
Research article

Diagnosis of attention deficit hyperactivity disorder: A deep learning approach

Nizar Alsharif^{1,2}, Mosleh Hmoud Al-Adhaileh^{1,3}, and Mohammed Al-Yaari^{1,4,*}

¹ King Salman Center for Disability Research, Riyadh 11614, Saudi Arabia

² Department of Computer Engineering and Science, Albaха University, Albaха 42331, Saudi Arabia

³ Deanship of e-learning and information technology, King Faisal University, P.O. Box 4000, Al-Ahsa 31982, Saudi Arabia

⁴ Chemical Engineering Department, King Faisal University, P.O. Box 380, Al-Ahsa 31982, Saudi Arabia

* Correspondence: Email: malyaari@kfu.edu.sa; Tel: +966135898583.

Abstract: In recent years, there has been significant interest in the analysis and classification of brain dis-orders using electroencephalography (EEG). We presented machine learning and deep learning (DL) frameworks that integrate an EEG-based brain network with various DL models to diagnose attention deficit hyperactivity disorder (ADHD). By incorporating an objective biomarker into the diagnostic process, the accuracy and effectiveness of diagnosis could be enhanced. We used public EEG datasets from 61 ADHD youngsters and 60 normally developing children. The raw EEG data underwent preprocessing, including the application of filters in clinically relevant frequency bands and notch filters. From the preprocessed EEG segments, statistical features (e.g., standard deviation, kurtosis) and spectral features (e.g., entropy) were extracted. Principal component analysis (PCA) and chi-square with PCA were used as feature selection methods to obtain the most useful features and keep them. The machine learning models achieved the highest accuracy result of 94.86% by utilizing support vector machines (SVM) with PCA features. Furthermore, integrating models combining a convolutional neural network (CNN) with bidirectional long short-term memory (BiLSTM) networks, and gated recurrent unit-Transformer (GRU-Transformer block) with Chi-square and PCA features achieved accuracies of 94.50% and 95.59%, respectively. The suggested framework demonstrated a wide range of applicability in addressing the identification of ADHD. To evaluate the performance of the proposed models, comparisons were made with existing models, and the proposed system exhibited superior performance. We enhanced EEG-based analysis and categorization of ADHD by

demonstrating the capabilities of advanced artificial intelligence models in enhancing diagnostic accuracy and efficacy.

Keywords: deep learning; machine learning; attention deficit hyperactivity disorder; electroencephalogram

Mathematics Subject Classification: 68Q32

1. Introduction

Attention deficit hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders, estimated to affect 5–7% of children worldwide [1]. Impaired functioning is characterized by major symptoms such as inattention, hyperactivity, impulsivity, and emotional dysregulation. A youngster with this illness also has trouble regulating their impulses and focusing on a single activity at a time. The timely and accurate diagnosis of ADHD is critical for early intervention and treatment access, which can improve long-term outcomes [2]. However, diagnosis remains challenging, often based on subjective behavioral observations due to the lack of reliable biomarkers [3]. Emerging research has sought to identify neurophysiological features that could serve as objective indicators of ADHD pathology. Electroencephalography (EEG) is a non-invasive technique that directly measures electrical brain activity and has shown promise for probing neural dysfunction in ADHD [4]. Prior studies have reported EEG abnormalities in ADHD children, including altered theta/beta ratio [5], reduced P300 event-related potential, and EEG power differences [6]. Advanced analytical approaches, such as machine learning applied to EEG data, may enable robust classification of ADHD.

The significances of a delayed diagnosis and treatment for ADHD may have detrimental effects on individuals, potentially leading to the development of more extensive mental health disorders, difficulties in interpersonal relationships and work, engagement in criminal behaviors, and the abuse of substances. The detrimental consequences of untreated ADHD have been extensively reported, exhibiting poor impacts on scholastic achievements [7], social interactions [8], occupational prospects [9], and overall mortality rates [10].

The use of ML approaches for the diagnosis of ADHD in individuals aged 17 years and above is a contemporary strategy for addressing this particular concern. Knowledge-based systems are frequently employed in medical environments where there is a substantial need for interpretability. These systems strive to explicitly represent knowledge by utilizing tools such as production or if-then rules. This enables the system to engage in reasoning processes to arrive at conclusions and offer explanations of its rationale to the user [11]. A hybrid approach was employed to leverage the advantages of machine learning-based methodologies while maintaining the interpretability of knowledge-based systems [12]. This approach integrates patterns derived from machine-learning algorithms with the expertise provided by clinicians, resulting in a unified framework that optimally combines both approaches.

1.1. Contribution

ADHD is a complex disorder characterized by a diverse array of symptoms, as mentioned previously. Timely intervention and accurate identification offer the potential to modify neural connections and enhance symptomatology. However, due to the multifaceted nature of ADHD, the presence of co-occurring disorders, and a global shortage of diagnostic professionals, the identification

of ADHD is frequently delayed. Therefore, it is crucial to explore alternative approaches to enhance the effectiveness of early detection, such as leveraging deep learning techniques. These techniques have the potential to augment existing diagnostic methods and contribute to more efficient and timely identification of ADHD.

This system aims to automate ADHD classification using EEG. The method employs neural signal analysis to discriminate ADHD from typically developing children through rigorous automated classification of ADHD. The successful implementation of this framework has the potential to significantly aid clinical diagnosis and facilitate early access to treatment for individuals with ADHD.

Furthermore, this work contributes to the ongoing efforts to enhance ADHD diagnostics and deepen our understanding of the underlying neural mechanisms of the disorder. To address the research gaps identified in previous investigations, an integrated system was devised, combining a CNN with BiLSTM networks, along with a gated recurrent unit-transformer (GRU-Transformer block). This integrated system aims to improve upon existing approaches and advance the field of ADHD classification.

By utilizing deep learning techniques and developing an EEG-based framework, we strive to enhance the accuracy and efficiency of ADHD identification, ultimately benefiting individuals with ADHD and the clinical community.

1.2. Background of research

The National Survey on Children's Health (NSCH) during 2016–2017 found that 6,630 children (ranging in age from 3 to 17) were officially diagnosed with ADHD. It is worth noting that children, on average, are diagnosed with ADHD at the age of 12.4 years. In the realm of research, scientists have employed diverse machine-learning classifiers [13–18] to identify children who might be at risk of developing ADHD.

Uluyagmur-Ozturk and colleagues [19] conducted a study investigating the relationship between psychological health and diagnoses of autism spectrum disorder (ASD) and ADHD among young people in Turkey. The study included 61 participants sourced from the Marmara University Medical Center. This diverse group consisted of 30 children diagnosed with ADHD, 18 classified with ASD, and 13 typically developing children, with ages ranging from 9.22 to 10.50 years. To classify individuals into the respective groups, the researchers employed five machine-learning approaches: decision tree (DT), random forest (RF), SVM, k-nearest neighbor (KNN), and AdaBoost (AB) algorithms. Remarkably, the AB algorithm achieved an accuracy of 80%, demonstrating its effectiveness in classification tasks.

Slobodin et al. [20] conducted a study utilizing a continuous performance test (CPT) to examine signs of ADHD in children and adolescents. The study involved 458 participants aged 6 to 12 years, who were divided into different age groups. The mean age of the group consisting of children and adolescents was 8.7 years, with a standard deviation of 1.8 years. Among the children, 46.51% were diagnosed with ADHD. Interestingly, statistical analysis revealed no significant age difference between individuals with ADHD and those without the condition (p -value = 0.94). Various machine learning classifiers were employed to generate predictions for ADHD. The training set, which constituted 60% of the data, was used to train the classifiers, while the remaining 40% served as a test set to evaluate classifier performance. The machine learning-based classifiers demonstrated impressive levels of accuracy (87.0%), sensitivity (89.0%), and specificity (84.0%).

Morrow et al. [21] conducted a research study focusing on examining the impact of therapy on children and adolescents diagnosed with ADHD. The study employed four distinct machine learning classifiers to identify relevant characteristics of children seeking medical treatment for ADHD.

Specifically, Classification and Regression Trees (CART), Logistic Regression (LR), Endpoint Detection and Response (EDR), and Deep Net classifiers were utilized. Notably, the Deep Net-based classifier demonstrated superior performance with a robust area under the curve (AUC) value of 0.72, outperforming CART, EDR, and LR classifiers.

The utilization of EEG data as a diagnostic tool for ADHD remains a topic of debate, primarily due to the lack of standardized criteria for its usage and interpretation. Moghaddari et al. [22] addressed this issue by implementing a CNN model to diagnose ADHD in children based on EEG readings. A total of 61 participants took part in the study, with 31 diagnosed with ADHD and 30 considered to be typically developing. Prior to analysis, the collected EEG signals underwent a preprocessing step to eliminate unwanted artifacts and background noise. Subsequently, the EEG data was transformed into RGB pictures using a conversion process, which was then processed by a 13-layer CNN. The model's performance was evaluated using a 5-fold cross-validation procedure, resulting in an impressive average accuracy of 99.06% for the model validation.

Tosun et al. [23] employed LSTM, power spectral density (PSD), and spectral entropy (SE). The suggested method for conducting research was evaluated using an 80:20 hold-out validation strategy, which resulted in an accuracy of 92.15%.

In their study, Khoshnoud et al. [24] investigated a group of young individuals diagnosed with ADHD, employing nonlinear EEG analytic techniques for their analysis. The researchers utilized two measures, namely the maximum value of the Lyapunov exponent (LLE) and the approximation entropy (ApEn), to evaluate the nonlinear characteristics of the EEG signals. To facilitate the evaluation process, a probabilistic neural network (PNN) was employed, leading to a classification accuracy of 87.5%.

Chen et al. [25] undertook the development of a deep learning model with the purpose of accurately identifying individuals with ADHD within the pediatric population. The study involved the translation of raw EEG data into visual representations, which were subsequently inputted into a CNN model. The dataset utilized in their work consisted of EEG recordings collected from a representative group of children and adolescents. Specifically, the dataset included a total of 102 participants, comprising 51 typically developing children and 51 children diagnosed with ADHD. Notably, the deep learning model achieved an impressive accuracy rate of 94.67%.

Tenev et al. [26] presented a technique to speed up the process of detecting and classifying individuals who have been diagnosed with ADHD. Their study included 117 individuals: 67 diagnosed with ADHD, and 50 healthy controls. The EEG signals of the participants were analyzed. To conduct the data analysis, they made use of SVM in addition to a voting mechanism. The suggested approach had a significant overall effectiveness of 82.3%.

Saini et al. [27] used EEG data for testing ML models for predicting ADHD. The featured selection in the given model was done using PCA approach, while the data classification was done using KNN. With EEG data collected from 77 ADHD children and 80 typically developing children.

Dubreuil-Vall et al. [28] conducted a study involving EEG data collection from a group of 40 volunteers, including 20 individuals diagnosed with ADHD and 20 healthy controls. The study aimed to develop a procedure for identifying ADHD using EEG data. The researchers utilized spectrogram visuals generated from the EEG data as inputs for a CNN model. The findings of the study revealed an accuracy of 88% for data classification.

Tor et al. [29] used a mix of nonlinear features, empirical mode decomposition (EMD), and discrete wavelet transform (DWT) decomposition methods for predicting ADHD. They analyzed EEG data obtained from a group of 123 children and teenagers. A subgroup representing 45% exhibited

symptoms of ADHD. By leveraging the aforementioned analytical approaches, the researchers aimed to differentiate and classify individuals based on their specific diagnoses.

In a recent study conducted by Loh et al. [30], a comprehensive analysis was undertaken to explore and compare various computerized diagnostic approaches for ADHD. The researchers delved into published research encompassing these different techniques, aiming to gain a thorough understanding of their strengths and limitations. By conducting a comparative analysis, the study shed light on the distinctions and similarities among the evaluated approaches, providing valuable insights into the advancements made in the field of computerized ADHD.

