Computers in Biology and Medicine 155 (2023) 106676

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals

Smith K. Khare [a,*], U. Rajendra Acharya [b,c,d,e,f]

[a] Electrical and Computer Engineering Department, Aarhus University, 8200, Aarhus, Denmark
[b] School of Mathematics, Physics, and Computing, University of Southern Queensland, Springfield , Australia
[c] Department of Biomedical Engineering, School of Science and Technology, Singapore University of Social Sciences, Singapore
[d] Department of Biomedical Informatics and Medical Engineering, Asia University, Taiwan
[e] Kumamoto University, Japan
[f] University of Malaya, Malaysia

ARTICLE INFO

Keywords:
Attention deficit hyperactivity disorder
Electroencephalography
Variational mode decomposition
Explainable machine learning
Interpretable machine learning

ABSTRACT

Background: Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder that affects a person's sleep, mood, anxiety, and learning. Early diagnosis and timely medication can help individuals with ADHD perform daily tasks without difficulty. Electroencephalogram (EEG) signals can help neurologists to detect ADHD by examining the changes occurring in it. The EEG signals are complex, non-linear, and non-stationary. It is difficult to find the subtle differences between ADHD and healthy control EEG signals visually. Also, making decisions from existing machine learning (ML) models do not guarantee similar performance (unreliable).

Method: The paper explores a combination of variational mode decomposition (VMD), and Hilbert transform (HT) called VMD-HT to extract hidden information from EEG signals. Forty-one statistical parameters extracted from the absolute value of analytical mode functions (AMF) have been classified using the explainable boosted machine (EBM) model. The interpretability of the model is tested using statistical analysis and performance measurement. The importance of the features, channels and brain regions has been identified using the glass-box and black-box approach. The model's local and global explainability has been visualized using Local Interpretable Model-agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), Partial Dependence Plot (PDP), and Morris sensitivity. To the best of our knowledge, this is the first work that explores the explainability of the model prediction in ADHD detection, particularly for children.

Results: Our results show that the explainable model has provided an accuracy of 99.81%, a sensitivity of 99.78%, 99.84% specificity, an F-1 measure of 99.83%, the precision of 99.87%, a false detection rate of 0.13%, and Mathew's correlation coefficient, negative predicted value, and critical success index of 99.61%, 99.73%, and 99.66%, respectively in detecting the ADHD automatically with ten-fold cross-validation. The model has provided an area under the curve of 100% while the detection rate of 99.87% and 99.73% has been obtained for ADHD and HC, respectively.

Conclusions: The model show that the interpretability and explainability of frontal region is highest compared to pre-frontal, central, parietal, occipital, and temporal regions. Our findings has provided important insight into the developed model which is highly reliable, robust, interpretable, and explainable for the clinicians to detect ADHD in children. Early and rapid ADHD diagnosis using robust explainable technologies may reduce the cost of treatment and lessen the number of patients undergoing lengthy diagnosis procedures.

1. Introduction

Attention deficit hyperactivity disorder (ADHD) or hyper-kinetic disorder is one of the most common neurodevelopmental disorders of childhood and also affects adults. Global pervasiveness of ADHD in children and adolescents is approximately 2.2% (0.1–8.1%) while in adults it is 2.8% (0.6–7.3%) [1]. The pervasiveness of ADHD is 9.4% in the USA, 7% in western countries, and 1.6 to 17.9% in India [2,3]. The pervasiveness of ADHD is higher among boys than girls during childhood and males than females during adulthood [2]. In India, the

* Corresponding author.
E-mail address: smith7khare@gmail.com (S.K. Khare).

https://doi.org/10.1016/j.compbiomed.2023.106676
Received 5 October 2022; Received in revised form 9 January 2023; Accepted 13 February 2023
Available online 18 February 2023

prevalence of ADHD in school children was found to be 11.3%, whereas it was 6% in 9–10 and 10–15 year age groups, respectively [2]. Children and adults who have ADHD may experience symptoms like impulse control (oppositional defiant, borderline personality, and antisocial personality), mood (depressive, dysthymic, bipolar, and cyclothymic), anxiety (generalized anxiety, agoraphobia, post-traumatic stress, and social phobia), substance use (alcohol and drug dependence), learning problems (reading, written expression, and mathematics), and sleep disorders (rhythm disturbance, excessive daytime sleepiness, obstructive sleep apnea, restless legs/periodic limbs, etc.) [2]. The demographic variable risk factors of ADHD in children include parents' educational level, gender, mother's occupation, living with a single parent, and birth order [4,5]. The factors related to a family are brain injury, consumption of alcohol and tobacco during pregnancy, premature delivery, low birth weight, and exposure to chemicals like lead during pregnancy or at a young age [1,3,6].

Late or untreated ADHD increases risk factors like poor social skills, delayed learning, poor academic performances, low self-confidence and esteem, increased probability of committing crimes, and susceptibility to injury [7,8]. Timely detection, effective diagnosis, and proper medication help children with ADHD perform routine tasks and learn . Over time, different techniques like a one-to-one interview with an expert, continuous performance tests (CPTs), neuroimaging techniques, electrocardiogram (ECG) signals, magnetoencephalogram (MEG) signals, and electroencephalogram (EEG) signals have been suggested for the detection of ADHD [9–12]. The CPTs and one-to-one interviews with an expert are time-consuming, error-prone, and have a high probability of false detection. Neuroimaging techniques are radioactive, bulky, costly, and require additional recordings [9]. Koh et al. [11] have obtained low performance in the detection of ADHD using ECG signals. The MEG signals acquisition is bulky, costly, and requires isolated room to suppress surrounding noise which makes their usage in ADHD detection very limited [11]. The EEG signals are capable to track the changes occurring in the brain during ADHD. The EEG signals have been used to detect brain abnormalities and physiological disorders like motor imagery tasks, seizures, emotions, and drowsiness [13–16]. Also, EEG acquisition is non-invasive, portable, and low-cost solution.

Over the years, many EEG-related studies have been presented for the automatic detection of ADHD. The analysis of event-related potential (ERP) investigated the neurological changes happening during auditory tasks of GO and NOGO in different brain regions in ADHD subjects [17]. Their study showed the brain's frontal portion has differences in ERP during pre and post-treatment. The EEG signals acquired during the CPTs of GO and NOGO sessions have been investigated using bandpass filtering, independent component analysis (ICA), and K-nearest neighbour (KNN) classifier [18]. The study of brain functional dynamics has been done using different non-linear features, frequency band-power features, radial basis function neural network (RBFNN), and support vector machine (SVM) [19]. In [20], the classification of EEG signals using an SVM classifier and logical Karnaugh map during visual CPT, emotional CPT, eyes open, and eyes closed have been performed.

A fast Fourier transform (FFT) has been investigated to study the absolute and relative activity of EEG signals from delta (1–3 Hz), theta (4–7 Hz), alpha1 (8–9 Hz), alpha2 (10–13 Hz), total alpha (8–13 Hz), and beta (14–30 Hz) rhythms. Their study found increased theta activities in anterior regions whereas the reduced posterior beta activity in ADHD EEG segments [21]. The use of FFT with the Hanning window to evaluate relative power spectral density (PSD), absolute PSD of rhythms, and total PSD of EEG signals have been classified using the SVM classifier [22]. The filtering techniques have been analyzed to study spectral analysis of mean and relative mean power of rhythms to detect ADHD [23]. Their study reported higher average power of delta and theta rhythms and increased delta to a beta ratio in ADHD than that of HC children [23]. Various nonlinear features based on correlation dimension (CorrD), fractal dimension (FD), entropies, and Lyapunov exponent (LE) have been classified using the SVM classifier [24]. Many nonlinear features (LE, Katz, Higuchi, and Sevcik FD) are extracted and classified using the multilayer perceptron (MLP) model [25,26]. The SVM and KNN classifiers have been studied to classify covariance, Burg covariance, and relative power features selected using the covariance technique [27]. The SVM classifier has been presented to classify the features extracted from the preprocessed EEG signals using ICA [28]. An adaptive model called VHERS has been explored to detect ADHD in children [29]. Different rhythms' mean power and ratios have been studied to detect ADHD [30,31]. The convolutional neural networks (CNN) model capable of extracting and classifying in-depth features has been used to detect ADHD [32,33]. In [34], empirical mode decomposition (EMD) and discrete wavelet transform techniques have been used to extract different bands. The nonlinear features from these bands have been classified using various classifiers. In [35], nonlinear causal relationship estimation by an artificial neural network (nCREANN) and direct directed transfer function has been used to extract the spectral information of connectivity patterns for ADHD detection. A hybrid model composed of long–short term memory and CNN (ConvLSTM) with an attention mechanism has been developed for encoding the spatiotemporal representation for ADHD detection [36]. In [37], a combination of directed phase transfer entropy, feature selection by genetic algorithm, and the artificial neural network has been used to detect ADHD. An eight-point start pattern to construct a direct graph combined with tunable q wavelet transform and wavelet packet decomposition has been used for feature extraction classified with KNN classifiers [38].

Studies discussed in the literature have shown limited performance in the extraction of hidden information from EEG signals [10]. Few methods used direct feature extraction from EEG signals, whereas techniques like FD, entropy and LE produced degraded performance due to noise and inappropriate selection of scaling range. The filtering techniques that use rhythmic analysis require proper filter coefficients for sharp boundaries of filters. FFT suffers from noise sensitivity, localization in time–frequency domain, poor spectral estimation, and improper localized spikes [39]. Although CNN-based techniques simultaneously provide feature extraction and classification but have higher memory requirements. In addition, the existing techniques use an empirical choice of ML models. One ML model does not guarantee similar performance when multiple testing combinations are used. The current ML lacks robustness and local and global explainability of the models in making predictions [40]. This urgently demands an effective, robust, explainable model for detecting ADHD and HC EEG segments. To address this, an explainable and interpretable model combining variational mode decomposition (VMD) and Hilbert transform (HT)-based two-level signal analysis method VMD-HT and explainable boosting machine (EBM) classifier is used to detect ADHD in children. To the best of our knowledge, this is the first work that explores the explainability of the model prediction in ADHD detection, particularly for children. The model was evaluated using multiple classifier performance using a ten-fold cross-validation technique. The effectiveness of the explainable model was evaluated by computing nine performance measures (PM) such as accuracy (ACC), precision (PRS), negative predictive value (NPV), critical success index (CSI), sensitivity (SNS), false detection rate (FDR), specificity (SPF), F-1 score (F-1), and Mathew's correlation coefficient (MCC). Further, we have used a glass-box and black-box model interpretations using an EBM model with four explainable characteristics i.e. Local Interpretable Model-Agnostic Explanations (LIME), Partial Dependence Plot (PDP), Shapley Additive explanations (SHAP), and Morris sensitivity (MS) to find local and global explainability of the model. Finally, we have tested our developed explainable model using the ten-fold cross-validation technique, and compared the model performance with existing state-of-the-art (SOTA) classifications and ADHD detection techniques to study the model robustness. The remaining part of the paper is arranged as follows: Section 2 presents materials and methods, Section 3 gives results and performance analysis, discussion on existing SOTA is provided in Section 4, and Section 5 concludes the paper.

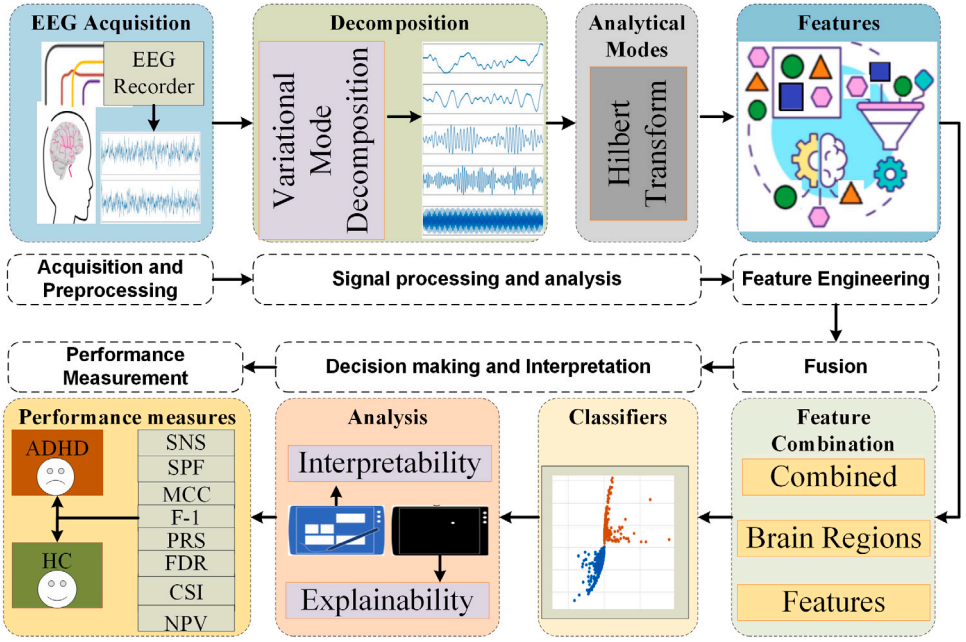

Fig. 1. Graphical representation of the proposed explainable and interpretable ADHD model.

