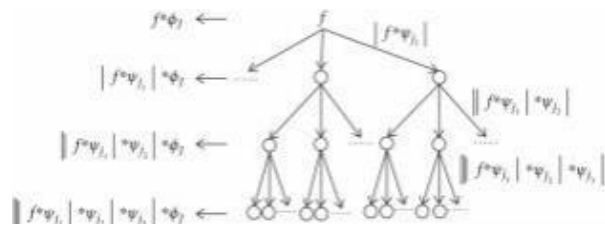

Figure 5. A tree-based visualisation of the scattering decomposition network. [48]

The wavelet scattering decomposition retains characteristics of both the CNN and the wavelet transform itself; for example, the wavelet scattering exhibits translation invariance and is also stable to local deformations [48]. In terms of the key differences between the CNN and the wavelet scattering decomposition, in addition to not needing to learn the filters (as they are pre-set), the output features are not only the output from the last layer but also a combination of all preceding layers [48]. The energy of the scatter coefficients is seen to dissipate with an increasing number of layers, with the majority of the energy hosted in the first two layers; thus in this work, a two-order scattering network is used for the extraction of features [48]. Other parameters used for the wavelet scattering decomposition in this work include the utilisation of the Gabor wavelet for decomposition purposes, where the invariance scale was set to 1 second as inspired by a related work, and the default value for the filter banks of 8 wavelets per octave in the first filter bank and 1 wavelet per octave in the second filter bank [48]. A visualisation of the filter banks from the two network layers and the low pass filter with a 1 second invariance scale can be seen in Figures 6 and 7.

3.2. Feature Extraction

Inspired by work done in physiological signal processing, the features used in this work include a concatenation of various feature groups in order to effectively model the nonlinear EEG signals-these features are described as follows [25,50–52]:

● Linear Features: these are commonly used low order and computationally effective features. The list of features in this category is as follows: mean absolute value, waveform length, zero crossing, root mean square, 4th order autoregressive coeffi-cient, number of signal peaks, simple squared integral, and variance. For the features which require a threshold, a value of 1μv was used, while for the number of signal peaks, a peak can be defined as

$$\begin{cases} x_{peak.n'} x_{n-1 \, and \, X_{n+1}} \\ 0, Otherwise \end{cases}$$

● Frequency Features: these features are extracted from a frequency representation of a candidate signal. The frequency features used were the maximum cepstrum coefficient, and median frequency.
● Nonlinear Features: these features represent a selection taken from areas such as chaos and complexity theory and generally display a good capability towards charactering a set of nonlinear signals. The list of features from this category is as follows: sample entropy, maximum fractal length, Higuchi fractal dimension, and detrended fluctuation analysis. The parameters used in the computation of the various features were 2 and 0.2 for the values of m and r for the sample entropy calculation, and then K as 10 for the Higuchi fractal dimension.
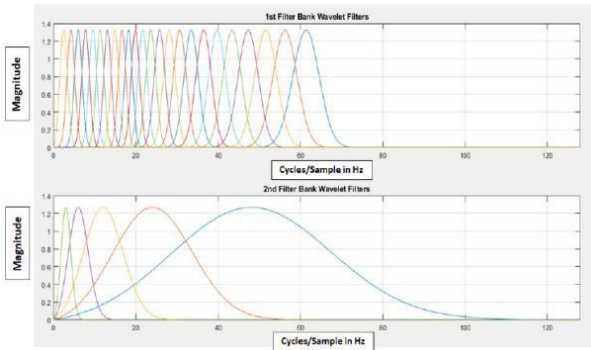

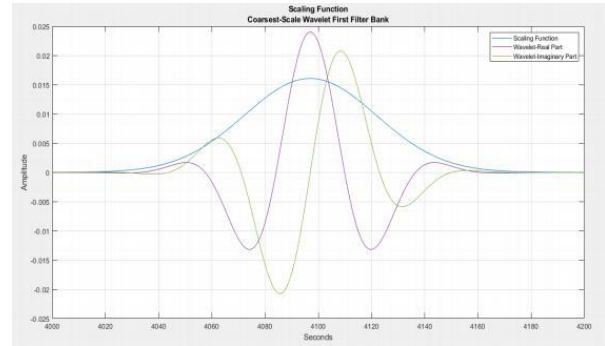**Figure 6.** The first filter bank with 8 wavelets per octave.