The aforementioned techniques were devised to automate the diagnosis of ADHD via the analysis of physiological data and photographic documentation. Table 1 presents a summary of the previous study focusing on the detection of ADHD.

Table 1. Previous studies for detecting ADHD.

Author /Year	Models	Dataset size	Accuracy %
Christiansen et al. (2020) [31]	Decision tree Logistic regression (LR)	385 samples 200 samples	80.0% 67.0% cross validation
Sato et al. (2012) [32]	SVM	200 samples	68.6% cross validation
Tan et al. (2017) [33]	SVM	200 samples	81.0% training and testing
Khan et al. (2021) [34]	SVM	659 samples	85.3% LOOCV
Sun et al. (2020) [35]	SVM	781 samples	90.2% LOOCV
Peng et al. (2013) [36]	ELM	153 samples	88.9% training and testing
Vaidya et al. (2019) [37]	SVM	1012 samples (ADHD, and ASD)	97.6% LOOCV
Tang et al. (2020) [38]	DT	152 samples	94.0% cross-validation
Chen et al. (2019) [39]	CNN	101 samples	81.0% LOOCV
Kim et al. (2021) [40]	SVM	79 samples	

2. Materials and methods

We implement a rigorous methodology pipeline to classify ADHD from healthy and controls using EEG data. EEG provides a direct measure of cortical activity and has shown utility as a biomarker for various neurological and psychiatric conditions.

The proposed framework includes critical steps for robust EEG-based classification, including preprocessing, feature extraction, feature selection, classifier optimization, and performance evaluation. Preprocessing involves filtering and artifact removal to isolate neural signals. Informative features that capture distinguishing characteristics are then extracted from the clean EEG data, as shown in Figure 1. Feature selection serves to eliminate redundant variables and improve

generalizability. The chosen features were utilized to train a machine learning classifier with the objective of accurately classifying subjects as either ADHD or healthy.

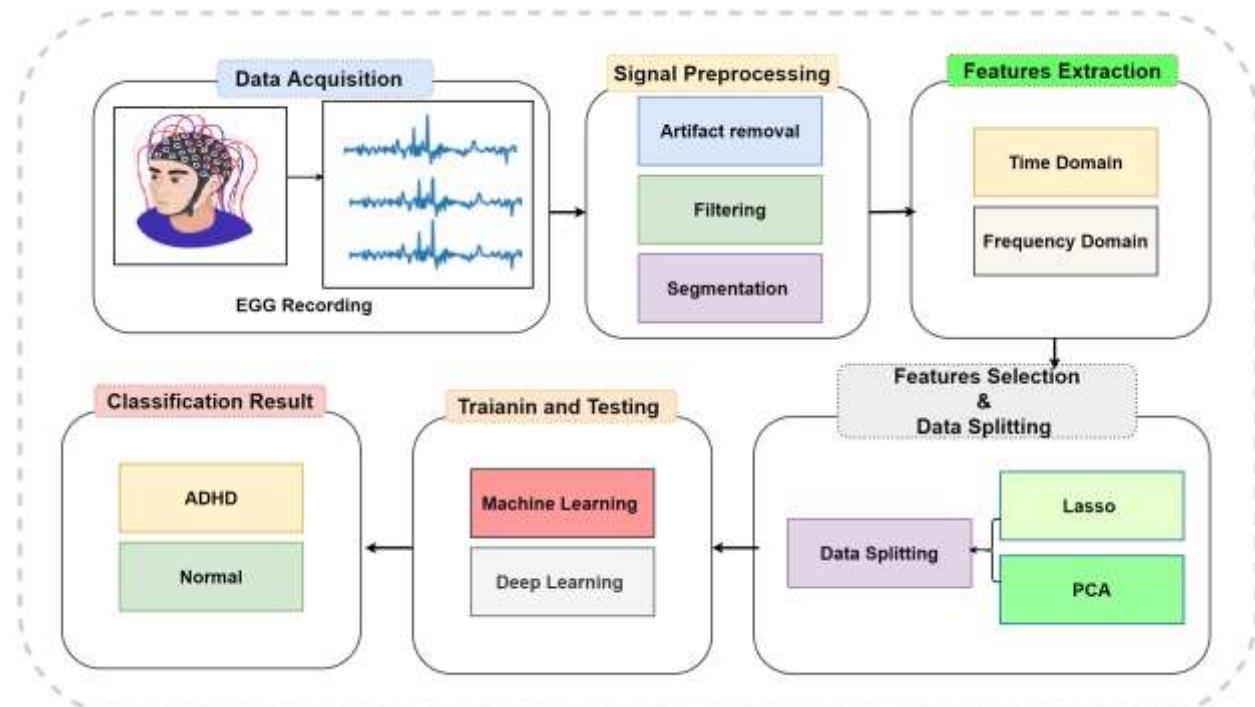


Figure 1. The proposed methodology for ADHD detection.

2.1. EEG Dataset

The EEG data employed in this study were sourced from a standard dataset, which consisted of 61 children diagnosed with ADHD and 60 typically developing controls. Further details regarding the dataset can be found in this section.

The dataset, collected by Nasrabadi, A.M., is accessible online at the following link: "<https://ieeexplore.ieee.org/open-access/eeg-data-adhd-control-children>" (accessed on 25 August 2023).

2.1.1. Participants

The dataset consisted of 61 children diagnosed with ADHD and 60 typically developing children. Additional details regarding the dataset can be found in Table 2. The ADHD cohort was recruited from referrals to the psychiatric clinic. The diagnosis of ADHD was made by an experienced child and adolescent psychiatrist, following the DSM-IV criteria. The control group, comprising 50 participants from an all-boys school and 10 from an all-girls school in Tehran, was carefully selected. A psychiatrist evaluated the control group to ensure the absence of any neurological or psychiatric disorders [41].

Table 2. The participants' details.

Group	Number of Participants	Gender	Age (years)
ADHD	61	48 Male, 13 Female	9.62 ± 1.75
Control	60	50 Male, 10 Female	9.85 ± 1.77

2.1.2. EEG data acquisition

The collection of EEG signals was performed utilizing a 19-channel digital acquisition equipment known as the SD-C24. The signals were acquired at a sampling rate of 128 Hz, with a resolution of 16 bits. Recording was conducted based on a visual attention task comprising 20 images with varying numbers of characters (5–16 per image). The images were sufficiently large and randomly distributed to elicit sustained visual attention. Each image was displayed immediately after the participant's response to maintain continuous visual stimuli. The task had performance penalties associated with performance. The globally established 10–20 technique was used to put the 19 scalp EEG electrodes [41]. This system ensures standardized coverage of all regions of the brain. The anterior region electrodes included Fp1, Fp2, F7, F3, Fz, F4, and F8. The posterior region was covered by electrodes T5, P3, Pz, P4, T6, O1, and O2. The central region is comprised of C3 and T3 in the left hemisphere and C4 and T4 in the right hemisphere. Finally, reference electrodes A1 and A2 were placed on the left and right earlobes, respectively, as shown in Figure 2. This spatial arrangement and color coding allowed for a clear visualization of the electrode positions across the scalp and their correspondence to the underlying cortical regions. The 10–20 placement optimized the recording of brain electrical activity relevant for analysis and classification.

The EEG signal activity of both the Control and ADHD groups was visualized, with each subplot depicting 5 seconds of data from three randomly chosen electrodes: Fp1, Cz, and Pz. Figure 3 presents these plots, showcasing the temporal variation in signal amplitude for the selected electrodes. The left column corresponds to the Control group, while the right column corresponds to the ADHD group.

Figure 4 displays the visualization of raw EEG signals for a subset of channels (Fp1, Fp2, C3, C4, O1, and O2). The EEG time series data is presented over a 10-second window, providing a visual representation of the recorded EEG activity.

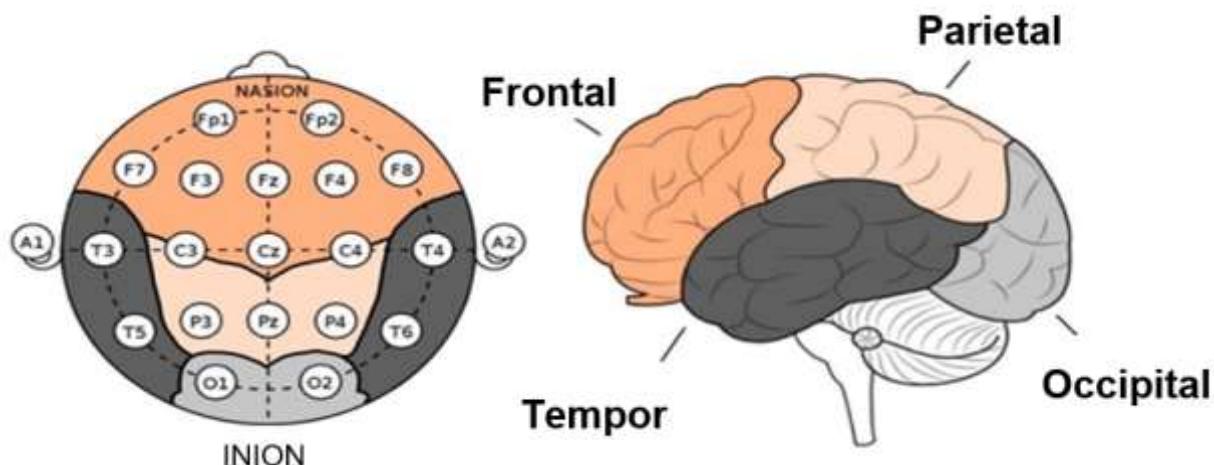


Figure 2. The positions of earlobes of the 10–20 system.

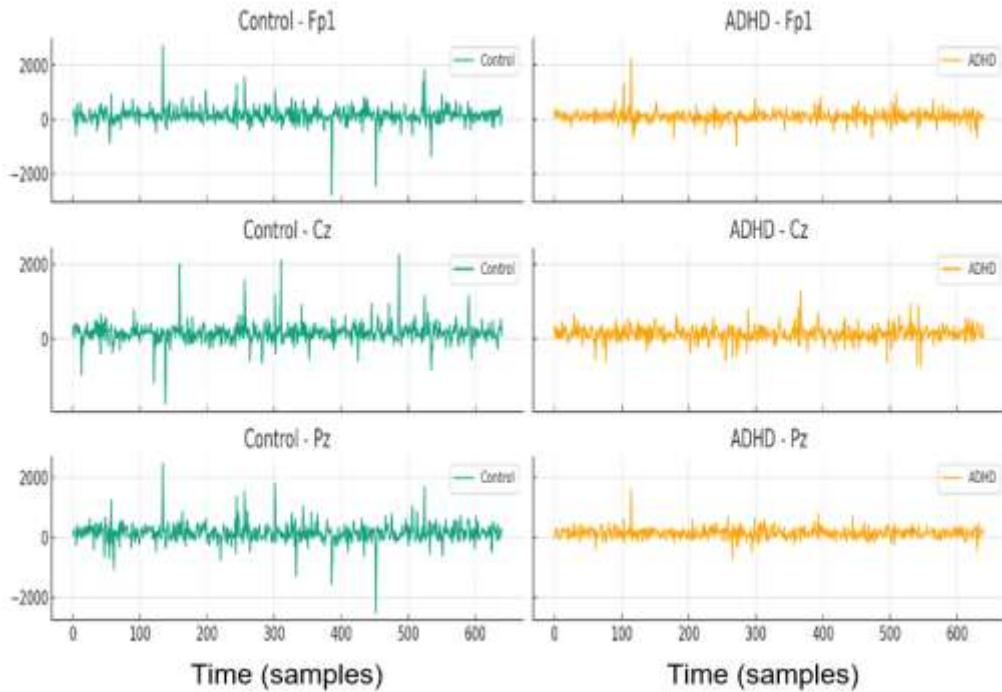


Figure 3. EEG signal activity for control and ADHD groups.

