



OPEN Early attention-deficit/hyperactivity disorder (ADHD) with NeuroDCT-ICA and rhinofish optimization (RFO) algorithm based optimized ADHD-AttentionNet

Ahmed Alhussen¹, Ahmed Ibrahim Alutaibi¹, Sunil Kumar Sharma^{2✉}, Ahmad Raza Khan³, Fuzail Ahmad⁴ & Ghanshyam G. Tejani^{5,6✉}

The ADHD detector analyzes behavioral, cognitive, or physiological data (e.g., EEG, eye-tracking, or surveys) to identify patterns associated with ADHD symptoms. This work offers a more sophisticated method of detecting ADHD by overcoming the main drawbacks of existing approaches in terms of data processing, detection accuracy, and computational time. The work is inspired by the fact that Deep Learning (DL) frameworks could transform the existing detection systems of ADHD. In the proposed framework, there is a new NeuroDCT-ICA module for the preprocessing of raw EEG data, which guarantees the elimination of noise and extraction of informative features. Moreover, the method introduces a novel RhinoFish Optimization (RFO) algorithm for selecting optimal features, which enhance the data processing capacity and the stability of the system. As a core of the approach, there is the ADHD-AttentionNet – the deep learning-based model aimed at improving the accuracy and confidence of ADHD identification. The model is validated with the standard metrics, and the performance of the model is outstanding as it has high accuracy of 98.52%, F-score of 98.26% and specificity of 98.16%. These outcomes show that the proposed model yields better accuracy in detecting ADHD related patterns.

Keywords ADHD detection, Deep learning, ADHD-AttentionNet, NeuroDCT-ICA, Hybrid optimization, RhinoFish optimization

ADHD is one of the most prevalent neurodevelopmental disorders in the world with prevalence of 5–7% in children and 2–5% in adults¹. Currently, it is estimated that over 365 million people globally suffer from ADHD, and the condition is more rampant in the US, Canada, and some regions of Western Europe². The vulnerable population is children aged 4–17 years, and the boys are nearly three more likely to be diagnosed than girls³. ADHD symptoms might require early intervention and the best practices to minimize risk factors are important to prevent⁴. Examples include sedentary lifestyles due to spending prolonged hours in front of screens, irregular schedules, and stressful conditions that tend to worsen the symptoms of ADHD, which means that children who are at risk should be given adequate attention and support⁵.

ADHD is identified as a condition of deficits in the areas of inattention, hyperactivity, and impulsivity as well as influencing cognitive and social functioning⁶. The causes of ADHD are considered to be not entirely known;

¹Department of Computer Engineering, College of Computer and Information Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia. ²Department of Information Systems, College of Computer and Information Sciences, Majmaah University, Majmaah 11952, Saudi Arabia. ³Information Technology Department, College of Computer and Information Sciences, Majmaah University, Majmaah 11952, Saudi Arabia. ⁴Respiratory Care Department, College of Applied Sciences, Almaarefa University, Diriyah, Riyadh, Saudi Arabia. ⁵Department of Industrial Engineering and Management, Yuan Ze University, Taoyuan 320315, Taiwan. ⁶Applied Science Research Center, Applied Science Private University, Amman 11937, Jordan. ✉email: s.sharma@mu.edu.sa; p.shyam23@gmail.com; gtejani@saturn.yzu.edu.tw

indeed, it is thought to be due to an interplay of genetics, environmental, and neurological factors. Genetic predisposition plays a big role, and other environmental factors like exposure to toxins during the prenatal period, low birth weight, and childhood trauma may all have been involved in increasing their risk^{7,8}. ADHD often interferes with educational advancement, work and social interactions, all thereby reducing quality of life. Without treatment, ADHD develops comorbid conditions like anxiety, depression, and substance abuse, so diagnosis and intervention are fundamentally important at an early stage⁹.

Identifying ADHD is important for its management, early intercession and interventions as well as support that may enhance the life span of the patients¹⁰. Diagnosis in its own turn is not very accurate since ADHD's symptoms are quite vague and can be accompanied by other diseases. Age, gender, cultural beliefs, and perception of symptoms differ from male to female, and the method and criteria used in assessment also affect the detection^{11,12}. Neuroimaging and Machine Learning (ML) are two fields that provide hope in increasing the specificity in ADHD, which in turn, will lead to better treatment and prognosis¹³.

There are numerous diagnostic methods for ADHD, and each has its pros and cons. The main procedures include the traditional behavioral evaluations-based questionnaires and semi-structured interviews among others. However, these commonly used methods often suffer from lack of objectivity and systematic errors since it often depends on information provided by the patients' parents, teachers, or even the patients themselves¹⁴. This risks over-diagnosis or failure to diagnose, since some symptoms can overlap with others. Neuropsychological testing, comprising assessment tests for attention and impulse regulation, provides more standardized measurements but is time-consuming, expensive, and resource-intensive, thereby not easily accessible. Neuroimaging techniques like fMRI and EEG were promising for determining patterns of brain activity that may be related to ADHD, but expensive and quite impractical for more routine clinical utility from a practical and technical standpoint^{15,16}. Due to these limitations, a novel detection technique is introduced to bridge the gaps in ADHD diagnosis.

The following highlights the research's primary contribution:

- To utilize two databases such as EEG and diagnostic databases in the ADHD detection for enabling comprehensive, accurate, and reliable ADHD detection.
- To introduce a novel NeuroDCT-ICA for EEG data preprocessing which combines Improved Discrete Cosine Transform (I-DCT) with Independent Component Analysis (ICA) to eliminate noise and extract meaningful features from raw EEG signals, ensuring high-quality data preprocessing.
- To utilize Multi-Modal Feature Fusion using an attention network to integrate EEG features and diagnostic data, ensuring robust feature representation for ADHD detection.
- To develop a new RFO for feature selection which is a hybrid algorithm, blending Bitterling Fish Optimization (BFO) and Rhinopithecus Swarm Optimization (RSO) for optimal feature selection, enhancing processing stability and reducing computational complexity.
- To introduce a novel ADHD AttentionNet, DL model designed for accurate and efficient ADHD identification, offering superior performance with minimal computational cost.

The organization of the work is as follows: Sect "Introduction" gives an introduction related to the given title, and Sect "Related work" mentions recent literature related to the title. Furthermore, the suggested approach will be deliberated in Sect "Proposed methodology". Lastly, the result and discussion of the suggested model are given in Sect "Result and discussion". Finally, the work will be concluded with the conclusion Sect "Conclusion".

Related work

In 2022, Liu et al.¹⁷ presents a DL model for the diagnosis of ADHD through rs-fMRI. For spatial feature extraction, the model uses Nested Residual Convolutional Denoising Autoencoder (NRCDAE) and for temporal feature extraction, a 3D Convolutional Gated Recurrent Unit (GRU) is used. These features are then classified using a sigmoid classifier: The method demonstrates enhanced classification accuracy when distinguishing between ADHD and control subjects and generalization to other sites more effectively than other models.

In 2023, Chen et al.¹⁸ suggested an attention auto-encoding network (Att-AENet) under binary hypothesis testing (BHT) paradigm. This model takes brain functional connectivities (FCs) as input and dynamically computes FC weights using the attention mechanism to quantify the importance of FCs concerning ADHD. The Att-AENet consists of two attention subnetworks: one with dense structure and the other with two-dimensional convolution structures and their performance on the classification of ADHD is compared. The dense-based subnetwork is particularly employed for biomarker detection analysis because of its higher classification accuracy.

In 2023, Liu et al.¹⁹ suggested integrating radiomics features from resting-state functional magnetic resonance imaging (rs-fMRI). From the four kinds of preprocessed rs-fMRI data: ReHo, ALFF, VMHC, and DC, 93 features of the brain areas were obtained. The next step was to choose the most appropriate features and after this a SVM model was developed with accuracy of 76.3% in training and 77.0% in testing. The results indicate that rs-fMRI based radiomics features can be used as useful neuroimaging biomarkers for differentiating ADHD from HC.

In 2022, Hamed et al.²⁰ presented a brain connectomics model based on eyes-open resting state magnetoencephalography (rs-MEG) for the diagnosis of ADHD from HC. It involves computing Coherence (COH) between MEG sensors for different frequency bands, picking the best features for COH using Neighborhood Component Analysis (NCA) and using three classifiers, Support Vector Machines (SVMs), k-Nearest Neighbors (KNN) and Decision Tree for classification.

In 2023, Chandra and Rana²¹ presented an idea of an automatic diagnostic tool for ADHD based on structural MRI and PC data for classification. It includes feature extraction from GM volumes and CT; feature ranking using mRMR and EFS; and classifier training KNN, Logistic Regression (LR), linear SVM, RBSVM, and Random Forest (RF). The system performance is measured by accuracy, recall and specificity using ten-fold cross validation, with maximum accuracy of 75% using CT and PC features with RBSVM and SVM.

In 2022, Uyulan et al.²² proposed an automatic system for detecting ADHD through transfer learning with a ResNet-50 type 2D-Convolutional Neural Network (CNN). This method resolves issues of data acquisition variation and class distribution. The system attained a classification accuracy of 93.45% via 10-CV. The results are analyzed using the Class Activation Map (CAM) which shows the regions of the brain involved in the frontal, parietal, and temporal lobes where children with ADHD are different from the normal controls.

In 2024, Gülşah and Özmen²³ introduced a new DL model in the form of a 3D CNN architecture for the classification of people with ADHD using fMRI. The model employed fractional amplitude of low-frequency fluctuations (fALFF) and regional homogeneity (ReHo) data and balances the datasets through five-fold cross-validation. The proposed 3D CNN is compared with a fully connected neural network (FCNN), which clearly depicts higher accuracy rates in classifying ADHD especially on fALFF data. It is found that the proposed 3D CNNs are useful for the diagnosis of ADHD and are better than the FCNN models in this regard.

In 2022, Maniruzzaman et al.²⁴ presented an ML algorithm for classification of ADHD from normal children based on EEG signal. It uses morphological and time-domain features from the EEG signals and t-test and LASSO for feature selection. The selected features are then classified using four ML algorithms: SVM, KNN, MLP, and LR algorithms were used in the study. Comparing the results of LASSO and SVM, it was possible to conclude that the integrated method has the highest accuracy of 94.2%, which indicated the possibility of using this tool in the early diagnosis of ADHD.

In 2021, Zhou et al.²⁵ proposed a novel ML approach based on Boruta feature selection and Multiple Kernel Learning (MKL) to incorporate structural and functional MRI and DTI data to diagnose early adolescent ADHD. Overall, the framework synthesizes macrostructural characteristics, microstructural parameters, and functional connection density at the kernel network level and a SVM classifier for differentiating ADHD from non-ADHD children. The experiment results demonstrated that, the kernel-level fusion of multimodal features can yield the classification accuracy of 64.3%, and the AUC of 0.698, which are higher than those of unimodal and early feature fusion methods.