Figure 7. The low pass filter with 1 second invariance scale.

3.3. Machine Learning Models

3.3.1 Unsupervised Learning

The K-means unsupervised learning method was adopted as a mode towards an unsupervised partitioning of the unlabelled feature vector belonging to the two data classes; a description of the K-means algorithm can be seen as follows [53]:

K-means: this represents an iterative clustering method which separates data into a defined K number of clusters based on the Euclidean distance performance index [53]. The clustering method works with the expectation-maximisation (E-M) method where the E step assigns data points into various groups after a random initialisation using the following objective function $\sum_{i=1}^{m} \sum_{k}^{K} ar \, gmln_j \, |x^2-u_k|^2$ is a specific data point and represents a cluster centroid mean; the M-step rep-resents a form of a recursive update stage for the cluster centroid via $\mu_k = \frac{\sum_{i=1}^{m} w_{ik} x^i}{\sum_{i=1}^{m} w_{ik}}$ where w_{ik} represents a binary variable used to indicate whether a particular data point belongs in a specific class [53]. As the K-means works with a random initialisation, the model was run five times with K selected as 2, where the model with the least error was chosen as the final designated K-means model.

It can be noted that the justification for the selection of the K-means algorithm (ahead of similar models such as the Gaussian Mixture Models) stems from the advantages of the algorithm, such as being easy to implement and carryinglow computational complexity.

3.3.2 Unsupervised Learning

Four machine learning classification models with a low complexity and an "easy" interpretability, as highlighted via various sources, were chosen for the supervised learning aspect of the work done in this paper, and are described as follows [14]:

● Logistic Regression (LR): this is a classification model which emanates from statistics and uses a sigmoidal activation function and an assigned threshold towards distinguishing between data from two different classes in a binary-like fashion. A mathematical underpinning of the LR classifier and a previous use case can be seen inNsugbe and Sanusi [24].

● Naïve Bayes (NB): this model works with the assumption that the underlying structure to the data is Gaussian and assigns data to various classes while utilising the Bayes probabilistic framework for class sorting.

● Discriminant Analysis (Linear and Quadratic, LDA and QDA): this is a statistically driven and computationally effective method towards classification of data via the projection into a lower dimensional space, and shortly followed by the instillation of class boundaries, both linear and quadratic class boundaries were utilised for the work done in this paper. A mathematical framework for both the LDA and QDA can be seen inNsugbe et al. [25,50].

● Support Vector Machine (L-SVM): these are iterative classification models based around an optimal separation boundary between data classes, using a subset of the data referred to as support vectors, in a process that involves the projection of the data into a higher dimensional space, which maximises class separability and where class boundaries are set [24] . This is followed by a preservation of the structure whilst the data is projected back down into a lower dimensional space and the class boundaries are pre- served in a process known as a kernel trick [24].

Three feature vectors were formed and used as part of the initial unsupervised learning machine learning exercise, namely, the raw data feature vector comprising 5400 samples (15 electrodes × 18 features × 20 patients), the LSDL feature vector comprising 5400 samples (15 electrodes × 18 features × 20 patients), and the wavelet scattering feature vector comprising 330,240 samples (7,680 multiscale features × 43 windows)

All supervised learning models were validated using a holdout dataset comprising of 25% unseen data, which was used to validate the models and obtain the final classification metrics.

A flow diagram of the various steps involved in the full prediction process in this work can be seen in Figure 8.
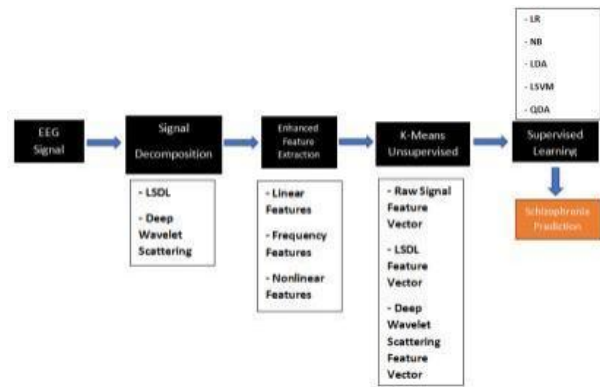

Figure 8. Flow diagram of the full prediction process.