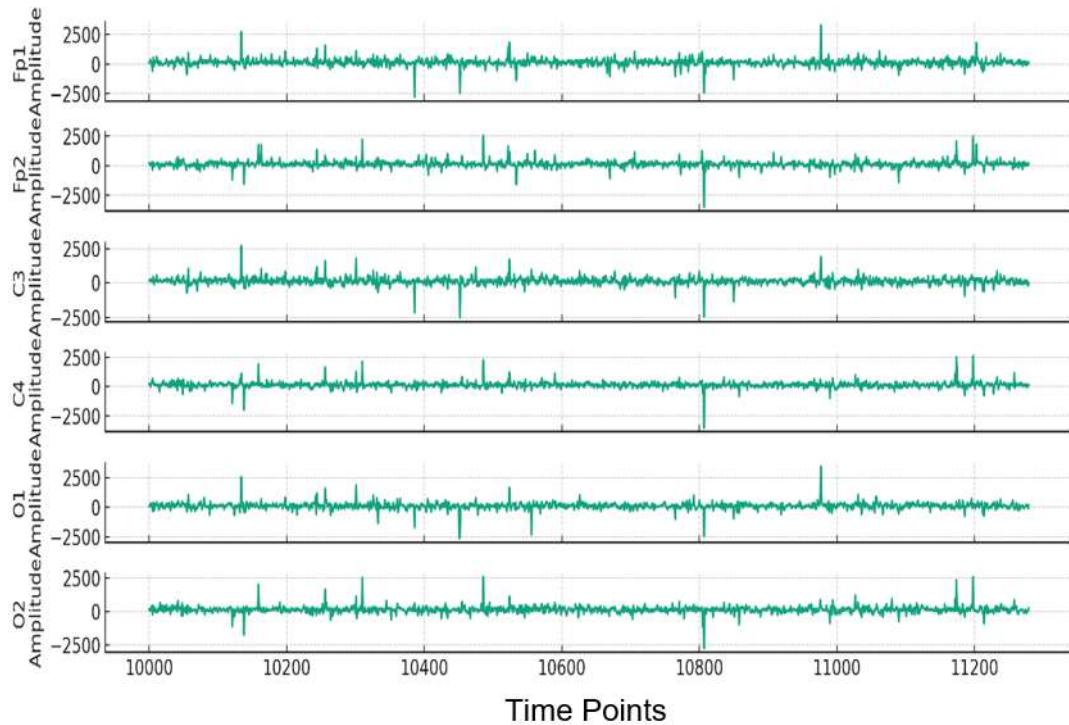


Figure 4. EEG time series data over 10-second window.

Short-Time Fourier Transform (STFT) time-frequency analysis was performed on selected EEG channels. STFT analysis in ADHD research has several benefits. First, it has the capability to investigate non-stationary signals in both the temporal and frequency domains. This unique approach

allows for the detection of transient events and facilitates the understanding of changes in frequency bands over time, which is crucial for gaining insights into the neurological aspects of ADHD. Second, the STFT enables the examination of event-related potentials, spectral analysis, and abnormality detection. These capabilities are valuable in differentiating ADHD patients from individuals with typical development. The fine-grained time-frequency resolution of the STFT is particularly valuable, as it enhances the localization of neuronal events. This makes it useful for tasks such as artifact removal and the development of sophisticated analysis techniques like machine learning models for ADHD classification. Overall, the STFT excels in managing complex EEG data and provides comprehensive insights into the brain pathways associated with ADHD. It is preferred over other techniques such as the Wavelet Transform and Multitaper in ADHD research due to its balanced time-frequency resolution and easy interpretability. The STFT's uniform analytic grid is well-suited for capturing the temporal dynamics of EEG signals, which is essential for investigating the neurological features of ADHD. Furthermore, the STFT is recommended for its computing efficiency and the clarity it offers in time-frequency analysis. These qualities make it a practical and efficient solution for the sophisticated examination of EEG data in ADHD research. This analysis generates spectrograms for each channel, providing insights into the dynamics of frequency content over time. Figure 5 illustrates the STFT representations of EEG channels.

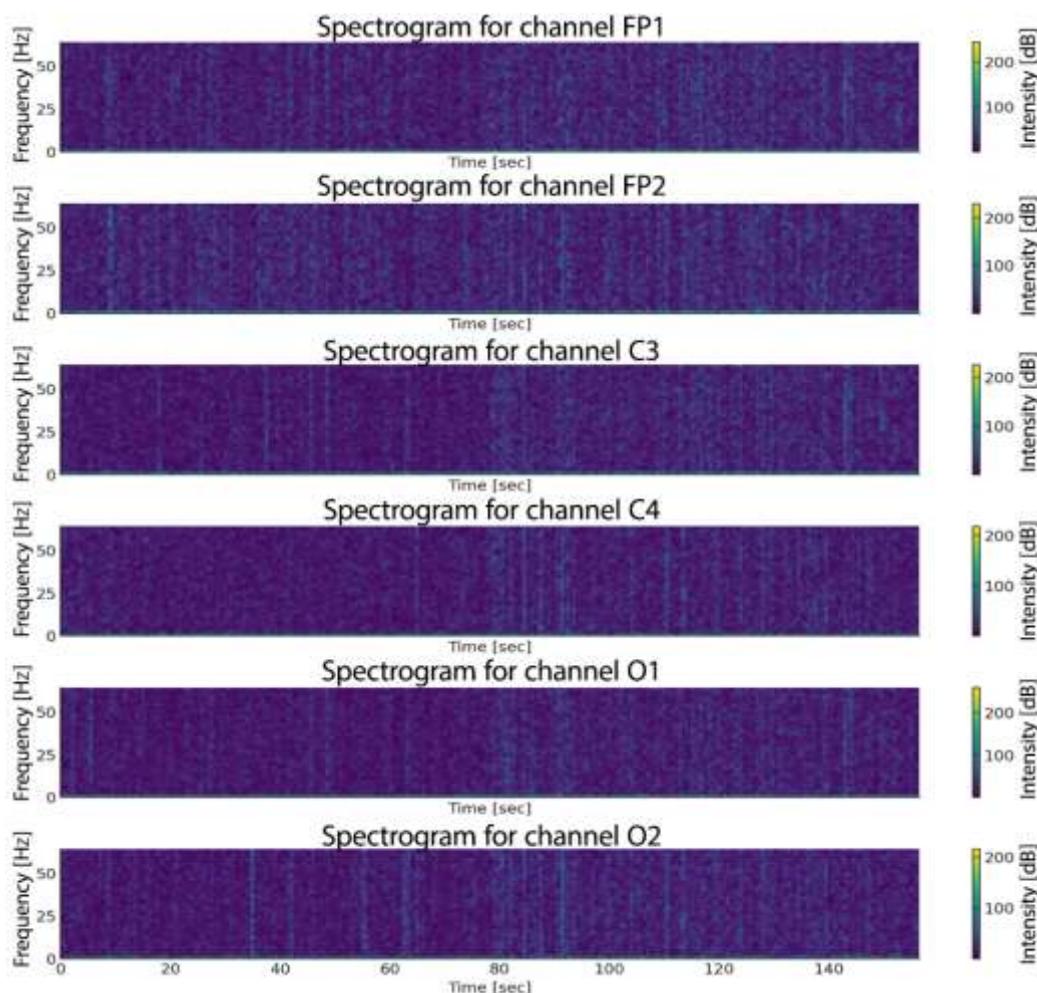


Figure 5. STFT approach.

The spectrograms obtained from the analysis depict the power distribution of each frequency band across different time points, enabling the identification of EEG patterns associated with cognitive or neurological events. For instance, fluctuations in the alpha band (8–13 Hz) may indicate a state of relaxation, while alterations in the beta band (13–30 Hz) could be indicative of active thinking or focused attention.

2.2. Preprocessing

We implement a rigorous data preprocessing and feature engineering pipeline to develop a robust EEG-based diagnostic system for ADHD. Raw EEG signals were filtered to isolate clinically relevant frequency bands, including delta, theta, alpha, and beta using fourth-order Butterworth band pass filters. Notch filters were also applied to eliminate power line interference.

Filtering enabled the extraction of band-specific features corresponding to neuronal oscillations associated with ADHD pathophysiology. ADHD neuro biomarkers include heightened theta and lower beta activity. Preserving frequency-specific information is critical for modeling the complex spectral signatures of ADHD versus healthy brains.

2.2.1. Feature extraction approach

The preprocessed EEG signals were divided into segments of 2-second duration using a sliding window technique with an overlap of 50% between consecutive windows. This segmentation allows for the extraction of discriminative features localized over time. Descriptive features capturing the characteristics of the signal in both the time and frequency domains were extracted for each 2-second window. Statistical features based on the raw signal values in the time domain were calculated to capture variability, distribution, and complexity. Measures such as standard deviation, skewness, kurtosis, and Hjorth parameters represent the variance, symmetry, tailed ness, and complexity of the distribution, respectively. This combination of statistical and spectral features provides a multidimensional profile, encapsulating both time-based morphology and frequency-based brain activity relevant to EEG analysis. The values calculated on short windows enable the tracking of dynamic fluctuations over time.

2.2.2. Feature selection approach

Robust feature selection was critical for extracting the most discriminative biomarkers from the complex, multidimensional EEG data and for enhancing the performance of our machine- and deep-learning models. A two-stage feature-screening pipeline, incorporating both filter-based and wrapper-based techniques to identify optimal ADHD-relevant features, was implemented.

In the first stage, PCA was applied as a filter-based approach to reduce dimensionality and derive a lower-dimensional feature subspace that captured 95% of the variance. By eliminating redundant and irrelevant features, PCA enhanced the signal-to-noise ratio in the input data for more efficient modeling.

A wrapper-based feature selector, using Chi-square testing to refine the PCA-filtered features, was subsequently implemented. Chi-square assessed the correlation between each feature and the ADHD/control classes to rank features by discriminative power. Only the top-ranked Chi-square features, that exhibited the strongest diagnostic relevance, were retained.

This two-step filtering approach combines the power of a supervised wrapper technique (Chi-square) with an unsupervised filter method (PCA) to extract both generalized and class-specific predictive information from EEG data. The research experiments demonstrated significant performance gains from coupling Chi-square and PCA, rather than using either in isolation.

2.2.3. Data splitting

The final dataset was split into training and testing sets using a ratio of 80:20. While 80% of the data was allocated to the training set, which is used to fit the machine-learning models, the remaining 20% was assigned to the testing set, which provided an unbiased evaluation of the model's performance on new unseen data.

2.2.4. Handling imbalanced class

In the original dataset, the ADHD class constituted the majority class, with significantly more samples than the control group. This imbalance could lead the learning algorithms to focus primarily on the ADHD class and compromise the accuracy of the control group. Figure 6 shows class distribution of the original dataset.

To mitigate this issue, the Synthetic Minority Oversampling Technique (SMOTE) algorithm was applied to address the class imbalance in the training data after data preprocessing. SMOTE approach generates new synthetic samples from the minority class (control group) to improve its representation. Oversampling occurs by providing synthetic instances along line segments to nearest neighbors for each minority class sample [8].

SMOTE oversampling was implemented on the training data to generate new synthetic control samples. This achieved a more balanced class distribution between the ADHD and control groups in the training set. Balanced representation will enable the machine- and deep-learning models to learn the nuances of both classes more effectively. The testing set was kept untouched to provide an unbiased estimate of real-world performance. The balanced training data is expected to build more robust classifiers with equal emphasis on each class and improved sensitivity. This approach ensures a meticulous and unbiased evaluation of the model's capabilities, promising reliable performance assessments in real-world scenarios.

