

BRAIN. Broad Research in Artificial Intelligence and Neuroscience

e-ISSN: 2067-3957 | p-ISSN: 2068-0473

Covered in: Web of Science (ESCI); EBSCO; JERIH PLUS (hkdir.no); IndexCopernicus; Google Scholar; SHERPA/RoMEO; ArticleReach Direct; WorldCat; CrossRef; Peeref; Bridge of Knowledge (mostwiedzy.pl); abcdindex.com; Editage; Ingenta Connect Publication; OALib; scite.ai; Scholar9; Scientific and Technical Information Portal; FID Move; ADVANCED SCIENCES INDEX (European Science Evaluation Center, neredataltics.org); ivySCI; exaly.com; Journal Selector Tool (letpub.com); Citefactor.org; fatcat!; ZDB catalogue; Catalogue SUDOC (abes.fr); OpenAlex; Wikidata; The ISSN Portal; Socolar; KVK-Volltitel (kit.edu) 2025, Volume 16, Issue 1, pages: 286-305.

Submitted: December 12th, 2024 | Accepted for publication: January 2nd, 2025

A Novel End-To-End Learning Framework Based on Optimised Residual Gated Units and Stacked Pre-Trained Layers for Detection of Autism Spectrum Disorders (ASD)

Parameswaran Sarvalingam

Research Scholar, Faculty of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.
ge230006@student.uthm.edu.my

Ashok Vajravelu*

Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.
ashok@uthm.edu.my
<https://orcid.org/0000-0001-6642-8366>

Janani Selvam

Faculty of Civil and Built Environment Engineering, Lincoln University College, Selangore, Malaysia.
vijayanani.s@gmail.com
<https://orcid.org/0000-0002-1041-9003>

Asmarashid Bin Ponniran

Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.
asmar@uthm.edu.my
<https://orcid.org/0000-0003-0397-0566>

Wan Suhaimizan Bin Wan Zaki

Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.
mizan@uthm.edu.my
<https://orcid.org/0000-0002-7431-7911>

Kathambari P

Associate software lead/ Research, ABE Semiconductor Designs, LLP, Tamilnadu, India.
ktpkathambari@gmail.com

Abstract: Autism Spectrum Disorder (ASD) is distinguished in detailed with diverse assemblage neural behaviours, diagnosed over the recent time of years. These ASD are developed using the increased internal connectivity among neural units of functional brain. Since Autism characteristics tend to occur over the months or even years, cognitive system for an early diagnosis of ASD is badly required to overcome the neural complexities of ASD with an accurate prediction. Electroencephalogram (EEG) is common technique for analysing brain's electrical activity and widely used as the major tool for the gathering the neural information. Recent ensemble of EEG and Artificial Intelligence (AI) has thrown the brighter light in the ASD early prediction, but still computational overhead hurdles in achieving the highest performance. In this research article, novel deep learning framework that consist of hybrid combination of multiple tiers of pre-trained convolutional models and residual gated recurrent units with the hyper-parameters are tuned by the chaotic Cat swarm optimisation. As the first step, EEG signals are collected and translated into two-dimensional (2D) spectrograms followed by the pre-processing and feature engineering using the multi-scale stacked pre-trained models. Finally, the residual gated units are entangled for extracting the temporal features which are used to train the optimised classification networks. The proposed model is validated using EEG signals extracted from the 43 subjects and performance including accuracy, precision, recall and F1-score are measured then juxtaposed through the other cutting-edge learning models. The accuracy of detection is found to be 98.5%, precision was 98.2% recall was 98% and F1-score was 98% for the real time EEG subjects and also statistically proved its stability in handling the non-linear EEG datasets.

Keywords: autism spectrum disorder; electroencephalogram; residual gated units; optimised classification networks; chaotic cat swarm.

How to cite: Sarvalingam, P., Vajravelu, A., Selvam, J., Bin Ponniran, A., Bin Wan Zaki, W.S., & P. Kathambari (2025). A novel end-to-end learning framework based on optimised residual gated units and stacked pre-trained layers for detection of Autism Spectrum Disorders (ASD). *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 16(1), 286-305.
<https://doi.org/10.70594/brain/16.1/21>

1. Introduction

ASD or autism arises from numerous influences within the brain, leading to atypical behaviours and communication challenges, along with restricted repetitive actions. It is classified as a spectrum disorder due to the broad range of indications and their varying degrees of severity, which include repetitive behaviours, difficulties with language, sensory issues, cognitive impairments, anxiety, social withdrawal, and more. ASD typically manifests during childhood and continues to evolve into adolescence and adulthood. Most of the Indications of Autism symptoms happen within the initial five years of life. Certain signs can be observed through facial features. Currently, the identification of ASD is a crucial concern, as it affects both the mental and physical growth of children.

Over the past few decades, research has revealed that approximately ASD affects approximately 1 in 68 children. However, more recent statistics suggest that the rate of diagnosis has increased to 1 in 59 children. Studies from various regions around the world also support the notion that ASD prevalence is on the rise. A critical task for developmental clinical neuroscience is identifying the most relevant combinations of consistent measures and scalp regions for detecting and tracking ASD-related traits. Utilising a data-driven approach could be instrumental in discovering valuable clinical biomarkers. Furthermore, this approach might highlight the neural correlates that are associated with important behavioural characteristics or cognitive patterns. A surge of attention has emerged in the last few years regarding the use of Electroencephalogram (EEG) for ASD prediction. EEG signals render more insight data about the brain functionality which is used to study the ASD EEG activities. The feature extraction with the respective to the ASD was considered to differentiate from the normal brain activity.

