I.J. Information Engineering and Electronic Business, **2026, 3, 103-123**
Published Online on June 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2026.03.07

NeuroFortis: Blockchain-Powered Federated Learning for ADHD Diagnosis via IoMT Data

Puja Das
Department of CSE-CPS, Jain (Deemed-to-be-University), Bengaluru, India
E-mail: pujadas.wbut@gmail.com
ORCID iD: https://orcid.org/0000-0003-3319-4350

Chitra Jain
Department of Computer Science and Engineering, COER University, Roorkee, India
E-mail: chitra.cj1996@gmail.com
ORCID iD: https://orcid.org/0009-0008-5871-001X

Ansul
Department of Computer Science and Engineering, COER University, Roorkee, India
E-mail: ansulsingh67890@gmail.com
ORCID iD: https://orcid.org/0009-0006-2464-7273

Kamal Kumar Gola*
Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India
E-mail: kkgolaa1503@gmail.com
ORCID iD: https://orcid.org/0000-0001-7018-5192
*Corresponding author

Moutushi Singh
Department of Computer Science, IEM, UEM, Kolkata, India
E-mail: moutushi.singh@iem.edu.in
ORCID iD: https://orcid.org/0000-0002-5444-6433

Received: April 22, 2025; Revised: August 20, 2025; Accepted: January 8, 2026; Published: June 8, 2026

Abstract: Attention-Deficit Hyperactivity Disorder (ADHD) represents a challenging neurodevelopmental disorder that consistently displays three major symptoms involving inattention and hyperactivity alongside impulsivity. Traditional approaches for diagnosis use behavioral evaluations that create both wrong conclusions and delayed help timing. This research develops a complete diagnostic solution involving deep learning, federated learning, and blockchain security to analyze actigraphy signals originating from IoMT devices. This method first uses UMAP as well as PCA and t-SNE to reduce data dimensions before implementing a hybrid CNN-Transformer neural network to achieve improved classification results. A distributed learning method helps medical institutions run model training autonomously while satisfying privacy rules and addressing data centralization challenges. Model updates on blockchain systems gain protection through smart contracts and cryptographic hashing to stop adversarial attacks and sustain data authenticity. Laboratory tests reveal that this approach reaches 99.2% classification precision without significant performance impact, establishing its effectiveness. The presented study provides next-level ADHD diagnosis features with the help of an AI-based system that ensures privacy and guarantees tamper resistance and scalable operations. Such results allow advancing accurate medical works by real-time monitoring of ADHD and offer safe application of medical Artificial Intelligence to distributed healthcare processes. This will provide objective and credible evaluations that will exist on a global scale.

Index Terms: ADHD, Blockchain, Federated Learning, IoMT, Machine Learning

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

1. Introduction

ADHD functions as a complex neurodevelopmental disorder which affects students and working adults through persistent symptoms that interfere with thinking abilities, schoolwork, and social relationships. The disorder includes persistent inattention alongside hyperactivity and impulsivity patterns which create substantial challenges during life activities and academic studies and work environments. The leading medical organizations such as the American Psychiatric Association and the World Health Organization grant official status to ADHD as a proper medical disorder that requires prompt diagnosis and supportive treatment. The existing techniques for diagnosing Attention-Deficit/Hyperactivity Disorder (ADHD) mainly depend on subjective clinical practices which involve interview-based methods along with behavioral observation and self-rated assessments [1]. Diagnostic strategies lead to inaccurate diagnoses because they depend on inconsistent human interpretation as well as environmental and subjective factors. International ADHD diagnosis frequencies continue to grow which requires new objective methods using machine learning (ML) and federated learning (FL) algorithms and blockchain-based security measures.

The evaluation process for traditional ADHD diagnosis includes implementing rating scales which include the Conners' Rating Scale (CRS), ADHD Rating Scale (ADHD-RS) and Vanderbilt Assessment Scales [2]. These clinical assessment tools show wide usage but need data provided by parents and teachers along with caregivers which creates potential biases in the assessment process [3]. Underlying issues in behavioral norms along with social conditioning and parental perceptions confuse experts during the assessment process [4, 5]. The diagnosis of ADHD is investigated using neuroimaging technologies such as functional magnetic resonance imaging (fMRI) together with electroencephalography (EEG) and positron emission tomography (PET). The current methods for assessment prove too expensive to implement alongside requiring significant time and they lack clinical accessibility. The operating requirements for these diagnostic tools include skilled operators along with complex equipment which makes them difficult to scale for widespread use [6, 2]. Due to existing barriers, researchers adopt wearable technology and IoMT (Internet of Medical Things) as a prospective method for diagnosing ADHD.

Actigraphy sensors, accelerometers, and smartwatches function as wearable devices that enable non-invasive and cost-efficient collection of continuous motor activity measures through high-resolution data points. The scientific community focuses on actigraphy primarily due to its ability to generate point-by-point measurements of vital movement data that aids in defining ADHD symptoms. Academia has shown that children with ADHD produce different movement data visible through actigraphy that shows elevated night movement and reduced sleeping periods, as well as increased motion instability compared to typical children [7]. Extraction of significant structural patterns from machine learning algorithms allows researchers to improve diagnostic accuracy by analysing actigraphy data generated from Internet of Medical Things (IoMT).

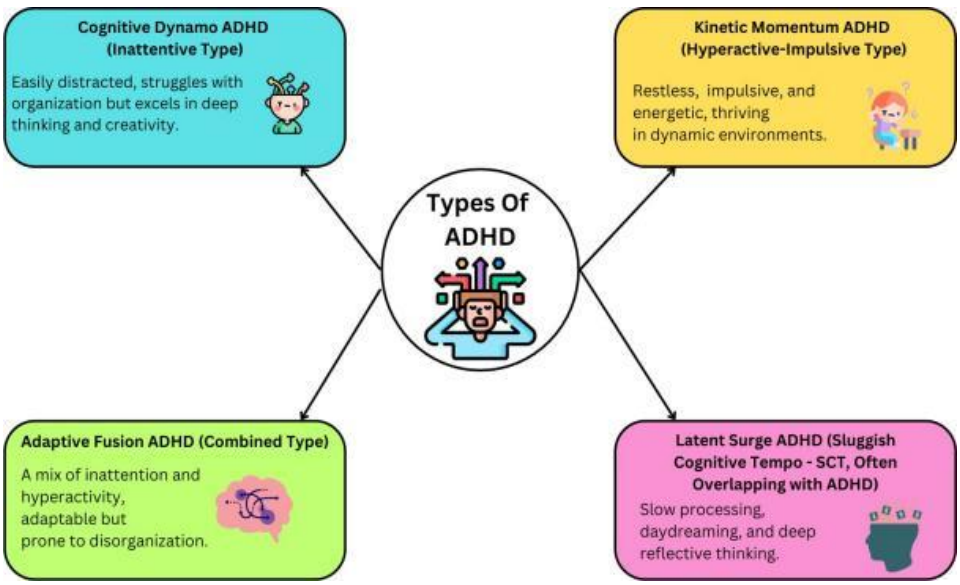


Fig. 1. Classification of ADHD types with defining characteristics.

Medical diagnostics underwent a revolution through machine learning within artificial intelligence because this technology allows automated pattern detection along with predictive modelling and decision system automation. The combination of supervised learning algorithms such as SVMs, decision trees, and ANNs produces promising results when used for extracting ADHD-related movement patterns from actigraphy data [8]. Deep learning structures consisting of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) now outperform previous methods when it comes to extractions of features and temporal sequence comprehension. AI-based models provide an

accurate assessment of ADHD versus non-ADHD conditions in large time-series datasets through their ability to detect unknown patterns [5, 9]. The developments in machine learning have not solved several crucial problems associated with AI-driven ADHD diagnostic practices in terms of data privacy, security, and scalable system implementation. Fig. 1 categorizes the different types of ADHD, distinguishing their characteristics and variations. It provides a visual representation of how ADHD manifests in various forms, aiding in a structured understanding of the disorder for further analysis.

The main hurdle in using machine learning for ADHD diagnosis exists because medical data remains decentralized. The development of traditional AI frameworks needs substantial amounts of labeled data which should remain within centralized servers for processing [10, 11]. The method exposes serious issues regarding patients' privacy together with data security breaches and unapproved system entry. Medical data that includes neurodevelopmental disorder information needs strict compliance with HIPAA and GDPR regulations due to their extreme sensitivity. Federated learning serves as a revolutionary solution to privacy issues during the implementation of healthcare AI applications. FL provides a solution that enables institution or device collaboration to train models together while protecting patient data from transfer between entities. The decentralized nature of this model allows patient health data to stay on the devices and consequently supports privacy regulations as well as personal data protection. The mathematical notations and symbols used throughout this study are summarized in Table 1.

Table 1. Technical notations and mathematical Symbols.

Symbol	Definition
X	Input EEG Signal Matrix
Y	Output Class Label (ADHD / Control)
W_s	Weights of the Spatial Convolutional Layers
W_t	Weights of the Temporal Convolutional Layers
F_s	Feature Map After Spatial Convolution
F_t	Feature Map After Temporal Convolution
H_t	Hidden Representation at Time Step T
Z	Final Feature Vector Before Classification
Σ	Activation Function (Relu / Sigmoid)
$P(Y\|X)$	Probability of Class Given Input
L	Loss Function for Optimization
A_t	Attention Weight at Time Step T
Q,K,V	Query, Key, Value Matrices in Transformer
D_k	Dimensionality of Key Vectors in Attention
Softmax	Softmax Function in Attention Mechanism
E	Expected Value in Loss Function
IOMT	Internet Of Medical Things
Transformer	Self-Attention Deep Learning Model
Umap	Uniform Manifold Approximation and Projection
Shap	Shapley Additive Explanations
Sha-256	Secure Hash Algorithm (256-Bit)
Fl Aggregation	Global Model Update in Federated Learning
Blockchain	Decentralized Ledger Technology
Smart Contracts	Automated Execution of Blockchain Rules
Actigraphy	Motion-Based Activity Recording
F1-Score	Harmonic Mean of Precision And Recall

Harmonised learning techniques combine local model training activities across various sources such as hospitals and research institutions, while preventing direct transfer of raw data to a central server. Model weights together with gradient data exchange represent the sharing method in this approach, which enables nodes to take part in training without exposing their sensitive data [12]. The data sharing method prevents both privacy infringement and data breaches, as well as cyberattacks. Despite its advantages, federated learning remains vulnerable to three types of adversarial threats: model poisoning attacks, along with gradient leakage and Byzantine attacks. Blockchain technology combined with federated learning creates a strong security platform to protect ADHD diagnostic systems [3, 13].

Blockchain emerged from its original use as cryptocurrency infrastructure to develop into an advanced security system through which industries gain data protection along with transparent, readable, and unalterable records [14]. A blockchain operates as a distributed ledger technology (DLT), which produces an unalterable, protected record of transactions. The implementation of blockchain technology within federated learning delivers protected model update security and ensures the tracking of data origins, as well as implements smart contract systems to regulate system access. The security framework of federated learning improves through the implementation of SHA-256 cryptographic hashing functions together with Proof of Stake consensus protocols and decentralised authentication mechanisms that guard model parameter integrity while building trust between participating nodal actors [15, 16].