4. Results and Discussion

The plots in Figure 9 represent the fast Fourier transform (FFT) of the LSDL decomposed data and raw data from electrode channel 1 of a Normal patient and a Schizophrenic patient [54]. Note that the FFT of the deep wavelet scattering was not plotted since the method decomposes the signal into separate frequency bands as part of the feature vector which it returns.

Interpreting the FFT from the LSDL, it can be seen that the bulk of the frequency content resides within the 0 Hz-5 Hz region, which constitutes the delta cognitive state, and accounts for unconscious and deeply relaxed states. Thus, the implication is that this frequency band is where the schizophrenia psychosis manifests itself in the adolescent case. The remainder of the frequency spectrum consists of what appears to be a form of broadband noise, with the Normal patient's EEG recording showing a slightly higher noise magnitude in comparison.

It can be noted that, in adult schizophrenic psychotic patients, the optimal frequency band for the diagnosis of the schizophrenia psychosis has been previously noted to be in the gamma-range, as reported by Baradits et al. [55]. These results obtained in this paper, echo the differences in the configuration of the neural circuitry of adolescents in contrast to adults, whose brains are no longer in the developmental stage, and therein forms further insight as to why means of psychosis diagnosis developed in studies involving mostly adult patients should ideally not be generalised towards adolescent patients [55] .

In the case of the raw data, which is effectively without any form of decomposition, more activity can be seen in the spectrum, albeit largely contaminated by noise. The underlying pattern is similar to the LSDL but, as mentioned, a lot noisier. The visualisation via the FFT showcases the benefits of the LSDL decomposition. It can be seen that the bulk of the energy in the spectrum of the raw signal is also present in the very low frequency region and exhibits an exponential decay towards the

higher frequency. A principal component analysis plot of the various feature vectors can be seen in Figure 10.

From the PCA plots in Figure 10, the first subplot (which represents the raw data) shows some degree of separation between the clusters with some degree of over - lap. The middle plot (which represents the LSDL) shows a high separability between clusters with a linear decision boundary and a few outliers overlapping in an opposite cluster. Finally, the plot for the deep wavelet scatter shows a substantial amount of cluster overlap, although it is worth being mindful that only the first two principal components from the deep wavelet scatter were plotted and account for only 57% of the total variability of the data, while the prior two plots contained up to 99%, hence poor visualisation.

4.1 K-means Clustering Results

The K-means clustering results for the various feature vectors can be seen in Tables 4-6. The clustering result for the feature vector from the raw data showed good clustering prowess for the Normal patient cluster, but was only sparsely able to accurately cluster the data from the Schizophrenic patient, hence an overall clustering accuracy of 46%. The results of the LSDL clustering showed a good accuracy for both the
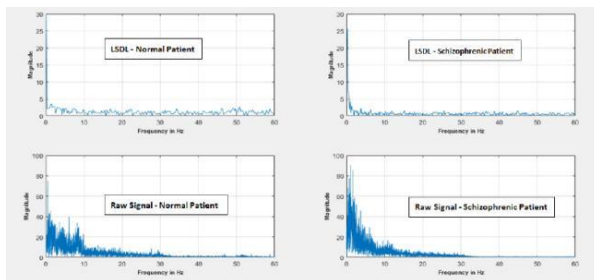

Figure 9. FFTplots of the LSDL signal from electrode channel1 of a Normal patient and a Schizophrenic patient.
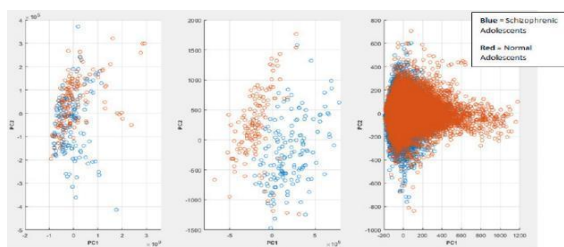

Figure 10. Left plot: raw data with 99% variability. Middle plot: LSDL data with 99% variability. Right plot: deep wavelet scattering with 57% variability.