2. Materials and methods

This section consists of details of the EEG dataset, analysis to obtain analytic modes using VMD-HT, features, and decision-making. The steps of the proposed explainable and interpretable model are shown in Fig. 1.

2.1. Dataset

This work used a publicly available EEG dataset from IEEE data port [41,42]. It comprises 60 HC and 61 ADHD children, including boys and girls with age ranging from 7–12 years. An experienced psychiatrist diagnosed subjects with ADHD as per DSM-IV criteria [43]. ADHD subjects have taken Ritalin (is a central nervous system stimulant medication used to treat ADHD) for about six months. HC children had no history of hard head injury, epilepsy, drug abuse, psychiatric disorders, or any reported high-risk behaviors. A visual attention task was used to record EEG since one of the deficits in ADHD is visual attention. Children were asked to count the number of cartoon characters from pictures shown during the task. For a better visibility, the picture size was kept large enough to be countable by children. The image shown to them was randomly selected, consisting of five to sixteen characters. Images were displayed immediately and uninterrupted after the child's response to a continuous stimulus during EEG recordings. The duration of EEG recordings depended on a child's response speed during the cognitive visual task. A 19-channel EEG recorder built per a 10–20 international system was used to record EEG of frontal, parietal, occipital, and central regions (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2). Two electrodes were below and above the right eye to record eye movement. Each EEG recording was recorded using a sampling frequency of 128 Hz. A notch filter of 50 Hz removes the noise and power line disturbances. A sixth-order Butterworth bandpass filter (0.1–60 Hz) was used to remove the unwanted frequencies and artifacts. After pre-processing, the EEG signals were segmented into 4-sec durations. Finally, we have obtained total EEG segments of 2289 and 1843 for ADHD and HC subjects, respectively. To perform the brain region analysis, we have considered six different brain regions: prefrontal, frontal, central, parietal, occipital, and temporal [44]. The prefrontal is composed of Fp1 and Fp2, channels F3, F4, F7, F8, and Fz make up the frontal region, the central region is formed by C3,

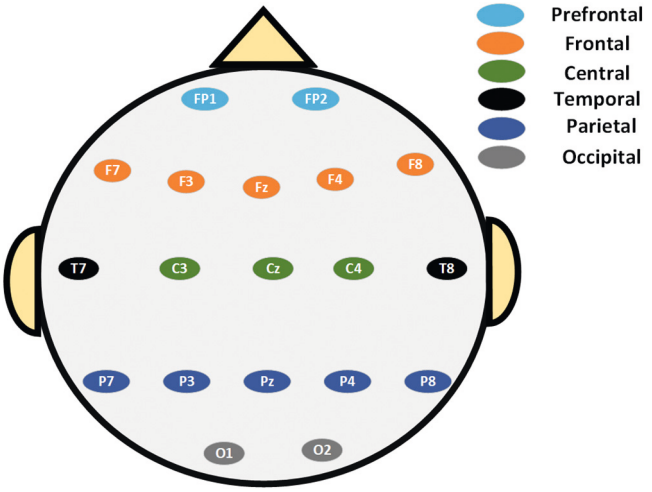

Fig. 2. Brain region configuration for different channels.

C4, and Cz channels, the parietal is formed by P3, P4, P7, P8, and Pz channels, occipital is composed of O1 and O2 while temporal is from T7 and T8 channels, respectively. The brain region formation with its corresponding channel configuration is shown in Fig. 2.

2.2. Variational Mode Decomposition-Hilbert Transform (VMD-HT)

EEG signal in its raw form is difficult to analyze and classify due to its non-linearity, complexity, and non-stationary nature. To get detailed information and representative characteristics of a signal it is required to decompose it into multi-components. However, the existing methods of decomposition yielded limited performance. For example (i) empirical mode decomposition (EMD): ad-hoc nature due to lack of mathematical theory and lacks the ability for backward error correction due to recursive shifting, (ii) empirical wavelet transform (EWT): require predefined filter-bank boundaries, (iii) EMD and EWT: inability to cope with noise properly, and (iv) wavelets: suffers hard-band limits. The ability of VMD to concurrently estimate the corresponding modes

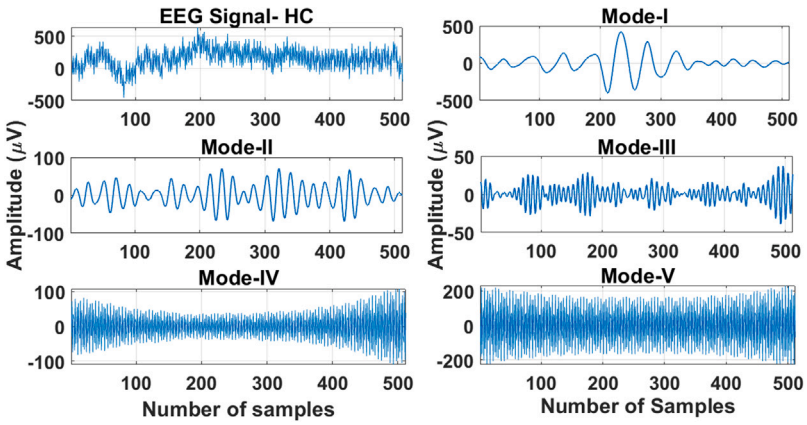

Fig. 3. Typical EEG signal and its corresponding modes obtained from VMD.

and adaptively determine the relevant band and hence, balance the error between them [45]. The modes generated by VMD are such that they are band-limited about a center frequency and collectively reconstruct an input signal optimally either perfectly or in a least-square sense. Moreover, the modes generated using VMD have some optimality in dealing with noise. It is due to the above reason that VMD has been used in the analysis of mechanical gearbox fault diagnosis and biomedical signal processing [43,46–48].

VMD decomposes a real-valued input signal into a fixed number of components with specific sparsity properties. Bandwidth in the spectral domain is chosen as a sparsity for each mode. The bandwidth of each mode can be accessed by following steps: (a) the unilateral frequency spectrum of each mode is computed by applying HT; (b) the frequency spectrum of each mode is shifted to "baseband" by respective estimated center frequency; and (c) Finally, squared $L2$ norm is applied to estimate the bandwidth which results into following constraint problem [45]:

$$\min_{\{h_m\},\{w_m\}} \sum_m \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * h_m(t) \right] e^{-jw_m t} \right\|_2^2 \quad \text{s.t.} \quad \sum_m h_m = z \tag{1}$$

where $z(t)$ is the original signal, h_m is the mth mode, M is total number of modes, and w_m is the corresponding frequencies of mth mode. The constraint problem is converted into an unconstrained one by using Lagrangian multipliers ($\vec{\lambda}$) and quadratic penalty factor ($\hat{a}$). Thus, an introduction of $\hat{a}$ and $\vec{\lambda}$ for a better convergence property. The augmented Lagrangian $\mathcal{L}$ is denoted as [45]:

$$\begin{aligned} &\mathcal{L}(\{h_m\},\{w_m\},\vec{\lambda}) = \\ &\hat{a} \sum_m \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * h_m(t) \right] e^{-jw_m t} \right\|_2^2 \\ &+ \left\| z(t) - \sum_m h_m(t) \right\|_2^2 + \left\langle \vec{\lambda}(t), z(t) - \sum_m h_m(t) \right\rangle \end{aligned} \tag{2}$$

A typical example of modes obtained after VMD decomposition is illustrated in Fig. 3.

In the second stage, HT is applied to modes of VMD. One-dimensional (1-D) HT ($\mathcal{H}$) is a shift-invariant, and linear operator that maps all 1-D sine functions into their corresponding sine functions and vice-versa [49]. HT is assumed as an all-pass filter characterized by a transfer function denoted as:

$$\hat{g}(w) = -j\,sgn(w) = \frac{-jw}{|w|} \tag{3}$$

Thus, HT in spectral domain is a multiplier operator whose impulse response is denoted as $g(t) = 1/(\pi t)$. The HT of any real-valued function ($z(t)$) is defined as Cauchy principal value ($p.v.$) of convolutional integral [45]:

$$\begin{aligned} \mathcal{H}(z(t)) &= z(t) * \frac{1}{\pi t} \\ &= \frac{1}{\pi} p.v. \int_{-\infty}^{\infty} \frac{z(T)}{t-T} dT \end{aligned} \tag{4}$$

A purely real signal ($z(t)$) in its analytical form can be represented as:

$$\begin{aligned} z_a(t) &= z(t) + j\mathcal{H}(z(t)) \\ &= A_m(t) e^{j\phi(t)} \end{aligned} \tag{5}$$

where $A_m(t) = \sqrt{z(t)^2 + \mathcal{H}(z(t))^2}$ is the magnitude of signal $z(t)$, and $\pi(t)$ is an instantaneous phase $= arctan(\mathcal{H}(z(t))/z(t))$. The analytical modes obtained after applying HT called an AMF if and only if it satisfies the following conditions [50].

1. The mean value must be zero for an envelope defined by its local minima and maxima at any time.
2. The number of zero crossings and extremes must be equal or, at most, differ by one.
3. The weighted bandwidth of L_2 norm must be small.
4. Analytic mode is $\mathcal{H}(z(t)) = A_m(t)e^{j\phi(t)}$, such that A_m and $\phi(t)$ must be smoothed functions and $\phi(t) > 0$.

The AMF obtained from VMD-HT extracts the following properties of EEG signals:

1. AMF rotates around a unique center in a specific direction.
2. Due to the multi-resolution properties of generated AMF, it will extract hidden and representative information of EEG signals.
3. The variations of EEG signals are captured in terms of time-dependent frequency and amplitude components.

Amplitude envelope (AE) is obtained from different AMFs. Figs. 4 and 5 show an example of a typical amplitude envelope obtained using VMD-HT for HC and ADHD subjects, respectively. It is evident from Figs. 4 and 5 amplitudes of EEG signals of ADHD and HC children are in the same amplitude range, due to which getting representative information is challenging. But, there is a significant difference in amplitude envelope fifth and sixth AMF of ADHD and HC subjects. Hence, AE obtained from VMD-HT provide discriminant information compared to the EEG signals.

2.3. Features

In this work, we have extracted many features from the analytical modes obtained from VMD-HT. A total of forty-one features including entropy measures, statistical measures, and nonlinear measures. These features are denoted in Table 1. The details of these features are discussed in [14,51–58].

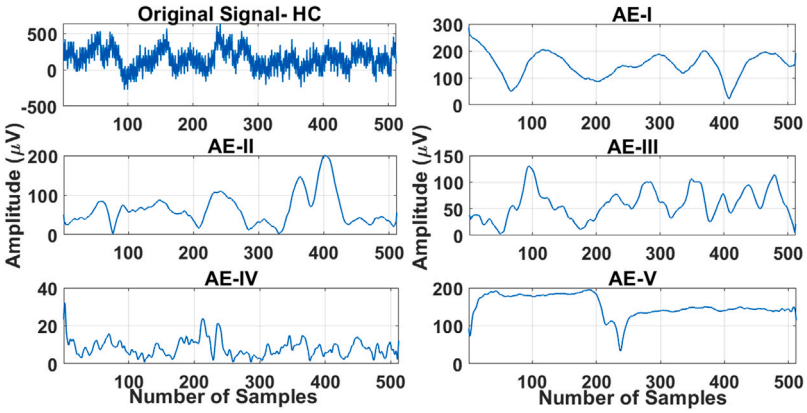

Fig. 4. Examples of an absolute amplitude envelope obtained for HC.

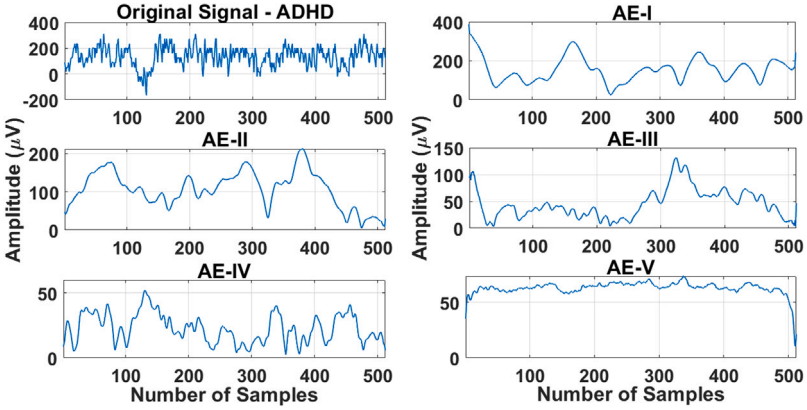

Fig. 5. Examples of an absolute amplitude envelope obtained for ADHD.