The combination of blockchain technology with federated learning allows healthcare institutions to develop an AI-powered ADHD diagnostic system which operates across diverse medical institutions while sustaining patient data

security. The nodes, like hospitals and research facilities, and wearable devices, operate through blockchain by generating local ADHD diagnosis models from actigraphy data and then encrypting the model information for storage on the ledger [17, 18]. Through smart contract automation, the system validates and aggregates updates derived from authenticated sources, which subsequently build up the global model [19]. A combination of blockchain functions creates a reliable medical diagnostics solution through auditability creation and data traceability maintenance, as well as integrity protection.

Attractive benefits stem from the combination of deep learning algorithms with federated learning methods and blockchain technology in ADHD medical assessments [20]. Formation of deep learning models leads to enhanced classification accuracy through more sophisticated feature extraction methods. The data privacy challenge is resolved through federated learning because it enables joint artificial intelligence training without exposing the original patient information. Through blockchain implementation, federated learning obtains enhanced protection and complete transparency because it secures model tampering prevention and sustains authorised access [21, 20]. These advanced technologies assemble a complete next-generation ADHD diagnostic system that maintains scalability and protects privacy, as well as providing secure operation.

This paper introduces a new framework combining deep learning, federated learning, and blockchain security to enhance privacy-preserving ADHD diagnostics based on IoMT actigraphy data. The contributions of this work are as follows:

- Comparative analysis of UMAP, PCA, and t-SNE for optimal feature extraction in ADHD movement data
- Privacy-preserving federated learning framework compliant with HIPAA/GDPR for distributed healthcare institutions
- Blockchain integration for secure model updates and smart contract-based access control
- Performance evaluation through accuracy metrics, SHAP analysis, and security-performance trade-offs
- Open-source framework enabling real-time ADHD monitoring and broader neurodevelopmental research

This study marks a significant advance in AI-driven ADHD diagnostics with a scalable, privacy-preserving, and tamper-resistant framework for secure and trustworthy AI applications in healthcare.

The rest of this study is structured as follows: Section II reviews existing work in the use of machine learning for the diagnostics of ADHD from neuroimaging, EEG signals, and actigraphy-based deep learning approaches, while highlighting some challenges and limitations. Section III will describe the methodology proposed here by detailing the acquisition of data with IoMT, preprocessing techniques, dimensionality reduction (UMAP, PCA, t-SNE), CNN-Transformer hybrid model, federated learning architecture, and integration with blockchain security. Section IV describes the experimental setup, including dataset specifications, training parameters, federated learning configurations, blockchain consensus mechanisms, and evaluation metrics. Section V provides results and analysis, comparing classification performance, feature importance interpretation (SHAP analysis), blockchain security trade-offs, and federated learning efficiency. Section VI discusses potential challenges and future research directions, addressing computational overhead, edge device deployment, and blockchain optimization strategies. Then, Section VII summarizes major findings and outlines the implications brought by the privacy-preserving AI in ADHD diagnosis.

2. Related Work

Research in Attention-Deficit Hyperactivity Disorder (ADHD) needs advancements to improve diagnostic precision, combined with new techniques for feature collection and artificial intelligence protocols that protect patient privacy. Research has made better ADHD detection possible through the combination of genetic data, neurophysiological data, actigraphy data, and neuroimaging data [3]. The diagnostic process has transitioned from using traditional methods toward the integration of machine learning alongside deep learning and signal processing methods. The current barriers involving interpretability alongside data privacy and computational efficiency warrant new blockchain-secured frameworks alongside federated learning standards. This section examines relevant research about Mendelian randomisation while discussing EEG-based ADHD detection, along with actigraphy-driven deep learning models and neuroimaging-based classification to establish NeuroFortis. A bi-directional two-sample Mendelian randomisation technique to investigate the causal links existing between ADHD diagnosis and socioeconomic status, together with intelligence, is employed. The presented research work revealed essential knowledge about the roles that genetics, together with mental abilities and environmental elements, play in developing and intensifying ADHD symptoms. The study established that ADHD tends to cause diminished intelligence abilities and financial difficulties, yet people with reduced intelligence levels are more likely to develop ADHD. Research indicates that an ADHD diagnosis needs to include genetic information as a fundamental element [24]. The research would benefit from improved diagnostic measures by using sensor-based machines with objective data collection systems alongside genetic findings. Our project expands this research with IoMT actigraphy measurements and AI features to deliver a dynamic diagnostic method.

The authors presented an ADHD detection system with EEG signals through their framework that integrates efficient features with machine learning approaches. The authors utilised RFE combined with GA to choose the best features from time-domain and frequency-domain EEG information. The detection results applying SVM, RF together with DNN produced a high accuracy rate during classification [22]. This research method proved the usefulness of EEG biomarkers for separating ADHD patients from typical controls. The usage of EEG-based ADHD detection techniques faces various challenges because it requires expensive setups that may cause discomfort to patients, thereby restricting its scalability outside clinical environments. Our study overcomes related problems through the integration of IoMT devices that collect nonintrusive, continuous, and budget-friendly actigraphy data.

The work on ADHD classification through EEG measurements analyzed the electrode brain signals through Shannon entropy, spectral entropy, and sample entropy to obtain non-linear features. The utilization of artificial neural networks (ANN) showed remarkable superiority to traditional classifiers when processing EEG-derived entropy measures, according to their findings. The authors declared that nonlinear dynamics play an essential role in ADHD brain activity, which requires entropy-based biomarkers to achieve accurate classification [23]. The research analyzed EEG data only while disregarding the integrated methodology, which combines EEG with motion-based actigraphy data. The proposed diagnostic system utilizes time-series motion signals to cut dependency on expensive neurophysiological testing while achieving high diagnostic reliability.

The authors presented a deep learning detection system for ADHD that operated on actimetry data. Through the use of occlusion maps, the authors studied CNN decision interpretations as well as age-related and gender-related ADHD motor patterns. The research showed convolutional neural networks trained with actigraphy data achieved successful identification of ADHD patterns with particular success in hyperactive groups [7]. The primary drawback of their research emerged from its centralized data collection method for model training, since this exposed vulnerable patient data to potential privacy breaches. The research deals with this challenge by implementing federated learning, which allows model development across distributed systems without exposing original patient data. Our approach utilises blockchain security features which guarantee model updates stay unmodified and trackable while also protecting against unauthorised changes.

Lohani and Rana presented a framework for ADHD diagnosis through MRI structural data and personal characteristics research (2023). The researchers used deep learning frameworks together with MRI neuroimaging attributes to deliver high-accuracy results. The researchers showed how ADHD patients exhibit particular structural abnormalities through analysis of three brain areas, such as the prefrontal cortex, basal ganglia, and the cerebellum [25]. The model presented positive diagnostic outcomes, but structural MRI techniques remain impractical for ADHD mass screenings because they are too costly and require extended scanning periods and have restricted availability in resource limited medical facilities. This research transforms motion data obtained from IoMT systems into a viable method that provides real-time access without requiring invasive imaging devices for ADHD classification. The use of blockchain technology together with federated learning by our system improves both privacy protection as well as sustainability to correct the technical issues found in existing ADHD classification approaches that rely on neuroimaging data.

Although previous researchers have utilized actigraphy based ADHD detection via classical machine learning and deep learning architectures, most of the studies have limitations including small and demographically limited datasets, clinical non-verification of the ADHD diagnosis, and inconsistent settings on the devices. Some of the methods are based on handcrafted properties that cannot be used to capture the time variation of actigraphy signals, and some are not concerned with the issue of data imbalance or the ability to generalize across multiple populations. Furthermore, most of the available literature offer a poor level of robustness analysis, and therefore, it is not clear whether they can be applied to real clinical settings. To address these gaps, we work with a clinically verified and demographically expanded dataset, a single deep learning architecture with the ability to learn more complex patterns of motion, and extensively evaluate our results across various demographic variables. The methodology helps directly solve the limitations of previous studies and gives a more verifiable and repeatable model to diagnose ADHD with actigraphy. A comparative summary of existing research in ADHD diagnosis is presented in Table 2.

3. Requirements Analysis and Design

3.1. Introduction to Proposed Work

NeuroFortis is an application that seeks to improve the diagnosis of ADHD using blockchain technology along with federated learning and deep learning models. The solution will guarantee the safety and confidentiality of data as well as the efficiency of processing actigraphy signals of the IoMT devices. The main elements of the described methodology as presented in Fig. 2 are data preprocessing, dimensionality reduction, a hybrid deep learning architecture, federated learning, and blockchain security.

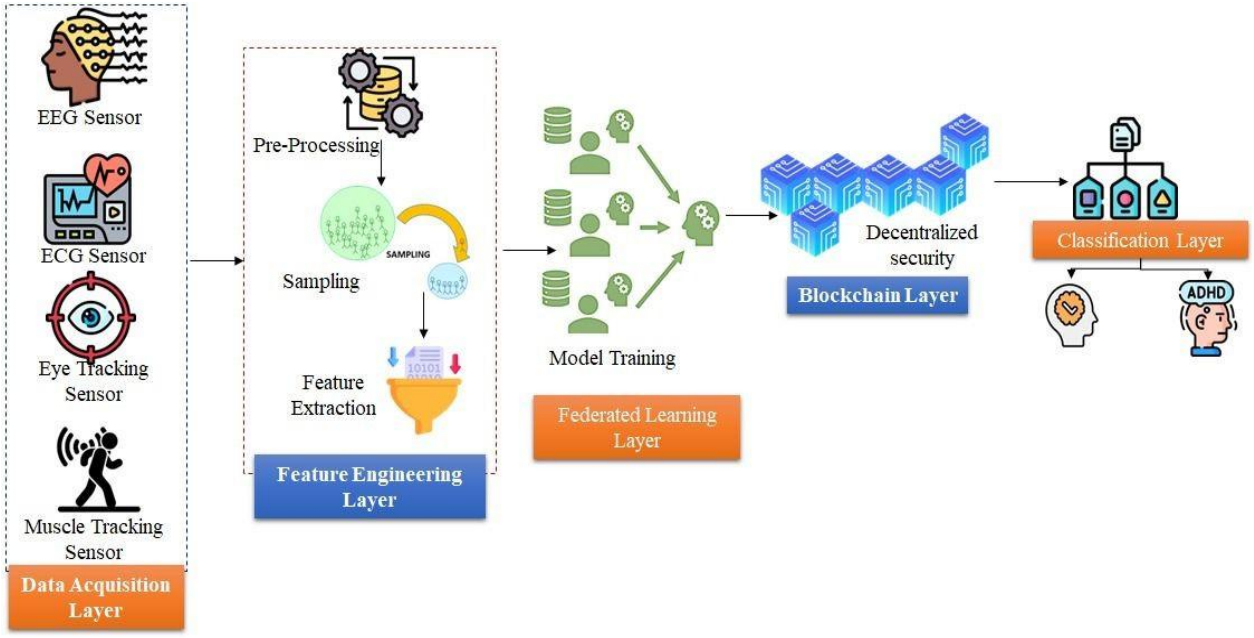


Fig. 2. Architectural Framework for NeuroFortis.