Schizophrenic and Normal patients clusters, recording an overall clustering accuracy of 83%. The results for the clustering exercise for the deep wavelet scattering can be seen to comprise more samples due to the multiscale decomposition pattern of the wavelet scattering network, even though it was fed the same number of samples as the LSDL. Although the deep wavelet scattering provided the best accuracy in the clustering of the Schizophrenic patient cluster, the clustering accuracy for the Normal patients was relatively low in comparison, thus the final clustering accuracy was seen to be 55%. Due to this result, only the LSDL data and the associated clustering labels were used for the supervised learning exercise in the subsequent section.

Table 4. Clustering results with the raw data.

Schizophrenia Patient Cluster	Normal Patient Cluster
Correctly Clustered/Total data points: 8/150	Correctly Clustered/Total data points: 131/150

Clustering Accuracy: 46%

Table 5. Clustering results with LSDL.

Schizophrenia Patient Cluster	Normal Patient Cluster
Correctly Clustered/Total data points: 109/150	Correctly Clustered/Total data points: 139/150

Clustering Accuracy: 83%

Table 6. Clustering results with the deep wavelet scattering.

Schizophrenia Patient Cluster	Normal Patient Cluster
Correctly Clustered/Total data points: 32807/38400	Correctly Clustered/Total data points: 9666/38400

Clustering Accuracy: 55%

A plot showing the K-means cluster partitions can be seen in Figure 11.

It can be said that the reason the K-means was able to cluster the data from the LSDL is largely due to theminimal overlap between the respective clusters, therein implying that the K-means is primarily suited with data that has a good degree of linearseparability.

4.2 Supervised Learning with Clustering Labels

To effectively characterise the performance of the trained model the following classification metrics were adopted, as utilised in prior studies: accuracy (ACC),

sensitivity (SEN), specificity (SPEC) and area under the curve (AUC) [52].

The results of the supervised learning exercise are shown in Table 7 across all five candidate classification models, from which it can be seen that all models produced high accuracies in the range of 95%+. Comparatively speaking, the LDA produced the lowest metrics, which was followed by a closely matched performance from both the LR and NB, while the LSVM and QDA jointly produced the best model performances. This is thought to be due to the complexity of the classification boundary and computational power of the latter two models when compared to the other models.
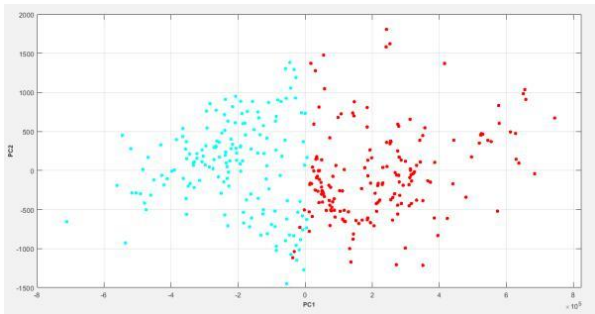

Figure 11. Image showing the K-means LSDL clustered data visualised as PCA with 99% variability explained.

Table 7. Supervised learning results with 25% holdout validation.

	Acc	SEN	SPEC	AUC
LR	96.7	97.7	95.6	96.7
NB	96.7	97.7	95.6	96.7
LDA	95.6	97.7	93.6	95.7
LSVM	98.9	97.8	100	98.9
QDA	98.9	97.8	100	98.9

These results thus showcase how the LSDL decomposition, alongside an unsupervised learning driven platform, can contribute towards an automated schizophrenia classification platform using EEG signals, with interpretable models. Furthermore, they provide a quantitative validation of the visual interpretation of the LSDL, as the data clusters can be effectively classified with the use of a linear boundary and a low complexity classifier.