Table 1
Details of the features used for our proposed model.

Nonlinear features	Statistical features
Renyi entropy (RE) [14,51]	Minima (MIN) [53]
Corrected conditional entropy (CCEN) [55]	Second quartile (Q2) [56]
Conditional entropy (CEN) [55]	Third quartile (Q3) [14]
Tsallis entropy (TEN) [14,55]	Maxima (MAX) [53]
Fuzzy entropy (FEN) [52]	Median (MED) [53]
Sure entropy (SuE) [56]	Mean (MEAN) [53]
Shannon entropy (SHEN) [14,55]	Second moment (MOM2) [56]
Permutation entropy (PEN) [14]	Third moment (MOM3) [59]
Kraskov entropy (KEN) [56]	Standard deviation (SD) [53]
Log entropy (LEN) [14]	Peak value (PEAK) [56]
Average amplitude change (AAC) [59]	Integral EEG (IEEG) [59]
Wavelength (WL) [59]	First quartile (Q1) [59]
Hjorth activity (HA) [56]	Inter quartile range (IQR) [14]
Hjorth mobility (HM) [14]	Root mean square (RMS) [14]
Higuchi fractal dimensions (HFD) [52]	Mean absolute value (MAV) [56]
Skewness (SKEW) [56]	Clearance factor (CF) [60]
Kurtosis (KURT) [56]	Mean deviation from max (M1)
Spectral flatness (SPF) [60]	Mean deviation from min (COV)
Hjorth complexity (HjC) [14]	Mean deviation from range (R1)
Hurst exponent (HE) [52,54]	–
Modified mean absolute value type1 (MAV1) [59]	–
Modified mean absolute value type2 (MAV2) [59]	–

2.4. Decision making

Nowadays, many artificial intelligence (AI) models have been increasingly used for accurate prediction in healthcare, finance, industries, etc. These models are facilitated and eased human lives in one or the other way, especially in healthcare data analytics. As the decisions made by the underlying ML or DL model directly affect the life and death of an individual, it becomes necessary for the researchers to make a predictable, interpretable, robust, and explainable decision-making model. Therefore, the predictions and decisions made by such models are fair with the least bias or with the least discrimination. Thus, this paper presents an explainable and interpretable ML model for the decision-making of ADHD in children. For this, an EBM model in combination with black-box explainers like LIME, SHAP, PDP, and MS is explored [61].

2.4.1. Glass-box model

The models designed to be interpretable are termed glass-box or intrinsic models. The glass-box models make model-specific interpretations with an exact explanation of losses or gains. In this work, an EBM model is a generalized additive model (with automatic interaction detection) is used to predict ADHD and HC classes [62]. The EBM model provides detection accuracy comparable to state-of-the-art black-box models while being perfectly interpretable. The EBM models provide slow training of the model compared to other glass-box models with highly compact and fast at prediction time. We have also compared the ACC of EBM model with various benchmark classifiers like decision tree (DT), KNN, random forest (RF), and narrow neural network (NNN) models to show the effectiveness and robustness of our developed model. The generalized additive model for EBM can be expressed as [62]

$$f(E[y]) = \beta_0 + \sum g_i(x_i) \tag{6}$$

where β_0 is intercept, y is observations and x_i are dependent variables. The learning or training of each feature function g_i of EBM is performed

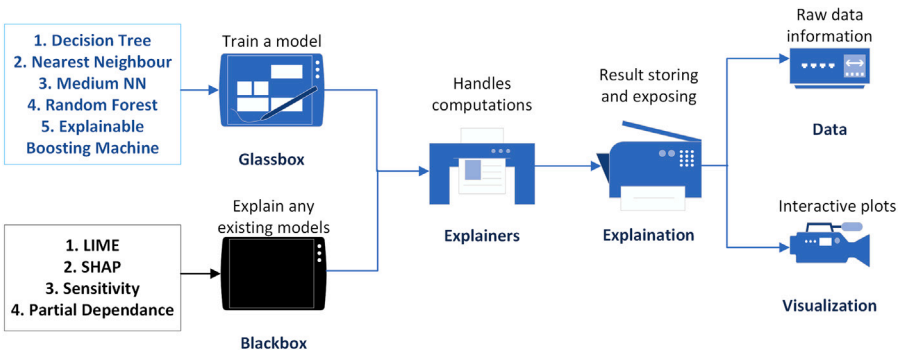

Fig. 6. Illustrative interpretations and explanations procedure using glass-box and black-box models.

by using either boosted gradient or bagging algorithm. The training is performed on each feature function in a round-robin fashion with a very low learning rate. The round-robin provides an insight scenario of how each feature function is contributing in making predictions. EBM can also detect and include pairwise interaction terms by [62]

$$f(E[y]) = \beta_0 + \sum g_i(x_i) + \sum g_{j,i}(x_j, x_i) \tag{7}$$

2.4.2. Black-box model

The black-box explainers (post hoc technique) are model agnostic, which provide model analysis after training. It assumes access to the model's inputs and output, thus, only providing approximate explainability. The models are trained on existing techniques. The black-box models offer local as well as global explanations. Local explanations aim to explain individual predictions, while global descriptions explain model behavior. In our work, we have used LIME and SHAP for predicting local explanation whereas PDP and MS for global explainability.

- ***LIME***: creates a brand-new dataset made up of altered samples and the related black-box model predictions. The weighting of the interpretable model that LIME trains on this new dataset is based on how close the sampled examples are to the instance of interest. Any model like DT, may be used as the interpretable model. Locally, the learned model should be a good approximation of the predictions made by the ML model; however, a good global approximation is not required. This kind of accuracy is also called local fidelity. The LIME explanations can be denoted as [63]

$$explanations(y) = \arg\min_{f \epsilon F} L(g, f, \pi_y) + \Omega(f) \tag{8}$$

  where y is the instances in the model f, L is the loss function measuring the closeness of explanation w.r.t. predicted model g, $\Omega(f)$ is the complexity measured by number of feature matrix or variables, and π_y is proximity measure.
- ***SHAP***: Shapley values are calculated using the SHAP explanation approach from coalitional game theory [64]. The feature values of a data instance act as players in a coalition. We can equally distribute the "payout" (i.e., the prediction) among the characteristics by using Shapley values. The Shapley value explanation is portrayed as an additive feature attribution approach, a linear model, which is its novelty. SHAP specifies the explanation as [64]

$$g(y') = \phi_0 + \sum_{j=1}^{M} \phi_j y'_j \tag{9}$$

  where explanation model is denoted by $g(y')$, $y' = \{0,1\}^M$ is coalition vector, M denotes maximum coalition size, and ϕ_j is the feature attribution.
- ***MORRIS***: It is often known as the one-step-at-a-time (OAT) global sensitivity analysis, modifies the level (discretized value) of just one input every run [65,66]. In comparison to other sensitivity analysis algorithms, it is quicker (needs fewer model iterations), but at the cost of being unable to discern non-linearities from interactions. This is used to assess whether inputs are important enough for further analysis.
- ***PDP***: It displays the small influence of one or two features on an ML model's anticipated result [67]. A PDP can demonstrate whether there is a linear, monotonic, or complicated relationship between the target and a feature. Partial dependence graphs, for instance, always display a linear connection when used with a linear regression model. It is denoted as [67]

$$\hat{f}_s(x_s) = E_{X_c}[\hat{f}_s(x_s, X_C)] = \int \hat{f}_s(x_s, X_C) d\,P(X_C) \tag{10}$$

  where x_s are the features and X_C is the features used in the ML model $\hat{f}_s$

The graphical representation of glass-box and black-box interpretations and explanations steps are illustrated in Fig. 6 [61].

3. Results

An EEG-stratified signal of 4s is given as input to VMD. Several decomposition parameters of VMD must be tuned to decompose the signals into modes. In this work, penalty factor $\hat{a} = 2000$; time-step of the dual ascent $\tau = 0$; modes $M = 5$; to distribute all omegas start uniformly $init = 1$; and tolerance of convergence criterion = 1e-3; are selected. It is noteworthy that all the parameters are selected empirically and kept uniform throughout the experiment. Modes obtained from VMD are converted to their analytical form using HT to evaluate amplitude envelope from its corresponding AMF. Different features are computed to reduce amplitude envelope dimensionality and extract representative information forms. Forty-one attributes have been evaluated from the 19 channels to form a combined feature matrix. The combined feature matrix of size 779 × 4132 is applied to each mode's EBM and other glass-box models. The feature matrix for HC EEG segments is 779 × 1843, while for ADHD, it is 779 × 2289. The exemplary histogram distribution for ADHD and HC EEG segments obtained for Hjorth mobility, Shannon entropy, fuzzy entropy, and Renyi entropy is denoted in Fig. 7 while the histogram distribution obtained for ADHD and HC EEG segments of T7, P3, F7, and Fp1 channel is shown in Fig. 8. It is evident from the histogram plots that all the features provide discriminable ability due to the non-uniform distribution of both the classes and the channels.

After obtaining the feature matrix, it is fed as an input to the glass-box model, i.e. EBM, to check the detection ability. The classification is performed using a tenfold cross-validation technique. The boosted tree algorithm is used for making predictions with 50 learners. The learning rate is set to default within the initialization. The feature matrix is

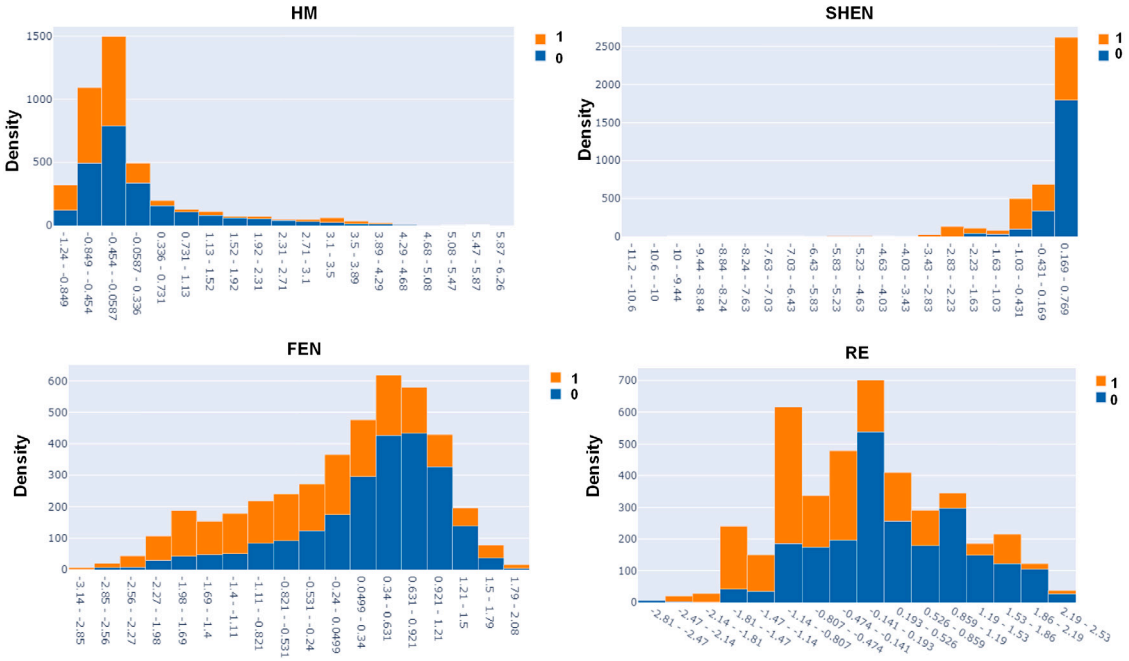

Fig. 7. Histogram of features obtained for ADHD (denoted by 0) and HC (denoted by 1) EEG segments.