In 2021, Khullar et al.²⁶ presented a ML approach for the discrimination of ADHD from healthy controls based on resting state functional magnetic resonance imaging (rs-fMRI) data. The proposed method uses a 2D CNN and a 2D CNN-LSTM model as the base network architecture. The performance of the system is compared with the existing methods with reference to accuracy, specificity, sensitivity, F1-score, and AUC, and the results demonstrate the enhanced performance of the proposed method in differentiating ADHD from healthy control subjects.

In 2022, Barua et al.³³ suggested a ternary motif pattern (TMP) for ADHD detection using EEG signals. To create wavelet subbands, the Tunable Q Wavelet Transform (TQWT) was employed. The TMP and statistics were employed to create meaningful features from both raw EEG signals and wavelet bands by producing TQWT. Here, the original EEG signal and 18 subbands had produced characteristics. This model was hence called TMP19. Neighborhood component analysis (NCA) was used to choose the most informative features, and the kNN classifier was used to classify the features that were chosen.

In 2022, TaghiBeyglou et al.⁶ suggested a novel CNN structure in combination with traditional ML models to diagnose attention deficit hyperactivity disorder (ADHD) in children. The raw electroencephalography (EEG) output was used as the input. Neither transformation nor artifact rejection procedures were needed for the EEG-based method. To diagnose ADHD, this model first trains a CNN using raw EEG data. Then, several traditional classifiers, including SVM, LR, RF, etc., were trained using the feature maps that were derived from the various layers of the learned CNN.

In 2023, Maniruzzaman et al.³⁴ proposed two distinct methods such as SVMs and t-tests to choose the best channels separately. A hybrid channel selection strategy was then suggested, which combined the two methods to choose the best channels. The key features from the chosen channels were then chosen using a model based on LASSO LR. Lastly, six ML (ML)-based classifiers such as LR, multilayer perceptrons, KNNs, RF, and Gaussian process classification (GPC) were used to identify children with ADHD. Accuracy and area under the curve (AUC) were used to assess each classifier's performance. Table 1 represents the comparison of related works of ADHD detection.

Problem statement

Proposed methodology

The proposed enhanced DL framework for ADHD detection aims at improving the accuracy of early-stage detection to enable effective handling of the condition. This approach utilizes two key databases namely the EEG database and the Diagnostic database. In case of EEG database, raw data are pre-processed with NeuroDCT-ICA which is a combination of ICA and I-DCT. This new kind of preprocessing helps to filter noise and extract useful features from EEG signals successfully. After that, several feature extraction techniques are applied on the pre-processed EEG signals, including time domain, frequency domain, and connectivity. The diagnostic database, on the other hand, is pre-processed by one-hot encoding, outlier detection, and correction, and mean imputation. Subsequently, feature extraction is done through the Pearson correlation coefficient and statistical measures, which give useful information from the clinical data. These two data modalities are combined through multi-modal feature fusion using an attention-based network to improve the system's reliability. One of the new developments in this work is the RFO algorithm that is a combination of BFO and RSO, which enhances the process of feature selection, thus enhancing the efficiency in data processing and the stability of the system. Lastly, the newly proposed ADHD-AttentionNet is used for the ADHD detection with high accuracy and low computational complexity. This all-encompassing approach presents a major leap forward in the diagnosis of ADHD and presents a practical, effective, and accurate system to assist the clinician in enhancing the diagnostic

Authors, Year	Techniques	Databases	Advantage	Disadvantage	Outcomes
Liu et al., 2022	Deep spatio-temporal features of functional Magnetic Resonance Imaging	ADHD-200	reduces spatial dimension of rs-fMRI, uses spatio-temporal features	Cross-site testing may involve data variability	Accuracy = 71.3%
Chen et al., 2023	Attention auto-encoding neural network	ADHD-200	High Classification Accuracy	Lack of Interpretability of DL Features	Accuracy = 98.9%
Liu et al., 2023	Radiomics-based feature extraction, SVM	ADHD-200	Uses four preprocessed rs-fMRI images, feature selection for better classification	High-dimensional feature space, requires dimensionality reduction	Accuracy = 76.3% (training), 77.0% (testing), AUC = 0.811 (training), 0.797 (testing)
Hamed et al., 2022	Coherence (COH) analysis, Neighbourhood Component Analysis (NCA), SVM, KNN, Decision Tree	Eyes-open rs-MEG	Novel use of rs-MEG for ADHD classification	Small sample size (25 ADHD, 25 HC)	Accuracy = 91.1%
Chandra and Rana, 2023	Structural brain MRI and personal characteristic data with ML framework	Publicly available dataset (316 ADHD, 316 TDC)	Uses anatomical and cortical thickness features for diagnosis	Limited generalization with age-balanced data	Accuracy = 75%
Uyulan et al., 2022	Transfer learning with ResNet-50 pre-trained 2D-CNN	ADHD-200	Solves dimensionality issues, robust to data acquisition variances	Requires a large amount of training data	Accuracy = 93.45%
Gülşah and Özmen, 2024	3D CNN	ADHD-200 (NI, NYU, PU)	Utilizes 3D CNN for feature extraction, improves generalization	Inconsistent dataset sizes, lower accuracy on some datasets	Accuracy = 76.19% (NI), 69.92% (NYU), 70.77% (PU)
Maniruzzaman et al., 2022	Machine Learning	EEG (61 ADHD, 60 HC)	High classification accuracy with LASSO and SVM combination	May not generalize well for other types of ADHD data	Accuracy = 94.2%, Sensitivity = 93.3%
Zhou et al., 2021	Boruta-based feature selection, Multiple Kernel Learning (MKL), SVM	ADHD-200, Adolescent Brain and Cognitive Development (ABCD)	Combines structural and functional MRI with DTI for early diagnosis	Complex feature fusion process, limited to early adolescents	Accuracy = 64.3%, AUC = 0.698
Khullar et al., 2021	2D CNN, Hybrid 2D CNN-LSTM, Deep Learning	ADHD-200	Uses DL for ADHD classification, hybrid approach	Requires large labeled dataset for training	Accuracy = 95.32%, Specificity = 93.54%
Barua et al., 2022	Ternary Motif Pattern (TMP) using TQWT (Tunable Q Wavelet Transform) and NCA (Neighborhood Component Analysis)	ADHD EEG dataset	Captures meaningful features from EEG signals and wavelet subbands, effective in feature extraction.	Dependent on feature selection process and quality of data.	Accuracy = 95.57%
TaghiBeyglou et al., 2022	CNN combined with traditional ML models (SVM, LR, RF) for ADHD diagnosis using raw EEG data	EEG competition by the National Brain Mapping Laboratory (NBML).	No need for transformation or artifact rejection, straightforward approach to using raw EEG data	CNN model may require large datasets for effective learning.	Accuracy = 95.83%
Maniruzzaman et al., 2023	Hybrid channel selection (SVM + t-tests) and LASSO LR for feature selection	EEG-based dataset	Hybrid approach combines SVM and t-test for better feature selection; effective in choosing key features.	May require significant computational effort for feature selection.	Accuracy of 97.53% and AUC of 0.999.

Table 1. Comparison of related works.

process and subsequently the care of the patient. Figure 1 presents the overall architecture of the suggested approach.

ADHD is often associated with abnormal brain activity, particularly in areas related to attention, working memory, and impulse control. These irregularities can manifest in complex brain signals, making it hard to interpret them directly. NeuroDCT-ICA optimizes the separation of brain activity into independent sources, helping to identify ADHD-related features, such as abnormal neural oscillations, more accurately. The combination of DCT and ICA allows the model to capture both high-level features (via DCT) and separate underlying sources of brain activity (via ICA). This ensures that only the most relevant and noise-free features are considered when diagnosing ADHD, improving the accuracy of detection. This framework incorporates DL components like an attention mechanism is likely being used in the feature-level fusion of pre-extracted features, which suggests that the overall system is a hybrid that still aligns with ML.

Data collection

The proposed approach is evaluated based on two databases specifically Diagnostic database (<https://www.kaggle.com/datasets/arashnic/adhd-diagnosis-data>) which is a public dataset containing health, activity, and heart rate data from adult patients diagnosed with ADHD. It includes the results and sum scores of many diagnostic assessment tools, time series of heart rate and motor activity, the output of a neuropsychological computer test, the participant's sex, age, and medications. The numbers for sex are zero for females and one for males. Four groups of participants' ages are shown: 1 represents those aged 17–29, 2 represents those aged 30–39, 3 represents those aged 40–49, and 4 represents those aged 50–67. Of the 85 patients whose motor activity was recorded, 23 were in age group 1, 26 were in age group 2, 24 were in age group 3, and 12 were in age group 4. The majority of the individuals did not use any drugs. Among those with an ADHD diagnosis who recorded their motor activities, 73% did not take medication, and only one person received a stimulant prescription.

The dataset consists of data collected from 51 patients with ADHD and 52 clinical-controls and EEG database (<https://www.kaggle.com/datasets/inancigdem/eeg-data-for-mental-attention-state-detection>), collection of 34