Further advantages of the LSDL as a signal decomposition tool are that it can allow for an unsupervised decomposition of the signal, while it requires only a small decomposed subset of the signal and thus encourages sparse and parsimonious signal modelling. It can also be implemented with an analogue circuitry which can allow for further computational efficiency. Furthermore, the LSDL can also serve as an adaptive decomposition scheme in the case of a degrading or varying signal source. The main

shortcoming of the method remains a lack of insight into the frequency content associated with the various signal decompositions.

In terms of comparison with the current literature, the state of the art appears to suggest that the use of the raw data for signal processing purposes is the commonality amongst authors, and is surpassed by the LSDL, which can be seen from Tables 4 and 5. In terms of further comparisons, Piryatinska et al. [29] utilised the same dataset alongside complexity modelling for the prediction of adolescent schizophrenia with a prediction accuracy of 79.7%-83.6%, dependent on the classifier, the results of which are below the prediction accuracy obtained by the proposed method in this manuscript, as shown in Table 7.

5. Conclusions

Schizophrenia is a psychotic disorder which is characterised by an array of symptoms that make it difficult for an individual to fit into standard settings. The psychosis can be said to involve four various stages where the most common form of the disorder is the paranoid variant. The effective diagnosis of schizophrenia has historically proven to be challenging as the disorder also shares commonalities with other forms of psychiatric disorders. In addition to this, the diagnosis of schizophrenia in adolescents is even more tedious due to the naïve psychology of the age cohort.

EEG signals provide surface manifestations of brain neurons firing in synchronism and have proved to be a useful tool in the non-invasive detection of brain-related disorders. This paper explored the prediction of schizophrenia from acquired EEG signals from adolescent patients as a means of eliminating bias and shortcomings associated with standard psychiatric-based interview style diagnosis methods. EEG signals are nonlinear manifestations of neural oscillations within the brain and tend to benefit from pre-processing and signal decomposition. As part of the work done, this paper contrasts a metaheuristic decomposition with a deep wavelet scattering decomposition, followed by unsupervised and supervised learning for an automated schizophrenia prediction platform. Following the extraction of an enhanced group of features, the clustering work done with K-means showed that the LSDL decomposition method provided the best clustering accuracy, and its associated labels were used to train five candidate supervised learning models with an "easy" interpretability, where the results showed high metrics in the region of 95% for the various classification metrics which were investigated, therein showcasing the capability of the LSDL alongside unsupervised learning to serve as an automated clinical support tool for the diagnosis of schizophrenia in adolescents. The main limitation of the proposed work is tied down to the long and arduous process involved in the tuning and optimisation of the LSDL in order to find

the optimal decomposition region. Hence, subsequent work would likely involve the potential streamlining of the heuristics used as part of the algorithm in order to minimise this.

Further work would also now involve the use of deep learning architectures such as long short-term memory, and convolution neural networks for classification of the EEG signals. In addition, further work would also explore parsimonious modelling based on the application of electrode selection methods towards reducing the amount of channels used to make predictions of the presence of the psychosis [56].

Acknowledgments

Statement and Declarations

References

1. Arciniegas, D.B., 2015. Psychosis. Continuum (Minneap Minn). 21, 715–736. DOI:https://doi.org/10.1212/01.CON.0000466662.89908.

2. Read, J., van Os, J., Morrison, A.P., et al., 2005. Childhood trauma, psychosis and schizophrenia: a literature review with theoretical and clinical implications. Acta Psychiatr Scand. 112, 330–350. DOI: https://doi.org/10.1111/j.1600-0447.2005.00634.x.

3. Prabhakar, S.K., Rajaguru, H., Kim, S.H., 2020. Schizophrenia EEG Signal Classification Based on Swarm Intelligence Computing. Computational Intelligence and Neuroscience. pp. e8853835. DOI: https://doi.org/10.1155/2020/8853835

4. Gottesman, I.I., Shields, J., 1973. Genetic theorizing and schizophrenia. British Journal of Psychiatry. 122, 15–30. DOI: https://doi.org/10.1192/bjp.122.1.15

5. World Health Organization (editor), 1992. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Gene-va: World Health Organization.