Brain activities and role in machine, deep, and other learning algorithms have attracted many researchers to implement these models for predicting ASD using the EEG input datasets

1. Since the EEG signals are often lengthy, one-dimensional, complicated, and nonlinear time signals, training these sequences in deep learning models suffers from a gradient explosion problem that affects the performance of ASD detection.
2. The long-dependence nature of the EEG signals tends to the failure of learning models to obtain the most pivotal hierarchical attributes that aid the refined classification.
3. Computational overhead is the most important complexity that hurdles the learning framework to achieve a better ASD detection process.

1.1. Research Contribution

Motivated by the previously outlined challenges, this study article proposes the ensemble learning model using hybrid stacked pre-trained models and a Bidirectional Residual Gated unit whose hyper-parameters are tuned using chaotic sand cat optimisation technique achieving better results compared to traditional deep learning networks. The main findings related to ASD prediction include the organisation of EEG data, the design of hybrid deep learning models, and the system's assessment based on accuracy, precision, recall, and F1-score.

- a. **Thorough EEG Dataset Organisation:** This research has assisted in EEG dataset formation via the conversion of raw data into spectrograms with uniform dimensions. This technique has helped standardise the data and verify that all signals contain the necessary information, which plays a key role in training and validating the prediction algorithm. This study has given the bright light in pre-processing which successfully eradicates the noise and artifacts from EEG signals.
- b. **Hybrid Deep Learning Model:** The research study has developed the hybrid DL algorithm for ASD prediction in EEG signals. The suggested model uses the stacked convolutional pre-trained layer, Bi-directional residual GRU layers, and Chaotic Cat Swarm Optimised fully connected layers(CCSO-FL) to predict EEG signals into different ASD levels.
- c. **Enhanced Evaluation Performance:** The research has assessed the suggested framework, and the findings have demonstrated that the recommended model attained superior

performance in comparison to the current cutting-edge learning model for EEG analysis. This proposed study's contributions are considered to be significant and reported to have high prediction performance that can aid the high fuel for the diagnosis and treatment process.

1.2 Organisation of the Research Article

The leftover parts are summarised as follows: Section 2 discusses relevant studies conducted by various authors. The dataset preparation, pre-processing, and proposed methodology with a detailed description are provided in Section 3. The experimentation outcomes of the recommended network and relative assessments are demonstrated in Section 4. In conclusion, the paper discusses future enhancements in Segment 5.

2. Related Works

2.1. Machine Learning-Based ASD Classification

Abdolzadegan, Moattar, & Ghoshuni (2020) proposed that ASD detection systems using the non-linear and linear feature extraction, and selection coupled with the shallow ML Model are tested with 34 subjects and deploy techniques such as power spectrum extraction, fractal dimension, correlation dimensions for obtaining the EEG data information. The proposed system has shown a performance of 71.4% and 90.54% in detecting ASD using EEG patterns.

Ghembaza, & Djebbari (2023) explored electroencephalogram signals divided into different distinct ranges, including Eta and Gamma, according to respective frequency measures. Researchers explored a thorough examination of the obtained signal information data. They extracted features within the time-frequency domain by its contours of the spatio-temporal area. These parameters employed further analysis. Based on features obtained from different frequencies, bands are used for further classifications.

Abbas et al. (2020) introduced a soft computing-based assessment for ASD comprising a couple of different modules to deliver a prominent outcome for a four-minute questionnaire collected via an easy technological system, a bio-fundamental observed features through the videos regarding the mobility of the children in the living place, then a two minutes assessment conducted by the medical people during the evaluation. Their results improved the performance of 0.69 in specificity while maintaining a sensitivity of 90%.

Goel et al. (2020) introduce the Modified Grasshopper Optimisation integrated with different ML algorithms including SVM, KNN, Decision Trees (DT), and Random Forests (RF). This model exploits advantages like the grasshopper model and produces a promising performance in early detection of ASD in children from age groups of 3-12. The model achieved 100% accuracy, as well as 100% specificity and recall, in detecting ASD in children.

2.2. Deep Learning-Based ASD Classification

Ali et al. (2020) conducted an ASD diagnosis utilising a CNN approach with EEG signals. The data were gathered utilising a sixteen channel data system network sampled at 256 Hertz. Each collected signal amplitude was transformed as the product of each channel with 256 sized two-dimensional matrix format and, subsequently, input the detailed information to CNN towards identification of ASD. Finally, 80% classification accuracy was achieved.

Xu et al. (2020) investigated the use of short-term instant Dynamic blood flow variations to distinguish between children with autism as well as those without it. This work implemented a framework, which integrates a convolutional neural network with a gated recurrent unit (GRU). Analysing the brain signal with twenty-five data information of abnormal (affected), furthermore, 22 without, the CGRNN attained an accuracy rate of 92.2%, owing to its exceptional ability to identify distinctive features and capture inherent correlations within time series data.