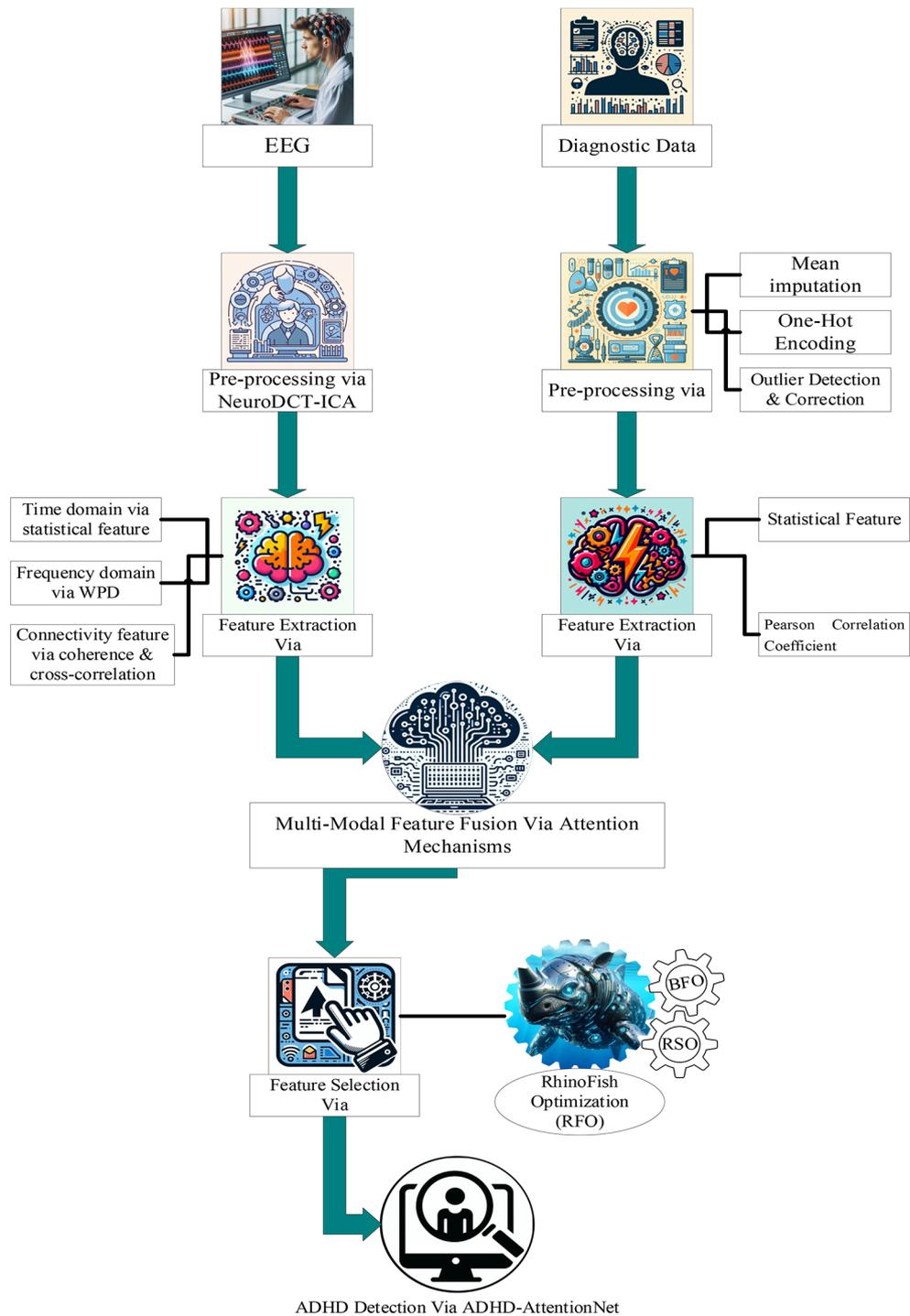


Fig. 1. Architecture of the suggested ADHD detection approach.

experiments for monitoring of attention state in human individuals using passive EEG BCI. The data collected from an EMOTIV device during a single experiment is contained in each file. O.data, an array of size {number-of-samples} \times 25, contains the raw data; therefore, o.data(:,i) consists of a single data channel. 128 Hz is the sampling frequency.

Pre-processing

EEG data

The NeuroDCT-ICA blends ICA with the I-DCT allowing for a better preprocessing of EEG signals. This method introduces in the new dimension of data enhancement by addressing the issues of cleaning of data and transforming it to the frequency domain. This has the effect of increasing the quality of the obtained EEG data. The technique prepares the data for further analysis – classification and feature extraction – by successfully

removing recordings of eye blinks, movements of muscles, electrical noise or any other such artifacts while maintaining the desired data of brain activities. NeuroDCT-ICA does not simply apply filtering or ICA at one stage as in many other existing solutions, but rather incorporates stencil two step sequence of operations.

Mathematical model Step 1: Savitzky-Golay filters.

Noise in high frequencies is reduced using the Savitzky-Golay filter²⁷, but the essential properties of the EEG signal are preserved. This filter functions by applying a polynomial fit to a set of points from within a certain area of the signal that is translated to the right. In other words, it smoothed the signal out while not changing the frequencies contained in it too much. This method is applied to the EEG. The next steps are preparatory to the cleaning process, facilitating onward processing of the data by offering cleaner data for aiding in artifact removal and the analysis of the data. Mathematically, this can be given as per the Eq. (1).

$$s(t) = \frac{1}{N} \sum_{k=-N}^N \psi_k \cdot r(t+k) \quad (1)$$

Where, $r(t+k)$ and $s(t)$ denotes the raw EEG signal and smoothed EEG signal, respectively. In addition, the weight function is denoted as ψ_k and the window size is denoted as N .

Step 2: Normalization.

To address participant variability, the reduced EEG data is normalized across individuals. Techniques such as Z-score normalization²⁸ are applied, rescaling the data to a standard range or a zero-centered normal distribution. The mathematical expression is expressed in the Eq. (2).

$$z(t) = \frac{s(t) - \zeta_r}{\beta_r} \quad (2)$$

Where, $z(t)$ represents the normalized EEG signal, ζ_r and β_r denotes the mean and standard deviation of the smoothed $s(t)$ signal.

Step 3: Artifact removal using ICA.

One of the crucial procedures in EEG preprocessing is the removal of artifacts or any undesired elements such as eye blinks, movements of muscles, and electric noise, which can affect the evaluation of brain activities. ICA is a computational technique used to separate mixed signals into independent components. In brain signal processing, this means identifying distinct sources of brain activity from complex, overlapping signals, such as EEG (electroencephalography) or fMRI (functional magnetic resonance imaging). ICA helps extract relevant features from these signals, crucial for detecting patterns related to mental health conditions like ADHD. It is used to decompose the EEG data into separate spatially independent components, thus these artifacts can be placed and excluded from analysis. The mathematical representation is given in the Eq. (3).

$$I(t) = A \cdot Q \quad (3)$$

where, $I(t)$ denoted as the decomposed the EEG data, A represents the mixing matrix and Q represents the independent sources acquired after the artifact removal of $z(t)$. Also, dot (\cdot) usually represents matrix multiplication.

Step 4: Signal segmentation via epoching.

The cleaned EEG signal $I(t)$ is segmented into epochs $E(t)$ that are related to certain events or actions. Epoching is done by splitting the uninterrupted EEG signal into brief time-locked segments which are coordinate with particular psychological events or stimuli. This facilitates understanding of the signal at various time instants enhancing its application in time series analysis and classification. Usually, the lengths of epochs are between 1 and 5 s but this varies with the depth of analysis that is needed.

Step 5: Frequency domain transformation using I-DCT.

NeuroDCT refers to a Neuro-Domain-Specific Transform that applies domain-specific signal processing techniques, like the Discrete Cosine Transform (DCT), to further refine the feature extraction process. DCT helps convert a signal into a frequency domain, making it easier to capture specific patterns of brain activity that are difficult to detect in the time domain. The Improved-DCT is performed on a cleaned EEG signal in each segmented ROI to transform it from the time domain to the frequency domain. This transformation emphasizes the low and high frequency components of the signal, which are important for the following analysis. When considering the standard DCT, the I-DCT gives higher importance to the trajectories of the system. Here, 2D-IDCT is applied to operate on the entire 2D data structure such as matrices representing spatial relationships. As a result, the system manages to differentiate even frequency components clearly. Low frequencies that represent delta and theta, slow brain wave patterns are captured well together with compact high frequency features that are characterized by sharp transients or rapid brain activities. The vital steps in the processing of EEG signals are the employment of the I-DCT which allows for feature extraction, classification and pattern recognition. The constructed method enables the operating distortion of the analysis focused on the raw signal time gaps by performing this signal shifting into the frequency domain. This improvement allows the better-quality analysis of the brain activity as especially weak brain activity refined changes could become visible that were not present in the time analysis.

In an 8×8 DCT matrix, low frequencies reside in the upper left, mid-frequencies in the middle, and high frequencies in the lower right. This transformation is applied to $E(t)$ (epochs) $f(x, y)$ of size $m \times n$ using an equation that rearranges pixel values to highlight different frequency components, a vital step in image compression and processing as per Eq. (4) and Eq. (5).

$$f(u, v) = \frac{2}{\sqrt{mn}} c(u) c(v) \sum_{x=0}^{n-1} \sum_{y=0}^{m-1} f(x, y) \times w(x, y) \times \cos \left[\frac{\pi (2x + 1) u}{2n} \right] \cos \left[\frac{\pi (2y + 1) v}{2m} \right] \quad (4)$$

for $u = 0, 1, \dots, n - 1$ and $v = 0, 1, \dots, m - 1$,

$$c(A) = \left\{ \frac{1}{\sqrt{2}}, \text{ for } a = 0, 1, \text{ otherwise} \right. \quad (5)$$

Inverse DCT reverts transformed $E(t)$, vital for signal reconstruction as per Eq. (6).

$$f(x, y) = \frac{2}{\sqrt{mn}} \sum_{x=0}^{n-1} \sum_{y=0}^{m-1} c(u) c(v) f(x, y) \times w(x, y) \times \cos \left[\frac{\pi (2x + 1) u}{2n} \right] \cos \left[\frac{\pi (2y + 1) v}{2m} \right] \quad (6)$$

Here, $f(x, y)$ is represented as the Frequency Domain Transformation parameter. When processing EEG signals, the coefficient $F(0,0)$, is known to carry a major portion of the energy of the signal and, thus, depicts the average strength of the signal over a period of time. This coefficient plays a very critical role in determining the overall background activity of the brain. The high-frequency coefficients usually found at the lower right section of the transformed matrix $w(x, y)$ contain information regarding a short period of time and the detailed neural activities such as rapid spikes, oscillations, or brain information which is associated with advanced cognitive processes. The frequency domain signal (pre-processed EEG signal) is represented as E_i^{pre} .

Diagnostic data preprocessing

The processes listed below describe the methods for preprocessing clinical data such as standardized, clean, and fit for applications employing ML or analysis.

Step 1: mean imputation for Missing Data Handling.

In order to ensure a complete dataset and prevent the loss of essential data, handling missing data is vital²⁹. This step is especially essential in clinical datasets, where missing values can arise from incomplete patient records, equipment malfunctions, or human errors. This method maintains the dataset's overall distribution and enables seamless analysis in the Eq. (7).

$$Y_i = \frac{1}{m} \sum_{i=1}^m X_i \quad (7)$$

where, Y_i represents the missing observations in the input data and the count of observations (samples) are denoted as n .

Step 2: One-Hot Encoding is a technique used for converting categorical variables (such as gender, symptoms, and severity) into a numeric form which can be understood by the ML models³⁰. Let the categorical values within the categories of Y_i be represented as C_i . The one-hot encoding is applied to each of the instance in the Eq. (8).

$$OneHot(C_i) = [0, ., 1, 0, 0] \quad (8)$$

Where, 1 denotes the active category and all other positions are 0.

Step 3: Outlier Detection and Correction.

The dataset has undergone one-hot encoding, and now comprises only numerical dimensions (continuous and binary features). For the continuous dimensions, Z-score normalization is carried out. Z-score is employed in outlier detection. It quantifies an element's distance from the mean of its feature in terms of standard deviations. If a Z-score is greater than 3 or less than -3, the corresponding measurement is considered an anomaly and discarded.