6. Insel, T.R., 2010. Rethinking schizophrenia. Nature. 468, 187–193. DOI: https://doi.org/10.1038/nature09552

7. Schmitt, A., Hasan, A., Gruber, O., et al., 2011. Schizophrenia as a disorder of disconnectivity. European Archives of Psychiatry and Clinical. 261, 150–154. DOI: https://doi.org/10.1007/s00406-011-0242-2

8. Devia, C., Mayol-Troncoso, R., Parrini, J., et al., 2019. EEG Classification During Scene Free-Viewing for Schizophrenia Detection. IEEE Transactions on Neural Systems & Rehabilitation. 27, 1193–1199. DOI: https://doi.org/10.1109/TNSRE.2019.2913799

9. Hollis, C., 2000. Adolescent schizophrenia. Advances in Psychiatric Treatment. 6, 83–92. DOI: https://doi.org/10.1192/apt.6.2.83

10. Gibbs, F.A., Gibbs, E.L., 1941. Atlas of electroencephalography. Oxford, England: F. A. Gibbs, Boston City Hospital.

11. Darkhovsky, B., Piryatinska, A., Kaplan, A., 2016. Binary classification of multi-channel EEG records based on the ε -complexity of continuous vector functions. ArXiv:161001633 [Stat].

12. Rickles, D., Hawe, P., Shiell, A., 2007. A simple guide to chaos and complexity. Journal of Epidemiology & Community Health. 61, 933–937. DOI: https://doi.org/10.1136/jech.2006.054254.

13. Arce, G.R., 2005. Nonlinear Signal Processing: A Statistical Approach. John Wiley & Sons.

14. Nsugbe, E. Enhanced Recognition of Adolescent Schizophrenia and a Computational Contrast of their Neuroanatomy with Healthy Patients using Brain-wave Signals In peer review.

15. Mallat, S.G., 1999. A wavelet tour of signal processing. 2nd ed. San Diego: Academic Press.

16. Nsugbe, E., 2017. Particle size distribution estimation of a powder agglomeration process using acousticemissions. PhD. Cranfield University.

17. Addison, P.S., 2005. Wavelet transforms and the ECG: a review. Physiological Measurement. 26, R155–199. DOI: https://doi.org/10.1088/0967-3334/26/5/R01

18. Sardy, S., Tseng, P., Bruce, A., 2001. Robust wavelet denoising. IEEE Transactions on Signal Processing. 49, 1146–1152. DOI: https://doi.org/10.1109/78.923297

19. Adeli, H., Zhou, Z., Dadmehr, N., 2003. Analysis of EEG records in an epileptic patient using wavelet transform. Journal of Neuroscience Methods. 12 3, 69–87. DOI: https://doi.org/10.1016/S0165-0270(02)00340-0.

20. Bracewell, R.N., 2000. The Fourier transform and its applications. 3rd ed. Boston: McGraw Hill.

21. Nsugbe, E., Starr, A., Foote, P., et al., 2016. Size differentiation of a continuous stream of particles using acoustic emissions. 20th Innovative Manufacturing Engineering and Energy Conference. 161, 012090. DOI: https://doi.org/10.1088/1757-899X/161/1/012090

22. Nsugbe, E., Starr, A., Jennions, I., et al., 2017. Particle size distribution estimation of a mixture of regular and irregular sized particles using acoustic emissions. Procedia Manufacturing. 11, 2252-2259. DOI: https://doi.org/10.1016/j.promfg.2017.07.373.

23. Nsugbe, E., Ruiz-Carcel, C., Starr, A., et al., 2018. Estimation of fine and oversize particle ratio in a heterogeneous compound with acoustic emissions. Sensors. 18, 851. DOI: https://doi.org/10.3390/s18030851.

24. Nsugbe, E., Sanusi, I., 2021. Towards an affordable magnetomyography instrumentation and low model complexity approach for labour imminency prediction using a novel multiresolution analysis. Applied AI Letters. ail2, 34. DOI: https://doi.org/10.1002/ail2.34.