We found a class imbalance in the original dataset, with ADHD and control group samples distributed unevenly. Each trial lasted varying amounts of time depending on how long each youngster took to count the animals and enter their response. The 50 seconds were the shortest trial for a control and the longest was 285 seconds for an ADHD patient. This bias had to be addressed to enable thorough model training due to this imbalance. To address the imbalance in the training data, SMOTE was carefully applied to the control class, which was underrepresented. The SMOTE method finds each minority class sample's k nearest neighbors based on feature space closeness. Then, along the chosen neighbor's line segment that connects the minority sample, synthetic samples are created. This will be carried out until the representation of minority classes achieves the necessary level of domination. To prevent bias in the test data, SMOTE was exclusively applied to the training data. To evaluate the model's generalizability to real-world unbalanced circumstances, the untouched test subset preserved the fundamental class imbalance. This imbalance gave an honest assessment of the model's capacity to manage dataset issues. SMOTE approach was applied to balance just the EEG training dataset allowed robust model tweaking while leaving the EEG test dataset.

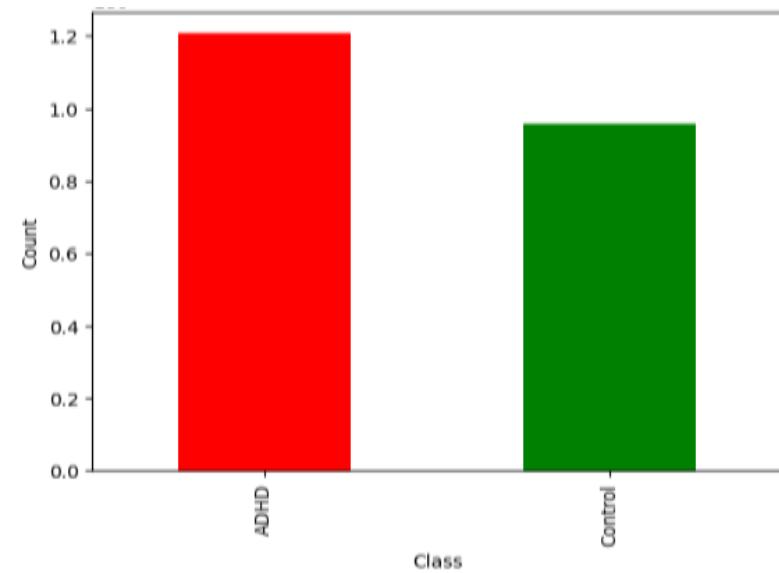


Figure 6. Class distribution of the original dataset.

2.3. Proposed model

The machine-learning algorithms, including naive Bayes, support vector machines, ensemble methods, and deep neural networks, such as convolutional and gated recurrent units, to EEG biomarker datasets were applied.

2.3.1. Machine learning

Machine-learning algorithms build mathematical models from sample data, known as training data, to make predictions or decisions without being specifically coded for the task [42]. Machine-learning approaches are commonly categorized as supervised, unsupervised, or reinforcement learning based on the nature of the problem and data labeling. In supervised learning, the training data comprise examples with known output labels, and the algorithms learn to map inputs to outputs [43].

A range of standard supervised learning algorithms were implemented for comparative benchmarking on the ADHD classification task. The models encompass both linear classifiers, such as probabilistic models including naive Bayes, support vector machines, and ensemble methods like random forest, gradient boosting, and model stacking. Naive Bayes provides a probabilistic framework for classification, assuming feature independence to estimate class probabilities via Bayes' theorem [44]. Support vector machines are commonly used for classification and regression tasks [45]. The goal of SVMs is to find the optimal separating hyperplane that maximizes the margin between classes in a high-dimensional space. SVMs do not use all training points to define the hyperplane; rather, they select a subset of points near the class boundaries called support vectors. SVMs are the critical elements that determine the hyperplane and have the greatest influence.

A stacking ensemble was implemented by training a high-level Cat Boost classifier on the combined predictions from base random forest and Light Gradient Boosting Machines (GBM) models to improve predictive performance. These diverse algorithms provide complementary modeling strengths for assessing a comprehensive set of approaches for robust ADHD classification from EEG data.

2.3.2. Deep learning algorithm

Deep learning constitutes a subfield of machine learning focused on architecting artificial neural networks with multiple layers to automatically learn representations and extract hierarchical features from data [46]. This facilitates the development of highly capable predictive models for complex tasks, including medical diagnoses.

In this study, CNN architecture used 1D convolutions to learn a hierarchical feature representation directly from the raw EEG signals and trained end-to-end on the labeled data for the diagnostic classification task.

The model comprises convolutional layers to automatically learn spatial feature representations interspersed with max pooling, dropout, and fully connected layers [47,48,49,50]. The input EEG signals are input into 1D convolutions. Temporal correlations were identified using two consecutive Conv1D layers with 64 filters and a kernel size of 3. Rectified linear unit (ReLU) activation was employed for nonlinear transformations. To improve generalization, dropout with a rate of 0.5 was implemented between convolutional blocks. Max pooling layers reduced dimensionality while retaining significant features. The convolutional feature maps were flattened into a 1D vector and connected to a series of dense layers for high-level reasoning, with a progressively decreasing number of units. ReLU activation was again applied to nonlinear combinations. Two nodes in the final Softmax output layer allowed binary categorization of ADHD and control groups. Figure 7 depicts model architecture.

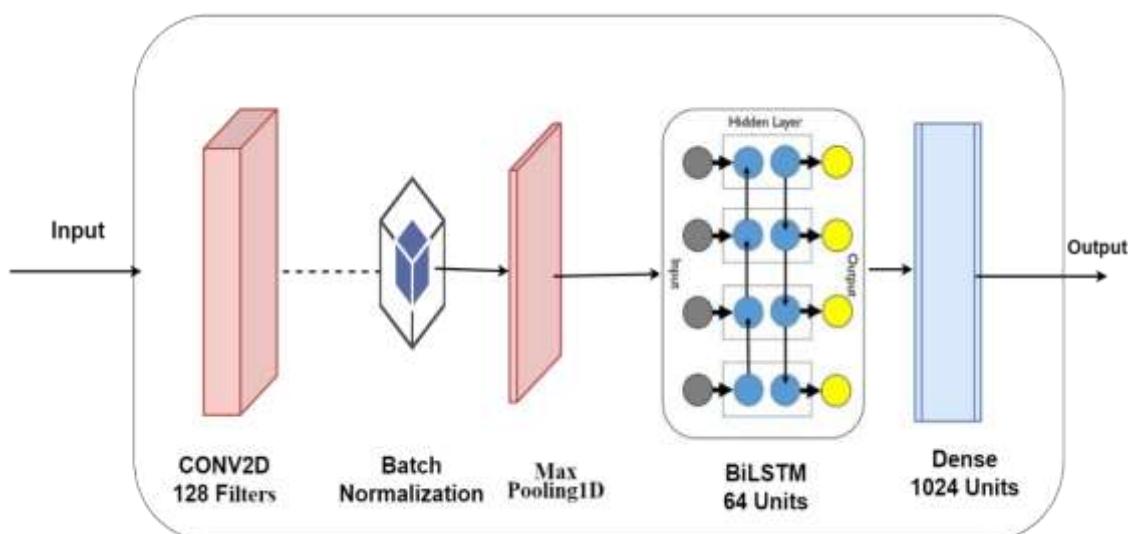


Figure 7. CNN-BiLSTM architecture.

The convolutional-LSTM path also applies 1D convolutions to obtain local features. A hundred twenty eight filters in the Conv1D layer were used. This is followed by max pooling and a bidirectional LSTM with 64 units to learn temporal dependencies in both directions. The LSTM output is processed by a 1024-unit dense layer and dropout before being concatenated with the other paths' outputs.

$$C = \sum_1^i \sum_1^j I_{ij} F_{ij} \quad (1)$$

Within the context of CNNs, the symbol F denotes a convolution kernel or filter, while i , and j pertain to the specific rows and columns of an image I , respectively. The input picture is convolved

with the kernel, yielding a novel two-dimensional output. The procedure involves decomposing the image into individual neurons, followed by flattening these neurons along the y and z dimensions. Each layer inside the network is equipped with a set of x filters that are designed to detect and identify traits. Feature maps of size X are generated by Layer L, and these feature maps are annotated appropriately.

$$C_i^L = B_i^L + \sum_{j=1}^{x^{(L-1)}} F_{i,j}^L * C_j^{(L-1)} \quad (2)$$

The term B_i^L denotes the bias matrix, whereas $F_{i,j}^L$ signifies the filter that connects the jth feature map inside the layer.

$$f_t = \sigma(W_{ef}X_t + W_{ef}h_{t-1} + W_{cf}C_{t-1} + U_f) \quad (3)$$

$$i_t = \sigma(W_{xi}X_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + U_i) \quad (4)$$

$$C_t = \sigma(f_t c_{t-1} + i_t \tanh(W_{xc}X_t + W_{hc}h_{t-1} + U)) \quad (5)$$

$$o_t = \sigma(W_{xo}X_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + U_o), \quad (6)$$

$$h_t = o_t \times \tanh(C_t) \quad (7)$$

The equations above are often used in the sequential forward and sequential backward procedures. They denote the equations that encapsulate the BiLSTM model. The BiLSTM network may be conceptualized as a gated cell that assesses input data and determines its retention depending on its significance or weight. The BiLSTM model is composed of three fundamental components: The input gate, the forget gate, and the output gate. The forget gate, denoted as f_t , is responsible for determining the states that should be retained in memory or discarded. The input gate i_t adjusts the value by considering the incoming signals. The output gate, denoted as o_t , facilitates the transmission of the cell state to adjacent neurons. The architecture comprises of a logistic layer and an additional layer responsible for generating a novel vector that is then combined with the existing state. In the context of a recurrent neural network (RNN), the input X_t is processed by the hidden layer using the weight matrix W , resulting in the generation of the final output y_t . The LSTM model incorporates a memory cell denoted as h_t , which serves as a pivotal component regulated by three distinct gates.

2.3.3. GRU-Transformer approach

The model implements GRU layers for sequence modeling [51]. The input layer is designed to accommodate the time series EEG feature matrix, which consists of a single channel. Two sequential GRU layers with 1,024 and 128 units were stacked to learn temporal relationships. The second GRU layer returns the full sequence for further modeling. GRUs contain gating units that modulate information flow, enabling them to better capture long-range dependencies compared to vanilla RNNs. The final Softmax output layer produced normalized predicted probabilities for the two target classes: ADHD and the healthy control. The model's architecture is shown in Figure 8.

$$\mu_t = \sigma(V_\mu x_t + W_\mu o_{t-1} + b_\mu) \quad (8)$$

$$r_t = \sigma(V_r x_t + W_r o_{t-1} + b_r) \quad (9)$$

$$i_t = \tanh(V_o x_t + W_o(r_t \odot o_{t-1}) + b_0) \quad (10)$$

$$o_t = \sigma(\mu_t \odot o_{t-1} (1 - \mu_t) \odot i_t) \quad (11)$$

The symbol x_t denotes the input, o_t represents the output, μ_t signifies the output of the update gate, r_t denotes the output of the reset gate, and \odot denotes the Hadamard product. The parameters or weight matrices are denoted as V, W, and b.

The GRU encoder and Transformer path uses a recurrent GRU layer to produce embeddings of the input sequence. Thirty-two GRU units were applied to encode 32-dimensional vectors at each timestep. Multi-head self-attention with 2 heads is then applied, which allows the GRU embeddings to attend to each other based on learned relationships. Residual connections and layer normalization stabilize the training. The attention outputs are flattened to a 1D vector.