Feature extraction

EEG feature extraction

Time-domain features For EEG signal processing, specifically in the analysis of EEG signals, feature extraction is an important step in representing the raw data in a more useful form. Time features, which are the features taken directly from the time-series data without any transformation into frequency data are one of the most common. Such features, which include mean, variance and peak amplitude. As one of the features extraction methods in EEG analysis that is commonly employed, it is known that there is a simple way to present the features of the signal which describe the activity of the nervous system. Table 2 provides time features and their descriptions.

Frequency-domain features - wavelet packet decomposition (WPD) for detailed frequency bands With Wavelet Packet Decomposition (WPD), an advanced frequency-domain feature extraction technique, EEG data are separated into multiple frequency bands. While the standard wavelet transform only analyzes the low-frequency (approximation) components, WPD analyzes the low-frequency and high-frequency components at all levels. Specifically, it is helpful for studying the various brain rhythms (delta, theta, alpha, beta, and gamma) associated with different neurological and cognitive activities since it provides a more detailed representation of the EEG signal.

Key steps in WPD for EEG analysis

Feature	Description	Interpretation
Mean	The average value of the EEG signal over a time segment. It shows the overall level of brain activity.	Higher mean values recommend sharp brain activity, like during tasks or arousal. Lower mean values specify relaxed states, such as deep sleep or rest.
Variance	Measures how much the EEG signal varies around the mean. It designates the variability of brain activity.	High variance replicates irregular signals, seen in active cognitive states or neurological events. Low variance recommends stable, calm brain states or absence of stimulation.
Peak amplitude	The maximum absolute value of the EEG signal in the segment, representing the intensity of neural events.	High peak amplitude could specify brain events like seizures or pathological phenomena. Low peak amplitude is characteristic in resting states or low cognitive/sensory demand.

Table 2. Time features and their description.

1. Decomposition Process: To enable complete analysis of an EEG signal's E_i^{pre} frequency components, WPD progressively split the signal up into smaller sub-bands.
2. There are two components to the signal at each stage of decomposition:
 - Approximation (low-frequency) components, which signify general signal patterns in the Eq. (9).

$$A_{i+1}[m] = \sum_k E_i^{pre}[k] \cdot h[2m - k] s \quad (9)$$

Where, $A_{i+1}[m]$ and $D_{i+1}[m]$ represents the Approximation coefficients and detailed at level $i + 1$, correspondingly. In addition, $h(k)$ and $g(k)$ symbolizes the Low-pass and high-pass filters, respectively.

- Detail (high-frequency) components, which capture finer oscillations and transient details in the Eq. (10).

$$D_{i+1}[m] = \sum_k E_i^{pre}[k] \cdot g[2m - k] \quad (10)$$

After several levels of decomposition, WPD offers a complete depiction of the signal across multiple frequency bands, allowing the analysis of both low-frequency rhythms and higher-frequency events.

3. Frequency Bands: WPD select the right number of decomposition stages D to isolate particular frequency bands in an EEG data. By concentrating on frequency components that are associated to standard EEG cycles, this method allows a complete examination of the data. The following are the key frequency bands that WPD identified:
 - Delta (0.5–4 Hz): Deep sleep, unconscious states.
 - Theta (4–8 Hz): Light sleep, relaxation, meditative states.
 - Alpha (8–13 Hz): Relaxed wakefulness, creativity, meditation.
 - Beta (13–30 Hz): Active thinking, problem-solving, focus.
 - Gamma (30–100 Hz): High-level cognitive functions, memory recall, learning.

Each deconstructed component inside these bands is then analyzed for features relevant to monitoring brain states and identifying anomalies, providing valuable data on neural and cognitive activity.

4. Power spectral density (PSD): First, we calculate the Power Spectral Density (PSD) for each EEG signal employing Fourier transforms which is a signal feature and reflects the frequency characteristics of a signal. For $a(t)$ and $b(t)$, the PSD can be represented as $P_{aa}(t)$, $P_{bb}(t)$, correspondingly. At the f^{th} frequency, the power shared between $a(t)$ and $b(t)$ be represented as $P_{ab}(t)$. The coherence function $E_{ab}(f)$ can be calculated in the Eq. (11).

$$E_{ab}(f) = \frac{|P_{ab}(f)|^2}{P_{aa}(f) \cdot P_{bb}(f)} \quad (11)$$

Where, 0 denotes no coherence, 1 represents perfect coherence and the value of $E_{ab}(f)$ ranges between [0,1],

5. Feature extraction: After decomposing the signal into frequency bands, various features can be calculated from each band. Frequency features and their descriptions are given in Table 3.

WPD based analysis for feature extraction from EEG signals in frequency domain permits to go beyond the limitation of conventional EEG bands. It successfully encompasses both the somnolent states, which are sustained patterns of activity and discrete short-firing high-frequency events, such as cognitive operations or seizures.

Connectivity features Measure inter-regional connectivity using coherence and cross-correlation.

Feature	Description
Energy	The energy e_i represents the power within the sub-band i and is calculated.
Entropy	Entropy provides a measure of the unpredictability in each frequency band
Mean	Average coefficient value for sub-band i , indicating central tendency.
Standard deviation	Spread of the coefficient values, indicating variability within the sub-band i .

Table 3. Frequency features and their description.

Features	Description
Geometric mean	The $n - th$ root of the product of n EEG signals is the geometric mean.
Kurtosis	Measures the “peakedness” or flatness of the EEG signal, which can be useful for detecting burst-like activities.
Skewness	Assesses the asymmetry of the EEG signal distribution, indicating extreme activity on one side.
Root Mean Square (RMS)	A measure of the energy content within the EEG signal, often associated with signal power.
Range	Difference between maximum and minimum values in the EEG data, showing the spread in measurements.

Table 4. Statistical features and their description.

- (1) Coherence: Coherence is an index of the frequency domain linear correlation between two EEG signals. It measures the degree of overlap in power in the signals at given frequencies, indicative of the interaction between different regions of the brain.
- (2) Cross-correlation.

Cross-correlation measures the similarity between two EEG signals as a function of the time-lag τ . This can indicate synchronous activity or temporal relationships between different brain regions. At τ^{th} lag, the Cross-Correlation Function $R_{ab}(\tau)$ can be given as per Eq. (12).

$$R_{ab}(\tau) = \sum_t a(t) \cdot b(t + \tau) \tag{12}$$

Diagnostic data feature extraction

Statistical features Pattern recognition in data is enhanced by the use of statistical features like mean, variance, correlation between symptoms or scores. Such features are exploited from EEG signals to represent the important statistics of the signal i.e. central tendency, dispersion and its distribution. Such features are useful for further treatment, for classification or for diagnostics. Statistical features and their descriptions are provided in Table 4.

Correlation analysis- Pearson correlation coefficient Identify the relationships between demographic data (e.g., age, gender) and clinical variables (e.g., blood pressure, cholesterol levels). The Pearson correlation coefficient (s_{YZ}), which is a statistical measure of the degree of linear correlation between two variables, Y and Z , is named after the English mathematician and biostatistician Karl Pearson. It is defined as shown in Eq. (13):

$$s_{YZ} = \frac{cov(Y, Z)}{\sigma_Y \cdot \sigma_Z} \tag{13}$$

Where $cov(Y, Z)$ covariance between Y and Z , and the mathematical expression of $cov(Y, Z)$ is shown in Eq. (14)

$$cov(Y, Z) = \frac{1}{o-1} \cdot \sum_{j=1}^o (Y_j - \bar{Y}) \cdot (Z_j - \bar{Z}) \tag{14}$$

$\sigma_Y \cdot \sigma_Z$ standard deviation of Y and standard deviation of Z . And their relative expression is shown in Eq. (15)

$$cov(Y, Z) = \frac{1}{o} \cdot \sum_{j=1}^o (Y_j - \bar{Y}) \cdot (Z_j - \bar{Z}) \tag{15}$$

By substituting (14) and (15) in Eq. (13), The following Eq. (16) can be used for the Pearson correlation coefficient:

$$S_{yz} = \frac{o \cdot \sum_{j=1}^o Y_j \cdot z_j - \left(\sum_{j=1}^o Y_j\right) \cdot \left(\sum_{j=1}^o z_j\right)}{\sqrt{o \cdot \sum_{j=1}^o Y_j^2 - \left(\sum_{j=1}^o Y_j\right)^2} \cdot \sqrt{o \cdot \sum_{j=1}^o z_j^2 - \left(\sum_{j=1}^o z_j\right)^2}} \tag{16}$$

Several sources refer to the Pearson correlation coefficient as Pearson's s rather than s_{YZ} . The Greek letter ρ , represented as ρ_{YZ} , stands for the Pearson correlation coefficient when it is applied to the entire population as opposed to a sample.

Multi-modal feature fusion with attention mechanisms

In order to distinguish the most useful elements from each source for improved diagnosis or prognosis, multi-modal fusion with attention mechanisms allows for the incorporation of both Diagnostic as well as EEG. Equation (17) shows the result of concatenating these features into a single, cohesive feature vector.

$$F_{can} = [f_{diagnostic} - f_{EEG}] \quad (17)$$

This concatenation combines all extracted features (diagnostic and EEG data features) into one cohesive representation, making them available for processing by the attention mechanism.

Attention score calculation: Use learnable weight matrix W_{att} to compute attention scores for each feature in the concatenated feature set. The attention score e_i for each feature is calculated by a dot product between the feature vector F_{concat} and the attention weight matrix W_{att} .

- Weighted feature vector: Once the attention weights α_i are calculated, it is used to re-weight the features in the concatenated vector. The weighted features are then used for the next step in the fusion process as per the Eqs. (18) and (19).

$$F_{att} = \sum_i \alpha_i F_{concat, i} \quad (18)$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_j \exp(e_j)} \quad (19)$$

Where, F_{att} represents the weighted feature vector after attention is applied.

In this step, the weighted features from both Diagnostic and EEG data are combined. The attention mechanism has already guaranteed that the most related features from both modalities are given higher rank. The final multi-modal fusion vector is obtained by adding the attention-weighted Diagnostic and clinical features together as per the Eq. (20).

$$F_{final} = F_{att}^{EEG} + F_{att}^{diagnostic} \quad (20)$$

Where, F_{final} denoted as the final multi-model feature set, which captures the essential data from both behavioural and clinical data (EEG), $F_{att}^{diagnostic}$ denoted as the attention-weighted behavioural features and F_{att}^{EEG} denoted as the attention-weighted clinical features.

Feature selection

Proposed RFO is a nature-inspired optimization algorithm based on the movement patterns of RhinoFish. It's part of a family of algorithms that mimic the behaviors of natural systems to solve complex optimization problems, especially those involving multi-dimensional data or high variability. RFO works by simulating the searching behavior of RhinoFish in a multi-dimensional environment, exploring the search space for optimal solutions. The algorithm balances exploration (searching for new areas of the solution space) and exploitation (focusing on promising areas identified earlier), ensuring that the search for optimal solutions is both thorough and efficient.