25. Nsugbe, E., Samuel, O.W., Asogbon, M.G., et al., 2021. Contrast of multi-resolution analysis approach to transhumeral phantom motion decoding. CAAI Transactions on Intelligence Technology. DOI: https://doi.org/10.1049/cit2.12039.

26. Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017. Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET). pp. 1–6. DOI:https://doi.org/10.1109/ICEngTechnol.2017.8308186.

27. Andén, J., Mallat, S., 2014. Deep Scattering Spectrum. IEEE Transactions on Signal Processing. 62, 4114–4128. DOI: https://doi.org/10.1109/TSP.2014.2326991.

28. Khare, S., Bajaj, V., Siuly, S., et al., 2020. Classification of schizophrenia patients through empirical wavelet transformation using electroencephalogram signals.

Modelling and Analysis of Active Biopotential Signals in Healthcare. IOP Science. 1, 1–26. DOI: https://doi.org/10.1088/978-0-7503-3279-8ch1.

29. Piryatinska, A., Darkhovsky, B., Kaplan, A., 2017. Binary classification of multichannel-EEG records based on the ϵ -complexity of continuous vector functions. Computer Methods and Programs in Biomedicine. 152, 131–139. DOI: https://doi.org/10.1016/j.cmpb.2017.09.001.

30. Prabhakar, S. , Rajaguru, H., Kim, S.H., 2020. Schizophrenia EEG Signal Classification Based on Swarm Intelligence Computing. Computational Intelligence and Neuroscience. pp. 1–14. DOI: https://doi.org/10.1155/2020/8853835.

31. Sabeti, M., Boostani, R., Katebi, S.D., et al., 2007. Selection of relevant features for EEG signal classification of schizophrenic patients. Biomedical Signal Processing and Control. 2, 122–134. DOI: https://doi.org/10.1016/j.bspc.2007.03.003

32. Prabhakar, S.K., Rajaguru, H., Lee, S.W., 2020. A Framework for Schizophrenia EEG Signal Classification With Nature Inspired Optimization Algorithms. IEEE Access. 8, 39875–39897. DOI: https://doi.org/10.1109/ACCESS.2020.2975848

33. Kutepov, I.E., Dobriyan, V.V., Zhigalov, M.V., et al., 2020. EEG analysis inpatients with schizophrenia based on Lyapunov exponents. Informatics in Medicine Unlocked. 18, 100289. DOI: https://doi.org/10.1016/j.imu.2020.100289.

34. Buettner, R., Beil, D., Scholtz, S., et al., 2020. Development of a Machine Learning Based Algorithm To Accurately Detect Schizophrenia based on One-minute EEG Recordings. DOI: https://doi.org/10.24251/HICSS.2020.393.

35. Parvinnia, E., Sabeti, M., Zolghadri Jahromi, M., et al., 2014. Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm. Journal of King Saud University-Computer and Information Sciences. 26, 1–6. DOI: https://doi.org/10.1016/j.jksuci.2013.01.001.

36. Sharma, M., Acharya, U.R., 2021. Automated detection of schizophrenia using optimal wave-let-based $$1_1$$ l_1 norm features extracted from single-channel EEG. Cognitive Neurodynamics. 15, 661-674. https://doi.org/10.1007/s11571-020-09655-w.

37. Darbas, M., Lohrengel, S., 2019. Review on Mathematical Modelling of Electroencephalography (EEG). Jahresbericht Der Deutschen Mathematiker. 121, 3–39. DOI: https://doi.org/10.1365/s13291-018-0183-z.

38. Siuly, S., Li, Y., Zhang, Y., 2016. Electroencephalogram (EEG) and Its Background. EEG Signal Analysis and Classification. pp. 3–21. DOI: https://doi.org/10.1007/978-3-319-47653-7_1.

39. EEG Database -Schizophrenia. MsuRu n.d. http://brain.bio.msu.ru/eeg_schizophrenia.htm (Accessed on 10 October 2021).

40. Phinyomark,A., Hirunviriya, S., Limsakul, C., et al., 2010. Evaluation of EMG feature extraction for hand movement recognition based on Euclidean distance and standard deviation. ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. pp. 856–860.