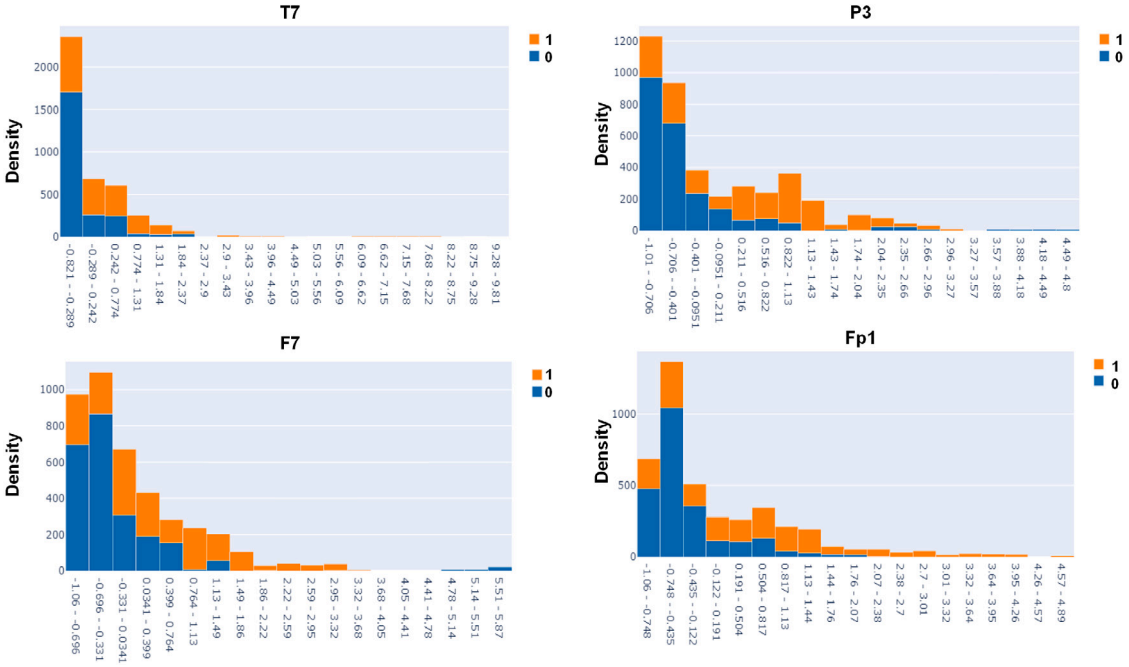

Fig. 8. Histogram of channels obtained for ADHD (denoted by 0) and HC (denoted by 1) EEG segments.

divided into ten equal parts, of which nine parts have been used for training and one for testing. To test the significance of the proposed EBM model, we have used benchmark glass-box models like DT, KNN, NNN, and RF to evaluate the performance measure. A flow diagram of ML interpretability and explainability using glass-box and black-box analysis is shown in Fig. 9. At first, the feature matrix (combining features and channels) is classified using EBM and benchmark glass-box models for all the modes. The best mode is identified and considered for further evaluation. Second, brain region analysis is performed for benchmark glass-box models including EBM for the best mode. The best performing region and classifier model are identified. Later, the EBM model analysis for interpretability (feature and channel rank using univariate analysis) and explainability (LIME, SHAP, PDP, and MS) analysis. Finally, performance measurement and comparison are performed for the combined feature matrix, frontal region, and third quartile feature.

Table 2 shows the ACC obtained using the benchmark glass-box model in different modes for combined features. The first mode yielded the least ACC, indicating that its characteristics overlap in a certain way. For each glass-box model, the fifth mode, in contrast, has offered the most significant identification rate. In mode five, the highest ACC with 99.81% is provided by the EBM glass-box model, while the least is obtained for the DT classifier. Table 2 reveals that the fifth mode has the most discriminable properties, which have been captured and

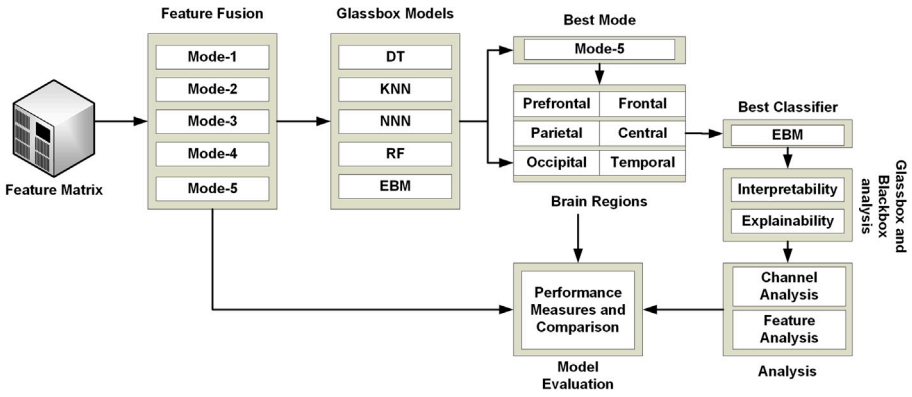

Fig. 9. Flow diagram of ML interpretability and explainability using glass-box and black-box analysis.

Table 2
Accuracy (%) obtained for glass-box models with combined feature matrix (Features × Channels).

Classifier	M-1	M-2	M-3	M-4	M-5
DT	69.96	67.02	69.75	73.57	97.23
KNN	71.92	69.36	74.45	82.95	98.97
NNN	74.47	71.75	77.48	81.73	98.21
RF	77.77	75.10	78.65	83.36	99.02
EBM	81.12	78.97	80.20	84.15	99.81

Table 3
Accuracy (%) obtained for different brain region with fifth mode using glass-box models.

Region	DT	KNN	MLP	RF	EBM
Prefrontal	86.56	86.99	84.81	93.10	94.39
Frontal	94.32	96.01	94.46	98.97	99.14
Central	90.76	90.09	95.53	94.57	96.36
Parietal	94.60	95.89	93.91	97.71	98.21
Occipital	82.83	85.03	84.36	92.31	92.93
Temporal	86.89	86.39	87.03	91.12	91.83

predicted accurately by the EBM model. This is because the amplitude and frequency contents in the fifth mode for ADHD and HC differs significantly, which resulted in the generation of non-overlapping features. As the fifth mode has extracted better time-dependent frequency and amplitude components of EEG signals, further analysis is performed in this mode. The effectiveness of ML models varies greatly. A model's superior performance on one type of analysis does not imply that it will do equally well on other forms of analysis. Therefore, to evaluate the effectiveness of the EBM model, we have performed the brain region analysis. The ACC obtained using DT, KNN, NNN, RF, and EBM models for various brain regions in the fifth mode is shown in Table 3. It is evident from Table 3, the frontal region has the highest distinguishable characteristics compared to other brain regions. The ACC of 99.14% for the frontal region using the EBM showed the effectiveness of the brain's frontal region in detecting ADHD. The analysis suggests that the temporal region captures the most subtle alterations, but the frontal region collects insight and minute information regarding changes occurring throughout ADHD. The second-best performance is provided by the parietal, followed by the central, prefrontal, and occipital regions. Again, the performance of the EBM model outperformed other models, making it practical and precise.

We have also performed a feature-level analysis on the channel to check the ability to classify ADHD and HC EEGs. The accuracy obtained for each of the forty-one features is shown in Table 4. The analysis shows that the prediction provided by the third quartile feature is the highest among all, with an ACC of 99.59%, whereas the second and third ACC is obtained for Renyi entropy and fuzzy entropy, respectively. On the other hand, permutation entropy and kurtosis have been the minor discriminable feature in detecting ADHD. Table 5 show the PM

Table 4
Featurewise accuracy (%) obtained for all the channels using EBM glass-box model.

Feature	ACC	Feature	ACC	Feature	ACC	Feature	ACC
RE	99.43	IEEG	98.87	MAV2	99.45	SSL	93.95
CCEN	86.55	AAC	96.49	MIN	90.68	SPF	88.68
CEN	88.85	Q1	96.12	MAX	99.09	SKEW	89.23
FEN	99.14	Q2	97.53	M1	84.75	KURT	74.65
TEN	98.16	Q3	**99.59**	MEAN	98.95	R1	99.02
SuE	76.55	IQR	94.16	HA	92.92	RMS	96.58
SHEN	99.37	WL	92.15	MOM2	95.85	HE	88.49
PEN	70.65	PEAK	94.81	MOM3	98.98	HjC	85.63
KEN	82.71	MAV	97.46	MED	92.85	HFD	76.55
LEN	98.49	MAV1	99.45	SD	88.44	HM	96.62
COV	88.78	–	–	–	–	–	–

Table 5
Performance measure (%) obtained for different combinations (All features, brain region, and feature level) using EBM model.

PM	Combined	Prefrontal	Frontal	Central	Parietal	Occipital	Temporal	Q3
ACC	99.81	94.39	99.14	96.35	98.21	92.93	91.83	99.59
SNS	99.78	93.62	98.97	96.88	98.36	91.42	90.51	99.44
SPF	99.84	95.40	99.36	95.70	98.02	95.03	93.67	99.79
F1	99.83	95.00	99.22	96.69	98.38	93.77	92.80	99.63
MCC	99.61	88.66	98.26	92.62	96.38	85.77	83.52	99.18
PRS	99.87	96.41	99.48	96.50	98.40	96.24	95.21	99.83
NPV	99.73	91.88	98.72	96.16	97.97	88.84	87.67	99.31
FDR	0.13	3.59	0.52	3.50	1.60	3.76	4.79	0.17
CSI	99.66	90.47	98.46	93.59	96.81	88.27	86.57	99.27

Table 6
Confusion matrix obtained using our proposed model.

	ADHD	HC
ADHD	99.87 (%)	0.13 (%)
HC	0.27 (%)	99.73 (%)

report of the proposed explainable model for different configurations. It can be noted that our developed model has provided an effective analysis in detecting ADHD and HC. The PM report reveals that our model captured detailed information for each PM report. Lastly, we have provided a percentage confusion matrix report shown in Table 6. Our model can correctly detect 99.87% and 99.73% of ADHD and HC EEGs. To evaluate the binary classification ability of our proposed model, we have evaluated the receiver operating characteristics (ROC) and the area under the curve (AUC) for all features, including the frontal region, and the third quartile feature shown in Fig. 10. It has been seen that an AUC of 100%, 99.95%, and 99.48%, for all features, Q3 feature, and frontal region, respectively. The graphical and numerical analysis shows that our model is accurate and effective.

Further, the interpretability of the glass-box model is analyzed using univariate analysis to check the channel and feature ranking. For this,

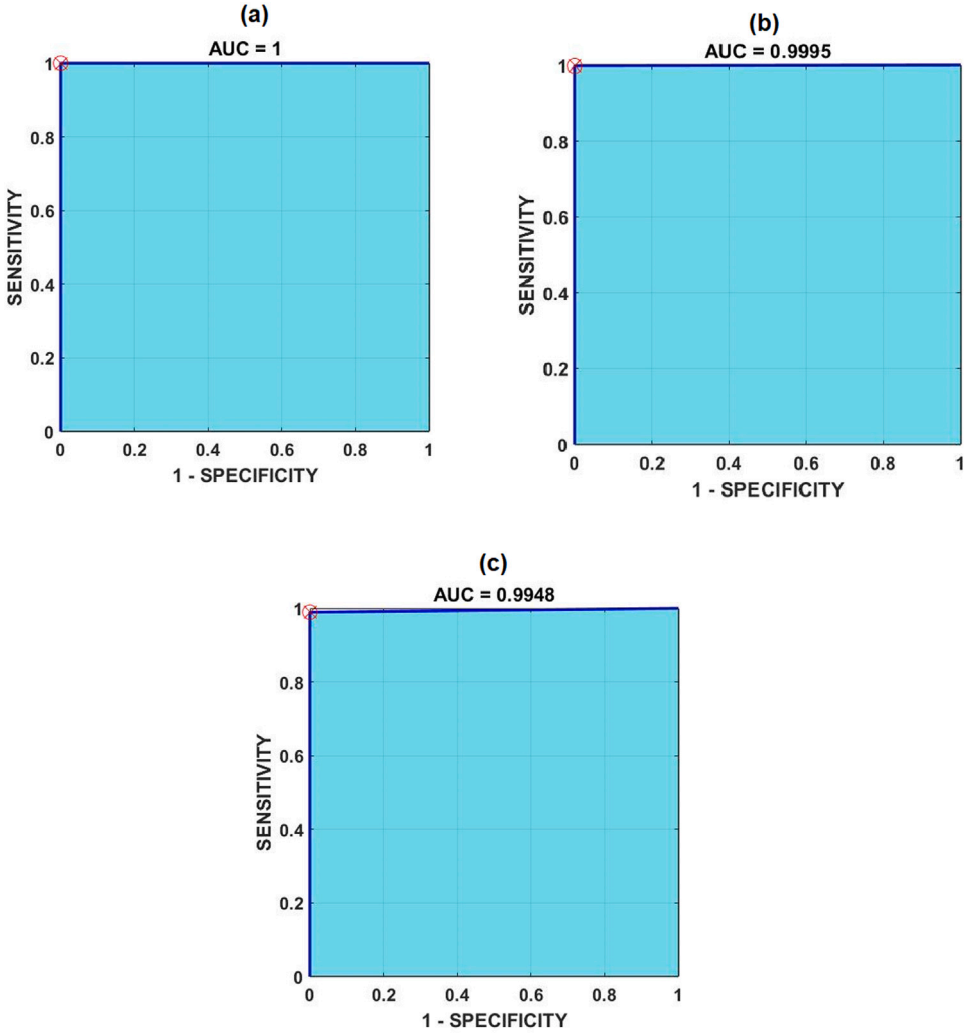

Fig. 10. ROC and AUC curve obtained for the proposed explainable and interpretable model: (a) All features, (b) Q3 feature, and (c) Frontal region.

the feature matrix of all the features for the frontal channel is used. Similarly, for evaluating the score of each channel, a feature matrix of channels is applied. The univariate score obtained for each feature and channel are shown in Figs. 11 and 12. It can be interpreted that, among entropy features, Shannon entropy and Tsallis entropy provided the highest discriminability. Average amplitude change, wavelength, mean deviation from max and mean deviation from range are among the top distinguishable nonlinear measures, whereas third quartile, median, and max are among the top-ranked statistical measures. Similarly, frontal channels provide the highest discriminability for channel analysis, specifically by Fz, followed by the parietal and central region channels. To realize the practical implementation of our proposed model in real-time, we have performed the accuracy analysis to get the optimal feature combination. To accomplish this, we have used feature and channel rank from univariate ranking. Based on the channel rank obtained from Fig. 12, the most discriminant channel (Fz) is used. For this channel (Fz), we have created a feature matrix for our developed classifier based on the feature rank obtained in Fig. 11. The accuracy provided by different combinations of features based on the univariate ranking is shown in Fig. 13. It has been observed that with an increase in the number of features, the accuracy tends to rise to a maximum point of 98.71% with 17 features. After this optimum number of feature combination, the accuracy of the model for a single channel either remain the same of tend to reduce. Thus, the feature dimensionality of our developed model is reduced significantly from 779 to only 17.