ADHD detection typically involves processing vast amounts of data from multiple sources (EEG, fMRI, behavioral data). This data can be noisy, high-dimensional, and often requires feature selection to determine the most informative inputs for accurate classification. RFO is used to optimize the feature selection process. It intelligently selects which features (or patterns in brain signals) should be prioritized for classification, refining the data inputs and improving the performance of ADHD detection models. By optimizing the classification process, RFO helps ensure that the most relevant brain signals and behavioral patterns are used for ADHD diagnosis, improving both accuracy and efficiency.

At this point, the optimal feature sets F_{final} are picked from extracted features using proposed RFO. It combines the aquatic behaviors of BFO and survival nature of RSO.

Proposed RFO: In the proposed RFO, the cooperative, adaptive search patterns of fish and the social and collaborative behaviors of Rhinopithecus monkeys are used to provide effective exploration and exploitation in solution spaces, and convergence stability. Firstly, the proposed RFO begins with defining the population of solutions (features), where each solution corresponds to a set of selected features from the dataset as referred in Eq. (21), where χ_i denotes i^{th} solution (feature subset) in the population, x_{ij} signifies whether feature j is selected, and x_{in} stands for group members count.

$$\chi_i = [x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in}] \quad (21)$$

The proposed Rhino Fish employs the aquatic nature of Bitterling fish and survival strategy of Rhinopithecus. The position for Rhino Fish is initialized in Eq. (22), where $\chi_{i,j}(t)$ point to Rhino Fish's position at iteration t , U and L represents upper and lower limits, n denotes population size, r indicates arbitrary number in $(0,1)$ and d points to dimension.

$$\chi_{i,j}(t) = L + r \times (U - L), i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, d \quad (22)$$

Each Rhino Fish specifies a candidate solution and the fitness for every solution is estimated via Eq. (23), in which χ_i denotes the set of optimal features and the objective is to reach maximum accuracy.

$$f(\chi_i) = (\text{Accuracy}_i) \quad (23)$$

In BFO, the Bitterling fish searches best mating oysters, on the other hand, Rhinopithecus aims to choose best leader among the swarm. Using these strategies, the proposed RFO aims to choose optimal features based on combining search patterns of best-mating oysters and leader selection. The mathematical model of such design is defined as follows. A parameter S is used to indicate the step rate of the Rhino Fish to approach the best feature and it minimizes over iteration so as to move the algorithm to local search from global search as represented in Eq. (24), where $S(1)$ and $S(t)$ indicate every Rhino Fish's step rate in the beginning, and iterations t over time, t and M_t denotes present and maximum iterations, $R(t)$ signifies arbitrary function to produce arbitrary sequences as given in Eq. (25).

$$S(t) = \left(S(1) - \frac{s(1) \cdot t}{M_t} \right) \cdot R(t) \quad (24)$$

$$R(t+1) = \text{coscos}(t \times (R(t))) \quad (25)$$

Assume $R(1) = 1$. Now, combine Eq. (24) with (25) which gives Eq. (26). It is performed to decrease the parameter M over iterations, and this M is used to decide the position update of Rhino Fish.

$$S(t+1) = \left(S(1) - \frac{s(1) \cdot t}{M_t} \right) \cdot \text{coscos}(t \times (R(t))) \quad (26)$$

Equation (27) gives the parameter M estimation, where a denotes reduction parameter and $a = \{0.1 \text{ or } 0.5 \text{ or } 0.9\}$. The lower the a value, the lower the M parameter.

$$M = \left| 1 - \frac{t}{\sqrt{1+t^2}} \right| + \frac{r}{t^a} \quad (27)$$

Based on these parameters, Eq. (28) computes the position update of Rhino Fish, where $\chi_i(t)$, and $\chi_i(t+1)$ refer to Rhino Fish present and new location at iteration t and $t+1$, D signifies an arbitrary number in $(0,1)$, and χ^+ stands for worth of the best features selected.

$$\chi_i(t+1) = S \cdot \chi_i(t) + (\chi^+ - S \cdot \chi_i(t)) \cdot D \quad (28)$$

On the other hand, the Rhino Fish chooses best survival position based on king (K) and mature (M) and adolescent A Rhino Fish. Here, each Rhino Fish update their position based on king (K) Rhino Fish using Eq. (29), in which *Gaussian* addresses Gaussian distribution having expectation a and variance b .

$$\chi_i(t+1) = \text{Gaussian}(a, b) \quad (29)$$

Based on M , the Rhino Fish location update is performed using Eq. (30).

$$\chi_i(t+1) = \{S \cdot \chi_i(t) + (\chi^+ - S \cdot \chi_i(t)) \cdot D \text{ if } \gamma > M \text{ Gaussian}(a, b) \text{ if } \gamma \leq M \quad (30)$$

Equations (31) and (32) gives the estimation of a and b , in which $a, b \in (0,2)$, K_α and M_β signify king and mature Rhino Fish locations.

$$a = \frac{K_\alpha + M_\beta}{2} \quad (31)$$

$$b = K_\alpha - M_\beta \quad (32)$$

In the exploitation phase, the proposed RFO update's location using escaping strategy as defined in Eq. (33).

$$\chi_i(t+1) = S \cdot \chi_i(t) + (\chi^* - S \cdot \rho) \cdot D \quad (33)$$

Here, ρ indicates population gravity point and is estimated in Eq. (34).

$$\rho = \sum_{i=1}^n \frac{\chi_i(t)}{n} \quad (34)$$

Equation (35) describes the location update of Rhino Fish based on learning adolescent Rhino Fish from mature ones, where (c, e) defines expectation, and (d, f) provides variance of Gaussian distribution *Gaussian*.

$$\chi_i(t+1) = \frac{\text{Gaussian}(c, d) + \text{Gaussian}(e, f)}{2} \quad (35)$$

$$c = \frac{K_\alpha + A_\delta}{2} \quad (36)$$

$$e = \frac{M_\beta + A_\delta}{2} \quad (37)$$

$$d = |K_\alpha - A_\delta| \quad (38)$$

$$f = |M_\beta - A_\delta| \quad (39)$$

Based on parameter M , the location update of Rhino Fish takes place as defined in Eq. (40).

$$\chi_i(t+1) = \begin{cases} \frac{\text{Gaussian}(c, d) + \text{Gaussian}(e, f)}{2} & \text{if } \gamma \geq M \\ S \cdot \chi_i(t) + (\chi^* - S \cdot \rho) \cdot D & \text{if } \gamma < M \end{cases} \quad (40)$$

After M_t iterations, the solution χ_i converges to the optimal set F_{opt} of refined features by minimizing the classification error and maximizing accuracy. Algorithm 1 demonstrates the pseudocode of proposed RFO.

ADHD-AttentionNet based ADHD detection

NeuroDCT-ICA is responsible for extracting clean, domain-specific features from complex brain signals. It enhances the signal processing and ensures that only the most relevant information is fed into the detection model. RFO works alongside NeuroDCT-ICA to fine-tune the feature selection process. It automatically

Step 1	Initialize parameters such as $t, U, L, n, d, M_t, S = 1, R = 1, a = 0.5$ Initialize random candidate solution For $i = 1:n$ For $i = 1:d$ Initialize $\chi_{i,j}(t)$ as per Eq. (30) End for End for
Step 2	Initial fitness estimation
Step 3	For $i = 1:M_t$ Update best Rhino Fish χ_{best} Set S, R , and M Update S and M
Step 4	For $i = 1:n$ If $rand < 0.5$ If $\gamma > M$ Update Rhino Fish new location as per Eq. (36) Else if $\gamma > M$ Update Rhino Fish new location as per Eq. (37) End if
Step 5	Else If $\gamma \geq M$ Update Rhino Fish new location as per Eq. (43) Else if $\gamma < M$ Update Rhino Fish via Eq. (41) End if End if Save the optimal solution χ_i End for

Algorithm 1. Pseudocode of proposed RFO.

identifies which features are most important for detecting ADHD and helps optimize the model's performance by ensuring that these features are correctly weighted in the final classification.

In this phase, the early ADHD detection is carried out via the proposed model which integrates Attention-Based DenseNet and Deep RCABP with soft attention mechanisms and GoogleNet and Squeezenet. This model is trained with the optimal features F_{opt} .

Proposed Channel-Aware DeepNet: In ADHD, differentiating various stages of disease is a complex task. In order to attain enhanced accuracy and minimize false positives, this research introduces Channel-Aware DeepNet which carefully analyzes the features using combined deep architecture. It uses DenseNet architecture with 3 dense blocks and 3 transition layers. Each Dense Block is made up of several convolutional layers, where each layer receives inputs from all previous layers in the block. This promotes feature reuse and helps mitigate the vanishing gradient problem. Each Dense Block consists of L convolutional layers where each layer l has k filters. The output of each layer is concatenated with the inputs to subsequent layers as defined in Eq. (41), in which χ_0 denotes input to the block (optimal features), and $H_l = Conv(X_{l-1})$.

$$X_{dense} = [\chi_0, H_1, H_2, \dots, H_L] \quad (41)$$

Transition Blocks are used to reduce the spatial dimensions (downsampling) and the number of feature maps. It comprised of convolution followed by a pooling layer. A Transition Block X_{tl} reduces feature maps C count and the spatial dimensions $H \times W$ using a 1×1 convolution followed by a 2×2 average pooling layer as shown in Eq. (42), in which $Conv$ refers to the 1×1 convolution.

$$X_{tl} = Pooling(Conv(X_{dense})) \quad (42)$$

At this point, the RCABP blocks enhance feature representation by focusing on the most informative channels using an attention mechanism while allowing residual connections for gradient flow. The RCABP enhances the feature representation using channel attention and spatial attention. Let X be the input to the RCABP. The attention weights for channels are computed in Eq. (43), where W_c means to learned weights, and σ represents sigmoid activation.

$$A_c = \sigma(W_c \cdot X) \quad (43)$$

Equation (44) expresses the output of the RCABP, where \odot represents element-wise multiplication (channel-wise scaling).

$$Y = X + A_c \odot X \quad (44)$$

Now, soft attention mechanism highlights important spatial regions in the feature maps and further refining the feature representation before the final classification. Soft attention refines the feature maps by generating attention weights based on spatial features as computed in Eq. (45), where W_s points to learned weights for spatial attention.

$$A_s = Sigmoid(W_s \cdot X) \quad (45)$$

Followed by soft attention, a GAP layer is used. GAP minimizes the feature maps spatial dimensions to a single vector per feature map, efficiently summarizing the information. Also, GAP converts the feature maps X into a vector by averaging each feature map as represented in Eq. (46), in which

$$X_{gap} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j) \quad (46)$$

Finally, the output from GAP is fed into an FC layer with softmax activation for classification that produces the final class probabilities as shown in Eq. (47), in which W_{fc} and b_{fc} signify weights and biases of the FC layer.