41. Yu, X., Gen, M., 2010. Introduction to evolutionary algorithms. London; New York: Springer.

42. Bansal, S., Sharma, K., 2018. Nature-Inspired-Based Modified Multi-Objective BB-BC Algorithm to Find Near-OGRs for Optical WDM Systems and Its Performance Comparison. Handbook of Research on Biomimicry in Information Retrieval and Knowledge Management. pp. 1–25. DOI: https://doi.org/10.4018/978-1-5225-3004-6.ch001

43. Bansal, S., Gupta, N., Singh,A.K., 2021. Application of Bat-Inspired Computing Algorithm and Its Variants in Search of Near-Optimal Golomb Rulers for WDM Systems: A Comparative Study. In: Dey N, Rajinikanth V, editors. Applications of Bat Algorithm and its Variants, Singapore: Springer. pp. 79–101. DOI: https://doi.org/10.1007/978-981-15-5097-3_5

44. Wavelet Scattering. MathworksCom n.d. https://uk.mathworks.com/help/wavelet/ug/wavelet scatter- ing.html.

45. Mallat, S., 2012. Group Invariant Scattering. Communications on Pure and Applied Mathematics. 65, 1331–1398. DOI: https://doi.org/10.1002/cpa.21413.

46. Bruna, J., Mallat, S., 2013. Invariant Scattering Convolution Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 35, 1872–1886. DOI: https://doi.org/10.1109/TPAMI.2012.230

47. Lostanlen, V., Scattering, M., 2021.

48. Liu, Z., Yao, G., Zhang, Q., et al., 2020. Wavelet Scattering Transform for ECG Beat Classification. Computational and Mathematical Methods in Medicine. pp. 3215681. DOI: https://doi.org/10.1155/2020/3215681.

49. Andén, J., Mallat, S., 2011. Multiscale Scattering for Audio Classification. pp. 657–662.

50. Nsugbe, E., Phillips, C., Fraser, M., et al., 2020. Gesture recognition for transhumeral prosthesis control using EMG and NIR. IET Cyber-Systems and Robotics. 2, 122–131. DOI: https://doi.org/10.1049/iet-csr.2020.0008.

51. Nsugbe, E., Samuel, O.W., Asogbon, M.G., et al., 2021. Phantom motion intent decoding fortranshu-meral prosthesis control with fused neuromuscular and brain wave signals. IET Cyber-Systems Robot-ics. 3, 77–88. DOI: https://doi.org/10.1049/csy2.12009.

52. Nsugbe, E., Obajemu, O., Samuel, O.W., et al., 2021. Application of noninvasive magnetomyography in labour imminency prediction for term and preterm pregnancies and ethnicity specific labour prediction. Machine Learning with Applications. 5, 100066. DOI: https://doi.org/10.1016/j.mlwa.2021.100066.

53. Nsugbe, E., Samuel, O.W., Asogbon, M.G., et al., 2020 . A self-learning and adaptive control scheme for phantom prosthesis control using combined neu-romuscular and brain-wave bio-signals. Engineering Proceedings. 2, 59. DOI: https://doi.org/10.3390/ecsa-7-08169.

54. Bochner, S., Chandrasekharan, K., 1950. Fourier Transforms. (AM- 19). Princeton: Princeton University Press.

55. Baradits, M., Kakuszi, B., Bálint, S., et al., 2019. Alterations in resting-state gamma activity inpatients with schizophrenia: a high-density EEG study. European Archives of Psychiatry and Clinical Neuroscience. pp. 269. DOI: https://doi.org/10.1007/s00406-018-0889-z.

56. Janani,A.S., Grummett, T.S., Bakhshayesh, H., et al., 2018. How many channels are enough?: 26th Euro- pean Signal Processing Conference, EUSIPCO 2018. 2018 26th European Signal Processing Conference, EUSIPCO 2018. pp. 101–105. DOI: https://doi.org/10.23919/EUSIPCO.2018.8553261.