Patel et al. (2024) employed the BCIAUT - P300 system for collecting EEG information measuring and analysing ASD persons. This dataset includes information collected from individuals with autism when exposed to stimuli in a controlled setting. Researchers developed a CNN-based model that relies on feature extraction for classification, utilising BCI Attention Classification to achieve their goal of predicting the responses of individuals with autism through detailed examination. The CNN structure incorporates a feature extraction network designed specifically for classification, aiming to forecast abnormal reactions to stimuli in individuals with autism.

Amrani et al. (2021) conducted a systematic literature review on EEG signal analysis using deep learning (DL), emphasising the growing importance of automated techniques over manual methods due to their complexity, time requirements, and reliance on user expertise. Their review identified 403 papers from databases like Web of Science, PubMed, and Science Direct, analysing EEG processing methods across multiple neurological disorders, including epilepsy, depression, schizophrenia, and autism. The review revealed that 57% of studies employed Convolutional Neural Networks (CNNs), 12% utilised Recurrent Neural Networks (RNNs), and 13% used a combination of CNNs and Long Short-Term Memory (LSTM) networks. Other architectures like Generative Adversarial Networks (GANs), Autoencoders (AEs), and Restricted Boltzmann Machines (RBMs) were less common, highlighting the dominance of CNN-based models. Despite its promise, the review pointed out that DL models require large datasets and suffer from interpretability issues due to their black-box nature. The authors suggested that advancements in Explainable Artificial Intelligence (XAI), transfer learning, and technologies like Big Data, IoT, and cloud computing could enhance the efficiency and accuracy of DL-based EEG analysis systems. However, the lack of confidence in results and the dependency on extensive data remain significant drawbacks, emphasising the need for further research to address these limitations.

Siddiqui et al. (2023) proposed a deep neural network (DNN) model for subject-independent mental task classification using electroencephalography (EEG) signals. This method is designed for patients with limited motor movements and focuses on identifying mental tasks without subject-specific training data. The raw EEG signals were preprocessed using spatial filtering with the Laplacian surface, followed by principal component analysis (PCA) for dimensionality reduction and extraction of discriminative features. The model utilised Power Spectrum Density (PSD) values for training, excluding one subject's data for testing. The framework achieved an accuracy of 77.62% on a benchmark dataset, outperforming state-of-the-art algorithms in subject-independent mental task classification. While the approach demonstrates significant potential, the accuracy achieved indicates room for improvement, particularly in handling inter-subject variability. Additionally, the reliance on pre-computed features may limit adaptability in real-time applications.

3. Suggested Methodology

The proposed architecture consists of four components: electroencephalogram database gathering, electroencephalogram database preparation, deep representation mining, and Optimised classification networks. The complete analysis of each component is presented below. The suggested algorithm has been developed utilising EEG datasets with 64 channels. The filter banks are utilised in every channel gathering the database through the signal collected. These signals are used to construct the spectrogram images utilising Short Time Fourier Transform (STFT) which serves as input for advanced stacked learning models, such as Mobilenet, Squeezenets, and Multi-scale Residual Gated Recurrent Units. The information-enriched features are fed to a chaotic cat swarm-optimised learning network for the prediction of ASD

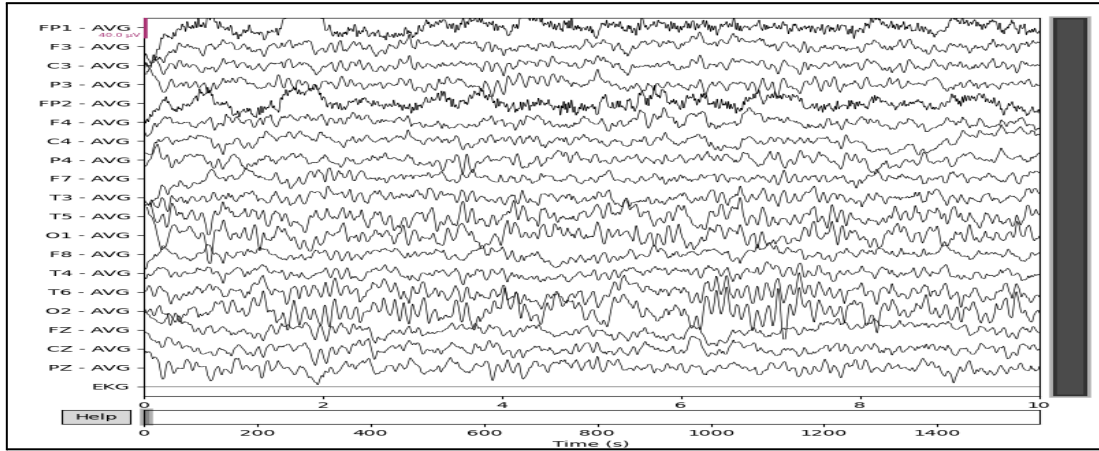
3.1. Materials and Methods

The gathered datasets consist of EEG signals obtained from 143 participants, recorded using 64 channels. These EEG signals were sourced from 72 individuals diagnosed with ASD and 71 healthy controls. The selection of patients with ASD was conducted at approved special education