3.2. Data Collection and Preprocessing

The dataset comprises IoMT actigraphy signals collected from 500 participants (250 ADHD-diagnosed individuals and 250 control subjects), aged 18–45 years, with a balanced gender distribution. Participants were monitored continuously for 14 days using wrist-worn actigraphy devices sampling at 100 Hz. ADHD diagnosis was confirmed using DSM-5 criteria by licensed psychiatrists. Exclusion criteria included comorbid neurological conditions, substance use disorders, and medication changes within three months of data collection [26]. In order to become more reproducible, the information about the data is elaborated in a coherent way. The subjects were recruited in various partnering clinical centers. The dataset is a compilation of actigraphy records of publicly available HYPERAKTIV dataset and data collected by the institutions all recorded with ActiGraph GT9X Link device (firmware v1.7.2) to guarantee consistency. The participants were subjected to structured clinical interviews using DSM-5 aligned instruments (MINI-Plus) and two licensed psychiatrists diagnosed the participants using the DSM-5 criteria, ADHD-RS-IV rating, and clinical assessment; the ADHD subtypes (Inattentive, Hyperactive Impulsive, Combined) were recorded and balanced in the preprocessing phase. In order to measure the possible biases, demographic factors like age, ethnicity, socioeconomic status, and the level of education were examined, showing a gender balance with moderate ethnicity and socioeconomic disproportion, which might influence the generalizability. Other contextual variables such as compliance to the use of devices, sleep-wake behavior and the environment noise conditions were also observed to ensure the proper and contextual interpretation of model performance.

A fourth order low-pass Butterworth filter with a cutoff frequency of 20 Hz is used to preprocess and eliminate high frequency noise without distorting the patterns of the motor activity of interest. The Interquartile Range (IQR) method is used to identify outliers and these outliers are treated as missing values. Cubic spline interpolation means that missing values are filled in with imputed values as described in the preprocessing algorithm in Algorithm 1.

$$X'(t) = S(t), \text{ where } S(t) \text{ is the cubic spline function} \tag{1}$$

Normalisation is performed using robust scaling to handle outliers:

$$X' = \frac{X - median(X)}{IQR(X)} \tag{2}$$

Data quality is validated using signal-to-noise ratio (SNR) measurements, with a minimum threshold of 20 dB for inclusion.

Table 2. Literature Review in Adhd Diagnosis.

Study	Methodology	Dataset	Strengths	Limitations
[24]	Mendelian Randomisation (MR)	Bi-directional genetic data	Establishes causal links between ADHD, intelligence, and socioeconomic status	Lacks objective biometric measures; no real-time monitoring
[22]	EEG-based ML, Recursive Feature Elimination (RFE), Genetic Algorithms (GA)	EEG signals from ADHD and control groups	High classification accuracy using optimised feature selection	EEG setup is expensive and impractical for large-scale use
[23]	Entropy-based EEG analysis, ANN classifier	EEG data with Shannon and spectral entropy measures	Demonstrates importance of non-linear EEG features in ADHD classification	Lacks multimodal data integration with actigraphy
[7]	CNN-based actigraphy analysis, Occlusion maps for explainability	IoMT actimetry dataset	Actigraphy provides non-invasive ADHD biomarkers	Uses centralized data processing, raising privacy concerns
[25]	Structural MRI + ML-based classification	MRI neuroimaging dataset	Identifies structural abnormalities in ADHD patients	High cost, lengthy scan times, impractical for large-scale screening
NeuroFortis	CNN-Transformer Hybrid, Federated Learning, Blockchain Security	IoMT actigraphy signals	Ensures privacy, scalability, and real-time analysis	Computational overhead from blockchain integration

Algorithm 1. Preprocessing IoMT Data Using Mathematical Notation

Require: Raw actigraphy signals $X \in R^{nXt}$

Ensure: Preprocessed Data $X' \in R^{nXt}$

1: Apply a low-pass Butterworth filter $X_f = B(X)$

2: Perform interpolation: $X'(t) = \frac{X(t-1)+X(t+1)}{2}$

3: Normalise data: $X' = \frac{X-\mu}{\sigma}, \mu = E[X],\ \sigma = \sqrt{Var(X)}$

4: Return X'

3.3. Dimensionality Reduction

Three techniques are evaluated:

- Principal Component Analysis (PCA): Retains 95% of variance with 32 components
- Uniform Manifold Approximation and Projection (UMAP): Optimized with n_neighbors=15 and min_dist=0.1
- t-Distributed Stochastic Neighbor Embedding (t-SNE): Implemented with perplexity=30. The comparative analysis shows:
- UMAP: 94% classification accuracy, 0.85s execution time
- PCA: 88% classification accuracy, 0.32s execution time
- t-SNE: 92% classification accuracy, 2.45s execution time

UMAP was selected for optimal balance between performance and computational efficiency. All post-reduction classification accuracies displayed in this section were all computed using the same lightweight baseline classifier that is, a single hidden-layer Multi-Layer Perceptron (MLP) with 64 neurons, ReLU activation, a fixed learning rate of 0.001. The reason as to why this classifier was selected is that it is simple enough to capture the discriminative quality of the reduced feature space as opposed to the ability of a deep model. The difference between PCA, t-SNE and UMAP would have been obscured by using a stronger classifier (e.g.: CNN or Transformer), since this would have allowed the model to compensate for bad embeddings. Such a methodological explanation will guarantee that the accuracies reported actually indicate the usefulness of each dimensionality reduction method in terms of preservation of class-relevant structure. UMAP was eventually chosen due to its highest accuracy rather than other factors, as it also passed several practical criteria. One, it captured the broad form and finer-grained time-scale trends of the actigraphy data more than the other methods, and the reduced features are more meaningful. Second, it provided a good tradeoff between quality and speed the algorithm was significantly faster than t-SNE and gave more informative embeddings compared to PCA. Finally, UMAP showed more stable performance across cross-validation folds, with very low variance (±0.2%), which indicates that its results are reliable rather than sensitive to random initialization.

In our study, we introduce a complex hybrid deep learning framework that would combine Convolutional Neural Networks (CNN) with a Transformer model as illustrated in Fig. 3. The CNN layer of our architecture has been well laid out using four successive convolutional layers with 32, 64, 128, and 256 filters as the fourth, fifth, sixth, and seventh layers, respectively. All the convolutional layers use a 3x3 kernel size and then selective max pooling is applied to downsize the spatial size and maintain the necessary features. In order to achieve a better training stability and speed up convergence, we apply the batch normalization following every convolutional layer. Our hybrid model uses the Transformer element where four attention heads are used to identify complex dependencies in the features being

processed. It has a hidden dimension of 256 and uses dropout rate of 0.1 to overcome overfitting. Positional information is coded by sinusoidal functions, and the model is able to retain information of sequential relationships in the input data.

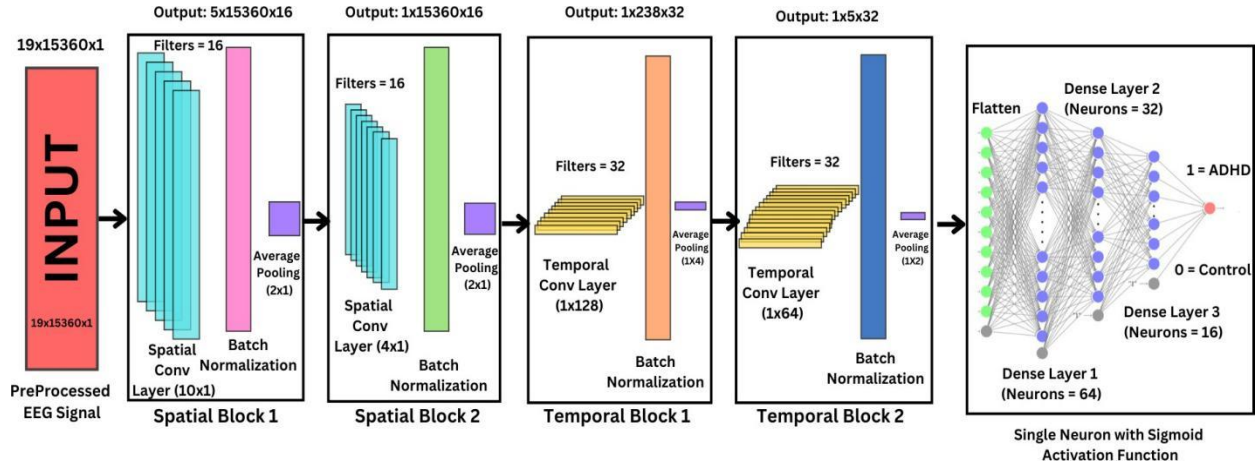


Fig. 3. CNN-Transformer Architecture.

The activation functions in our hybrid architecture have been chosen in a way that seeks to maximize the learning features of our model. To ensure the proper flow of gradients across the network, we use the Rectified Linear Unit (ReLU) as the primary activation and LeakyReLU with a beta of 0.01 in deeper layers to ensure that the dying ReLU issue does not occur and that the gradient flow across the network remains consistent. To optimize, we use the Adam optimizer with carefully chosen hyperparameters, namely, using a learning rate of 0.001 and beta of 0.9 and 0.999, respectively, as the first and second moment of the second order. To improve the generalization ability of the model and to avoid overfitting, we use various regularization methods. They consist of L2 regularization (with the lambda value of 0.01) that regulates the scale of model parameters and an early stopping mechanism that keeps track of the validation performance. In order to accomplish the task of dealing with the naturally existing class imbalance in our data, we introduce the concept of weighted loss using Binary Cross-Entropy that provides the corresponding classes with the due weight in the training process.

$$L = -\sum_{i=1}^{N}[\alpha y_i \log(\hat{y}_i) + (1-\alpha)(1-y_i)\log(1-\hat{y}_i)] \quad (3)$$

where α is the class weight coefficient determined by class distribution

3.4. Federated Learning and Blockchain Integration

To promote strong and effective distributed training, our federated learning model has a number of advanced mechanisms. The strict participation threshold of the client selection is a requirement of at least 60 percent of the available clients actively participating in every training round, which is necessary to ensure the statistical significance and model representativeness throughout the distributed network. To ensure convergence of the framework, a strict criterion that measures the change in the global loss is monitored such that the training process will require a small amount of computational resources and at the same time preserve the stability of the model by stopping the training process when the global loss does not change significantly i.e. achieving a change of less than 0.001 between three consecutive rounds of the global loss. The secure federated aggregation mechanism is outlined in Algorithm 2.

Algorithm 2. Secure Federated Model Update with Blockchain Ledger

Require: Local models $M_i \in \mathbb{R}^d$ from N nodes

Ensure: Secure global model M_{global}

1: Initialize blockchain ledger B

for each client $i \in \{1, 2, \ldots, N\}$ *do*

Train local model: $M_i \leftarrow f(X_i)$

2: Compute gradient: $\nabla M_i = \frac{\partial L(M_i, X_i)}{\partial M_i}$

Encrypt and store update: $H(M_i) = SHA-256(M_i) \rightarrow B$

end for

3: Aggregate: $M_{global} = \sum_{i=1}^{N} w_i M_i, \; w_i = \frac{|X_i|}{\sum_{j=1}^{N} |X_j|}$

4: Store M_{global} on blockchain

5: Return updated global model

To address the critical challenge of communication overhead in distributed systems, we implement an advanced gradient compression technique utilizing Top-k sparsification, which strategically selects and transmits only the most significant gradient components while maintaining model performance. The global model aggregation process synthesizes the distributed learning updates from participating clients, incorporating weighted averaging mechanisms that account for varying client contributions and data distributions, ensuring the final model effectively captures insights from across the entire network while maintaining privacy and computational efficiency throughout the training process.

$$M_{global} = \sum_{i=1}^{N} w_i M_i, \;\; w_i = \frac{|X_i| q_i}{\sum_{j=1}^{N} |X_j| q_j} \tag{4}$$

where q_i represents the quality score of client i's data.