$$Y_{output} = Sigmoid(W_{fc} \cdot X_{gap} + b_{fc}) \quad (47)$$

Furthermore, a custom loss function is used by combining dice and focal loss for effective classification. Dice loss function measures the overlap between the predicted and true class labels to make it ideal for classification tasks. This approach ensures that each class, especially minority classes contribute to the loss function as given in Eq. (48), where C denotes classes count, $\rho_{i,c}$ refers to predicted probability for instance i in class c , $\tau_{i,c}$ stands for true class labels (1 if instance i belongs to class c , 0 otherwise), and N signifies total instances count.

$$L_{dice} = 1 - \frac{2 \sum_{c=1}^C \sum_{i=1}^N \rho_{i,c} \cdot \tau_{i,c}}{\sum_{c=1}^C \sum_{i=1}^N \rho_{i,c} + \sum_{c=1}^C \sum_{i=1}^N \tau_{i,c}} \quad (48)$$

On the other hand, focal loss in a multi-class setup reduces the influence of easily classified samples and focuses on those that are harder to classify. For each class, it applies a scaling factor to down-weight well-classified examples to make it especially effective in handling class imbalance in a multi-class context as defined in Eq. (49),

in which $\rho_{t,c}$ denotes predicted probability of the true class for class c , α_c means to balancing factor for each class, and γ addresses focusing parameter.

$$L_{focal} = -\sum_{c=1}^C \alpha_c (1 - \rho_{t,c})^\gamma \log \log (\rho_{t,c}) \tag{49}$$

The combined loss function integrates multi-class Dice and Focal Loss to ensure the model pays attention to both classes overlap and hard-to-classify examples. This approach is especially valuable in medical imaging applications where each class requires distinct treatment. Equation (50) shows the combined loss function, in which λ_1 and λ_2 stands for weights controlling the contribution of each loss term. Figure 2 represents the architecture of suggested Channel-Aware DeepNet.

$$L_{combined} = \lambda_1 \times L_{dice} + \lambda_2 \times L_{focal} \tag{50}$$

Performance metrics

Several performance metrics are used in this research to verify the competence of the proposed in detecting ADHD as shown in Table 5.

Result and discussion

Experimental setup

The proposed approach has been implemented in Python. The proposed approach is validated through two databases, namely ADHD diagnosis database (<https://www.kaggle.com/datasets/arashnic/adhd-diagnosis-data>) and EEG database (<https://www.kaggle.com/datasets/inancigdem/eeeg-data-for-mental-attention-state-detectio-n>). The simultaneous input of EEG data and diagnostic data corresponding to the same patients and controls is vital for ensuring that the model can effectively identify and compare the unique patterns associated with ADHD. This alignment enables the model to leverage multimodal data in a way that maximizes its diagnostic capabilities, offering a more comprehensive and accurate method for ADHD detection. In both the databases, the proposed approach and existing approaches like Att-AENet¹⁸, SVM¹⁹, 2D-CNN²², and FCNN²³ are compared with respect to the metrics such as precision, NPV, FNR, recall, FPR, accuracy, F-score, MCC, and specificity to evaluate their performance in terms of ADHD detection.

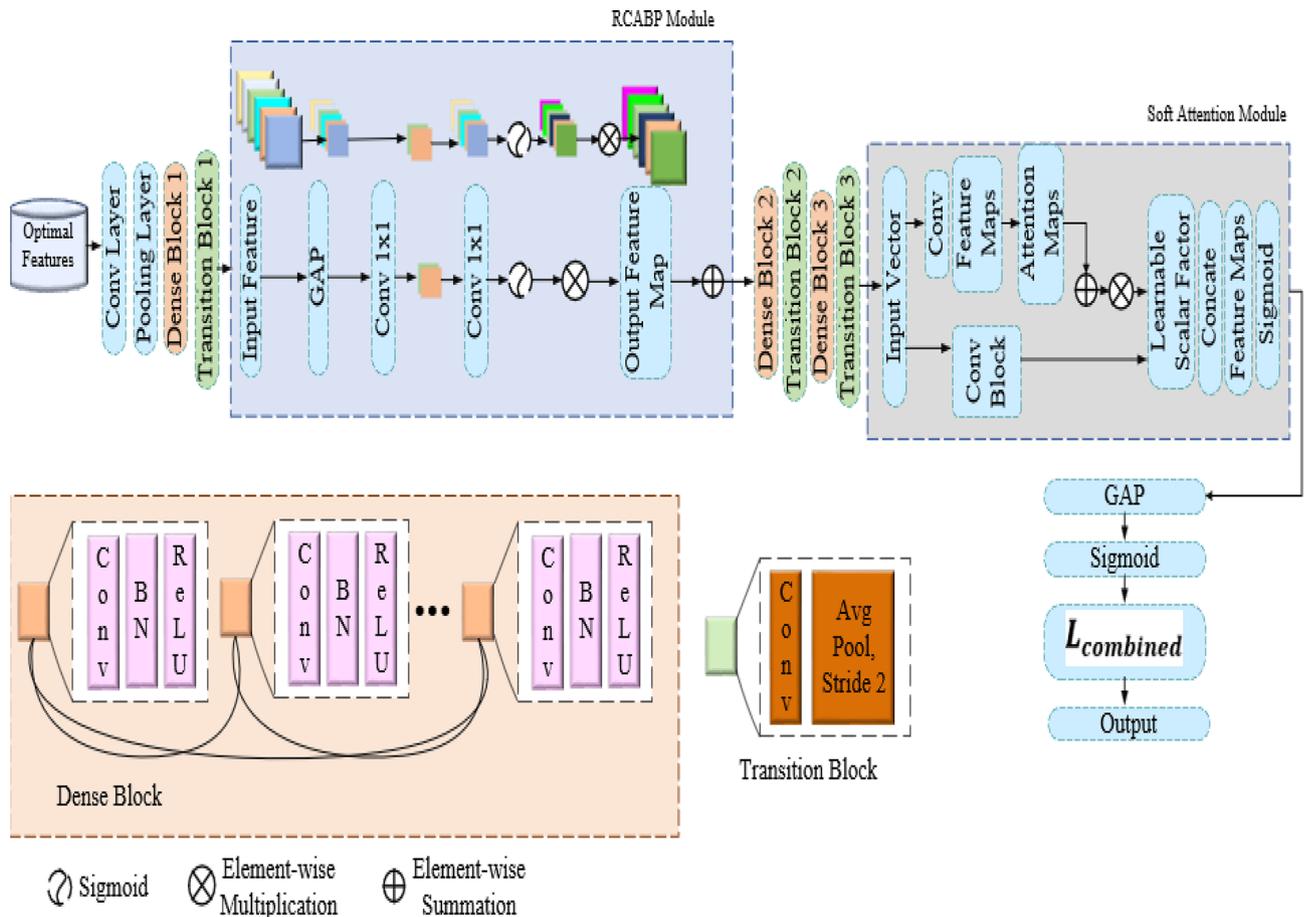


Fig. 2. displays the architecture of proposed Channel-Aware DeepNet.

Metrics	Formula	Description
Accuracy	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$	Percentage of true positive and true negative predictions that are accurate out of all the predictions
Precision	$Precision = \frac{TP}{TP+FP}$	Percentage of actual positive predictions among all of the model's positive predictions
Recall	$Recall = \frac{TP}{TP+FN}$	Percentage of accurate positive predictions among all real positive cases in the data
Specificity	$Specificity = \frac{TN}{FP+TN}$	Percentage of accurate negative predictions among all actual negative cases in the data
F1-Score	$F1 - Score = 2 \cdot (Precision \cdot Recall) / (Precision + Recall)$	Harmonic mean of recall and specificity
MCC	$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(FP+TP)(FN+TP)(TN+FP)(TN+FN)}}$	Accounts for both true and false positives and negatives, is a metric used to assess the validity of binary classifications
NPV	$NPV = \frac{TN}{FN+TN}$	Calculates the percentage of actual negative predictions among all of the model's negative predictions
FPR	$FPR = \frac{FP}{TN+FP}$	Percentage of true negative cases that the model mistakenly identified as positive
FNR	$FNR = \frac{FN}{TP+FN}$	Percentage of true positive cases that the model misclassified as negative

Table 5. Performance metrics.

Approaches	Accuracy	Precision	Recall	F1-Score	Specificity	MCC	NPV	FPR	FNR
Proposed	0.9852	0.9827	0.9805	0.9826	0.9816	0.9632	0.9805	0.0174	0.0195
Att-AENet ¹⁸	0.9691	0.9616	0.9772	0.9611	0.9693	0.9384	0.9769	0.0389	0.0228
SVM ¹⁹	0.9686	0.9594	0.9782	0.9591	0.9687	0.9374	0.978	0.0409	0.0218
2D-CNN ²²	0.9675	0.9594	0.9761	0.959	0.9677	0.9352	0.9759	0.041	0.0239
FCNN ²³	0.9697	0.9596	0.9804	0.959	0.9699	0.9396	0.9802	0.041	0.0196

Table 6. Performance analysis based on ADHD diagnosis database.

Approaches	Accuracy	Precision	Recall	Specificity	F1-Score	MCC	NPV	FPR	FNR
Att-AENet ¹⁸	0.9583	0.9478	0.9695	0.9473	0.9585	0.9169	0.9692	0.0527	0.0305
SVM ¹⁹	0.954	0.9457	0.9632	0.9448	0.9544	0.9081	0.9625	0.0552	0.0368
2D-CNN ²²	0.9491	0.9412	0.9576	0.9407	0.9494	0.8984	0.9572	0.0593	0.0424
FCNN ²³	0.9469	0.9411	0.9534	0.9405	0.9472	0.894	0.9529	0.0595	0.0466
Proposed	0.9789	0.9784	0.9794	0.9783	0.9789	0.9578	0.9794	0.0217	0.0206

Table 7. Performance analysis based on EEG database.

Performance analysis based on ADHD diagnosis database

The performance of the proposed method is compared with other techniques such as Att-AENet¹⁸, SVM¹⁹, 2D-CNN²², and FCNN²³ based on the ADHD diagnosis database is presented in Table 6. Precision, NPV, FNR, recall, FPR, accuracy, F-score, MCC, and specificity are the evaluation metrics used in this analysis. The accuracy results show that the proposed approach outperforms the existing ones, especially because of the newly developed RFO. The accuracy of the proposed model is 98.52%, which is higher than the existing techniques like FCNN²³ with 96.97% and Att-AENet¹⁸ with 96.91%. The suggested method also outshines the existing methods in terms of recall and MCC metrics as well. It obtained a recall of 98.05% and in the MCC metrics, the proposed model attained 96.32%, which is higher than the 93.96% of FCNN²³ and 93.52% of 2D-CNN²². Furthermore, the suggested approach attained the FNR of 0.0195, which is lower than the others. This analysis also shows that the proposed method has better results than the previous methods used in detecting ADHD due to the introduction of the novel ADHD-AttentionNet.