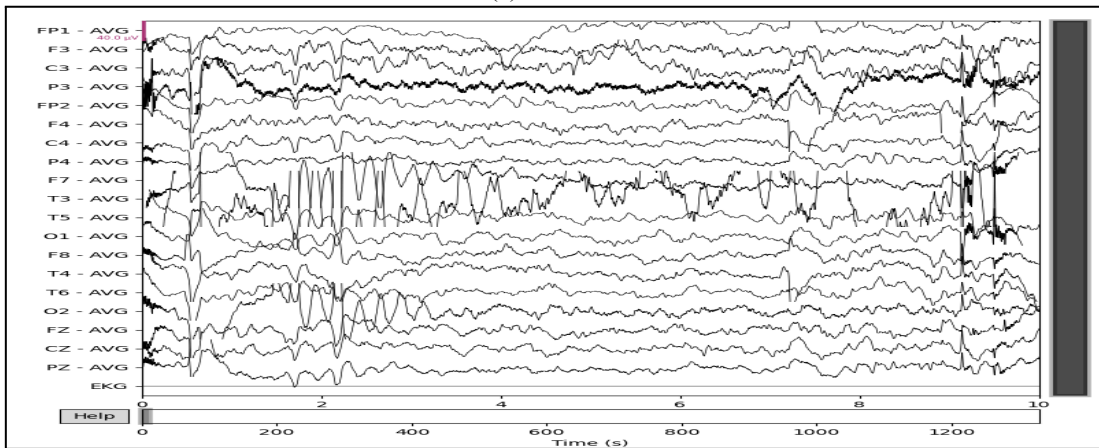
centers, guided by evaluations from clinical experts such as neurologists and neurorehabilitation specialists, as well as the patients' medical histories and standardised assessment scales. The process for selecting these patients aimed to: (1) eliminate co-morbid conditions such as epilepsy, complex genetic syndromes, sensory disorders, and a few more structural neurological issues (e.g., hydrocephalus) that could independently influence EEG readings; (2) exclude individuals who do not fulfill the DSM-5 standards towards autism (e.g., those exhibiting only few autism traits, like stereotypical behaviours); finally, (3) remove abnormal persons who undergo treatment with medications known to affect EEG results, ensuring a sufficient time gap before the recording. This evaluation aimed to form a starting cohort of “mild average person” who currently do not have or have not previously experienced significant issues like epilepsy or congenital syndromes). Further exclusions were made to filter out: (1) Cognitively advanced individuals, (2) those displaying Attention Deficit Hyperactivity Disorder, and (3) persons exhibiting mild average person with neural indicators. Table 1 presents the complete specification of the EEG data collection and Figure 1 (a-c) shows the sample EEG raw signals collected from normal healthy and ASD children.

Table 1. EEG Data Collection used for the Experimentation.

Specification	Description
Type of Signal	EEG
Number of Subjects	143
Frequency (Hz)	500
Details of ASD patients	71
Details of Healthy patients	72
EEG Electrodes used	Ag-AgCl Electrodes

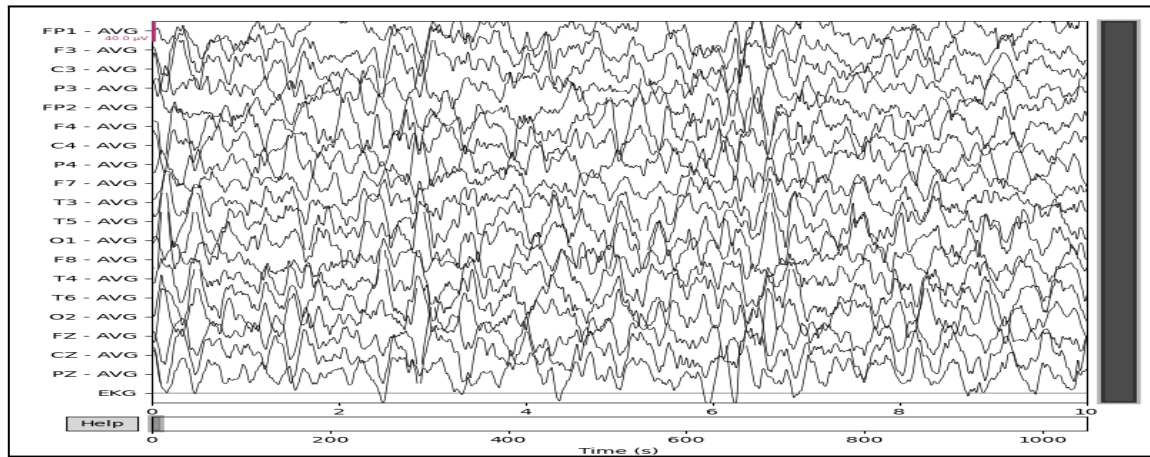


(a)



(b)

Figure 1 (a&b). Sample EEG Signals Collection a) ASD from Subject-3 b) Normal EEG signal from Subject-5



(c)

Figure 1 (c). Sample EEG Signals Collection - c) ASD signals from Subject-9.