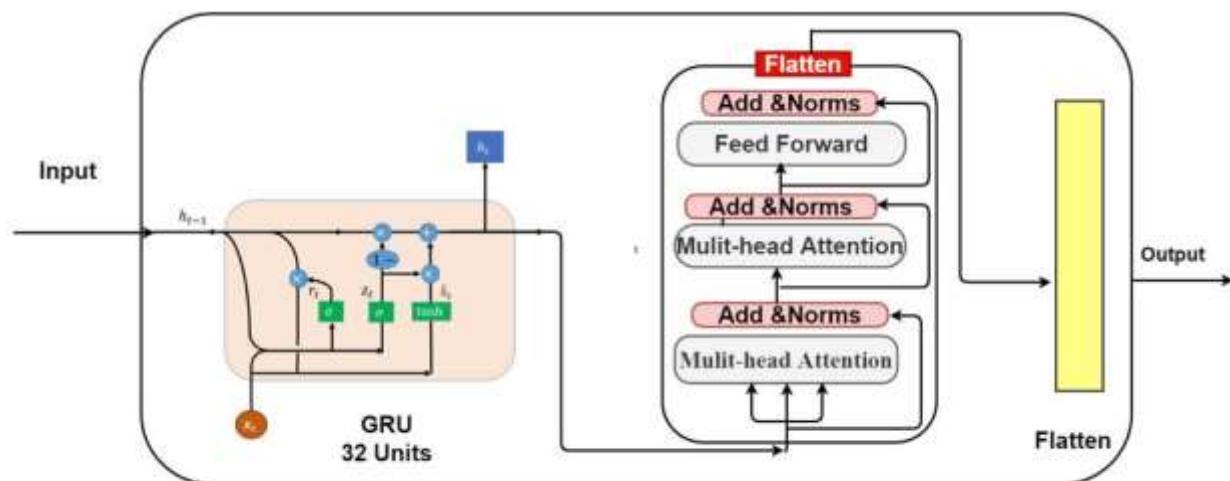


Figure 8. The architecture of the GRU-Transformer block.

2.3.4. Proposed system

A novel deep neural network architecture that combines convolutional, recurrent, and attention-based models for improved sequence classification performance has been developed. The core of the model is the integration of the CNN, CNN-LSTM, and GRU-Transformer blocks through concatenation of their outputs. This merging allows the model to leverage the diverse representations learned by each path in a unified architecture. The concatenated feature vector encapsulates local spatial-temporal correlations from the CNN, long-term dependencies from the LSTM, and global relationships from the Transformer's attention mechanism. This robust combined representation is fed into additional dense layers for final classification. A 1024-unit Dense layer learns nonlinear combinations of the concatenated features. Dropout regularization prevents overfitting. The final Softmax output layer predicts class probabilities. By concatenating the complementary outputs of the CNN, RNN, and Transformer paths, the developed model achieves robust sequence classification. The integrated architecture leverages the strengths of each block-convolutional features, recurrence, and attention - to represent the input data in multiple ways for improved accuracy. The goal of this multi-path, and concatenated design is to enhance classification performance compared to single-path models. By merging diverse spatial, temporal, and attention-based representations, the model can capture nuances in the data that might be missed by CNN, RNN, or Transformer architectures alone. Figure 9 displays the integrating model, and Table 3 shows the parameters of integrating models.

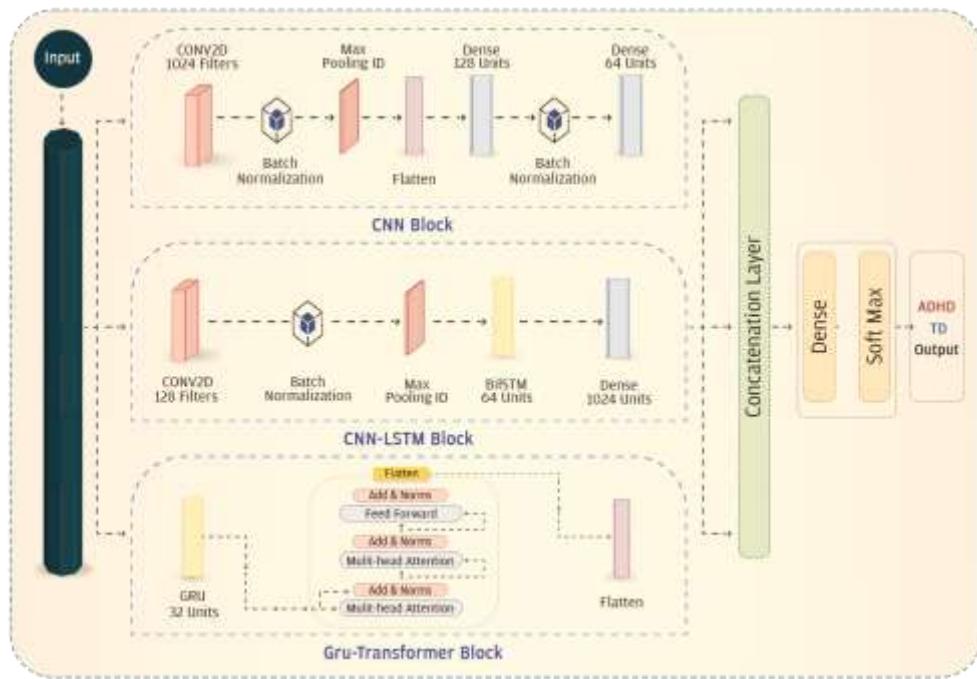


Figure 9. The proposed integrating model.

Table 3. Parameters of integrating model.

Model block	Layer	Parameters
CNN	Conv1D	1024 filters, kernel size 3
	BatchNorm	-
	MaxPooling1D	Pool size 2
	Flatten	-
	Dense	128 units
	Batch Norm	-
	Dropout	Rate 0.5
CNN-LSTM	Dense	64 units
	Conv1D	128 filters, kernel size 3
	Batch Norm	-
	MaxPooling1D	Pool size 2
	Bidirectional LSTM	64 units
	Dense	1024 units
GRU-Transformer	Dropout	Rate 0.5
	GRU	32 units
	Multi-Head Attention	2 heads
	Add	-
	Layer Norm	-
	Flatten	-
Model Integration	Concatenate	-
Output Layers	Dense	1024 units
	Dropout	Rate 0.5
	Dense	2 units, softmax

2.4. Experimental

In this section, we delineate the experimental framework used to develop and validate the developed EEG-based ADHD classification models. Detailed results are presented and analyzed to benchmark the efficacy of the proposed approach against state-of-the-art methods. The rigorous experimental pipeline provides critical insights into the real-world viability of using EEG data and machine learning for enhanced ADHD screening. The developed framework could inform best practices for applying these techniques to improve the diagnosis of neurological conditions.

2.4.1. System setup

The experiments were conducted using a laptop workstation with a Core i7 CPU, 8 GB of RAM. This provided sufficient computing capabilities for efficient model training and evaluation. Models were implemented using TensorFlow [15] and Scikit-learn, an open-source framework for model's designing, training, and testing. TensorFlow uses GPU acceleration, which significantly expedites neural network computations compared to CPU-only environments.

2.4.2. Evaluation metrics

Assessing the performance of models is crucial to comprehending their proficiency [16]. Several evaluation metrics exist, such as accuracy, sensitivity, precision, recall, F1-score, receiver operating characteristic (ROC) curve, and confusion matrix. Each metric provides unique insights into the strengths and weaknesses of the model. Thorough evaluation using diverse metrics offers a comprehensive portrayal of model efficacy.

2.4.3. Confusion matrix

The confusion matrix constitutes an essential evaluation tool for binary classification systems by summarizing the predictive performance across the test dataset. Its cardinal components encompass True Positives (TP): ADHD cases correctly classified by the model as positive; False Positives (FP): Control cases incorrectly predicted as ADHD (positive); True Negatives (TN): Control cases properly classified by the model as negative; and False Negatives (FN): ADHD cases incorrectly classified as controls (negative). By tabulating the true and false, positive ADHD, and negative control predictions, the confusion matrix facilitates the quantitative assessment of the classifier's discrimination proficiency between ADHD and control classes. It illuminates critical errors through false positives (controls predicted as ADHD) and false negatives (ADHD predicted as controls) to enable the identification of learning deficiencies.

2.4.4. Accuracy

Accuracy is calculated as the ratio of correct predictions to total predictions, as shown in Eq (12). It serves as a commonly used metric for assessing the performance of models.

$$\text{Accuracy} = \frac{TP+TN}{FP+FN+TP+TN} \times 100 \quad (12)$$

2.4.5. Sensitivity

Sensitivity is defined as the proportion of actual positives accurately detected as in Eq (13). It quantifies the rate of true positives in binary classification. It measures model proficiency in identifying positive cases (ADHD) without type II errors.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False positives}} \quad (13)$$

2.4.6. F1-Score

The F1-score is the harmonic mean of precision and recall. It synthesizes these metrics into a singular value, providing a balanced evaluation of model performance. A high F1-score indicates strong precision and recall, meaning few false positive and false negative predictions. Unlike accuracy alone, the F1-score offers a nuanced portrait of classification proficiency. Calculated via Eq (14), the F1-score furnishes crucial insights beyond accuracy.

$$F1 - score = 2 * \frac{\text{precision} \times \text{Sensitivity}}{\text{precision} + \text{Sensitivity}} \times 100 \quad (14)$$

2.4.7. Specificity

Specificity measures a binary classifier's ability to correctly identify negatives. It is quantified as the ratio of true negatives (TD) to total negatives. It evaluates model competence in avoiding false positives—negative instances mistakenly classified as positive. The specificity calculation was performed using Eq (15).

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}} \quad (15)$$

2.4.8. Receiver operating characteristics

The true positive rate (TPR) against false positive rate (FPR) ROC curve shows model discrimination abilities. The ROC curve shows the tradeoff between true and false positives as TPR measures proper identification and FPR measures negative misclassifications. This illuminates balanced model performance beyond accuracy.

3. Results

3.1. Machine learning classification results

We evaluated three machine-learning classifiers—Gaussian naïve Bayes, SVM, and a stacking ensemble to improve the accuracy of ADHD diagnosis from EEG biomarkers. PCA was first utilized for feature selection, retaining 95% of the data variance. With the PCA-reduced features alone, the SVM model achieved the highest diagnostic performance, attaining 94.86% accuracy, 96.33% sensitivity, 93.02% specificity, 95.42% F1-score, and 98.71% AUC curve. These results demonstrate SVM's robust capabilities for accurately categorizing both ADHD and non-ADHD cases based on

EEG data. Figure 10 shows the confusion matrix and ROC curve. Table 4 presents the SVM classification report.

Supplementing Chi-square with PCA for more targeted, variance-focused feature selection further improved outcomes. SVM attained 92.8% accuracy, 94.3% sensitivity, 91.0% specificity, 93.6% F1-score, and 97.8% AUC using the combined PCA Chi-square feature set. This highlights the utility of coupling PCA and Chi-square for honing the most diagnostically relevant features from complex EEG biomarkers. Figure 11 shows the confusion matrix and ROC curves, while Table 5 presents the SVM classification report.