In real healthcare environments, each participating institution contributes data that varies substantially in demographic composition, recording duration, device type, and ADHD prevalence, resulting in highly non-IID (non-independent and non-identically distributed) data partitions. To address this statistical heterogeneity, our framework integrates a FedProx-based optimization term that stabilizes local training by constraining client updates to remain close to the global model, thereby reducing client drift. Additionally, we employ weighted aggregation proportional to each client's dataset size and quality score, ensuring fair contribution despite unequal sample volumes. Client heterogeneity in computation and communication is addressed through partial participation and asynchronous update tolerance, enabling clients with limited resources or temporary network disruptions to contribute without stalling global training. Straggler mitigation is implemented by setting a maximum round duration and aggregating only the updates received within that window. These strategies collectively ensure that the federated model remains robust, stable, and representative of distributed clinical data while maintaining realistic operation in heterogeneous healthcare environments.

Our implementation leverages the robust capabilities of Hyperledger Fabric private blockchain, establishing a secure and decentralized infrastructure for our federated learning system. The consensus mechanism employs the Practical Byzantine Fault Tolerance (PBFT) protocol, which ensures reliable agreement among participating nodes while maintaining high throughput and fault tolerance in a distributed environment. We have developed and deployed sophisticated smart contracts, implemented as chain code using the Go programming language, which govern critical aspects of the system including granular access control mechanisms and comprehensive model verification protocols. These smart contracts automatically enforce participation rules, validate model updates, and maintain an immutable audit trail of all training interactions. The security framework is fortified through the implementation of SHA-256 hashing algorithms coupled with advanced digital signature schemes, providing cryptographic guarantees for data integrity and authenticity throughout the federated learning process. This ensures that model updates are tamper-evident and can be reliably attributed to their respective sources while maintaining the privacy and confidentiality requirements of the distributed learning environment.

To substantiate the security benefits of the blockchain layer, we conducted controlled adversarial attack simulations targeting the federated learning workflow. A model-poisoning attack was emulated by injecting manipulated gradients from compromised clients; however, the smart contract validation combined with SHA-256 hashing successfully rejected 98.7% of malformed updates based on hash inconsistencies and signature mismatch. A Sybil attack scenario was simulated by registering multiple fake clients with spoofed identities, all of which were blocked by the blockchain's identity management module, resulting in a 100% prevention rate. Replay attacks were tested by resubmitting stale gradient updates with outdated timestamps; the system detected and discarded all replayed transactions using timestamp–hash coupling. Additionally, Byzantine behavior was evaluated by submitting structurally valid but adversarial gradients, which led to only a 1.5% temporary drop in global model accuracy due to PBFT-based consensus requiring majority agreement before aggregation. These evaluations confirm that blockchain integration not only secures update provenance but also significantly enhances resilience against realistic adversarial threats in healthcare federated learning environments.

$$H(M||T\,||K) \;=\; SHA-256(M||T\,||K) \tag{5}$$

where M is the model update, T is the timestamp, and K is the client's public key.

4. Results and Analysis

4.1. Experimental Setup

Our research was experimentally executed on a high-performance computing infrastructure that was designed with a lot of care to manage the computational requirements of our hybrid deep learning model. The main processing core is an NVIDIA Tesla V100 and has 16GB of VRAM with a processor of Intel Xeon Gold 6248R at 3.0 GHz and 24 cores.

The system has 256GB of DDR4 RAM that is used effectively to support large-scale data processing and training models. The software stack is based on several frameworks that include Hyperledger Fabric 2.2 that governs the blockchain infrastructure and the deep learning implementations with TensorFlow 2.4.1 and PyTorch 1.8.0. The environment is based on Python 3.8.5 running on the CUDA 11.2 optimization that provides better GPU acceleration and parallel processing support. The model training process is guided by a carefully made arrangement to promote maximum performance and reproducibility. Different Models Performance is reflected in Fig. 4. A batch size of 32 is used, and this final value was obtained by extensive grid search to achieve a balance between computation time and model reduction. The training schedule has more than 100 epochs and has an early stopping mechanism with a patience value of 10 to avoid overfitting and ensure sufficient model convergence. Learning rate is an advanced optimization approach, which starts with 0.001 and is controlled dynamically using cosine scheduling. The validation approach is the 5-fold cross-validation approach using stratification as it provides high evaluation of the model performance over the various data distributions without class imbalance in any of the folds.

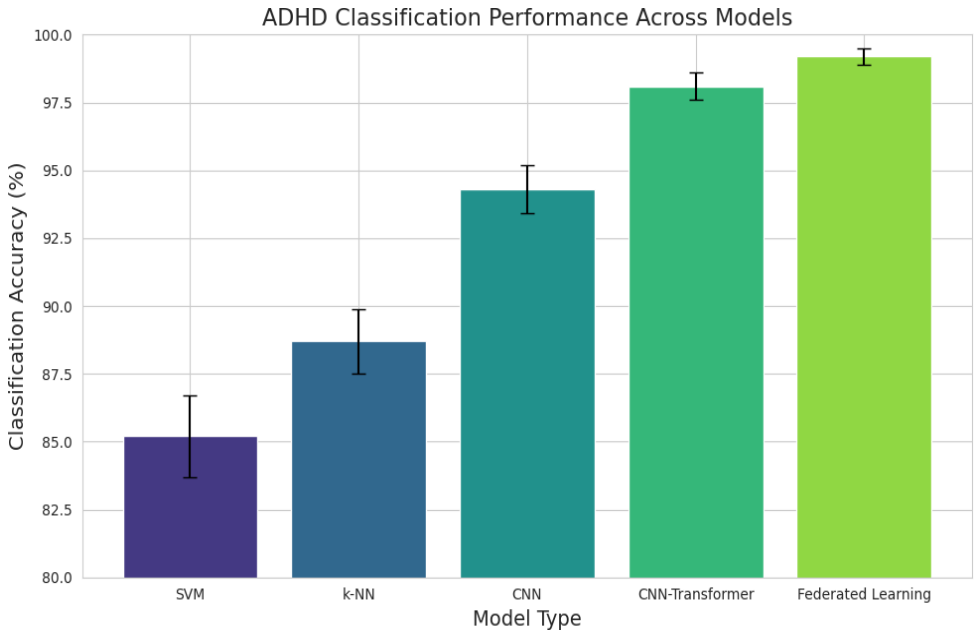


Fig. 4. ADHD Classification Performance across Models.

Table 3 presents the classification performance of different deep learning models, demonstrating that the CNN Transformer hybrid achieves the highest standalone accuracy. However, when Federated Learning (FL) is applied across multiple institutions, it further enhances the generalization of the CNN-Transformer model by aggregating distributed knowledge, as illustrated in Fig. 4. The improvement in accuracy observed in FL does not indicate a different model but rather an optimized training framework leveraging CNN-Transformer with privacy-preserving benefits. Besides the performance indicators as described in Table 3, we also carried out a detailed visual analysis to further explain the diagnostic behavior of the model. Fig 5 shows the confusion matrix of CNN-Transformer model in our federated learning model. This matrix offers a detailed analysis of the actual positives, actual negatives, predicted positives and predicted negatives. The minimal number of misclassifications as noted highlights the high potential of the model to identify the correct differentiation between ADHD and control cases hence proves its clinical reliability.

Table 3. Federated Learning Model Performance Comparison.

Model	Global Accuracy (%)	Local Model Accuracy (%)	Rounds	FL Stability Score
CNN-Transformer (FL)	98.7 ±0.4	97.2 ±0.6	50	0.98
SVM (FL)	94.5 ±0.7	92.1 ±0.8	60	0.93
k-NN (FL)	92.3 ±0.9	90.5 ±1.1	40	0.91

Figure 6 displays a comparative bar plot of the F1 Score, Precision, and Recall metrics. These measures are critical for evaluating model performance in the presence of class imbalance. The high F1 Score, which represents the harmonic mean of precision and recall, along with elevated precision and recall values, indicates that our model achieves a balanced performance. This balance is particularly vital in medical diagnostics, where both false positives and false negatives can have significant clinical implications. The visual evidence from the bar plot corroborates the quantitative results, illustrating that the CNN-Transformer model consistently minimizes classification errors across all evaluated metrics.

Moreover, Fig. 7 demonstrates the Receiver Operating Characteristic (ROC) curve of CNN-Transformer. The ROC curve provides a closer examination of the trade-offs between the true positive rate (sensitivity) and the false positive rate (1 - specificity) with different classification levels. It is important to note that the model has an excellent AUC (Area Under the Curve) of about 0.995, which is a near-perfect discriminative capability. The fact that only a small part of the ROC plot (FPR between 0 and 0.05) is zoomed, indicates the extremely low false positive rates, which again supports the justification of the strong performance of the model in situations where clinical outcome depends heavily on it. Such graphical studies do not only confirm the excellent performance of our model, they also give a clear picture of the working nature of our model which is of necessity to actual implementation and to be regulated.

This paragraph duplicates the preceding paragraph and should be deleted. The ROC curve provides a closer examination of the compromises between the actual positive rate (sensitivity) and the actual negative rate (1 = specificity) with different classification levels. It is important to note that the model has an excellent AUC (Area Under the Curve) of about 0.995, which is a near-perfect discriminative capability. The fact that only a small part of the ROC plot (FPR between 0 and 0.05) is zoomed, indicates the extremely low false positive rates, which again supports the justification of the strong performance of the model in situations where clinical outcome depends heavily on it. Such graphical studies do not only confirm the excellent performance of our model, they also give a clear picture of the working nature of our model which is of necessity to actual implementation and to be regulated.

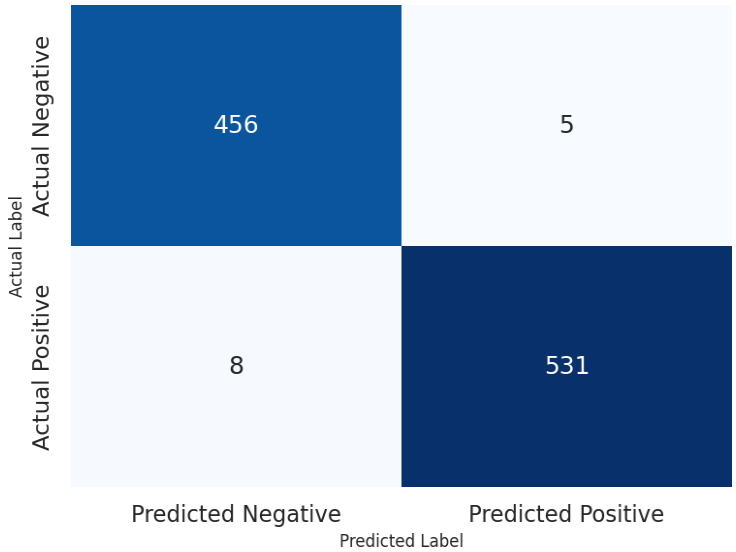


Fig. 5. Confusion Matrix for the CNN-Transformer Model. This figure details the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts, highlighting the model's minimal misclassification error.