3.2. EEG Data Pre-processing

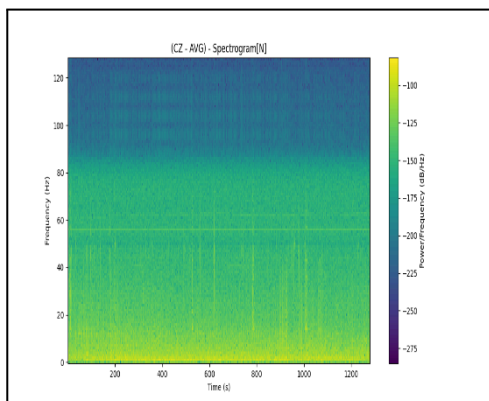
In this segment, the spectrogram image is generated by utilising the incorporation of notch filter banks and STFT. This pre-processing scheme's target is to segment the signals into one-dimensional features that aid in obtaining the spectrogram images. Six filter banks with a frequency range from 1 to 70 Hz are employed for obtaining the spectrogram images. Algorithm 1 provides the proposed EEG pre-processing. Figure 2 shows the spectrogram images.

Step

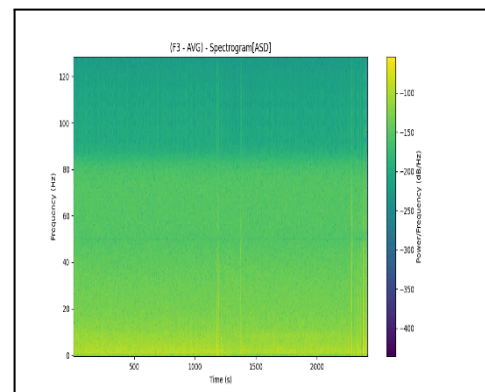
```

1
2
3
4
5
6
7
8
9
10
11
Algorithm-1 EEG Pre-processing technique
Input: EEG Signals
Output: Spectrogram Images
For n=1 to Max_iteration
    EEG_Signal_1 = Filter_1(EEG_raw)
    EEG_Signal_2 = Filter_2(EEG_raw)
    EEG_Signal_3 = Filter_3(EEG_raw)
    EEG_Signal_4 = Filter_4(EEG_raw)
    EEG_Signal_5 = Filter_5(EEG_raw)
    EEG_Signal_6 = Filter_6(EEG_raw)
    Spectrogram_images= STFT(EEG_Signal_1, ..... EEG_Signal_6 )
End

```



(a)



(b)

Figure 2. Spectrogram Images - (a) (b) Non-Autistic Subjects

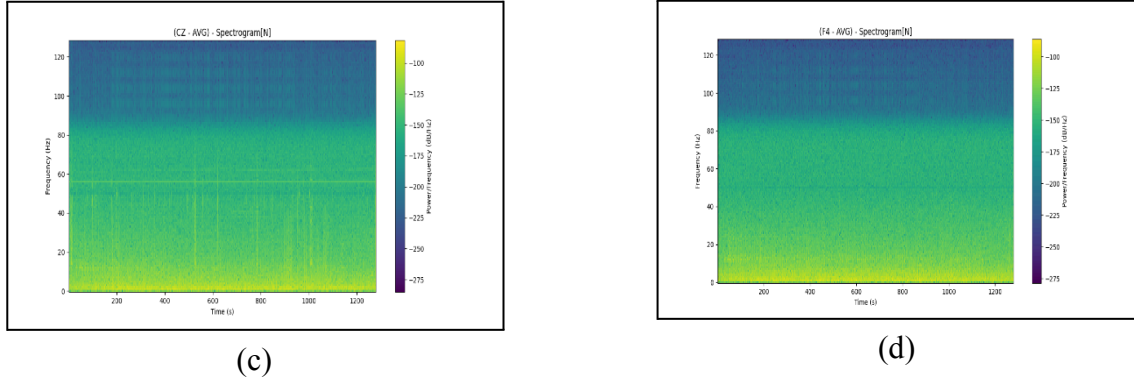


Figure 2. Spectrogram Images - (c) (d) Autistic Subjects.

3.3. Dense Feature Extraction Module

For effective feature extraction, a combination of lightweight pre-trained models with modified residual gated recurrent units is created to eradicate the problem of computational complexity from the traditional convolutional networks. In the first part, the transfer learning model is used to acquire the lightweight deep feature extraction. MobileNet and SqueezeNet are utilised for extracting the spatial features from the spectrogram images. The selected lightweight networks underwent training on the ImageNet datasets with millions of images categorised into 1000 classes. Thus, these learning networks are developed using Python 3.19 programming. The extracted images are fed to this pre-trained structure for deep information. Table 2 emphasises characteristics to pre-trained schemas leveraged for feature extraction.

Table 2. Pre-trained Models used in the Feature Extraction Module.

Types	MobileNet	Squeezenet
Version	V2	V1
Convolutional layer used	19	18
No of filters	32	21
Activation Function	ReLU	ReLU
Application	Image processing	Object detection and Image processing

From the above Table 2, the traditional ReLU activation function was replaced with the PRELU (parameterised ReLU) to obtain the lossless data transmission when compared to ReLU. The features extracted are then fed to modified residual gated recurrent units for the deep feature extraction.

3.3.1. Gated Recurrent Units (GRU)

GRUs are regarded as intriguing derivatives of Long Short-Term Memory (LSTM) networks. This concept was prescribed by Chung et al. (2014), aiming to merge unnoticed gate as well as the database available in the structure input matrix as a single unit. The GRU architecture is designed to handle extended sequences and maintain long-term memories effectively. Additionally, it is substantially less complicated than LSTM networks. Figure 3 shows the GRU network.