Lastly, we have performed a black-box analysis to study how the decisions and predictions have been performed by our best-performing glass-box model, i.e. the EBM classifier. The explanations for local predictions have been carried out with LIME and SHAP explanations. The LIME explanations for local predictions in channels are shown in Fig. 14, whereas for features, local descriptions are illustrated in Fig. 15. Fz has the greatest ability to collect insights into ADHD and HC EEGs, according to the local explainability of the channels provided by LIME. Similar to how channels like F8, Pz, Cz, and P7 often help to collect local characteristics, it makes sense to trust these channels while training the model. However, since they appear to be non-significant in one or more instances, channels like Fp2, F4, O2, T8, and C3 cannot be entirely relied on to make predictions. In contrast, the third quartile, average amplitude change, wavelength, fuzzy entropy, mean, and integral EEGs are some of the most reliable characteristics when investigating the capacity to generate local predictions produced in feature-level analysis. Another well-known SHAP technique has been applied to capture the details of local-level explainability. Figs. 16 and 17 illustrate the local-level explanations for channels and features. It has been observed that SHAP explainability shows that Fp2, Fz, F7, F4, Pz, Cz, P3, and P4 captures local predictions. These are the most trusted channels to be relied upon while building a trainable model for ADHD detection. Features like average amplitude change, wavelength, third quartile, second moment, mean, and maxima can often be used for ADHD detection by experts. Physicians and clinicians can also use the features like Shannon and fuzzy to capture local details to build a

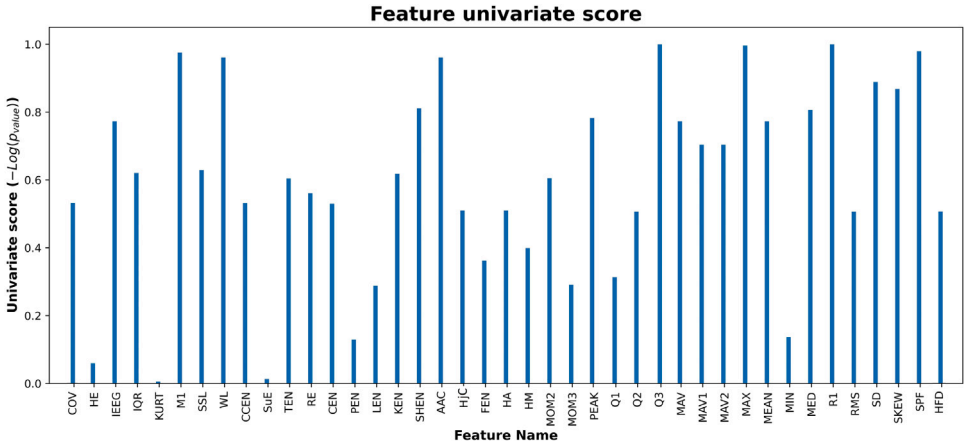

Fig. 11. Feature rank obtained for each feature using univariate ranking.

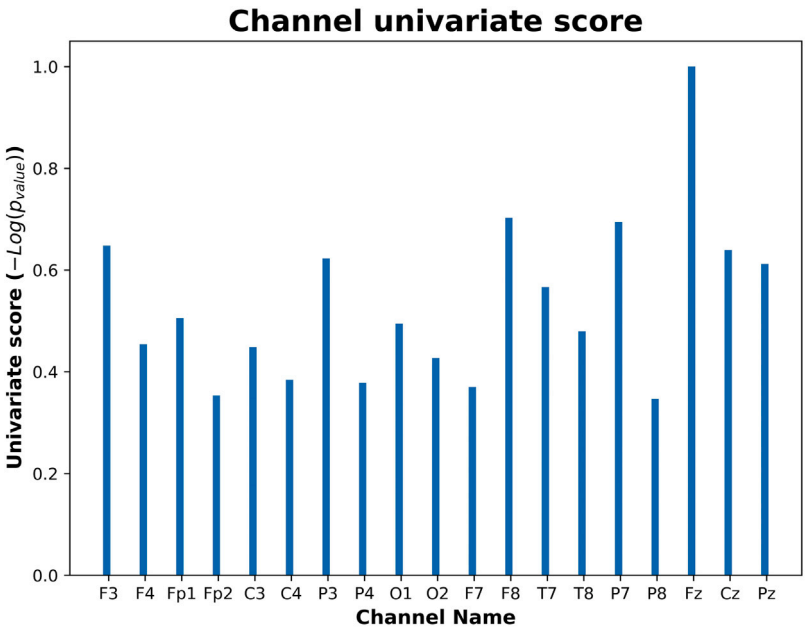

Fig. 12. Channel rank obtained for each channel using univariate ranking.

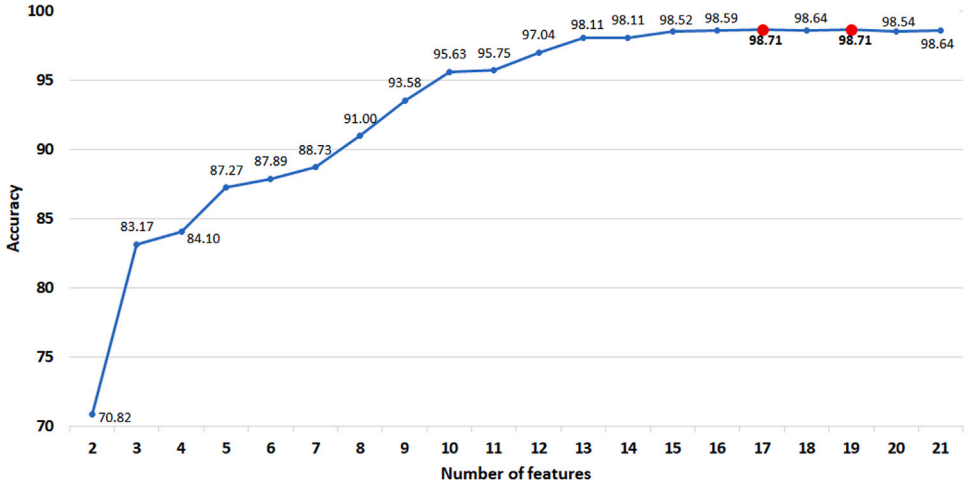

Fig. 13. Variations of accuracy w.r. to the number of features.

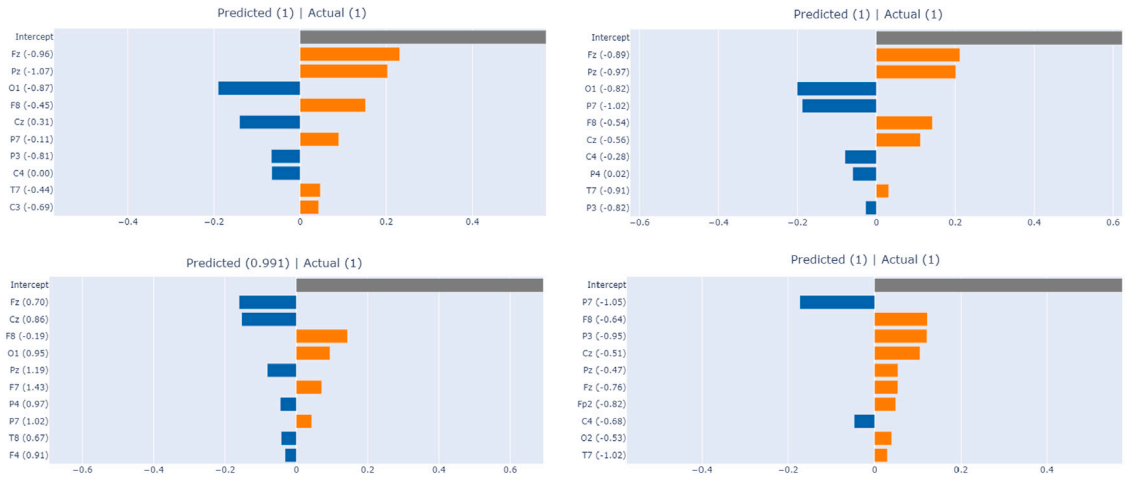

Fig. 14. Results obtained using LIME local explainability for different channels.

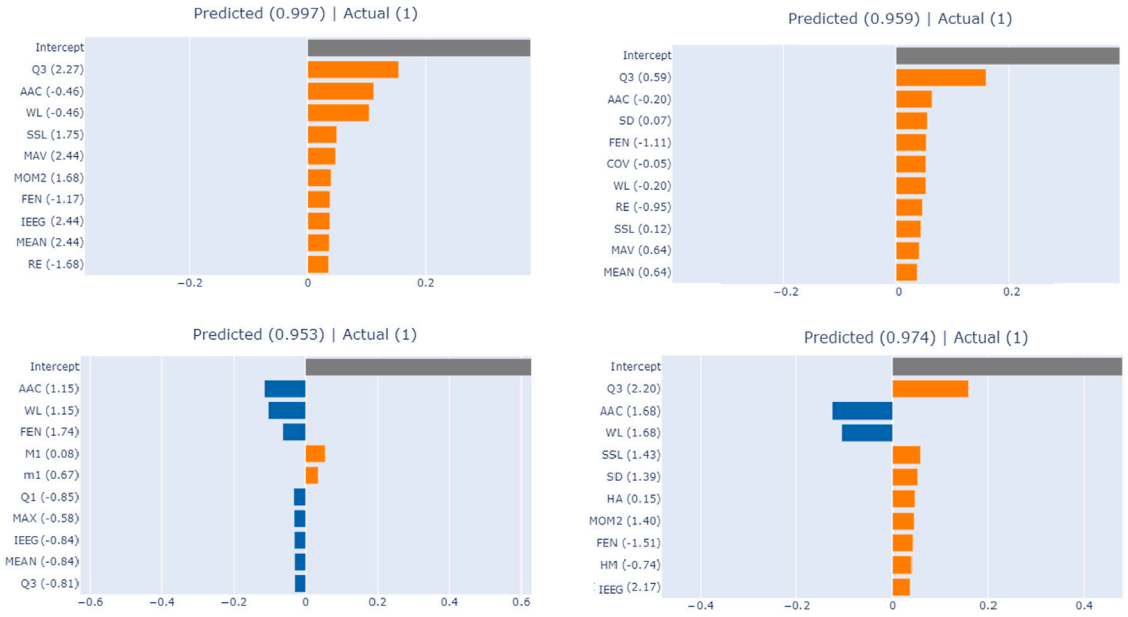

Fig. 15. Results obtained using LIME local explainability for various features.

trainable model. Similarly, the global explainability of the model analyzed by the PDP model shows that frontal channel and entropies (fuzzy and Shannon) provide global discrimination, as shown in Figs. 18 and 19. Similarly, MS has also been used to study the explanations for global prediction. Figs. 20 and 21 provide MS obtained for features and channels. It can be observed from the Figures that Fz provides contributions to overall predictions, Pz, F7, P3, P4, etc., contributes towards overall predictions. Similarly, fuzzy entropy, Shannon entropy, average amplitude change, wavelength, third quartile, Hjorth mobility, mean deviation from range, etc., contribute significantly to the model's overall predictions during training and testing.