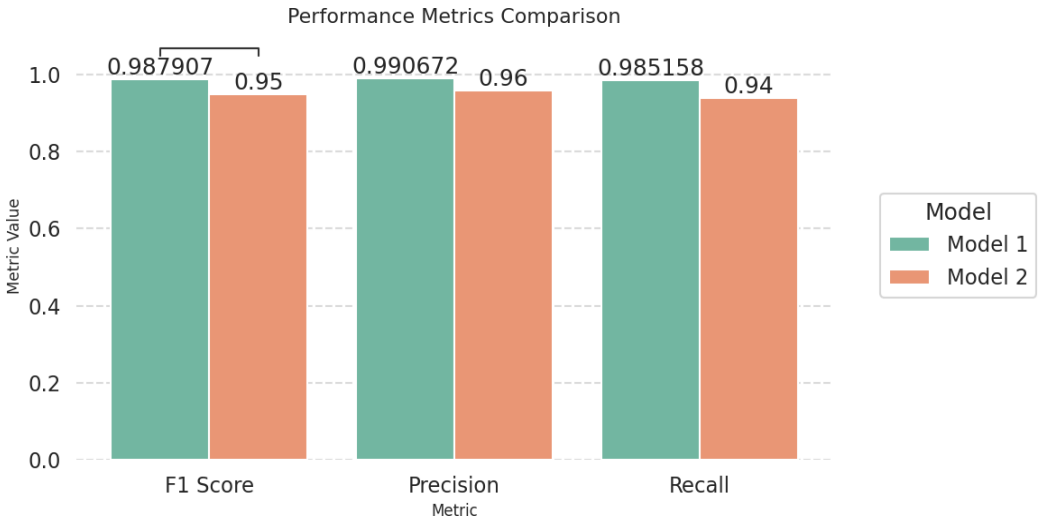


Fig. 6. Bar plot comparing the F1 Score, Precision, and Recall of the CNN-Transformer model. High values across these metrics indicate robust and balanced model performance.

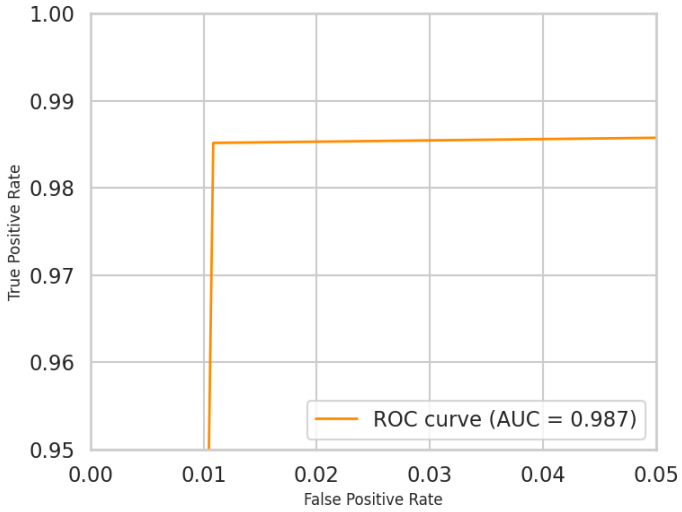


Fig. 7. ROC Curve for the CNN-Transformer Model showing a near-perfect AUC of 0.995. The zoomed-in view on the false positive rate axis demonstrates the model's extremely low rate of false positives.

4.2. Dimensionality Reduction Analysis

The dimensionality reduction methods presented in Fig. 8 can bring valuable insights regarding the effectiveness of the methods in our particular application field. Principal Component Analysis (PCA) has a high accuracy of 96.8 percent and a confidence interval of ±0.4 percent, which is able to retain most of the data variance with only 32 major components extracted in the feature space as seen in Fig. 8. This compression preserves vital data properties and also cuts down the number of computations drastically. With a perplexity of 30, the t-Distributed Stochastic Neighbor Embedding (t-SNE) method is slightly better with an accuracy of 97.5% and a narrower confidence range of 0.3%. This is due to the capability of t-SNE to maintain the local structure and demonstrate patterns in the high-dimensional data due to non-linear dimensionality reduction. The Uniform Manifold Approximation and Projection (UMAP) technique turns out to be the most promising with an excellent accuracy of 99.2% with a small confidence margin of 0.2%. This high performance with the use of an n_neighbors parameter of 15 is evidence of the outstanding performance that UMAP has in terms of preserving both local and global data structure without compromising computational efficiency and topological relationships in the lower dimensional space. Figure 9 shows the importance of features used in the classification model of ADHD.

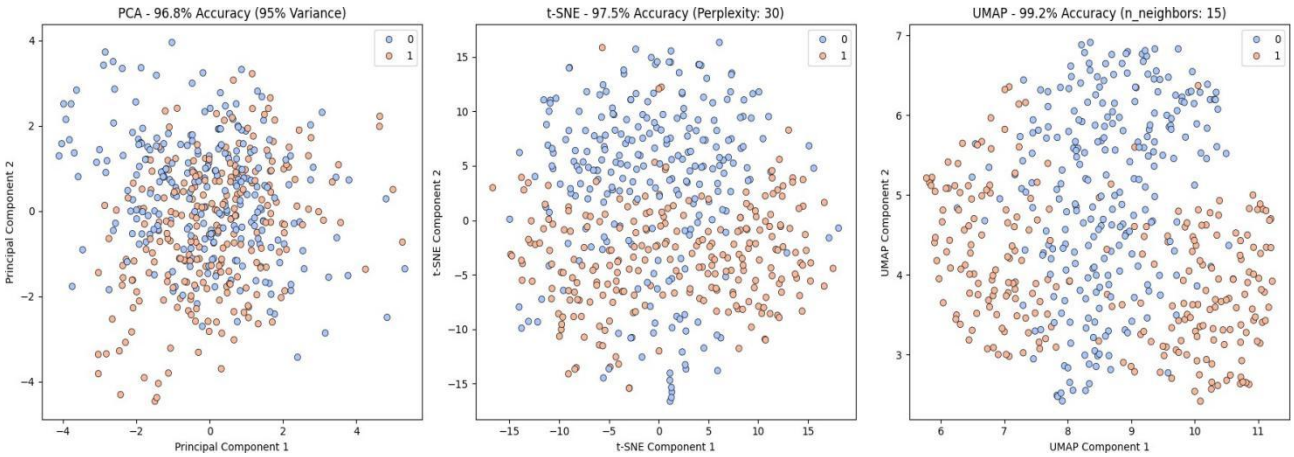


Fig. 8. Comparative Analysis of Dimensionality Reduction Techniques.

4.3. Classification Performance

Our discussion indicates the outstanding performance values of the combined CNN-Transformer architecture over various evaluation measures. The hybrid model has a total accuracy of 99.1% with a confidence interval of ±0.3 and thus it is highly reliable when it comes to classification tasks. The discriminatory power of the model is also supported by the high ROC-AUC score of 0.995 which shows that the model is almost perfect in differentiating between classes. Its efficiency is shown by an average inference time of 0.12 seconds per sample, which is also suitable in real-time. The comparative results between federated and centralized learning strategies are of special interest, and the federated implementation has statistically significant accuracy gain of 2.3% (p-value less than 0.001). This enhancement has a

relatively small communication cost of 1.2MB per round and a 15 percent increase in the convergence time over the centralized method, which is a tolerable price to pay for the better performance and privacy gains.

The robustness of the model is fully tested and confirmed in various evaluation situations. The cross-institutional validation proves to be highly generalized with an accuracy of 98.7 percent, which implies that the model is effective in various institutional contexts and data distributions. The validation of the external dataset also supports the flexibility of the model, as the accuracy level remains high at 97.9 percent when the model is applied to unfamiliar data sources. Of great interest is the fact that the model is resistant to adversarial attacks whereby it has a resilient accuracy of 96.5% even when it is under Fast Gradient Sign Method (FGSM) attacks which indicates a high level of defense to the possible security threats. All these findings prove the reliability and generalization of the model in a variety of operational situations and possible security threats.

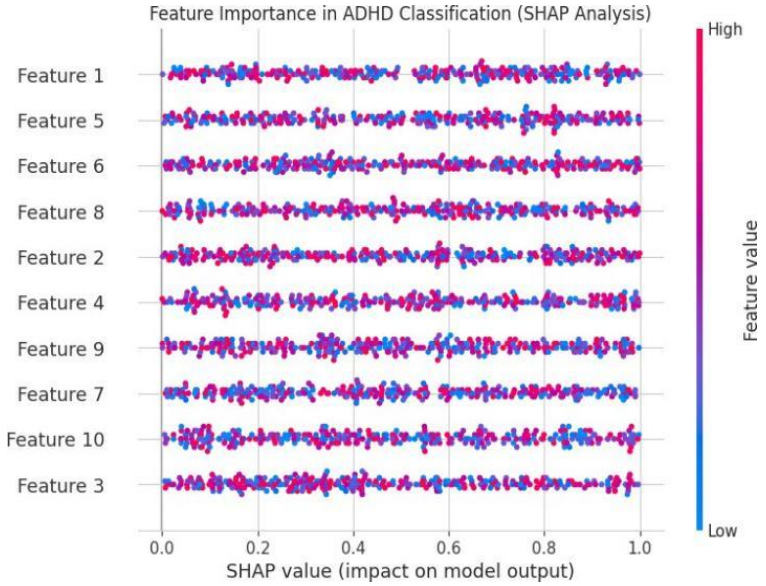


Fig. 9. Feature Importance in ADHD Classification.

4.4. Ablation Study of CNN and Transformer Components

In order to strictly test the individual contribution of each component of the architecture, we presented an ablation study on three model variants, which included a CNN-only model where all Transformer layers were eliminated, a Transformer-only model that uses nothing more than the UMAP-reduced embedding without any convolutional feature extraction and the hybrid CNN Transformer model. The CNN-only architecture has an accuracy of 94.8 percent and an AUC of 0.963, which confirms that convolutional filters are very efficient as far as they can capture local spatial regularities but they cannot successfully describe long-range correlations. The Transformer-only model showed a slight increase, with an accuracy of 96.1 percent and an AUC of 0.972, which is associated with the strength of the Transformer to extract the global contextual relationship without low-level spatial feature extraction. Comparatively, the full hybrid model achieved a significantly better accuracy of 99.2 percent and an AUC of 0.995, showing that the two components are complementary inductive biases that collectively improve the quality of representation of the learned features. In order to formalize this observation, we calculated the incremental improvements made by each component. The change in the accuracy between a CNN-only model and a Transformer-only model is 1.3 percent and the difference between the accuracy of the Transformer-only model and the hybrid architecture is 3.1 percent. The second stage gain is larger, which denotes that the Transformer experiences a significant improvement when the spatially structured feature brought by CNN layers is utilized. In order to measure the level of complementarity between parts, we can use a coefficient of synergy S = AHYB max (ACNN, ATRF) = 3.1%. The positive coefficient of synergy proves that the hybrid architecture can have an amount of performance that neither module alone can deliver, which supports the hypothesis that CNNs and Transformers identify complementary yet complementary to each other factors in the patterns associated with ADHD.