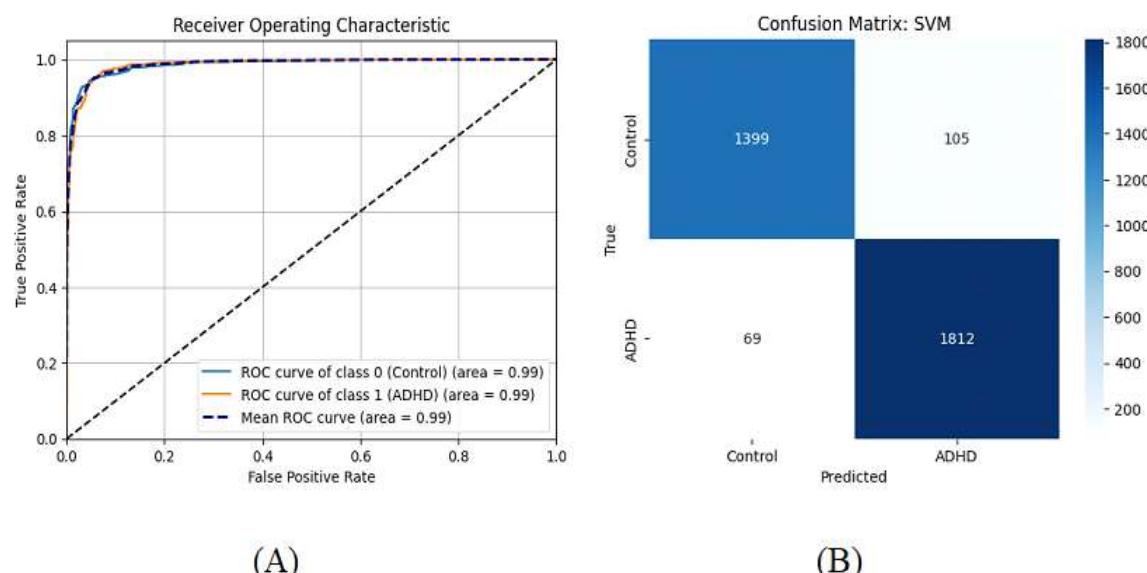


Figure 10. The ROC curves and confusion matrix of SVM using PCA: (A) ROC, and (B) confusion metrics.

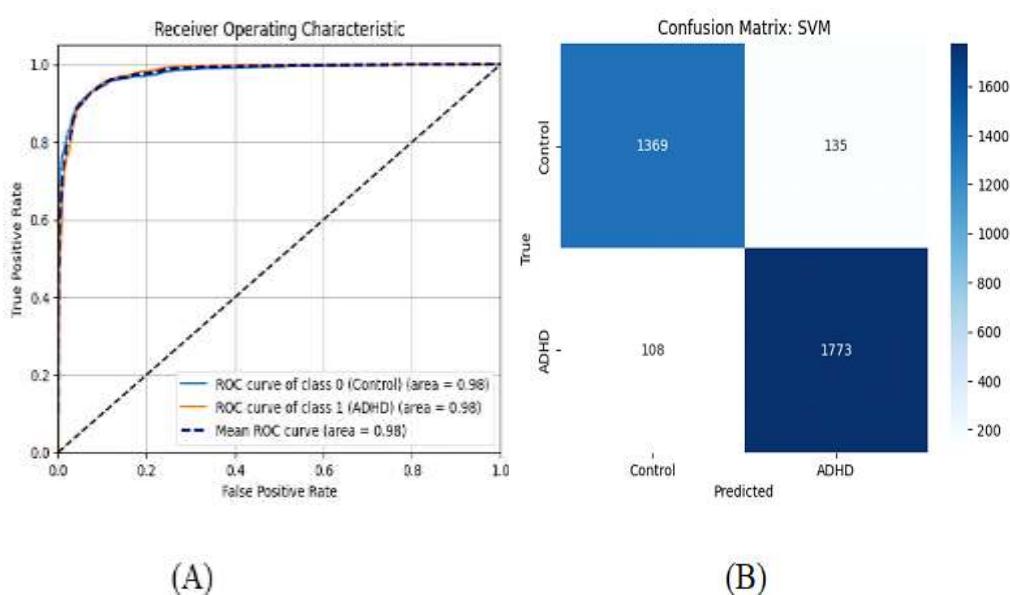


Figure 11. ROC curves and confusion matrix of SVM using Chi-square with PCA: (A) ROC, and (B) confusion metrics.

Table 4. The classification report of the SVM using PCA.

Class Name	Precision%	Recall%	F1-Score%	Support
Control	95	93	94	1504
ADHD	95	96	95	1881
Accuracy			95	3385
Macro Avg	95	95	95	3385
Weighted Avg	95	95	95	3385

Table 5. The classification report of the SVM using chi-square with PCA.

Class Name	Precision%	Recall%	F1-Score%	Support
Control	93	91	92	1504
ADHD	93	94	94	1881
Accuracy			93	3385
Macro Avg	93	93	93	3385
Weighted Avg	93	93	93	3385

However, SVM remained the top individual classifier, indicating it is presently the most effective standalone machine learning technique for EEG-based ADHD detection when combined with Chi-square and PCA feature selection. These results clearly demonstrate that machine learning, especially SVM architectures, can significantly enhance ADHD diagnostic accuracy compared to conventional methods (see Table 6).

Table 6. Results of the machine-learning classification.

Feature Selection Technique	Model Name	Accuracy%	Sensitivity%	Specificity%	F1 Score%	AUC%
PCA	GausnaïveNaive	71.7	81.4	59.6	76.2	78.8
	Bayes					
	Stacking Classifier	83.6	87.4	78.1	85.3	90.6
	SVM	94.9	96.3	93.0	95.4	98.7
Chi-Square with PCA	Anaïvesian Naive	71.8	81.6	59.5	76.3	78.1
	Bayes					
	Stacking Classifier					
	SVM					

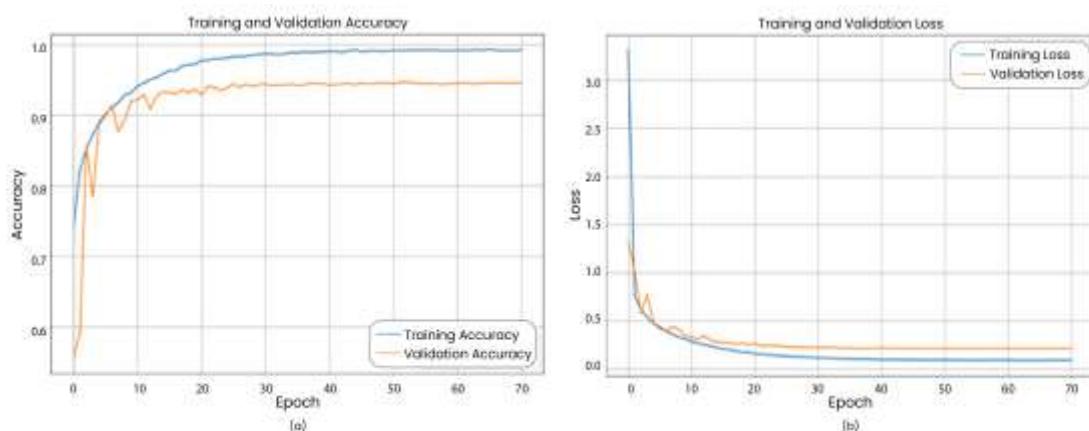
3.2. Results of integrating deep leaning with PCA

This study assessed combined deep leaning, known as CNN, CNN-BiLSTM block, and GRU-Transformer block, for the automated diagnosis of ADHD using neuroimaging biomarkers. Table 7 presents the classification report of integrating deep leaning model with PCA feature selection. The PCA initially reduced feature dimensionality, retaining 95% of the data variance. Utilizing just PCA-derived features, the integrating deep leaning attaining accuracy 95% accuracy, 95% recall, and 94% F1-score. This underscores the aptitude of integrating deep leaning for mining predictive ADHD patterns from neuroimaging data.

Table 7. Results of the proposed system with PCA.

Class	Precision %	Recall %	F1-Score %	Support
Control	95.1	93.2	94.0	1504
ADHD	94.2	96.3	95.1	1881
Accuracy	95.0			
Weighted Average	95.1	95.2	94.0	3385

Figure 12 shows accuracy of the training and testing of the integrating model with PCA. Throughout the training phase of the model, two significant metrics are often used to assess its performance: Training accuracy and validation accuracy. Moreover, the training and validation losses play a crucial role in assessing the model's learning progress. This investigation evaluated an instance where training accuracy increased from 75% to 99%. Concurrently, the validation accuracy started at 55% and had a moderate rise reaching an estimated value of 95%. Moreover, there was a decrease in the accuracy loss throughout the testing phase, from 1.3 to 0.4.

**Figure 12.** Training and testing performance of the proposed integrating deep learning system with PCA: (a) accuracy, (b) loss.

3.3. Results of integrating deep learning system with Chi-square

Add Chi-square for enhanced, variance-focused feature selection to increase suggested integrated deep learning models' accuracy, precision, recall, and F1-score to 96%. This highlights the utility of Chi-square to isolate the most diagnostically relevant imaging biomarkers. Table 8 presents the performance of the integrating deep learning system with Chi-square.

Table 8. Results of the integrating model with Chi-square.

Class	Precision %	Recall %	F1-Score %	Support
Control	96.2	94.1	95.3	1504
ADHD	95.0	97.2	96.4	1881
Accuracy	96.1			
Weighted Average	96.1	96.1	96.3	3385

The integrating deep learning model decisively surpassed the across all diagnostic metrics, indicating greater proficiency for extracting discriminative neuroimaging patterns. Coupling integrating deep learning system with Chi-square feature selection further bolsters detection accuracy. These results clearly demonstrate the viability of deep neural networks for enhancing ADHD diagnosis when combined with robust feature engineering. Figure 13 displays performance of Chi-square with integrating deep learning model.

To assess the efficacy of the proposed algorithms, the researchers chose to use the confusion matrix, which is a widely utilized metric for evaluating classification tasks. The deep learning model confusion matrix is shown in Figure 14.

To have a deeper understanding of the classification efficacy of the proposed integrating deep learning system, it is essential to examine its performance when trained and tested using PCA and Chi-square methodologies. The proposed integrating deep learning system model, when used with Chi-square, accurately identified 1395 participants as control and 1804 participants as ADHD. However, when integrating deep learning with PCA, 77 patients were misclassified. On the other hand, the integrating deep learning model with Chi-square classified 1413 participants as control and 1823 as ADHD, with 58 misclassifications.

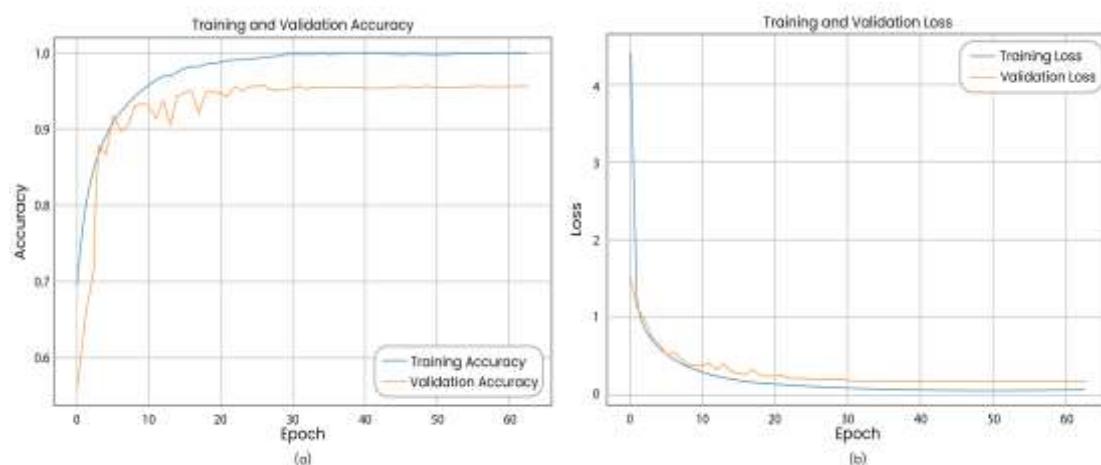


Figure 13. Training and testing performance of the integrating system with Chi-square: (a) accuracy and (b) loss.