Performance analysis based on EEG database

In Table 7, the performance analysis between the stated and current methods such as Att-AENet¹⁸, SVM¹⁹, 2D-CNN²², and FCNN²³ are evaluated based on the EEG database. The evaluation measures used in this evaluation includes precision, NPV, FNR, recall, FPR, accuracy, F-score, MCC, and specificity. The accuracy results reinforce the advantage of the proposed approach that attained accuracy of 97.89% compared to other techniques with DCNN¹⁶ having the nearest accuracy of 95.83%. The proposed model also performs well in precision metric with the highest precision rate of 97.84% and other approaches not even close to 95%, this major performance difference is due to the new NeuroDCT-ICA. In terms of specificity the stated approach achieved

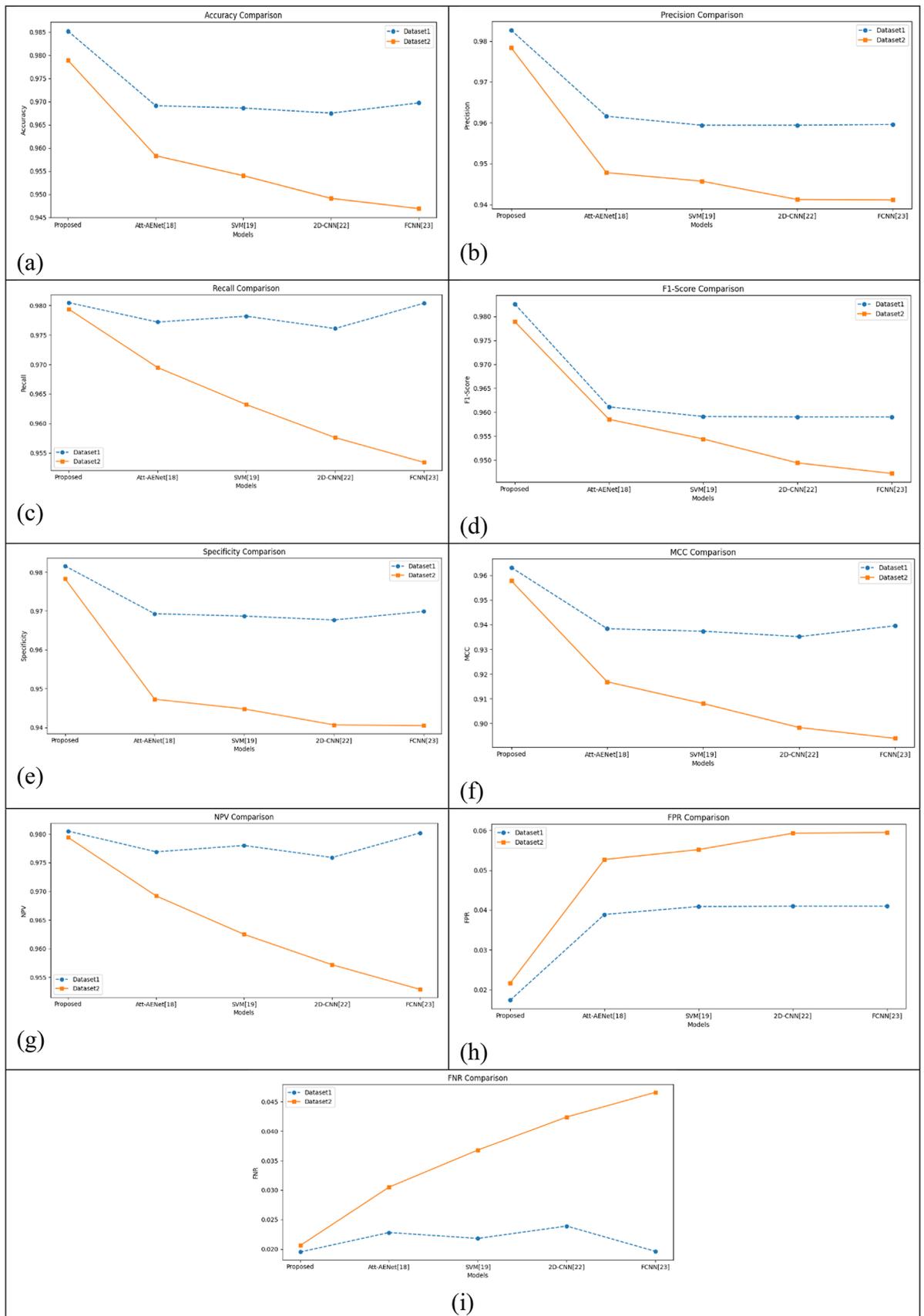


Fig. 3. Graphic representation of (a) Accuracy, (b) Precision, (c) Recall, (d) F1-score, (e) Specificity, (f) MCC, (g) NPV, (h) FPR, (i) FNR for proposed and existing models based on the two databases.

Approach	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean Accuracy
Dataset 1						
Proposed	0.9852	0.9835	0.985	0.986	0.987	0.9852
Att-AENet ¹⁸	0.9691	0.97	0.9672	0.9688	0.97	0.9691
SVM ¹⁹	0.9686	0.97	0.965	0.9678	0.9682	0.9686
2D-Convolutional CNN ²²	0.9675	0.965	0.966	0.968	0.967	0.9675
FCNN ²³	0.9697	0.9675	0.97	0.9689	0.9703	0.9697
Dataset 2						
Proposed	0.9789	0.9792	0.9794	0.979	0.9794	0.9792
Att-AENet ¹⁸	0.9583	0.96	0.959	0.9605	0.958	0.9592
SVM ¹⁹	0.954	0.953	0.9565	0.9548	0.9525	0.9542
2D-Convolutional CNN ²²	0.9491	0.9478	0.95	0.948	0.9492	0.949
FCNN ²³	0.9469	0.9475	0.9489	0.948	0.9478	0.9478

Table 8. Performance analysis on K-fold cross-validation (5 folds) for different datasets.

Approach	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean accuracy
Proposed	0.9852	0.9835	0.985	0.986	0.987	0.9852
VMD-HT ³¹	0.9691	0.97	0.9672	0.9688	0.97	0.9691
EMD-DWT ²⁸	0.9686	0.97	0.965	0.9678	0.9682	0.9686
CNN-AE ³²	0.9675	0.965	0.966	0.968	0.967	0.9675

Table 9. Performance analysis on K-fold cross-validation (5 folds) for dataset 1 on another existing models.

Approach	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean accuracy
Proposed	0.9789	0.9792	0.9794	0.979	0.9794	0.9792
VMD-HT ³¹	0.9583	0.96	0.959	0.9605	0.958	0.9592
EMD-DWT ²⁸	0.954	0.953	0.9565	0.9548	0.9525	0.9542
2CNN-AE ³²	0.9491	0.9478	0.95	0.948	0.9492	0.949

Table 10. Performance analysis on K-fold cross-validation (5 folds) for dataset 2 on another existing models.

highest with 97.83% and meanwhile SVM¹⁹ with 94.48%. Furthermore, the proposed approach provided the lowest FPR of 0.0217, while FCNN²³ had the highest FPR of 0.0595. These results demonstrate the superiority of the proposed model over the previous methods in the performance of ADHD detection. The comparison analysis between the suggested and current approaches are graphically represented in Fig. 3.

K-fold cross validation

K-Fold Cross Validation divides the dataset into ‘K’ equal subsets or folds in order to evaluate the performance of a ML model. The method is repeated ‘K’ times, each time using a different fold as the test set. The model is trained on ‘K-1’ folds and tested on the remaining fold. Over all ‘K’ iterations, the ultimate performance metric is averaged. This approach offers a more accurate measure of model performance while assisting in the reduction of overfitting.

The performance study of several methods employing K-fold cross-validation with five folds for two distinct datasets is shown in Table 8. The Proposed approach outperforms other models, including Att-AENet (mean accuracy 0.9691), SVM (mean accuracy 0.9686), 2D-Convolutional CNN (mean accuracy 0.9675), and FCNN (mean accuracy 0.9697), with the greatest mean accuracy of 0.9852 for Dataset (1) With a mean accuracy of 0.9792, the Proposed approach consistently outperforms Att-AENet(mean accuracy 0.9592), SVM (mean accuracy 0.9542), 2D-Convolutional CNN (mean accuracy 0.949), and FCNN (mean accuracy 0.9478) for Dataset (2) These outcomes demonstrate the Proposed method’s efficacy in comparison to other state-of-the-art techniques by highlighting its consistent and superior performance across both datasets.

Another comparative analysis of different existing models is provided with the K-fold cross validation for two datasets which is compared with several existing models such as variational mode decomposition-Hilbert transform (VMD-HT), Empirical mode decomposition-discrete wavelet Transform (EMD-DWT), CNN-Autoencoder (CNN-AE. The analysis using K-fold cross-validation is represented in Tables 9 and 10.

The performance study of several methods utilizing K-fold cross-validation with five folds on Dataset 1 is shown in Table 8. The accuracy of the Proposed approach, VMD-HT, EMD-DWT, CNN-AE is compared over the five folds in the table. The accuracy of the Proposed approach is consistently high; Fold 5 has the greatest accuracy of 0.987, with a mean accuracy of 0.9852. The EMD-DWT approach yields a mean accuracy

Dataset	Method	Processing time (s)	Memory usage (MB)
ADHD diagnosis database	Proposed	0.8	110
	Att-AENet ¹⁸	2.0	250
	SVM ¹⁹	2.5	300
	2D-Convolutional CNN ²²	2.3	280
	FCNN ²³	1.8	270
EEG database	Proposed	1.0	120
	Att-AENet ¹⁸	2.3	210
	SVM ¹⁹	1.5	150
	2D-Convolutional CNN ²²	2.6	220
	FCNN ²³	2.4	200

Table 11. Processing time and memory usage.

of 0.9686, whereas the VMD-HT method performs somewhat worse with a mean accuracy of 0.9691. The CNN-AE approach has a mean accuracy of 0.9675 and consistently performs well. A mean accuracy of 0.9697 indicates that the 3D-CNN model performs well, surpassing both VMD-HT and EMD-DWT. According to these findings, CNN-AE and EMD-DWT show somewhat lower accuracies, whereas the Proposed technique perform best overall across all folds. The data provided demonstrates that the suggested model provides the best accuracy and stability throughout the dataset's many folds.