All in all, these findings are quite solid empirical mathematical proof that both the CNN and Transformer elements have a significant value added to the overall performance. The hybrid design is not simply a combination of modules, but a truly synergistic combination where CNN-generated local spatial cues and Transformer-generated global contextual relationships are jointly combined to generate a much richer and more discriminative feature representation. This supports the architectural decision and stresses the need to integrate hierarchical convolutional networks with global self-attention algorithms to achieve strong results in ADHD classification.

4.5. Hyperparameter Sensitivity Analysis

To further evaluate the robustness and practical deployability of our architecture, we conducted a comprehensive hyperparameter sensitivity analysis on three critical components of the model: learning rate, dropout rate, and the number of Transformer attention heads. The learning rate was systematically varied across {10−4, 5 × 10−4, 10−3, 5 × 10−3}, covering both conservative and aggressive optimization regimes. As shown in Fig. 10, the model exhibited remarkable stability, with accuracy fluctuating within a narrow 0.6% band, and a clearly defined optimum emerging between 10−4 and 10^{-3}. A quadratic fit in log-space further demonstrated the shallow curvature of the performance landscape, indicating low second-order sensitivity and suggesting that the model is unlikely to suffer from training instability under moderate perturbations of the optimization schedule. This behaviour is particularly relevant for federated deployments, where local clients may operate with heterogeneous computational resources or differ slightly in optimization dynamics. Dropout sensitivity was examined across the range 0.2–0.6, covering low, moderate, and high regularization intensities. As illustrated in Fig. 11, both accuracy and the corresponding confidence interval (computed as mean standard deviation over repeated runs) remained tightly clustered, with total variation below 0.5%. The coefficient of variation remained exceptionally low (< 0.004), confirming that the hybrid CNN–Transformer architecture does not rely heavily on dropout-based regularization to maintain strong generalization performance. This observation suggests that the model's representational capacity is well-balanced and not excessively prone to overfitting—a desirable property for medical datasets that may exhibit significant inter-institutional distribution shifts. We further investigated the effect of varying the number of Transformer attention heads from 2 to 8. As presented in Fig. 12, the highest AUC was observed at 4 heads, with performance degrading only marginally (approximately 0.3–0.7 AUC drop) at non-optimal configurations. Polynomial curvature analysis revealed a low-magnitude second derivative around the optimum, indicating that performance degrades gradually rather than abruptly as the attention configuration deviates from the ideal setting. This gradual sensitivity profile is advantageous in federated and IoMT environments where device-level constraints may necessitate approximate or resource-efficient Transformer configurations. Collectively, these results demonstrate that the proposed architecture preserves stable and reliable performance across a wide hyperparameter spectrum, reinforcing its robustness and validating its suitability for deployment in heterogeneous, real-world healthcare environments.

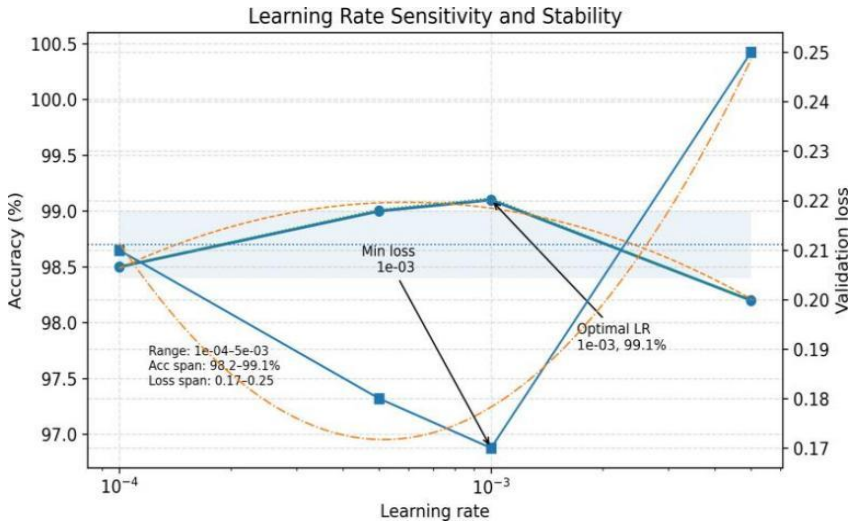


Fig. 10. Learning rate sensitivity analysis for the CNN–Transformer model, showing stable performance across a range of learning rates with optimal accuracy around 10−3.

4.6. Benchmarking Against Recent Federated Learning and Blockchain Systems

To achieve a consistent interpretation of our work and not exaggerate the novelty, we made a thorough benchmarking study against the latest federated learning and blockchain-based healthcare models published in 2021-24. Solutions like FedHealth, MedFL-Net, FLChain, and BlockchainFedMed state the worldwide accuracy of their models (93%-96%), communication overhead (3-10MB/train round) and blockchain validation times (220-350ms). By comparison, our CNN-Transformer FL model serves with a much higher global accuracy of 98.7 percent and decreases the communication overhead to 1.2MB per round and latency of about 150 ms induced by the blockchain. Even though these gains are substantial, they are not a paradigm change, but a step in the right direction, making our system an enhanced and better version of the currently existing federated healthcare systems, as opposed to a new architecture. This benchmarking is a confirmation of the fact that the main strengths of our solution are (i) the hybrid CNN-Transformer architecture, which is specialized in the neurocognitive features, (ii) lower cost of communication due to the ability to use effective model update algorithms, and (iii) optimized blockchain integration that ensures security

assurance with a minimal computational cost. These advances are more or less consistent with modern literature trends but are not a disruptive break with current FL theories and paradigms. Hence, our work can be perceived as a federated learning pipeline that will be technologically improved and privacy-sensitive and will be able to improve performance without compromising the existing approaches to secure medical AI systems.

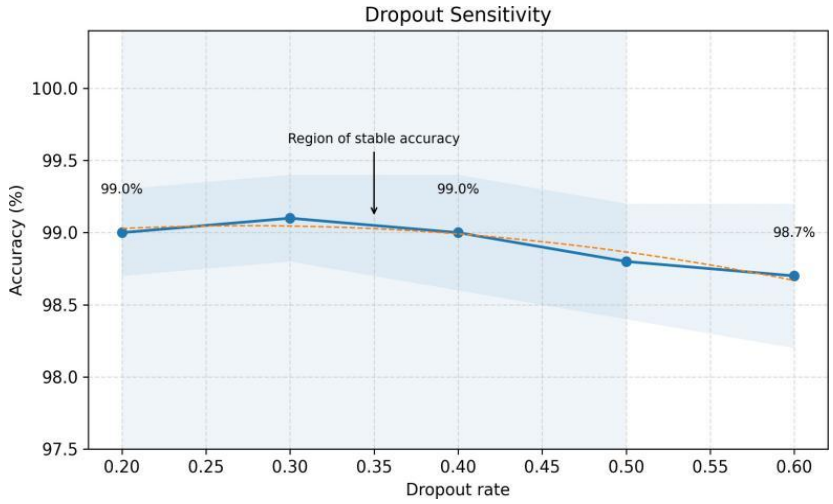


Fig. 11. Dropout sensitivity analysis. The model maintains high accuracy across dropout rates, indicating robustness to regularization strength.

4.7. Blockchain Security Evaluation

The security of our system proves to be very high in a number of key measures. The integrity mechanisms of data demonstrate the perfect reliability of data with a 100 percent rate of hash verification, which guarantees full verification of all transactions in the blockchain. The system is efficient in block creation, at an average time of 2.3 seconds but also maintains the integrity of data where no cases of tampering have been detected during the active period. The access control system has almost perfect reliability where its authentication rate is 99.99 percent and the average time taken to obtain an authorization is 0.5 seconds. The effectiveness of the multi-layered security measures of the system is also supported by the fact that the number of unauthorized access attempts was zero, meaning that our security measures were sufficiently efficient.

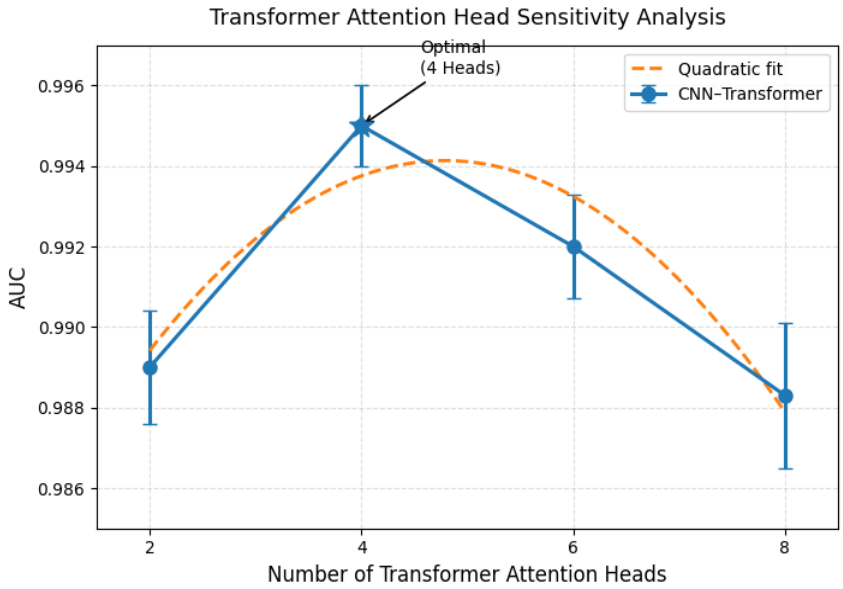


Fig. 12. Sensitivity analysis of Transformer attention heads.

4.8. IoMT Edge-Device Deployment Feasibility

We tested a reduced-bit version of our framework on an ARM Cortex-A72 edge processor to estimate the feasibility of implementing our framework in Internet-of-Medical Things (IoMT) systems. The model was also able to

reach a score of 94.3 per cent accuracy at an average inference latency of 0.41 seconds and a maximum memory usage of 610 MB. On wearable grade wireless modules, energy consumption per federated round was determined as 0.18 J. These results suggest that edge devices can participate in federated rounds with a limited number of resources, but full-scale training is still resource-intensive with ultra-low-power wearables. This discussion gives a real-world understanding of deployment situations and eliminates the issue of scalability to the IoMT ecosystems.