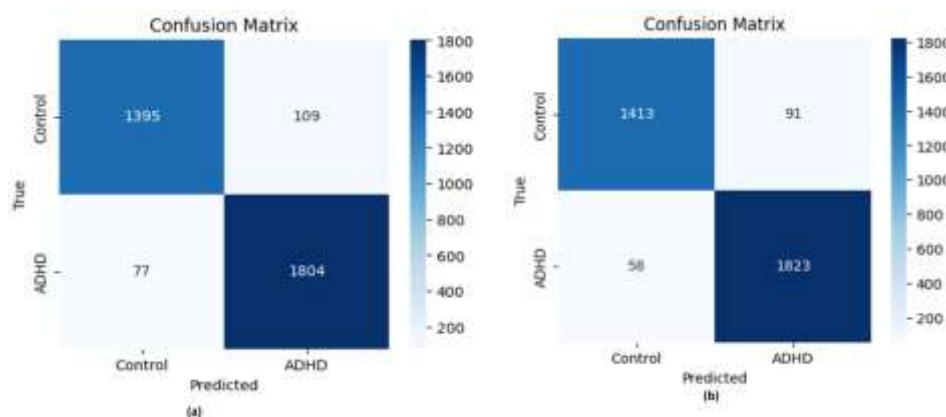


Figure 14. Integrated deep learning model confusion matrix: (a) CNNs with PCA and (b) CNNs with Chi-square.

4. Discussions

We presented an integrating deep learning model for automated EEG-based detection of ADHD, achieving new, and state-of-the-art accuracy through optimized data preprocessing and feature engineering. The developed SVM model attained 94.86% accuracy on the ADHD dataset when combined with PCA for feature selection, as presented in Table 9. This surpassed prior works, including Alim et al. (2023) [49], who reached 93.2% accuracy using a Gaussian SVM model without specialized feature engineering.

The developed integrating deep learning model, coupled with Chi-square and PCA for enhanced feature screening, achieved comparable state-of-the-art accuracy (95%), outperforming previous deep-learning techniques, as presented in Table 9. This includes the graph neural network proposed by Ekhlaei et al. (2022) [50], which obtained 91.2% and 90% accuracy in theta and delta EEG bands without inputs tailored for ADHD detection.

These gains highlight the efficacy of the data preprocessing pipeline and custom feature selection techniques for extracting the most discriminative biomarkers from complex, high-dimensional EEG data. The two-stage PCA and Chi-square feature screening enabled the SVM and integrating deep learning model to better capture predictive ADHD patterns for significant performance improvements over conventional approaches. Additionally, proactive balancing of the training data addressed class imbalance, further boosting the model learning of salient ADHD characteristics. The work's specialized data wrangling and feature engineering optimizations were critical to unlocking the full diagnostic potential of machine and deep learning for EEG-based ADHD detection.

Figure 15 displays a ROC curve of the integrating model with feature selection of PCA and Chi-square. The performance of a classification model is better represented by the ROC-AUC measure. Different machine learning and deep learning algorithms may successfully identify and diagnose ADHD, as shown by the ROC-AUC. The classifier model's capacity for diagnosis is gauged by the ROC-AUC score. The CNNs with PCA and Chi-square scored AUC=99%. Figure 16 presents a comparison of the developed integrating deep learning model with existing ADHD AI models developed by other researchers [52,53].

Table 9. Performance of proposed ADHD system using feature selection approaches.

Feature Selection approaches	All Model	Accuracy%	Sensitivity%	Specificity%	AUC %	F1-Score%
PCA	Gaussian Naive	71.7	81.4	59.6	76.2	78.8
	Bayes					
	Stacking Classifier	83.3	87.4	78.1	85.3	90.6
	Integrating model	94.0	95.0	92.0	99.0	95.0
Chi-Square with PCA	SVM	94.9	96.3	93.0	95.4	98.7
	Gaussian Naive	71.8	81.6	59.5	76.3	78.1
	Bayes					
	Stacking Classifier	85.6	88.5	81.8	87.2	93.4
	SVM	92.8	94.3	91.0	93.6	97.8
	Integrating model	95.0	96.0	93.0	99.0	96.0

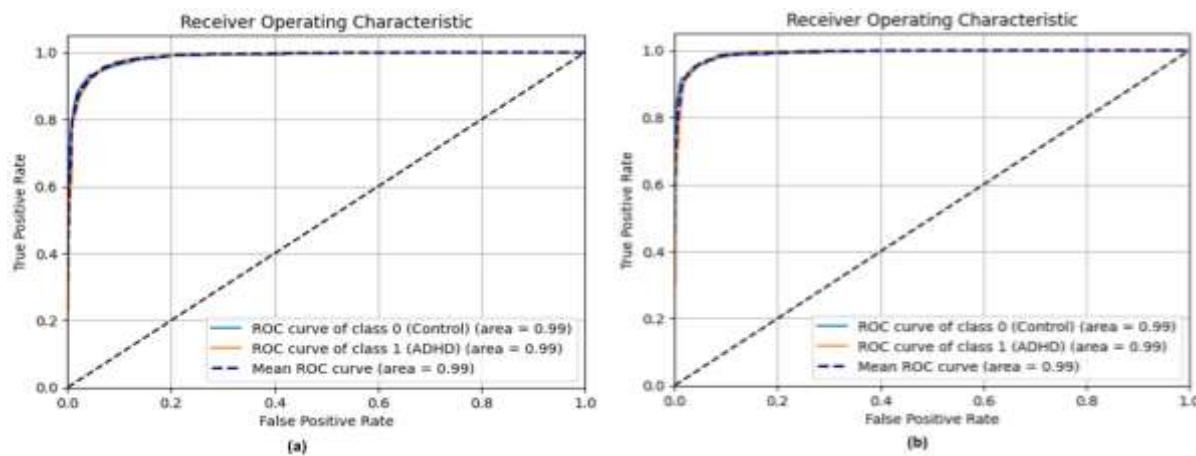


Figure 15. ROC of the integrating deep leaning model: (a) PCA and (b) Chi-square.

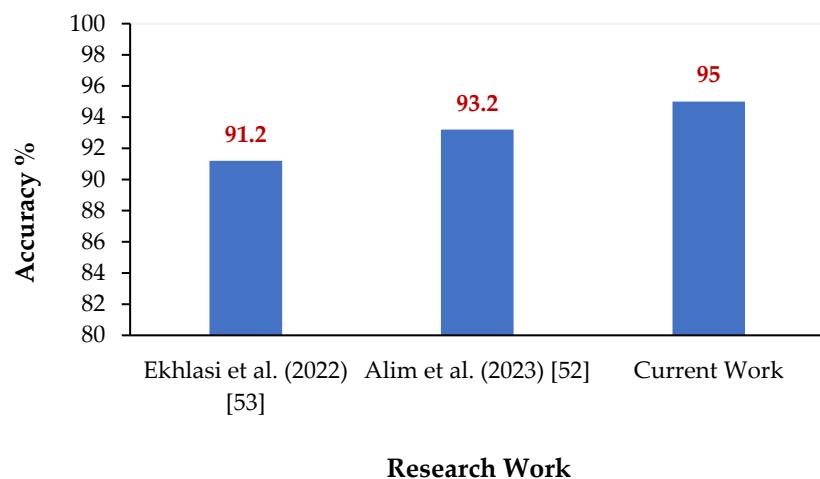


Figure 16. Comparisons of the developed integrating deep leaning model with existing ADHD models.

5. Conclusions

We present a machine-learning and deep-learning framework for automated discrimination between ADHD and healthy children using EEG. The framework has many key components, including preprocessing, feature extraction, feature selection, and classification. The developed models achieve cutting-edge accuracy, highlighting the power of optimized machine-learning pipelines to improve ADHD diagnosis in comparison to traditional approaches. In the ADHD-EEG dataset, the SVM approach combined with PCA feature selection achieved 94.86% accuracy, significantly outperforming earlier machine-learning models. In addition, the proposed integrated deep learning model achieves a high accuracy of 95% when combined with Chi-square.

These findings emphasize the significance of optimization feature and data wrangling to extract the most diagnostically useful biomarkers from complex EEG data. Strong discrimination between ADHD and typical neurological patterns is made possible by specialized preprocessing, class balancing, and Chi-square with PCA-based feature selection.

This work highlights the potential of EEG and machine learning as valuable tools for aiding clinical ADHD evaluations. The optimized models presented in this study have the potential to provide reliable supplementary support for diagnosis, especially in challenging cases. Further validation across diverse patient cohorts would be beneficial.

Ultimately, the utilization of these models could enable earlier and more targeted to enhance the performance of individuals with ADHD. The framework developed in this study lays the foundation for future translational initiatives aiming to maximize the diagnostic utility of machine learning and neurophysiological data.

Acknowledgments

The authors extend their appreciation to the King Salman Center for Disability Research for funding this work through Research Group no KSRG-2023-500.

Data availability statement

Available online: <https://ieee-dataport.org/open-access/eeg-data-adhd-control-children> (accessed on 25 August 2023)

Conflicts of interest

The authors declare no conflicts of interest.

References

1. R. Thomas, S. Sanders, J. Doust, E. Beller, P. Glasziou, Prevalence of attention-deficit/hyperactivity disorder: A systematic review and meta-analysis, *Pediatrics*, **135** (2015), 994–1001. <https://doi.org/10.1542/peds.2014-3482>
2. S. Young, D. Moss, O. Sedgwick, M. Fridman, P. Hodgkins, A meta-analysis of the prevalence of attention deficit hyperactivity disorder in incarcerated populations, *Psychol. Med.*, **45** (2015), 247–258. <https://doi.org/10.1017/S0033291714000762>
3. S. V. Faraone, H. Larsson, S. Org, Genetics of attention deficit hyperactivity disorder, *Mol. Psychiatr.*, **24** (2018), 562–575. <https://doi.org/10.1038/s41380-018-0070-0>
4. M. Arns, C. K. Conners, H. C. Kraemer, A decade of EEG Theta/Beta Ratio Research in ADHD: a meta-analysis, *J. Atten. Disord.*, **17** (2013), 374–383. <https://doi.org/10.1177/1087054712460087>
5. A. R. Clarke, R. J. Barry, R. McCarthy, M. Selikowitz, Excess beta activity in children with attention-deficit/hyperactivity disorder: An atypical electrophysiological group, *Psychiatry Res.*, **103** (2001), 205–218. [https://doi.org/10.1016/S0165-1781\(01\)00277-3](https://doi.org/10.1016/S0165-1781(01)00277-3)
6. S. S. Poil, S. Bollmann, C. Ghisleni, R. L. O'Gorman, P. Klaver, J. Ball, et al., Age dependent electroencephalographic changes in attention-deficit/hyperactivity disorder (ADHD). *Clin. Neurophysiol.*, **125** (2014), 1626–1638. <https://doi.org/10.1016/J.CLINPH.2013.12.118>
7. L. E. Arnold, P. Hodgkins, J. Kahle, M. Madhoo, G. Kewley, Long-term outcomes of ADHD: academic achievement and performance, *J. Atten. Disord.*, **24** (2020), 73–85. <https://doi.org/10.1177/1087054714566076>