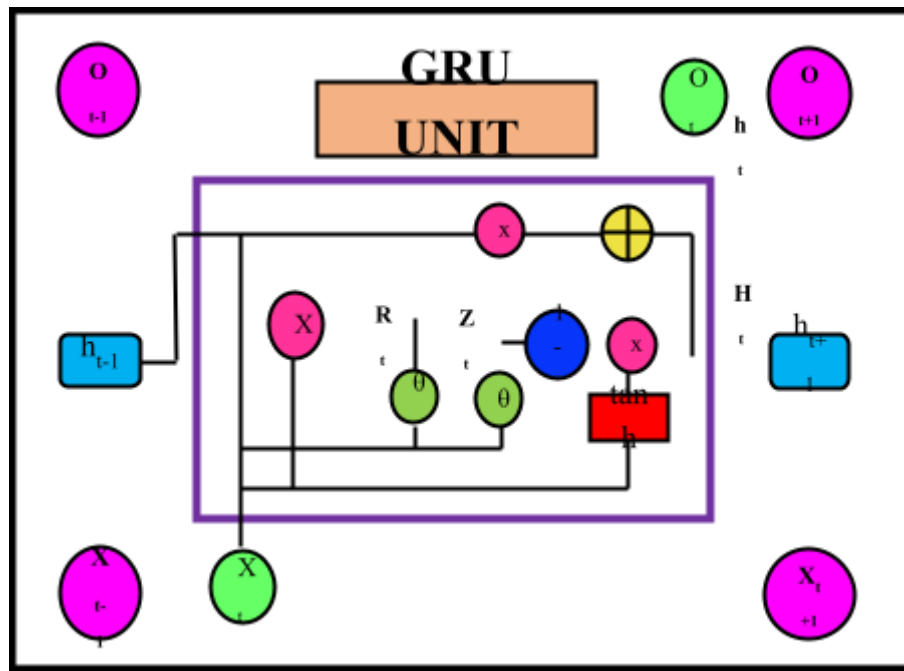


Figure 3. GRU-network Architecture.

Following equations are coined by Chung et al. (2014) to depict the attributes of GRU

$$l_t = (1 - f_t) \odot l_{t-1} + f_t \odot l_t \quad (1)$$

$$\tilde{l}_t = g(W_l f_t + U_l (r_t \odot l_{t-1}) + b_l) \quad (2)$$

$$z_t = \sigma(W_z f_t + U_z l_{t-1} + b_z) \quad (3)$$

$$r_t = \sigma(W_r f_t + U_r l_{t-1} + b_r) \quad (4)$$

The comprehensive GRU equation

$$P = GRU(\sum_{t=1}^n [f_t, l_t, z_t, r_t (W(t), B(t), \eta(tanh))]) \quad (5)$$

Here, f_t denotes the input attribute at present available information, so that y_t represents the output state. The value l_t signifies algorithm steps manipulated output in the present moment. Additionally, z_t is the update gate and r_t is the reset gate. $W(t)$ indicates the weights, as well as $B(t)$ signifies arm power updatation in the present moment. Moreover, the architecture of the GRU is simplified, whereby each gate is calculated solely by relying on the preceding hidden state and the bias. This modification contributes to a decrease in the overall number of parameters compared to the conventional GRU. The mathematical representation for the modified GRU is given as Equations (3) & (4) have been modified as

$$z_t = \sigma(U_z l_{t-1} + b_z) \quad (6)$$

$$r_t = \sigma(+ U_r l_{t-1} + b_r) \quad (7)$$

Hence, the Overall modified GRU's functionality is depicted as

$$F = GRU(\sum_{t=1}^n [f_t, l_t, z_t, r_t(B(t), \eta(tanh))]) \quad (8)$$

3.3.2. Modified Residual GRU Networks

Residual Connections are commonly used in the ResNETS, which is the class of deep neural networks that primarily utilised the image to categorise and identify applications. The role of residual connections in the recommended scheme is to surpass the vanishing gradient challenge which is then used to determine the redundant information and reserve the important feature in GRU networks. Figure 4 illustrates the proposed Modified-GRU architecture. The residual connection network is embedded in the standard GRU. In this model, the residual connection consists of 2 convolutional layers pursued by the batch normalisation (BN), residual block, and an activation function Leaky Relu layers (LReLU). The data needs to be input into the proposed skip connection and the R_k is calculated based on the blocks involved in designing the proposed Modified residual GRU networks.