4. Discussion

Finally, it is compared with existing state-of-the-art techniques to test the performance of VMD-HT. Table 7 shows the performance comparison of VMD-HT with state-of-the-art using children's ADHD EEG dataset. It can be noted from this table that existing ADHD detection models evaluate their model performance with limited performance measures. Also, many models evaluated their performance with a limited number of features, fewer subjects, and empirical choice classification techniques. In such scenarios, the decision made by these models is unreliable as it may not guarantee and yield similar performance in a clinical scenario. Also, existing models provide decent performance in ADHD detection when making decisions. But, users or experts do not get any visibility into the model or steps that have contributed to the final decision. Due to the absence of explanation in these situations, clinicians frequently express their reluctance to employ such systems in clinical settings. Explaining the decision-making of ML models may provide confidence to clinicians to employ such models for ADHD detection. Therefore, we have presented an explainable model for ADHD detection to overcome the above-mentioned issues. Explainable frameworks assist clinicians to understand the decision-making of the model. It provides insights into the model about the overall decision-making and the features that contributed to the detection of ADHD. Implementation of explainability has achieved three benefits:

Trust in the ML model: The local and global explanations using visual analysis provide transparent and interpretable reasoning about the individual and overall predictions.

Improved troubleshooting: The use of explainability also improves debugging of ML models. It also helps in troubleshooting the model's

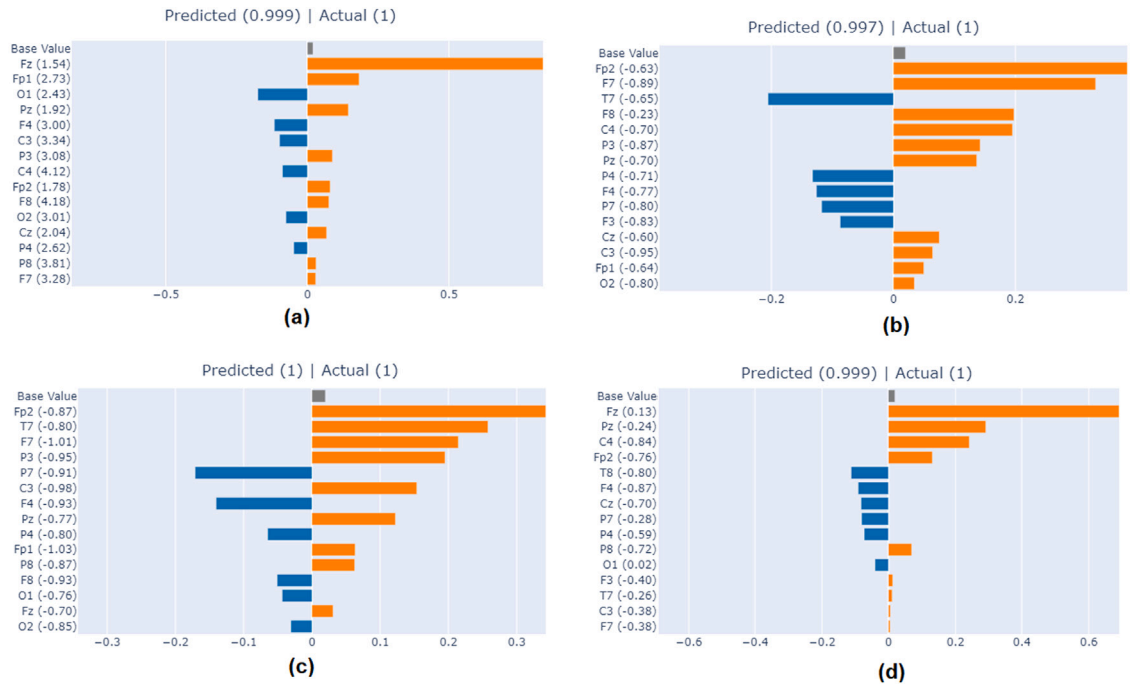

Fig. 16. Results obtained using SHAP local explainability for various channels.

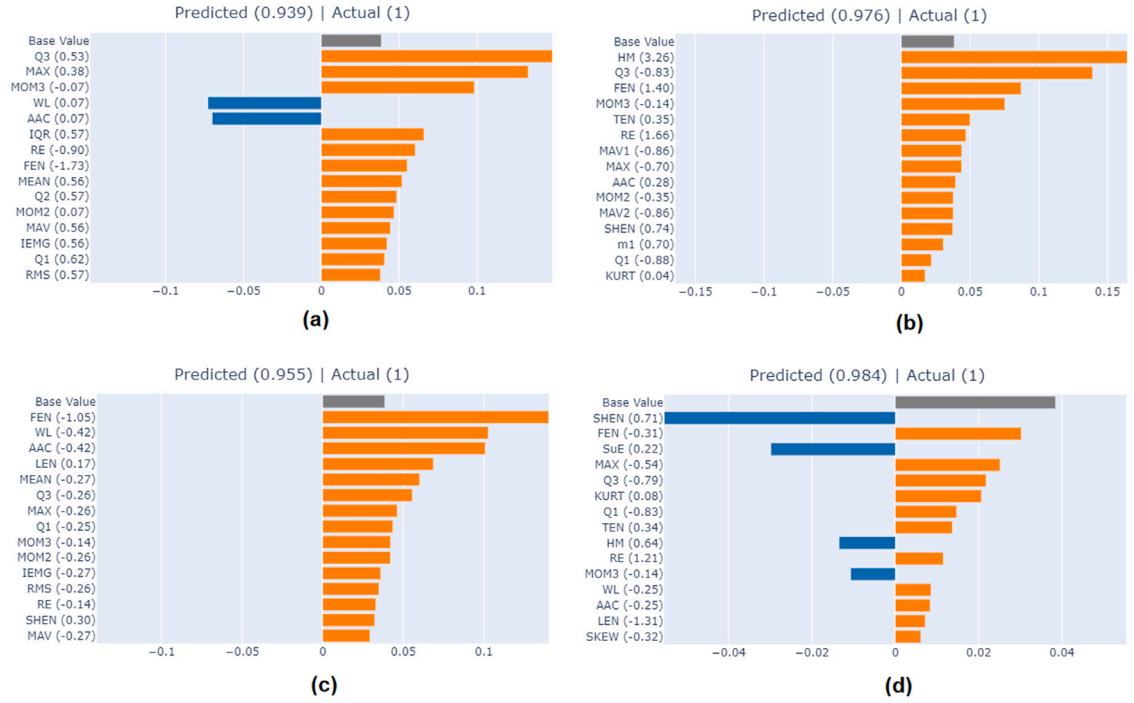

Fig. 17. Results obtained using SHAP local explainability for different features.

working. E.g. over time, if a model predicts HC EEGs as ADHD, troubleshooting helps to identify the reason for the decision and try to improve it.

Tackling bias and other potholes: With explainability, identifying and correcting bias in the source is easy. It also helps in detecting irregularities in the dataset.

Using explainability, our model has obtained the highest performance using a relatively larger dataset (61 ADHD and 60 HC children) as used by [25,26,32,35–38]. The combination of VMD-HT and an explainable glass-box black-box model proposed for ADHD detection has provided interpretable, effective, accurate, and explainable results. The performance measures of 99.81%, 99.78%, 99.84%, 99.83%, and 99.87% ACC, SNS, SPF, F-1, and PRS, respectively show that the proposed method is better than the existing state-of-the-art techniques using the same even with a relatively large dataset. Our explainable model has identified an optimal combination of brain regions, channel labels, and feature sets yielded the highest model performance. The explainable ADHD detection in this paper focused on explaining the significance of features and channels contributing to high performance. The explanations provided by our proposed model are limited to feature

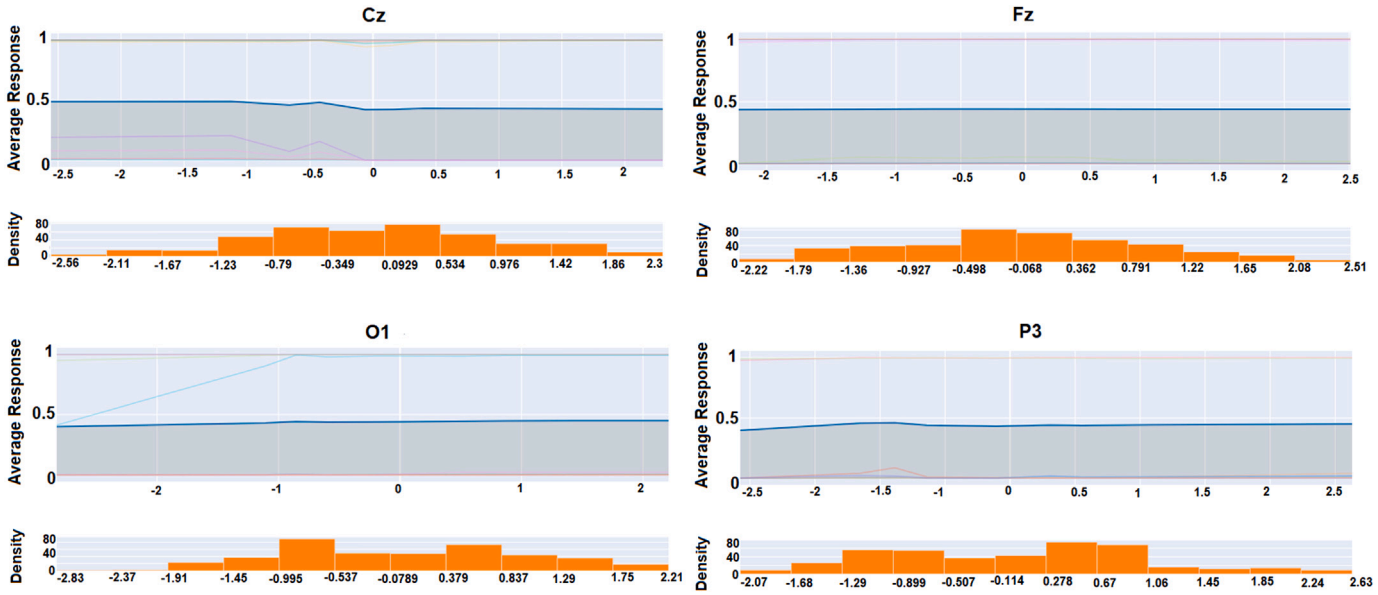

Fig. 18. Results obtained using PDP global explainability for various channels.

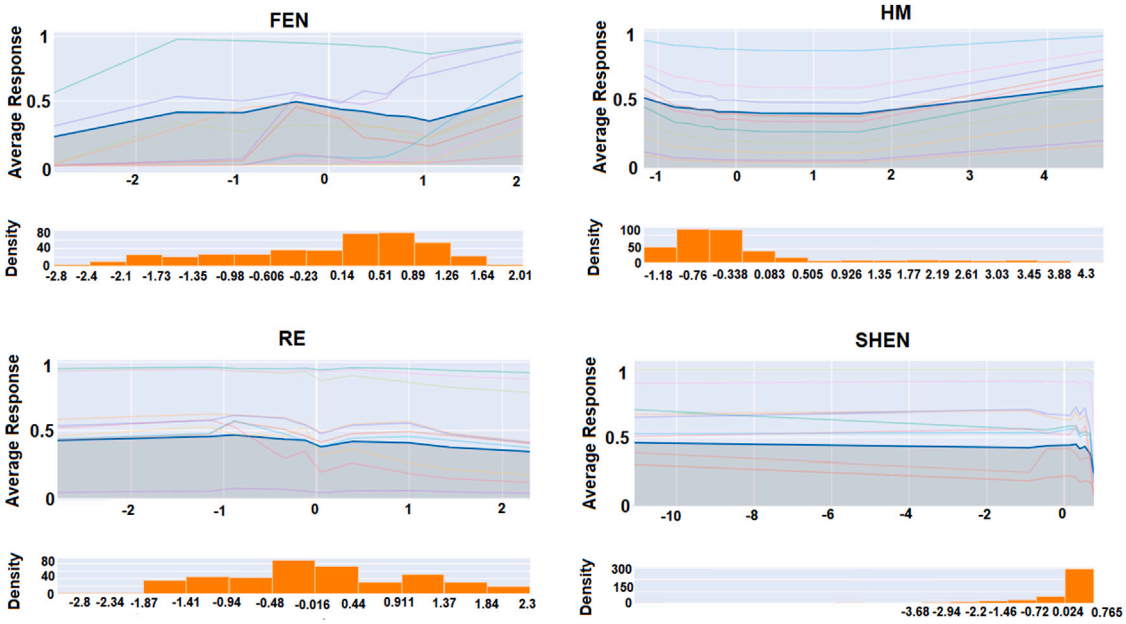

Fig. 19. Results obtained using PDP global explainability for different features.

engineering and channel selection for ADHD detection. However, our model fails to explain the clinical significance and characteristics of ADHD which are traditionally used by clinicians. It has also been observed that our developed model has yielded high performance in all the brain regions. Since we have tested our model on a single dataset, the accuracy may vary from one dataset to another, resulting in low accuracy on the same brain regions. The frontal area, specially Fz and F8 channel labels combined with third quartile, fuzzy, and Shannon entropies, capture minute changes in the irregularities of ADHD and HC EEGs and contributed significantly to high performance. The findings by our explainable model reported the significance of frontal regions, Fz channel, and nonlinear features for ADHD detection [11,24,26,35]. The key features of our developed model are listed below.