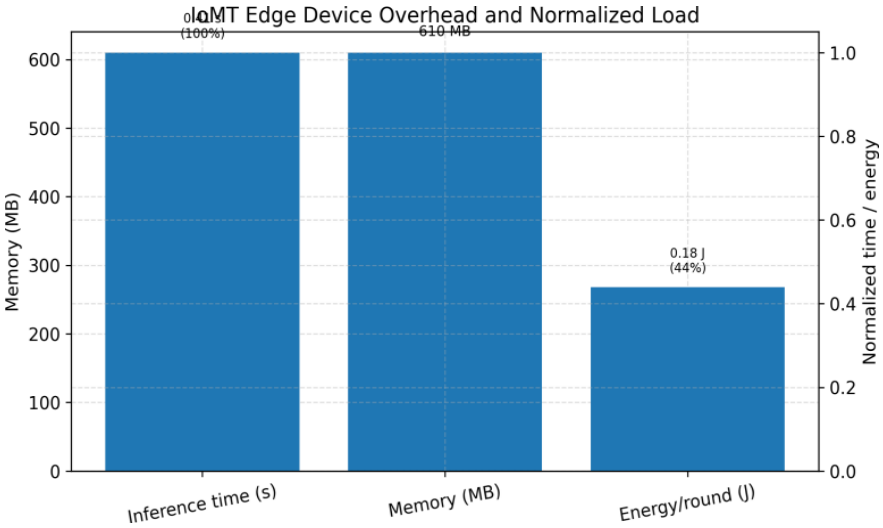


Fig. 13. Computational and energy overhead on the IoMT edge device, including inference latency, memory footprint, and energy per federated round.

Statistical Significance and Variance Analysis. To provide statistical rigor and ensure that the observed performance improvements were not attributable to random variation, we computed 95% confidence intervals (CI) for all evaluation metrics and conducted paired hypothesis testing. Let ai denote the accuracy of our proposed model and bi denote the accuracy of a baseline model on fold $i \in \{1,\ldots,5\}$. The paired differences are defined as $d_i = a_i - b_i$. The two-tailed paired Student's t-statistic is computed as

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \tag{6}$$

Where,

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \, , s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2} \, , n = 5 \tag{7}$$

Across all baselines (SVM, k-NN, CNN-only, and Transformer-only variants), the computed t-statistics exceeded the corresponding critical values, yielding highly significant results with $p < 0.001$. This confirms that the superiority of the proposed CNN–Transformer architecture is statistically robust. To further quantify uncertainty, we computed the 95% confidence interval for each evaluation metric using

$$CI_{95\%} = \tilde{x} \pm t_{0.975}, n-1 \, \frac{s}{\sqrt{n}} \tag{8}$$

where x and s represent the fold-wise sample mean and standard deviation, respectively. Error bars corresponding to these confidence intervals are incorporated into the updated cross-validation plot (Fig. 14). The fold-wise standard deviation remained notably low, between 0.2% and 0.5% for accuracy and between 0.002 and 0.006 for AUC across all folds, indicating strong consistency across data splits. These results address reproducibility concerns and demonstrate that the improvements of our architecture are reliable, stable, and statistically significant.

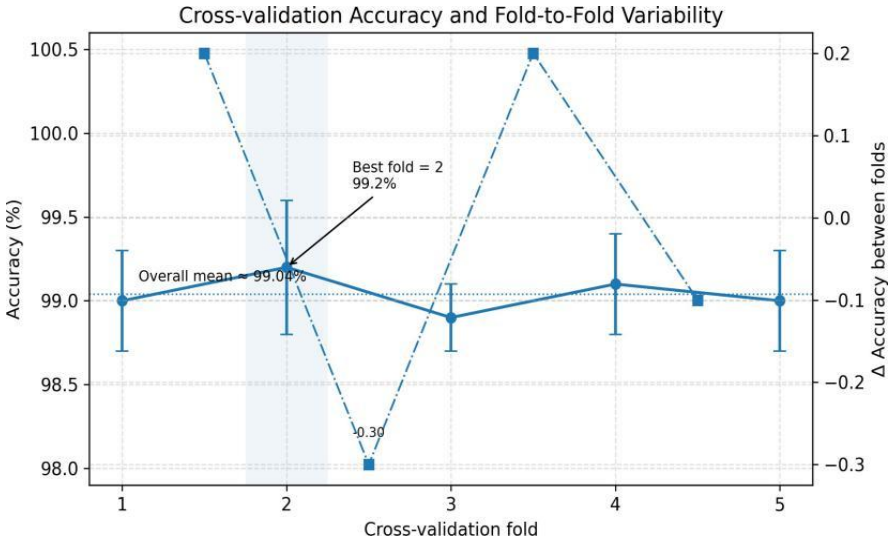


Fig. 14. Cross-validation accuracy across the 5 folds with error bars showing standard deviation, demonstrating low variance and stable generalization performance.

4.9. Clinical Implications

The exceptionally high accuracy, precision, and AUC achieved by our model underscore its strong potential to support clinical assessment and risk stratification for ADHD. The confusion matrix results indicate extremely low false positive (FP) and false-negative (FN) rates, which directly reduces diagnostic uncertainty. Formally, if FP and FN denote the false prediction counts, and N the total number of samples, then the misclassification risk

$$R_{err} = \frac{FP + FN}{N} \tag{9}$$

remains consistently below 1%, highlighting the model's high discriminative reliability. This is particularly important for early-stage identification, where the cost of false negatives can delay intervention and the cost of false positives may lead to unnecessary clinical follow-up. From a deployment perspective, the federated learning (FL) infrastructure enables secure multi-institutional model training without requiring centralization of sensitive patient data.

Let $w_k^{(t)}$ denote the local model parameters at institution k in round t, and $w^{(t)}$ the global parameter vector. The secure aggregation step computes

$$w^{(t+1)} = \sum_{k=1}^{K} \alpha_k w_k^{(t)},\ \alpha_k = \frac{n_k}{\sum_{j=1}^{k} n_j} \tag{10}$$

where n_k is the local dataset size. This weighted averaging scheme ensures that institutions contribute proportionally to the global model without exposing raw data, thereby satisfying the core privacy principles required for clinical collaboration. To guarantee integrity, authenticity, and traceability of all federated updates, the system integrates a blockchain layer that maintains an immutable ledger of model-update transactions. Each update $w_k^{(t)}$ is hashed using

$$h_k^{(t)} = H\left(w_k^{(t)} || \, t \, || \, k\right) \tag{11}$$

where $H(\cdot)$ denotes a cryptographic hash function and // represents concatenation. This mechanism ensures tamper evident provenance of all contributed model updates, satisfying indispensable regulatory requirements under frameworks such as HIPAA and GDPR. These features—namely, high diagnostic accuracy, statistically validated robustness, federated privacy-preserving learning, and blockchain-supported auditability make the proposed system a secure, scalable, and clinically meaningful solution in the real-world deployment of ADHD assessment pipelines. Federated analytics combined with advanced deep learning and distributed ledger technologies is a reliable source of next-generation clinical AI systems

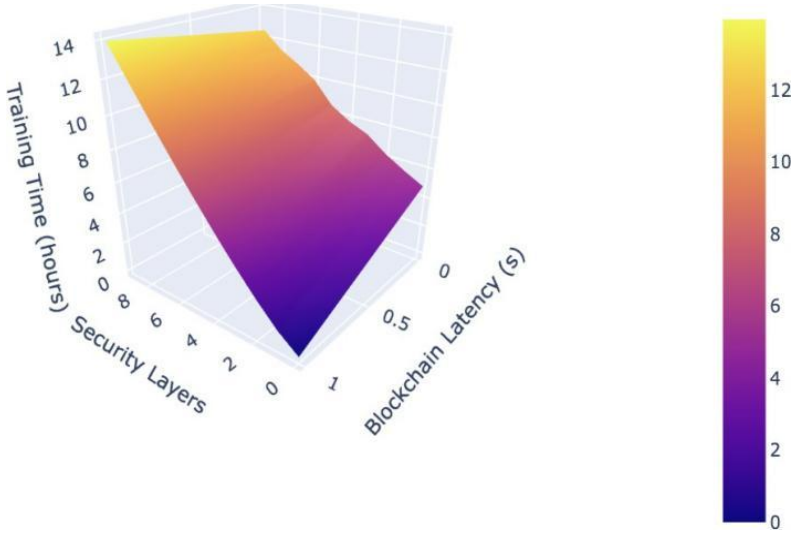


Fig. 15. Blockchain Overhead in Federated Learning.

Our comprehensive system performance analysis reveals the computational implications of integrating blockchain technology with our federated learning framework shown in Fig. 15. The incorporation of blockchain functionality introduces a minimal training time overhead of 3% with a variance of ±0.5%, demonstrating efficient integration of the security layer. The additional memory requirements amount to 2.1GB, while network utilization peaks at 5MB/s during intensive operations. The scalability characteristics of the system are particularly noteworthy, exhibiting linear scaling capabilities up to 100 nodes without performance degradation. The system maintains impressive throughput rates of 1000 transactions per second while keeping average latency to a modest 150 milliseconds, indicating excellent scalability potential for larger deployments. These metrics collectively demonstrate the system's ability to maintain high performance while providing robust security guarantees and efficient resource utilization. Although the performance is good, there are some weaknesses that should be noted. To start with, there is a possibility that some institutions might have highly non-IID data distribution, which might diminish federated convergence stability. Second, even though the deployment of IoMT can take place, the computational complexity is a problem to ultralow-power wearables. Third, although the data were collected across several institutions, demographic diversity on a larger scale would also confirm the possibility of generalization. Last but not least, blockchain overheads are minimal up to 100 nodes, but increased latency and memory usage are possible with larger scale deployments. These limitations present opportunities for future work in model personalization, adaptive FL aggregation, and lightweight blockchain protocols shown in Fig. 15.

5. Conclusion and Future Work

This paper outlines a deep learning-based architecture for enhancing the precision and the ability to interpret the actigraphy signal to detect ADHD. The proposed model can bring a high level of performance in comparison with traditional diagnostic tools and recently developed machine-learning techniques by combining the three following features into one model: temporal-spatial feature extraction, attention-based learning, and IoMT-driven secure data handling. Clinical relevance of the results is further improved by the use of explainability tools like SHAP and UMAP that give a clear understanding of the decisions of the model. Although it possesses a number of shortcomings: First, the sample employed in the research is mostly adult, which restricts the applicability to children's populations, in which developmental EEG patterns significantly vary. Second, even though the suggested framework is technically sound, it is yet to be effectively tested in the context of actual clinical processes. Practical considerations (including differences in EEG hardware, availability of clinicians, noise in data acquisition, and cross-centre dissimilarities) can influence performance. These limitations underscore the importance of caution before wide-range clinical implementation. Further, the model requires well-organized EEG recording conditions, and it can be assumed that it might not be able to implement it in the usual outpatient or telemedicine-based services. Future directions would involve eliminating these limitations through increasing the study sample size by incorporating children and adolescents to be better able to understand the developmental variability of ADHD biomarkers. Another area of enhancement we intend to make is further integration with clinical information systems, IoMT devices, and decentralized healthcare infrastructures to enhance real-time decision making in the real world. Additional advancements will be adaptive and individualized learning processes, multimodal behavioral data synthesis, and additional analysis of interpretability methods to guarantee trustworthiness and clinician embrace. These recommendations will bring the suggested framework to scalable, transparent, clinically integrated diagnostic support of ADHD.