8. J. Cook, E. Knight, I. Hume, A. Qureshi, The self-esteem of adults diagnosed with attention-deficit/hyperactivity disorder (ADHD): A systematic review of the literature, *Attent, Deficit Hyperact. Disord.*, **6** (2014), 249–68. <https://doi.org/10.1007/s12402-014-0133-2>
9. M. Adamou, M. Arif, P. Asherson, T. C. Aw, B. Bolea, D. Coghill, et al., Occupational issues of adults with ADHD, *BMC Psychiatry*, **13** (2013), 59. <https://doi.org/10.1186/1471-244X-13-59>
10. S. Dalsgaard, S. D. Østergaard, J. F. Leckman, P. B. Mortensen, M. G. Pedersen, Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: A nationwide cohort study, *Lancet*, **385** (2015), 2190–2196. [https://doi.org/10.1016/S0140-6736\(14\)61684-6](https://doi.org/10.1016/S0140-6736(14)61684-6)
11. T. Chen, C. Shang, P. Su, E. Keravnou-Papailiou, Y. Zhao, G. Antoniou, et al., A decision tree initialised neuro-fuzzy approach for clinical decision support, *Artif. Intell. Med.*, **111** (2021), 101986. <https://doi.org/10.1016/j.artmed.2020.101986>
12. I. Tachmazidis, T. Chen, M. Adamou, G. Antoniou, A hybrid AI approach for supporting clinical diagnosis of attention deficit hyperactivity disorder (ADHD) in adults, *Health Inform. Sci. Syst.*, **9** (2021), 1–8. <https://doi.org/10.1007/s13755-020-00123-7>
13. J. W. Kim, V. Sharma, N. D. Ryan, Predicting methylphenidate response in ADHD using machine learning approaches, *Int. J. Neuropsychoph.*, **18** (2015), 1–7. <https://doi.org/10.1093/ijnp/pyv052>
14. S. Kim, H. Lee, K. Lee, Can the MMPI Predict Adult ADHD? An approach using machine learning methods, *Diagnostics*, **11** (2021), 976. <https://doi.org/10.3390/diagnostics11060976>
15. Y. Zhang-James, E. C. Helminen, J. Liu, B. Franke, M. Hoogman, S. V. Faraone, Evidence for similar structural brain anomalies in youth and adult attention-deficit/hyperactivity disorder: A machine learning analysis, *Transl. Psychiatry*, **11** (2021), 1–9. <https://doi.org/10.1038/s41398-021-01201-4>
16. A. Yasumura, M. Omori, A. Fukuda, J. Takahashi, Y. Yasumura, E. Nakagawa, et al., Applied machine learning method to predict children with ADHD using prefrontal cortex activity: A multicenter study in Japan, *J. Atten. Disord.*, **24** (2020), 2012–2020. <https://doi.org/10.1177/1087054717740632>
17. M. Duda, R. Ma, N. Haber, D. Wall, Use of machine learning for behavioral distinction of autism and ADHD, *Transl. Psychiatry*, **6** (2016), 732. <https://doi.org/10.1038/tp.2015.221>
18. M. Duda, N. Haber, J. Daniels, D. Wall, D. Crowdsourced validation of a machine-learning classification system for autism and ADHD, *Transl. Psychiatry*, **7** (2017), 1133. <https://doi.org/10.1038/tp.2017.86>
19. M. Uluyagmur-Ozturk, A. R. Arman, S. S. Yilmaz, O. T. P. Findik, H. A. Genc, G. Carkaxhiu-Bulut, et al., ADHD and ASD classification based on emotion recognition data, In *Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, CA, USA, 2016, 810–813. <https://doi.org/10.1109/ICMLA.2016.0145>
20. O. Slobodin, I. Yahav, I. Berger, A Machine-Based Prediction Model of ADHD Using CPT Data, *Front. Hum. Neurosci.*, **14** (2020), 383. <https://doi.org/10.3389/fnhum.2020.560021>
21. A. S. Morrow, A. D. Campos Vega, X. Zhao, M. M. Liriano, Leveraging machine learning to identify predictors of receiving psychosocial treatment for Attention Deficit/Hyperactivity Disorder, *Adm. Policy Ment. Health*, **47** (2020), 680–692. <https://doi.org/10.1007/s10488-020-01045-y>
22. M. Moghaddari, M. Z. Lighvan, S. Danishvar, Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG, *Comput. Meth. Prog. Bio.*, **197** (2020), 105738. <https://doi.org/10.1016/j.cmpb.2020.105738>

23. M. Tosun, Effects of spectral features of EEG signals recorded with different channels and recording statuses on ADHD classification with deep learning, *Phys. Eng. Sci. Med.*, **44** (2021), 693–702. <https://doi.org/10.1007/s13246-021-01018-x>
24. S. Khoshnoud, M. Shamsi, M. A. Nazari, Non-linear EEG analysis in children with attention-deficit/hyperactivity disorder during the rest condition, In *Proceedings of the 2015 22nd Iranian Conference on Biomedical Engineering (ICBME)*, Tehran, Iran, 25–27 November 2015, 87–92. <https://doi.org/10.1109/ICBME.2015.7404122>
25. H. Chen, Y. Song, X. Li, A deep learning framework for identifying children with ADHD using an EEG-based brain network, *Neurocomputing*, **356** (2019), 83–96. <https://doi.org/10.1016/j.neucom.2019.04.058>
26. 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** (2014), 162–166. <https://doi.org/10.1016/j.ijpsycho.2013.01.008>
27. S. Saini, R. Rani, N. Kalra, Prediction of Attention Deficit Hyperactivity Disorder (ADHD) using machine learning Techniques based on classification of EEG signal, In *Proceedings of the 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 25–26 March 2022, 782–786. <https://doi.org/10.1109/ICACCS54159.2022.9785356>
28. L. Dubreuil-Vall, G. Ruffini, J. A. Camprodon, Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG, *Front. Neurosci.*, **14** (2020), 251. <https://doi.org/10.3389/fnins.2020.00251>
29. H. T. Tor, C. P. Ooi, N. S. Lim-Ashworth, J. K. E Wei, V. Jahmunah, S. L. Oh, et al., Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals, *Comput. Methods Prog. Bio.*, **200** (2021), 105941. <https://doi.org/10.1016/j.cmpb.2021.105941>
30. H. W. Loh, C. P. Ooi, P. D. Barua, E. E. Palmer, F. Molinari, U. Acharya, Automated detection of ADHD: Current trends and future perspective, *Comput. Biol. Med.*, **146** (2022), 105525. <https://doi.org/10.1016/j.combiomed.2022.105525>
31. H. Christiansen, M. L. Chavanon, O. Hirsch, M. H. Schmidt, C. Meyer, A. Müller, et al., Use of machine learning to classify adult ADHD and other conditions based on the Conners' Adult ADHD Rating Scales, *Sci. Rep.*, **10** (2020), 18871. <https://doi.org/10.1038/s41598-020-75868-y>
32. J. R. Sato, M. Q. Hoexter, A. Fujita, L. A. Rohde, Evaluation of pattern recognition and feature extraction methods in ADHD prediction, *Front. Syst. Neurosci.*, **6** (2012), 68. <https://doi.org/10.3389/fnsys.2012.00068>
33. L. Tan, X. Guo, S. Ren, J. N. Epstein, L. J. Lu, A computational model for the automatic diagnosis of attention deficit hyperactivity disorder based on functional brain volume, *Front. Comput. Neurosci.*, **11** (2017), 75. <https://doi.org/10.3389/fncom.2017.00075>
34. N. A. Khan, S. A. Waheed, A. Riaz, X. Shang, A novel knowledge distillation based feature selection for the classification of ADHD, *Biomolecules*, **11** (2021), 1093. <https://doi.org/10.3390/biom11081093>
35. Y. Sun, L. Zhao, Z. Lan, X. Z. Jia, S. W. Xue, Differentiating boys with ADHD from those with typical development based on whole-brain functional connections using a machine learning approach, *Neuropsychiatr. Dis. Treat.*, **16** (2020), 691–702. <https://doi.org/10.2147/NDT.S239013>

36. X. Peng, P. Lin, T. Zhang, J. Wang, Extreme learning machine-based classification of ADHD using brain structural MRI data, *PLoS One*, **8** (2013), 79476. <https://doi.org/10.1371/journal.pone.0079476>
37. C. J. Vaidya, X. You, S. Mostofsky, F. Pereira, M. M. Berl, L. Kenworthy, Data-driven identification of subtypes of executive function across typical development, attention deficit hyperactivity disorder, and autism spectrum disorders, *J. Child Psychol. Psychiatry*, **6** (2020), 51–61. <https://doi.org/10.1111/jcpp.13114>
38. Y. Tang, X. Li, Y. Chen, Y. Zhong, A. Jiang, C. Wang, High-accuracy classification of attention deficit hyperactivity disorder with 1 2,1-norm linear discriminant analysis and binary hypothesis testing, *IEEE Access*, **8** (2020), 56228–56237. <https://doi.org/10.1109/ACCESS.2020.2982401>
39. H. Chen, Y. Song, X. Li, Use of deep learning to detect personalized spatial frequency abnormalities in EEGs of children with ADHD, *J. Neural Eng.*, **16** (2019), 066046. <https://doi.org/10.1088/1741-2552/ab3a0a>
40. S. Kim, J. H. Baek, Y. J. Kwon, H. Y. Lee, J. H. Yoo, S. Shim, et al., Machine-learning-based diagnosis of drug-naive adult patients with attention-deficit hyperactivity disorder using mismatch negativity, *Transl. Psychiatry*, **11** (2021), 484. <https://doi.org/10.1038/s41398-021-01604-3>
41. A. Ekhlaei, A. M. Nasrabadi, M. R. Mohammadi, Direction of information flow between brain regions in ADHD and healthy children based on EEG by using directed phase transfer entropy, *Cogn. Neurodyn.*, **15** (2021), 975–986. <https://doi.org/10.1007/S11571-021-09680-3/METRICS>
42. M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, *Science*, **349** (1979), 255–260. <https://doi.org/10.1126/SCIENCE.AAA8415>
43. D. Bzdok, M. Krzywinski, N. Altman, Machine learning: supervised methods, *Nat. Methods*, **15** (2018), 5. <https://doi.org/10.1038/NMETH.4551>
44. I. Rish, An empirical study of the naive Bayes classifier, Computer Science, 2001, 41–46. <https://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf>
45. T. Joachims, Training linear SVMs in linear time. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, **2006** (2006), 217–226. <https://doi.org/10.1145/1150402.1150429>
46. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*, **521** (2015), 436–444. <https://doi.org/10.1038/nature14539>
47. K. O’Shea, R. Nash, An introduction to convolutional neural networks, *Neural Evol. Comput.*, **1** (2015). <https://doi.org/10.48550/arXiv.1511.08458>
48. H. Yu, X. Lei, Z. Song, C. Liu, J. Wang, Supervised network-based fuzzy learning of EEG signals for Alzheimer’s disease identification, *IEEE T. Fuzzy Syst.*, **28** (2020), 60–71. <https://doi.org/10.1109/TFUZZ.2019.2903753>
49. H. Yu, X. Wu, L. Cai, B. Deng, J. Wang, Modulation of spectral power and functional connectivity in human brain by acupuncture stimulation, *IEEE T. Neur. Syst. Reh.*, **26** (2018), 977–986. <https://doi.org/10.1109/TNSRE.2018.2828143>
50. K. Li, J. Wang, S. Li, H. Yu, L. Zhu, J. Liu, L. Wu, Feature extraction and identification of Alzheimer’s disease based on latent factor of multi-channel EEG, *IEEE T. Neural Syst. Reh.*, **29** (2021), 1557–1567. <https://doi.org/10.1109/TNSRE.2021.3101240>
51. J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, In *NIPS 2014 Workshop on Deep Learning*, December 2014. <https://doi.org/10.48550/arXiv.1412.3555>
52. A. Alim, M. H. Imtiaz, Automatic identification of children with ADHD from EEG brain waves, *Signals*, **4** (2023), 193–205. <https://doi.org/10.3390/signals4010010>

53. A. Ekhlaei, A. M. Nasrabadi, M. Mohammadi, Analysis of EEG brain connectivity of children with ADHD using graph theory and directional information transfer, *Biomed. Tech. (Berl.)*, **68** (2022), 133–146. <https://doi.org/10.1515/bmt-2022-0100>



AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)