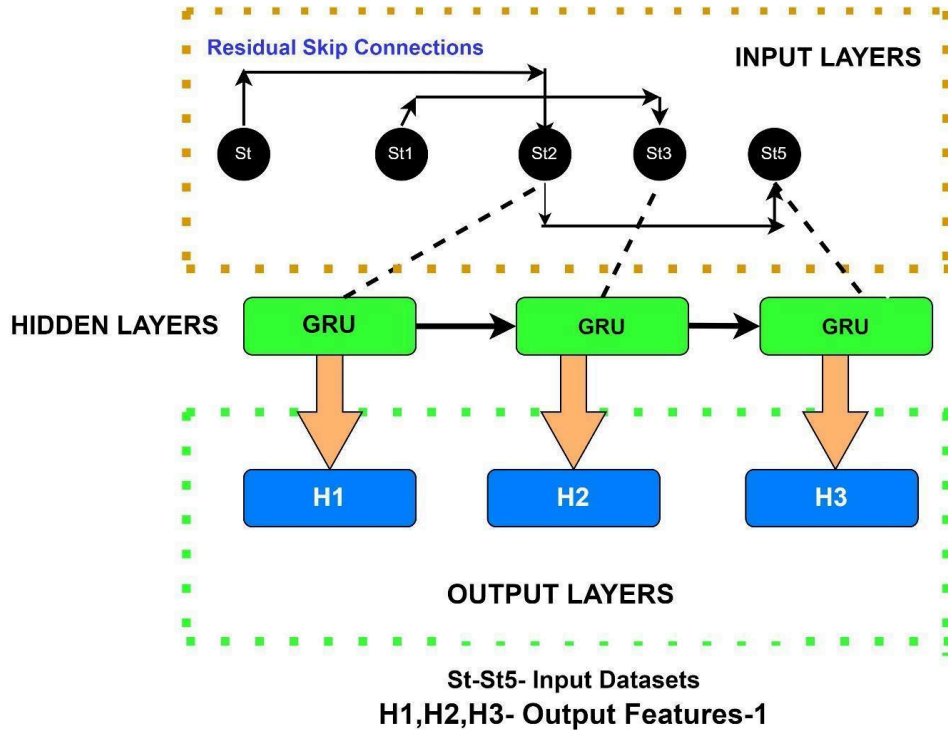


Figure 4. Recommended Residual-GRU structure for the Feature Extraction.

Mathematically, the residual-GRU networks are expressed as

$$H(t) = LReLU(Res(W1.S1 * b1)) \quad (9)$$

$$Rk = Softmax(Res(W2.H(t) * b2)) \quad (10)$$

Where $W1$, $W2$, $b1$, and $b2$ are weights and biases of the convolutional layers, $H(t)$, the state called the hidden state. Integration of residual and modified GRU offers the following advantages:

- Less Complexity Overhead
- Preventing the vanishing gradient problems
- Suitable to extract temporal details through non-linear electroencephalogram signals.

3.4. Dense Classification Layers

In this work, Chaotic CAT Swarm Optimised Deep learning networks are used for an effective prediction in ASD and mild average person. Training as well as testing Will be through 10 times cross-validated to diminish the vectors during the testing process. A detailed description of the model is explained below.

3.4.1. CAT Swarm Optimisation (CSO) Model

CSO algorithm is one of the optimisation techniques driven by bio-functional cats habits. Cats are known for their laziness, often spending long periods resting. However, during these times, they maintain a heightened awareness of their environment. They attentively observe their surroundings and, when a target catches their attention, they swiftly pursue it. The CSO algorithm models these two primary behaviours of cats: resting (tracing mode) and hunting (seeking mode). In this algorithm, each cat symbolises a potential solution, characterised by its position in an M-dimensional search space, a fitness value representing manipulated output. The first step involves determining the number of cats participating in the optimisation process. Throughout the iterations, the best-performing cat is recorded, and the solution represented by the best cat at the final iteration is considered manipulated output. The operational flow of the Cat Swarm Optimisation technic illustrated as displayed in Figure 5. The subsequent section will elaborate on the mechanisms of the seeking and tracing modes.