The performance study of various methods utilizing K-fold cross-validation with five folds on Dataset 2 is shown in Table 9. All folds show consistently high accuracy using the Proposed approach, with Fold 5 producing the highest value of 0.9794, for a mean accuracy of 0.9792. While the EMD-DWT method has a mean accuracy of 0.9542, suggesting slightly weaker performance, the VMD-HT approach performs slightly worse with a mean accuracy of 0.9592. The 2CNN-AE method performs more modestly with a mean accuracy of 0.949. These results reveal that the Proposed technique performs better than the other models on Dataset 2 in terms of overall accuracy and consistency, while the other methods, including VMD-HT and EMD-DWT, perform similarly but comparatively poorly. Table 11 shows the processing time and memory usage for both the datasets.

According to Table 10, the proposed method demonstrates the shortest processing time and lowest memory usage across both datasets, with times of 0.8 s (ADHD) and 1.0 s (EEG) and memory usage of 110 MB (ADHD) and 120 MB (EEG). In contrast, Att-AENet uses a lot more memory (250 MB and 210 MB, respectively) and takes a lot longer (2.0 s for ADHD and 2.3 s for EEG). SVM uses comparatively less memory (150 MB in EEG) but has the longest processing time (2.5 s) in the ADHD dataset. Additionally, compared to the suggested technique, the 2D-Convolutional CNN and FCNN exhibit longer processing times and higher memory usage. The outcomes demonstrate how effectively the suggested model uses time and memory.

Conclusion

In this research, DL based ADHD detection model has been introduction with a primary aim of an approach with higher accuracy, effective data handling, and minimal complexity. To attain these goals, the work majorly presents a new module called NeuroDCT-ICA which was aimed at preprocessing raw EEG data and filtering out noise as well as extracting necessary features. It also included a new RFO optimization approach for feature selection to improve data analysis speed and system reliability. At the center of this is the ADHD-AttentionNet, a DL model that enhances the detection accuracy of approach. The proposed approach was implemented in Python, the validation results stated the outperformance of the suggested approach. With accuracy of 98.52%, precision of 98.27% and recall of 98.04%. ADHD detection's future is in the use of wearables for continuous tracking, creating individualized and multiple approach diagnostic models, and the use of DL techniques to increase detection accuracy. Cultural models, detection tools and Explainable AI will advance the availability, clinical utilisation and dependability of ADHD diagnosis.

Data availability

Database 1: Diagnostic database (<https://www.kaggle.com/datasets/arashnic/adhd-diagnosis-data>) is a public dataset that includes health, activity, and heart rate data from adult patients diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). It contains data from 51 ADHD patients and 52 clinical controls. Database 2: EEG database (<https://www.kaggle.com/datasets/inancigdem/eeg-data-for-mental-attention-state-detectio-n>) is a collection of 34 experiments for monitoring the attention state in humans using passive EEG-based Brain-Computer Interface (BCI).

Received: 2 December 2024; Accepted: 14 February 2025

Published online: 26 February 2025

References

- Zhang, K. F. et al. Fusion of Multi-Task Neurophysiological Data to Enhance the Detection of Attention- Deficit/Hyperactivity Disorder, *IEEE J. Transl. Eng. Health Med.* **12**, 668–674. <https://doi.org/10.1109/JTEHM.2024.3435553>. (2024).
- Tawhid, M. N. A., Siuly, S., Wang, K. & Wang, H. GENet: A Generic Neural Network for Detecting Various Neurological Disorders From EEG, in *IEEE Trans. Cognit. Dev. Syst.* **16**(5), 1829–1842. <https://doi.org/10.1109/TCDS.2024.3386364> (2024).

3. Latifi, B. & Amini, A. and Ali Motie Nasrabadi. Siamese based deep neural network for ADHD detection using EEG signal. *Comput. Biol. Med.*, **182**. (2024).
4. Sanchis, J., García-Ponsoda, S., Teruel, M. A., Il-Yeol & Song Juan Trujillo, and A novel approach to identify the brain regions that best classify ADHD by means of EEG and deep learning. *Heliyon* **10** (4). (2024).
5. Bakhtyari, M. & Mirzaei, S. ADHD detection using dynamic connectivity patterns of EEG data and ConvLSTM with attention framework. *Biomed. Signal Process. Control*. **76**, (2022).
6. TaghiBeyglou, B., Shahbazi, A., Bagheri, F., Akbarian, S. & Jahed, M. Detection of ADHD cases using CNN and classical classifiers of raw EEG. *Comput. Methods Progr. Biomed. Update* **2**. (2022).
7. Cura, O., Karabiber, A., Akan & Sibel Kocaaslan, A. Detection of attention deficit hyperactivity disorder based on EEG feature maps and deep learning. *Biocybern. Biomed. Eng.* **44** (3), 450–460. (2024).
8. Ali, A. Y., Inyang, B., Koshy, F. S., George, K., Poudel, P., Chalasani, R., Mohammed, L. Elements that influence the development of attention deficit hyperactivity disorder (ADHD) in children. *Cureus* **14**(8). (2022).
9. Antoniou, E., Rigas, N., Orovou, E. & Papatrechas, A. ADHD and the importance of comorbid disorders in the psychosocial development of children and adolescents. *J. Biosci. Med.* **9** (04), 1 (2021).
10. Bakhtyari, M. & Mirzaei, S. ADHD detection using dynamic connectivity patterns of EEG data and ConvLSTM with attention framework. *Biomed. Signal Process. Control*. **76**, 103708 (2022).
11. Das, W. & Khanna, S. A robust machine learning based framework for the automated detection of ADHD using pupillometric biomarkers and time series analysis. *Sci. Rep.* **11** (1), 16370 (2021).
12. Joy, R. C., George, S. T., Rajan, A. A., Subathra, M. S. P., Sairamya, N. J., Prasanna, J., ... Al-Andoli, M. N. (2022). Detection and Classification of ADHD from EEG Signals Using Tunable Q-Factor Wavelet Transform. *J. Sens.* **2022**(1), 3590973. (2022).
13. Riaz, A., Asad, M., Alonso, E. & Slabaugh, G. DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI. *J. Neurosci. Methods*. **335**, 108506 (2020).
14. Kaimara, P. et al. Waking up in the morning (WUIM): A transmedia project for daily living skills training. *Technol. Disabil.* **33** (2), 137–161 (2021).
15. Wright, A. J. Psychological and neuropsychological underpinnings of attention-deficit/hyperactivity disorder assessment. *Clin. Child Psychol. Psychiatry*. **26** (3), 783–794 (2021).
16. Ariyaratne, G., De Silva, S., Dayarathna, S., Meedeniya, D. & Jayarathne, S. ADHD identification using convolutional neural network with seed-based approach for fMRI data. In *Proceedings of the 2020 9th international conference on software and computer applications* (31–35). (2020).
17. Liu, S., Zhao, L., Zhao, J., Li, B. & Wang, S. H. Attention deficit/hyperactivity disorder classification based on deep spatio-temporal features of functional magnetic resonance imaging. *Biomed. Signal Process. Control*. **71**, 103239 (2022).
18. Chen, Y., Gao, Y., Jiang, A., Tang, Y. & Wang, C. ADHD classification combining biomarker detection with attention auto-encoding neural network. *Biomed. Signal Process. Control*. **84**, 104733 (2023).
19. Liu, G., Lu, W., Qiu, J. & Shi, L. Identifying individuals with attention-deficit/hyperactivity disorder based on multisite resting-state functional magnetic resonance imaging: a radiomics analysis. *Hum. Brain. Mapp.* **44** (8), 3433–3445 (2023).
20. Hamed, N., Khadem, A., Delrobaei, M. & Babajani-Feremi, A. Detecting ADHD based on brain functional connectivity using resting-state MEG signals. *Front. Biomed. Technol.*. (2022).
21. Lohani, D. C. & Rana, B. ADHD diagnosis using structural brain MRI and personal characteristic data with machine learning framework. *Psychiatry Res. Neuroimaging*. **334**, 111689 (2023).
22. Uyulan, C. et al. A class activation map-based interpretable transfer learning model for automated detection of ADHD from fMRI data. *Clin. EEG Neurosci.* **54** (2), 151–159 (2023).
23. Gülhan, P. G. & Özmen, G. The use of FMRI regional analysis to automatically detect ADHD through a 3D CNN-based approach. *J. Imaging Inf. Med.* 1–14. (2024).
24. Maniruzzaman, M., Shin, J., Hasan, M. A. M. & Yasumura, A. Efficient feature selection and machine learning based ADHD detection using EEG signal. *Comput. Mater. Continua*. **72**(3). (2022).
25. Zhou, X. et al. Multimodal MR images-based diagnosis of early adolescent attention-deficit/hyperactivity disorder using multiple kernel learning. *Front. NeuroSci.* **15**, 710133 (2021).
26. Khullar, V., Salgotra, K., Singh, H. P. & Sharma, D. P. Deep learning-based binary classification of ADHD using resting state MR images. *Augment. Hum. Res.* **6** (1), 5 (2021).
27. Gajbhiye, P. et al. Wavelet domain optimized Savitzky–Golay filter for the removal of motion artifacts from EEG recordings. *IEEE Trans. Instrum. Meas.* **70**, 1–11 (2020).
28. Tor, H. T., Ooi, C. P., Lim-Ashworth, N. S., Wei, J. K. E., Jahmunah, V., Oh, S. L., Fung, D. S. S. Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals. *Comput. Methods Progr. Biomed.* **200**, 105941. (2021).
29. Wiebe, A., Aslan, B., Brockmann, C., Lepartz, A., Dudek, D., Kannen, K., Braun, N. Multimodal assessment of adult attention-deficit hyperactivity disorder: a controlled virtual seminar room study. *Clin. Psychol. Psychother.* **30**(5), 1111–1129. (2023).
30. Hamim, M. A., Tanmoy, F. M., Tasfia, O. & Juthi, F. A. Attention deficit hyperactivity disorder detection using deep learning approach. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* IEEE, 1–7. (2023).
31. Khare, S. K. & Acharya, U. R. An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals. *Comput. Biol. Med.* **155**, 106676 (2023).
32. Shoebi, A., Ghassemi, N., Khodatars, M., Moridian, P., Khosravi, A., Zare, A., Rajendra Acharya, U. Automatic diagnosis of schizophrenia and attention deficit hyperactivity disorder in rs-fMRI modality using convolutional autoencoder model and interval type-2 fuzzy regression. *Cognit. Neurodyn.* **17**(6), 1501–1523. (2023).
33. Barua, P. D. et al. Tmp19: a novel ternary motif pattern-based ADHD detection model using Eeg signals. *Diagnostics* **12** (10), 2544 (2022).
34. Maniruzzaman, M., Hasan, M. A. M., Asai, N. & Shin, J. Optimal channels and features selection based ADHD detection from EEG signal using statistical and machine learning techniques. *IEEE Access*. **11**, 33570–33583 (2023).

Acknowledgements

The authors extend their appreciation to the King Salman center For Disability Research for funding this work through Research Group no KSRG-2024-475.

Author contributions

All authors have contributed equally.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.K.S. or G.G.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025