- The model is accurate and effective as it has provided the highest classification rate.
- It is robust and precise since it has been tested on different scenarios (brain regions, feature analysis, and combined analysis).
- The model has obtained optimal performance with reduced dimensionality of only seventeen features and one channel.
- Our developed model is explainable and interpretable as it has provided insight into the key predictors.
- It is efficient because it has been developed using a tenfold cross-validation technique and surpassed SOTA techniques' performance.

However, the model has the following limitations:

- In this work, we have used EEG features instead of clinical features. Hence, neurologists may find it difficult to understand the model prediction results. Also, the explanation of EEG features with SHAP may have little use.

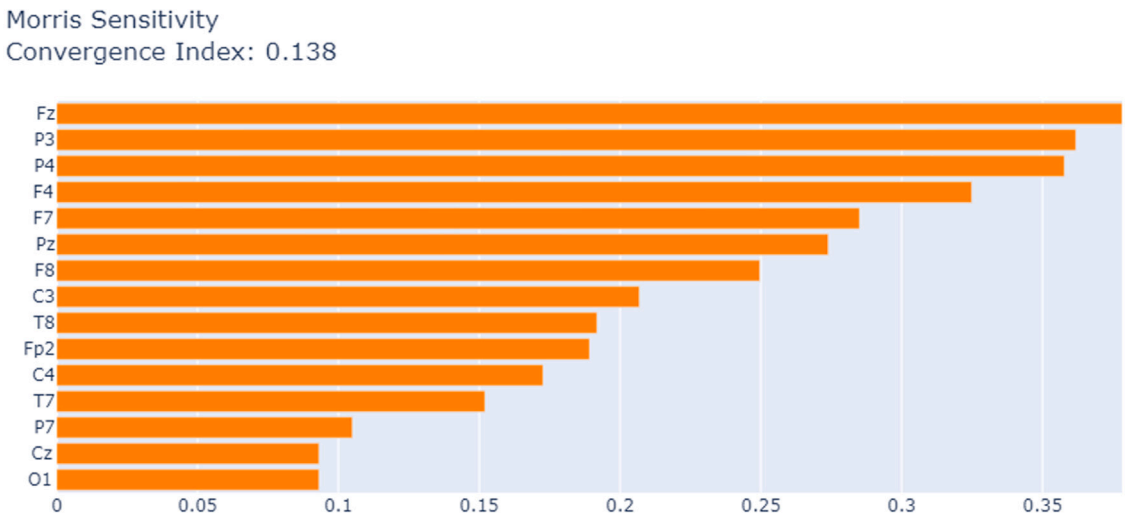

Fig. 20. Results obtained using Morris Sensitivity Convergence Index for global explainability for various channels.

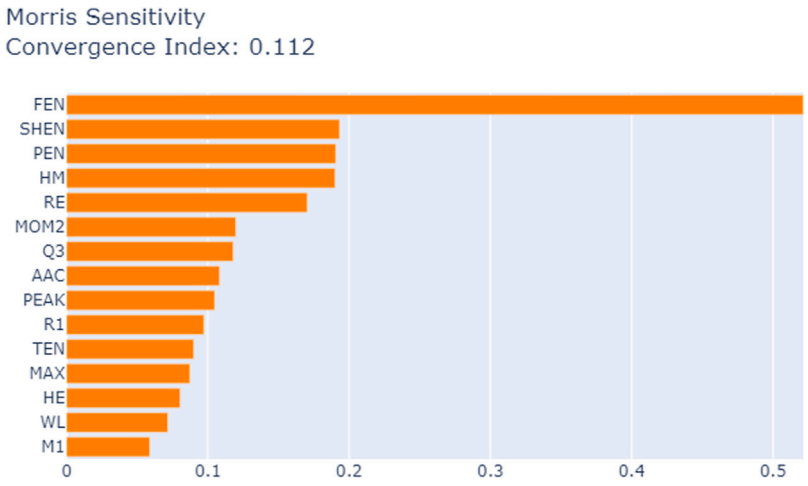

Fig. 21. Results obtained using Morris Sensitivity Convergence Index for global explainability for different features.

Table 7
Summary of performance (%) comparison with other state-of-the-art techniques developed for automated detection of ADHD using EEG signals.

Authors (Year)	Number of subjects	Method used	Performance measures (%)				
			ACC	SNS	SPF	F-1	PRS
Khoshnoud et al. [19] (2017)	12 ADHD and 12 HC	Band power, En, largest LE, and MSS	83.33	–	–	–	–
Abibullaev and An [22] (2011)	7 ADHD and 3 HC	PSD-based analysis	97	–	–	–	–
Boroujeni et al. [24] (2019)	50 ADHD and 26 HC	En, FD and CorrD, and LE	96.05	98	92.31	–	–
Ogrim et al. [30] (2012)	62 ADHD and 36 HC	Absolute power	85	85	85	–	–
Snyder et al. [31] (2008)	97 ADHD and 62 HC	Mean power and ratio of rhythms	89	87	94	–	–
Zhang et al. [33] (2019)	50 ADHD and 58 HC	BFN using PLI and CNN	94.39	97.83	91.8	–	–
Allahverdy et al. [26] (2016)	29 ADHD and 20 HC	Katz, Higuchi, Sevcik FD and LE	96.7	98.9	95.2	–	–
Mohammadi et al. [25] (2016)	30 ADHD and 30 HC	Non-linear features, DISR, and MLP	93.65	–	–	–	–
Moghaddari et al. [32] (2020)	31 ADHD and 30 HC	Frequency band separation and CNN	98.48	–	99	98.49	98.51
Khare et al. [29] (2022)	61 ADHD and 60 HC	VHERS	99.95	100	99.89	99.9	99.91
Talebi et al. [35] (2022)	61 ADHD and 60 HC	nCREANN	99				
Bakhtyari et al. [36] (2022)	46 ADHD and 45 HC	ConvLSTM	99.75	99.75	–	99.74	99.74
Ekhlasi et al. [37] (2021)	61 ADHD and 60 HC	connectivity matrices	89.1	–	–	–	–
Tanko et al. [38] (2022)	61 ADHD and 60 HC	EPSPatNet86	97.19	97.12	–	97.15	97.18
Proposed	**61 ADHD and 60 HC**	VMD-HT	**99.81**	**99.78**	**99.84**	**99.83**	**99.87**

- Due to the limited publicly available ADHD dataset in children, the proposed model is tested on a single EEG dataset.

In the future, we plan to focus our study on detecting ADHD at an early stage and also detecting other mental abnormalities like conduct disorder (CD) and ADHD + CD.

5. Conclusion

The EEG signal does not provide significant discriminant information. Hidden characteristics and representative data from EEG are obtained in terms of amplitude envelope using VMD-HT. Our analysis proves that the frontal region plays a crucial role in detecting ADHD

in children. Also, Fz, F7, Pz, P7, and Cz channels are significant in capturing local and global predictions. Also, entropy feature, specifically SHEN and FEN, nonlinear feature, i.e. AAC and WL, and Q3 and MAX statistical features, have yielded the best performance in terms of interpretability and explainability. These features capture not only local-level prediction explainers but also provide global explanations. We conclude that our developed model is not only accurate in ADHD detection but also effective, robust, predictive, and explainable. Our goal is to provide a model where the user can understand and create instances of the prediction process to obtain model performance. Healthcare organizations and research institutes constantly look for predictive and explainable models for various brain disorders. Our model is one such step to provide predictive and explainable for ADHD detection. Specialists may use our model to detect children with ADHD, transfer the knowledge about the predictions, and implement models in hospitals/research institutes. We conclude our paper with following major contributions

- Various brain regions have been analyzed to find the most suitable region for the detection of ADHD.
- Identification of most significant channel and optimal number of features required to yield maximum performance.
- Explored a glass-box and black-box model interpretations using an EBM model with four explainable characteristics (LIME, PDP, SHAP, and MS) to find local and global explainability of the model.
- Tested our developed explainable model using the ten-fold cross-validation technique and compared the model performance with existing SOTA classifications and ADHD detection techniques to study the model robustness.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] ADHD-Institute, 2021, https://adhd-institute.com/, (Accessed:25 Feb 2021).

[2] S. Pingali, J. Sunderajan, A study of comorbidities in attention deficit hyperactivity disorder: a retrospective analysis of case records, Arch. Ment. Health 15 (2) (2014) 206–210.

[3] Centers for Disease Control and Prevention, 2021, https://www.cdc.gov/ncbddd/adhd/data.html, (Accessed:25 Feb 2021).

[4] J. Hamed, A. Taha, A. Sabra, H. Bella, Attention deficit hyperactivity disorder (ADHD) among male primary school children in dammam, Saudi Arabia: Prevalence and associated factors, J. Egypt. Public Health Assoc. 83 (2008) 165–182.

[5] A. Meysamie, M. Daneshvar, M.-R. Mohammadi, Prevalence of attention-deficit/hyperactivity disorder symptoms in preschool-aged Iranian children, Iran. J. Ped. 21 (2011) 467–472.

[6] H.W. Loh, C.P. Ooi, P.D. Barua, E.E. Palmer, F. Molinari, U.R. Acharya, Automated detection of ADHD: Current trends and future perspective, Comput. Biol. Med. 146 (2022) 105525, http://dx.doi.org/10.1016/j.compbiomed.2022.105525, URL https://www.sciencedirect.com/science/article/pii/S0010482522003171.

[7] S. Sale, J. Joska, J.O. Abdulmalik, Correlates of attention/deficit hyperactivity disorder (ADHD) among children and adolescents in northern Nigeria, Neuropsychiatrie de l'Enfance et de l'Adolescence 60 (2012) S255, http://dx.doi.org/10.1016/j.neurenf.2012.04.650.

[8] H. Lola, H. Gobie, A. Worku, A. Zerihun, S. Yimer, K. Leta, Attention deficit hyperactivity disorder (ADHD) among children aged 6 to 17 years old living in girja district, rural ethiopia, Behav. Neurol. 2019 (2019) 1–8, http://dx.doi.org/10.1155/2019/1753580.

[9] M. Cabral, S. Liu, N. Soares, Attention-deficit/hyperactivity disorder: diagnostic criteria, epidemiology, risk factors and evaluation in youth, Transl. Pediatr. 9 (2020) S104–S113, http://dx.doi.org/10.21037/tp.2019.09.08.

[10] H.W. Loh, C.P. Ooi, P.D. Barua, E.E. Palmer, F. Molinari, U.R. Acharya, Automated detection of ADHD: Current trends and future perspective, Comput. Biol. Med. 146 (2022) 105525, http://dx.doi.org/10.1016/j.compbiomed.2022.105525.

[11] J. Koh, C.P. Ooi, N.S. Lim-Ashworth, J. Vicnesh, H.T. Tor, O.S. Lih, R.-S. Tan, U. Acharya, D.S.S. Fung, Asssutomated classification of attention deficit hyperactivity disorder and conduct disorder using entropy features with ECG signals, Comput. Biol. Med. 140 (C) (2022) http://dx.doi.org/10.1016/j.compbiomed.2021.105120.

[12] C. Sridhar, S. Bhat, U.R. Acharya, H. Adeli, G.M. Bairy, Diagnosis of attention deficit hyperactivity disorder using imaging and signal processing techniques, Comput. Biol. Med. 88 (2017) 93–99, http://dx.doi.org/10.1016/j.compbiomed.2017.07.009.

[13] Z. Liu, Q. Xie, M. Wu, W. Cao, D. Li, S. Li, Electroencephalogram emotion recognition based on empirical mode decomposition and optimal feature selection, IEEE Trans. Cogn. Dev. Syst. 11 (4) (2019) 517–526, http://dx.doi.org/10.1109/TCDS.2018.2868121.

[14] S.K. Khare, V. Bajaj, A facile and flexible motor imagery classification using electroencephalogram signals, Comput. Methods Programs Biomed. 197 (2020) 105722, http://dx.doi.org/10.1016/j.cmpb.2020.105722.

[15] S. Khare, V. Bajaj, Optimized tunable q wavelet transform based drowsiness detection from electroencephalogram signals, IRBM (2020) http://dx.doi.org/10.1016/j.irbm.2020.07.005.

[16] J. Cao, J. Zhu, W. Hu, A. Kummert, Epileptic signal classification with deep EEG features by stacked CNNs, IEEE Trans. Cogn. Dev. Syst. 12 (4) (2020) 709–722, http://dx.doi.org/10.1109/TCDS.2019.2936441.