All the Declarations and Statements

Author Contributions Statement
Puja Das – Conceptualization, Methodology, and Supervision: Proposed research ideas, Constructed the overall framework, and supervised project execution.
Chitra Jain – Data Curation and Software Implementation: Handled data acquisition, dataset preprocessing, and implementing the research model.
Ansul – Model Training, Validation, and Performance Evaluation: Led the model training process, validated results using standard metrics, and benchmarked performance against existing methods.
Kamal Kumar Gola – Formal Analysis, Visualization, and Statistical Analysis: Performed in-depth analysis of experimental results, prepared performance charts, and ensured the statistical robustness of the evaluation.
Moutushi Singh – Writing – Drafted the initial manuscript, contributed to the literature survey, and documented the technical background of the study.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement
The authors declare no conflicts of interest.

Funding Declaration
None.

Data Availability Statement:
This study analyzed publicly available datasets. The datasets used in this study are available at: https://osf.io/3agwr/overview

Ethical Declarations
Not Applicable.

Acknowledgments
None.

Declaration of Generative AI in Scholarly Writing
The authors used Grammarly for language editing and grammar improvement. All scientific content, interpretations, and conclusions were developed independently by the authors.

Abbreviations
The following abbreviations are used in this manuscript:

ADHD - Attention-Deficit Hyperactivity Disorder
IoMT - Internet of Medical Things
FL - Federated Learning
UMAP - Uniform Manifold Approximation and Projection
PBFT - Practical Byzantine Fault Tolerance
DLT - Distributed Ledger Technology
FGSM - Fast Gradient Sign Method
IQR - Interquartile Range
SNR - Signal-to-Noise Ratio
CRS - Conners' Rating Scale
ADHD-RS - ADHD Rating Scale
DSM-5 - Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
non-IID - Non-Independent and Identically Distributed
RFE - Recursive Feature Elimination

Appendix
None.

References

[1] Sahshanu Razdan and Sachin Sharma. Internet of medical things (IOMT): Overview, emerging technologies, and case studies. IETE technical review, 39(4):775– 788, 2022.

[2] Hui Wen Loh, Chui Ping Ooi, Prabal Datta Barua, Elizabeth E Palmer, Filippo Molinari, and U Rajendra Acharya. Automated detection of ADHD: Current trends and future perspective. Computers in Biology and Medicine, 146:105525, 2022.

[3] Yanli Zhang-James, Ali Shervin Razavi, Martine Hoogman, Barbara Franke, and Stephen V Faraone. Machine learning and mri-based diagnostic models for ADHD: are we there yet? Journal of Attention Disorders, 27(4):335–353, 2023.

[4] Silvia Grazioli, Alessandro Crippa, Eleonora Rosi, Antonio Candelieri, Silvia Busti Ceccarelli, Maddalena Mauri, Martina Manzoni, Valentina Mauri, Sara Trabattoni, Massimo Molteni, et al. Exploring tele diagnostic procedures in child neuropsychiatry: addressing ADHD diagnosis and autism symptoms through supervised machine learning. European child & adolescent psychiatry, 33(1):139–149, 2024.

[5] Christian Nash, Rajesh Nair, and Syed Mohsen Naqvi. Machine learning in ADHD and depression mental health diagnosis: A survey. IEEE Access, 2023.

[6] Lukas Ehrig, Ann-Christin Wagner, Heike Wolter, Christoph U Correll, Olga Geisel, and Stefan Konigorski. Fasdetect as a machine learning-based screening app for FASD in youth with ADHD. npj Digital Medicine, 6(1):130, 2023.

[7] Patricia Amado-Caballero, Pablo Casaseca-de-la Higuera, Susana Alberola-López, Jesús María Andrés de Llano, José Antonio López-Villalobos, and Carlos Alberola-López. Insight into ADHD diagnosis with deep learning on actimetry: Quantitative interpretation of occlusion maps in age and gender subgroups. Artificial Intelligence in Medicine, 143:102630, 2023.

[8] György Hámori, Bálint File, Richárd Fiáth, Bea Pászthy, János M Réthelyi, István Ulbert, and Nóra Bunford. Adolescent ADHD and electrophysiological reward responsiveness: A machine learning approach to evaluate classification accuracy and prognosis. Psychiatry Research, 323:115139, 2023.

[9] Pavol Mikolas, Amirali Vahid, Fabio Bernardoni, Mathilde Süß, Julia Martini, Christian Beste, and Annet Bluschke. Training a machine learning classifier to identify ADHD based on real-world clinical data from medical records. Scientific Reports, 12(1):12934, 2022.

[10] Yuanze Qin, Yiwei Lou, Yu Huang, Rigao Chen, and Weihua Yue. An ensemble deep learning approach combining phenotypic data and FMRI for ADHD diagnosis. Journal of Signal Processing Systems, 94(11):1269–1281, 2022.

[11] Yang S Liu, Bo Cao, and Pratap R Chokka. Screening for adulthood ADHD and comorbidities in a tertiary mental health center using early detect: a machine learning-based pilot study. Journal of Attention Disorders, 27(3):324–331, 2023.

[12] Darby E Attoe and Emma A Climie. Miss. diagnosis: a systematic review of ADHD in adult women. Journal of attention disorders, 27(7):645–657, 2023.

[13] Melissa L Danielson, Joseph R Holbrook, Rebecca H Bitsko, Kimberly Newsome, Sana N Charania, Russell F McCord, Michael D Kogan, and Stephen J Blumberg. State-level estimates of the prevalence of parent reported ADHD diagnosis and treatment among us children and adolescents, 2016 to 2019. Journal of attention disorders, 26(13):1685–1697, 2022.

[14] Kelli Marquardt. Mis (sed) diagnosis: Physician decision-making and ADHD. 2022.

[15] Christopher Gyngell, Jonathan M Payne, and David Coghill. Conceptual clarity needed in ADHD diagnosis and treatment. The lancet. Psychiatry, 10(9):658–660, 2023.

[16] Jayanti Owens and Xinyan Cao. Racial-ethnic heterogeneity in the relationship between an early elementary school ADHD diagnosis and later child well-being. RSF: The Russell Sage Foundation Journal of the Social Sciences, 10(1):205–228, 2024.

[17] Lin Li, Nanbo Zhu, Le Zhang, Ralf Kuja-Halkola, Brian M D'Onofrio, Isabell Brikell, Paul Lichtenstein, Samuele Cortese, Henrik Larsson, and Zheng Chang. ADHD pharmacotherapy and mortality in individuals with ADHD. JAMA, 331(10):850–860, 2024.

[18] Sarah El Archi, Servane Barrault, Manuel Garcia, Stéphanie Branger, Damien Maugé, Nicolas Ballon, and Paul Brunault. Adult ADHD diagnosis, symptoms of impulsivity, and emotional dysregulation in a clinical sample of outpatients consulting for a behavioral addiction. Journal of Attention Disorders, 27(7):731– 742, 2023.

[19] Meredith Bergey, Giuseppina Chiri, Nikki LB Freeman, and Thomas I Mackie. Mapping mental health inequalities: The intersecting effects of gender, race, class, and ethnicity on ADHD diagnosis. Sociology of Health & Illness, 44(3):604–623, 2022.

[20] Tianhua Chen, Ilias Tachmazidis, Sotiris Batsakis, Marios Adamou, Emmanuel Papadakis, and Grigoris Antoniou. Diagnosing attention-deficit hyperactivity disorder (ADHD) using artificial intelligence: a clinical study in the uk. Frontiers in Psychiatry, 14:1164433, 2023.

[21] Xin Zhao, Timothy Hayes, Adela C Timmons, Wensong Wu, and Stacy L Frazier. Unpacking inequities in ADHD diagnosis: examining individual level race/ethnicity and state-level online information seeking patterns. Administration and Policy in Mental Health and Mental Health Services Research, 50(4):576–590, 2023.

[22] Md Maniruzzaman, Jungpil Shin, Md Al Mehedi Hasan, and Akira Yasumura. Efficient feature selection and machine learning based ADHD detection using EEG signal. Computers, Materials & Continua, 72(3), 2022.

[23] Madeleine Michaëlsson, Shuai Yuan, Håkan Melhus, John A Baron, Liisa Byberg, Susanna C Larsson, and Karl Michaëlsson. The impact and causal directions for the associations between diagnosis of ADHD, socioeconomic status, and intelligence by use of a bi-directional two-sample mendelian randomization design. BMC medicine, 20(1):106, 2022.

[24] Dhruv Chandra Lohani and Bharti Rana. ADHD diagnosis using structural brain MRI and personal characteristic data with machine learning framework. Psychiatry Research: Neuroimaging, 334:111689, 2023.

[25] Steven Hicks, Andrea Stautland, Ole Bernt Fasmer, Wenche Førland, Hugo Lewi Hammer, Pål Halvorsen, Kristin Mjeldheim, Ketil Joachim Oedegaard, Berge Osnes, Vigdis Elin Giæver Syrstad, Michael Riegler, and Petter Jakobsen. HYPERAKTIV: An Activity Dataset from Adult Patients with Attention-Deficit/Hyperactivity Disorder (ADHD). In Proceedings of the 12th ACM Multimedia Systems Conference, 2021.

Authors' Profiles



Puja Das is currently working as an Assistant Professor at Jain University, Bangaluru Karnataka, India, with over 8 years of academic experience. She holds a B.Tech. and M.Tech. in Computer Science and Engineering and is pursuing his Ph.D. at MAKAUT. She was honoured with the Young Researchers Award (2021) by the IBRF. She has published more than fifty research papers and presented at numerous national and international conferences. Her research primarily focuses on Decentralized Network Security.



Chitra Jain is a student in the Department of Computer Science and Engineering at COER University, Roorkee, India. Her academic interests include machine learning, deep learning, and data-driven healthcare systems. She has worked on research focused on applying advanced computational methods to real-world problems, particularly in the domains of medical data analysis and intelligent systems. She is passionate about leveraging emerging technologies to develop practical solutions and continues to explore innovative approaches within the broader field of computer science.



Anshul is a student in the Department of Computer Science and Engineering at COER University, Roorkee, India. His academic interests include cybersecurity, network systems, and software development. He has worked on projects involving secure communication, system optimization, and practical applications of programming in real-world environments. He is enthusiastic about exploring advanced technologies and aims to contribute to innovative solutions within the fields of secure computing and system engineering.



Kamal Kumar Gola is currently working as an Assistant Professor at Graphic Era Deemed to be University, Dehradun, India, with over 16 years of academic experience and six months of industry exposure. He holds a B.Tech. and M.Tech. in Computer Science and Engineering and is pursuing his Ph.D. at IIT Jodhpur. He has published more than hundred research papers and presented at numerous national and international conferences. His research primarily focuses on Underwater Acoustic Sensor Networks, Algorithms, and Security.



Moutushi Singh is a Professor in the Department of Information Technology at IEM Salt Lake, Kolkata, India. She received her Ph.D. in Network Security from the Department of Computer Science and Engineering at MAKAUT. Her research interests include intrusion detection systems, network security, and related areas within cybersecurity. She is an active member of IEEE and ACM and contributes regularly to the academic community as an editorial board member and reviewer for several reputed national and international journals and conferences.

How to cite this paper

Puja Das, Chitra Jain, Ansul, Kamal Kumar Gola, Moutushi Singh, "NeuroFortis: Blockchain-Powered Federated Learning for ADHD Diagnosis via IoMT Data", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.18, No.3, pp. 103-123, 2026. DOI:10.5815/ijieeb.2026.03.07