[17] J.D. Kropotov, V.A. Grin-Yatsenko, V.A. Ponomarev, L.S. Chutko, E.A. Yakovenko, I.S. Nikishena, ERPs correlates of EEG relative beta training in ADHD children, Int. J. Psychophysiol. 55 (1) (2005) 23–34, http://dx.doi.org/10.1016/j.ijpsycho.2004.05.011.

[18] F. Ghassemi, M.H. Moradi, M.T. Doost, V. Abootalebi, Using non-linear features of EEG for ADHD/normal participants classification, Proc. - Soc. Behav. Sci. 32 (2012) 148–152, http://dx.doi.org/10.1016/j.sbspro.2012.01.024, The 4th International Conference of Cognitive Science.

[19] S. Khoshnoud, M.A. Nazari, M. Shamsi, Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals, J. Integr. Neurosci. 16 (2017) http://dx.doi.org/10.3233/JIN-170033.

[20] A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Müller, G. Candrian, Machine learning approach for classification of ADHD adults, Int. J. Psychophysiol. 93 (1) (2014) 162–166, http://dx.doi.org/10.1016/j.ijpsycho.2013.01.008, Applied Neuroscience: Functional enhancement, prevention, characterisation and methodology. (Hosting the Society of Applied Neuroscience).

[21] I. Lazzaro, E. Gordon, S. Whitmont, M. Plahn, W. Li, S. Clarke, A. Dosen, R. Meares, Quantified EEG activity in adolescent attention deficit hyperactivity disorder, Clin. Electroencephalogr. 29 (1) (1998) 37–42, http://dx.doi.org/10.1177/155005949802900111.

[22] B. Abibullaev, J. An, Decision support algorithm for diagnosis of ADHD using electroencephalograms, J. Med. Syst. 36 (2011) 2675–2688, http://dx.doi.org/10.1007/s10916-011-9742-x.

[23] R. Aldemir, E. Demirci, H. Per, M. Canpolat, S. Özmen, M. Tokmakçi, Investigation of attention deficit hyperactivity disorder (ADHD) sub-types in children via EEG frequency domain analysis, Int. J. Neurosci. 128 (4) (2018) 349–360, http://dx.doi.org/10.1080/00207454.2017.1382493.

[24] Y.K. Boroujeni, Diagnosis of attention deficit hyperactivity disorder using non-linear analysis of the EEG signal, IET Syst. Biol. 13 (2019) 260–266(6).

[25] M.-R. Mohammadi, A. Khaleghi, A. Motie Nasrabadi, S. Rafiei, M. Begol, H. Zarafshan, EEG classification of ADHD and normal children using non-linear features and neural network, Biomed. Eng. Lett. 6 (2016) 66–73, http://dx.doi.org/10.1007/s13534-016-0218-2.

[26] A. Allahverdy, A. Khorrami, M. Mohammadi, A. Motie Nasrabadi, Detecting ADHD children using the attention continuity as nonlinear feature of EEG, Front. Biomed. Technol. 3 (2016) 1–2.

[27] S. Kaur, P. Arun, S. Singh, D. Kaur, EEG based decision support system to diagnose adults with ADHD, in: 2018 IEEE Applied Signal Processing Conference, ASPCON, 2018, pp. 87–91, http://dx.doi.org/10.1109/ASPCON.2018.8748412.

[28] A. Mueller, G. Candrian, J. Kropotov, V. Ponomarev, G.-M. Baschera, Classification of ADHD patients on the basis of independent ERP components using a machine learning system, Nonlinear Biomed. Phys. 4 Suppl 1 (2010) S1, http://dx.doi.org/10.1186/1753-4631-4-S1-S1.

[29] S.K. Khare, N.B. Gaikwad, V. Bajaj, VHERS: A novel variational mode decomposition and Hilbert transform-based EEG rhythm separation for automatic ADHD detection, IEEE Trans. Instrum. Meas. 71 (2022) 1–10, http://dx.doi.org/10.1109/TIM.2022.3204076.

[30] G. Ogrim, J. Kropotov, K. Hestad, The quantitative EEG theta/beta ratio in attention deficit/hyperactivity disorder and normal controls: Sensitivity, specificity, and behavioral correlates, Psychiatry Res. 198 (3) (2012) 482–488, http://dx.doi.org/10.1016/j.psychres.2011.12.041.

[31] S.M. Snyder, H. Quintana, S.B. Sexson, P. Knott, A. Haque, D.A. Reynolds, Blinded, multi-center validation of EEG and rating scales in identifying ADHD within a clinical sample, Psychiatry Res. 159 (3) (2008) 346–358, http://dx.doi.org/10.1016/j.psychres.2007.05.006.

[32] M. Moghaddari, M.Z. Lighvan, S. Danishvar, Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG, Comput. Methods Programs Biomed. 197 (2020) 105738, http://dx.doi.org/10.1016/j.cmpb.2020.105738.

[33] Z. Zhang, X. Li, Use transfer learning to promote identification ADHD children with EEG recordings, in: 2019 Chinese Automation Congress, CAC, 2019, pp. 2809–2813, http://dx.doi.org/10.1109/CAC48633.2019.8997426.

[34] H.T. Tor, C.P. Ooi, N.S. Lim-Ashworth, J.K.E. Wei, V. Jahmunah, S.L. Oh, U.R. Acharya, D.S.S. Fung, Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals, Comput. Methods Programs Biomed. 200 (2021) 105941, http://dx.doi.org/10.1016/j.cmpb.2021.105941.

[35] N. Talebi, A.M. Nasrabadi, Investigating the discrimination of linear and non-linear effective connectivity patterns of EEG signals in children with attention-deficit/hyperactivity disorder and typically developing children, Comput. Biol. Med. 148 (2022) 105791.

[36] M. Bakhtyari, S. Mirzaei, ADHD detection using dynamic connectivity patterns of EEG data and ConvLSTM with attention framework, Biomed. Signal Process. Control 76 (2022) 103708.

[37] A. Ekhlasi, N.A. Motie, M.R. Mohammadi, Classification of the children with ADHD and healthy children based on the directed phase transfer entropy of EEG signals, Front. Biomed. Technol. (2021).

[38] D. Tanko, P.D. Barua, S. Dogan, T. Tuncer, E. Palmer, E.J. Ciaccio, U.R. Acharya, EPSPatNet86: eight-pointed star pattern learning network for detection ADHD disorder using EEG signals, Physiol. Meas. 43 (3) (2022) 035002.

[39] S. Parsons, A.M. Boonman, M.K. Obrist, Advantages and disadvantages of techniques for transforming and analyzing chiropteran echolocation calls, J. Mammal. 81 (4) (2000) 927–938, http://dx.doi.org/10.1644/1545-1542(2000)081<0927:AADOTF>2.0.CO;2.

[40] H.W. Loh, C.P. Ooi, S. Seoni, P.D. Barua, F. Molinari, U.R. Acharya, Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022), Comput. Methods Programs Biomed. 226 (2022) 107161, http://dx.doi.org/10.1016/j.cmpb.2022.107161, URL https://www.sciencedirect.com/science/article/pii/S0169260722005429.

[41] A.M. Nasrabadi, A. Allahverdy, M. Samavati, M.R. Mohammadi, EEG data for ADHD / control children, 2020, http://dx.doi.org/10.21227/rzfh-zn36.

[42] M. Samavati, A. Motie Nasrabadi, M.R. Mohammadi, Automatic minimization of eye blink artifacts using fractal dimension of independent components of multi-channel EEG, in: 20th Iranian Conference on Electrical Engineering (ICEE2012), 2012, pp. 1576–1578, http://dx.doi.org/10.1109/IranianCEE.2012.6292611.

[43] American Psychiatric Association, Diagnostic and statistical manual of mental disorders (DSM-5), 2013, URL https://www.psychiatry.org/psychiatrists/practice/dsm.

[44] C.A. Valentim, C.M.C. Inacio, S.A. David, Fractal methods and power spectral density as means to explore EEG patterns in patients undertaking mental tasks, Fractal Fract. 5 (4) (2021) http://dx.doi.org/10.3390/fractalfract5040225.

[45] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531–544, http://dx.doi.org/10.1109/TSP.2013.2288675.

[46] C. Zhang, Y. Zhang, C. Hu, Z. Liu, L. Cheng, Y. Zhou, A novel intelligent fault diagnosis method based on variational mode decomposition and ensemble deep belief network, IEEE Access 8 (2020) 36293–36312, http://dx.doi.org/10.1109/ACCESS.2020.2969412.

[47] S. Khare, V. Bajaj, An evolutionary optimized variational mode decomposition for emotion recognition, IEEE Sens. J. 21 (2) (2021) 2035–2042, http://dx.doi.org/10.1109/JSEN.2020.3020915.

[48] S.K. Khare, V. Bajaj, Entropy-based drowsiness detection using adaptive variational mode decomposition, IEEE Sens. J. 21 (5) (2021) 6421–6428, http://dx.doi.org/10.1109/JSEN.2020.3038440.

[49] M. Unser, D. Sage, D. Van De Ville, Multiresolution monogenic signal analysis using the Riesz-Laplace wavelet transform, IEEE Trans. Image Process. 18 (11) (2009) 2402–2418, http://dx.doi.org/10.1109/TIP.2009.2027628.

[50] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 454 (1971) (1998) 903–995, http://dx.doi.org/10.1098/rspa.1998.0193, arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1998.0193, URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1998.0193.

[51] N. Kannathal, M.L. Choo, U.R. Acharya, P. Sadasivan, Entropies for detection of epilepsy in EEG, Comput. Methods Programs Biomed. 80 (3) (2005) 187–194.

[52] C. Sridhar, O.S. Lih, V. Jahmunah, J.E. Koh, E.J. Ciaccio, T.R. San, N. Arunkumar, S. Kadry, U. Rajendra Acharya, Accurate detection of myocardial infarction using non linear features with ECG signals, J. Ambient Intell. Humaniz. Comput. 12 (3) (2021) 3227–3244.

[53] V.K. Sudarshan, U.R. Acharya, S.L. Oh, M. Adam, J.H. Tan, C.K. Chua, K.P. Chua, R. San Tan, Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2 s of ECG signals, Comput. Biol. Med. 83 (2017) 48–58.

[54] S.K. Khare, V. Bajaj, A self-learned decomposition and classification model for schizophrenia diagnosis, Comput. Methods Programs Biomed. 211 (2021) 106450.

[55] S.K. Khare, V. Bajaj, A hybrid decision support system for automatic detection of Schizophrenia using EEG signals, Comput. Biol. Med. 141 (2022) 105028.

[56] M. Baygin, An accurate automated schizophrenia detection using TQWT and statistical moment based feature extraction, Biomed. Signal Process. Control 68 (2021) 102777.

[57] M. Baygin, P.D. Barua, S. Dogan, T. Tuncer, S. Key, U.R. Acharya, K.H. Cheong, A hand-modeled feature extraction-based learning network to detect grasps using sEMG signal, Sensors 22 (5) (2022) http://dx.doi.org/10.3390/s22052007, URL https://www.mdpi.com/1424-8220/22/5/2007.

[58] K.C. Chua, V. Chandran, U.R. Acharya, C.M. Lim, Application of higher order statistics/spectra in biomedical signals—A review, Med. Eng. Phys. 32 (7) (2010) 679–689.

[59] A. Phinyomark, P. Phukpattaranont, C. Limsakul, Feature reduction and selection for EMG signal classification, Expert Syst. Appl. 39 (8) (2012) 7420–7431, http://dx.doi.org/10.1016/j.eswa.2012.01.102, URL https://www.sciencedirect.com/science/article/pii/S0957417412001200.

[60] S.K. Khare, V. Bajaj, A CACDSS for automatic detection of parkinson's disease using EEG signals, in: 2021 International Conference on Control, Automation, Power and Signal Processing, CAPS, 2021, pp. 1–5, http://dx.doi.org/10.1109/CAPS52117.2021.9730723.

[61] H. Nori, S. Jenkins, P. Koch, R. Caruana, Interpretml: A unified framework for machine learning interpretability, 2019, CoRR arXiv:1909.09223, URL http://arxiv.org/abs/1909.09223.

[62] Y. Lou, R. Caruana, J. Gehrke, G. Hooker, Accurate intelligible models with pairwise interactions, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 623–631, http://dx.doi.org/10.1145/2487575.2487579.

[63] M.T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 1135–1144, http://dx.doi.org/10.1145/2939672.2939778.

[64] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS '17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 4768–4777.

[65] M.D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (2) (1991) 161–174.

[66] J. Herman, W. Usher, SALib: An open-source python library for sensitivity analysis, J. Open Source Softw. 2 (2017) http://dx.doi.org/10.21105/joss.00097.

[67] J.H. Friedman, Greedy function approximation: A gradient boosting machine, Ann. Statist. 29 (5) (2001) 1189–1232.