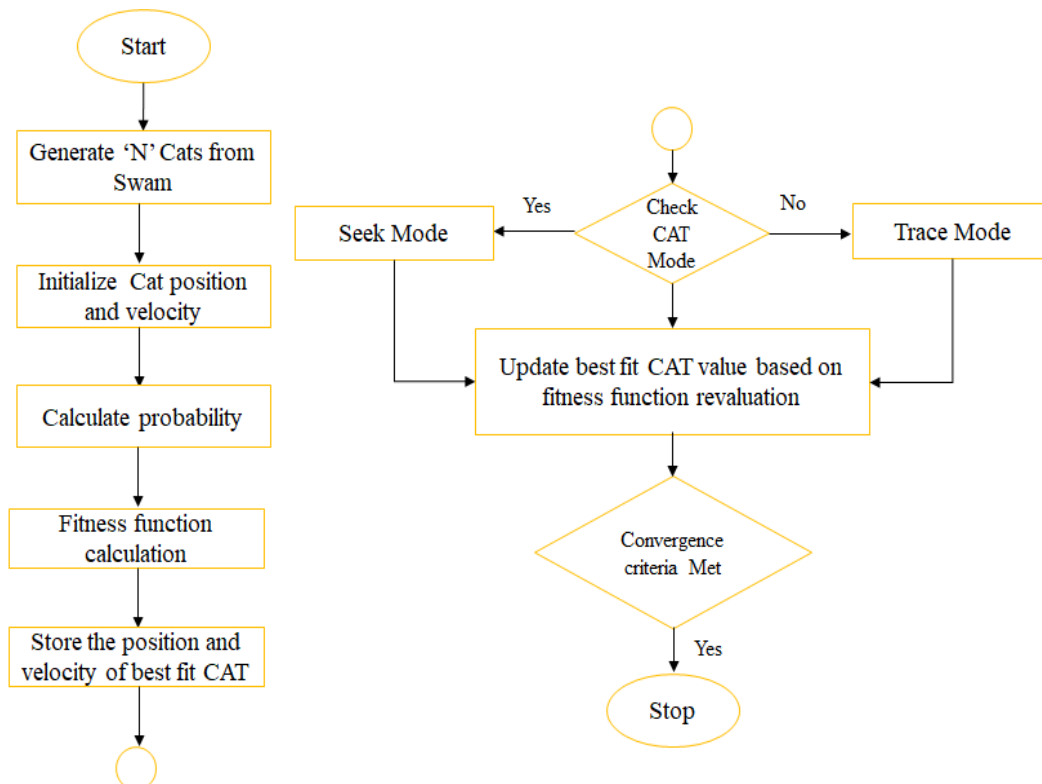


Figure 5. CAT Swarm Optimisation Algorithm work flow diagram.

3.4.1.1 Seeking Modes

Replicates bio-function of the cat, with 4 essential measures: seeking memory reservoir (SMP), seeking parameter scope (SRD), count of dimensions to modify (CDC), and self-location consideration (SPC). Users adjust and define these values through a trial-and-error approach.

Cats, SMP determines quantity potential locations, specifying how many candidate spots will be generated for selection by each cat. For instance, a set of SMP initiates the cat's numerical

strength produced, from which anyone will be chosen as the successor destination for the cat. The method of generating these new positions relies on two additional parameters: the Candidate Dimension Count (CDC) and the Search Rate Determinant (SRD). The CDC indicates the number of dimensions that will be altered, ranging from 0 to 1. An analogy, in a searching area with five measures, if the CDC is set to 0.2, then four random dimensions will be modified for each cat, while the remaining one will remain unchanged. The SRD represents the mutation ratio applied to the selected dimensions, dictating the extent of changes made to those dimensions identified by the CDC. Lastly, the Search Position Candidate (SPC) is a Boolean value that indicates the present location where the cat will be included as a potential candidate for the subsequent cycle.

If the SPC flag is activated, then for every category, we must create (SMP-1) candidates rather than the full SMP count, as the existing position counts as one of them. The following are the procedures for the seeking mode.

- (1) Create multiple identical copies of the present location of the Cat
- (2) For each instance, randomly choose a number of CDC dimensions to undergo mutation. Additionally, randomly either increase or decrease the SRD values from the current figures, which will then update the previous values according to the formula presented in equation (11).

$$x(\text{new_cat}) = (1 + \text{rand} + \text{SRD}) * x(\text{old_cat}) \quad (11)$$

Where $x(\text{new_cat})$ represents the updated location of the cat, $x(\text{old_cat})$ denotes the starting point and signifies a different schedule ranging from 0 to 1.

- (3) The Fitness Function (FF) is computed, subsequently, a person placement is chosen according to the maximum value of the FF according to the probability outlined in the following expression (12).

$$P(i) = |(FF(i) - FF(b)) / (FF_{\max} - FF_{\min})| \quad (12)$$

$FF(i)$ represents the fitness present individual cat, $FF(b)$ indicates cats numerical strength population, $FF_{\max}$ denotes the highest measures, $FF_{\min}$ signifies the lowest value.

3.4.1.2. Tracing Modes

The movement mimics the tracking behaviour observed in felines. Initially, random velocity measures are assigned to each dimension and cat location. In subsequent iterations, these velocity values must be revised. The movement of cats in the different situations occurs: (i) Upgrade velocities ($V(\text{CAT})$) attributes as per subsequent equation

$$V(\text{CAT}) = V(\text{CAT}) + a * c (x(\text{new_cat}) - x(\text{old_cat})) \quad (13)$$

Where a and c are constants.

3.5. Chaotic Tuned CAT Swarm Optimisation Model

Metaheuristic techniques are often advocated for minimising execution time and computational expenses due to the intricate and challenging nature of real-world engineering or design problems. Nonetheless, these methods may encounter issues like premature convergence, inconsistent search behaviour, local optimum traps, inefficient searching processes, and reduced population diversity, which can lead them through challenges are evident as the CAT Swarm technique addresses the shortcomings, this research introduces a novel hybrid algorithm that incorporates chaotic maps. These maps possess benefits such as the ability to escape local regions and enhance the search process through their dynamic characteristics. While the CAT Swarm

algorithm exhibits an unbiased and practical convergence rate, it doesn't consistently excel in locating the global optimum, which influences the overall convergence performance. Therefore, the theory of chaos is integrated into the proposed scheme, named CTCSSO, to mitigate this limitation and improve overall effectiveness.

In broad terms, chaos refers to a deterministic yet seemingly random process found within a nonlinear, dynamic system that lacks periodicity and does not converge, all while remaining finite. This concept is particularly sensitive to initial conditions. Chaos characterises a tangled system where behaviour is highly unreliable. From a quantitative perspective, chaos illustrates the randomness inherent in a straightforward system. This allows chaotic systems to use generators as various chaotic maps, defined by distinct quantitative formulations. These mappings are employed to ensure that the chaotic objective replicates chaotic behaviour in the optimisation algorithm, achieved by utilising specific parameters through a function. This approach facilitates a more dynamic and global exploration of the search space, thereby enhancing the schema's dynamic attributes. These maps display intricate and dynamic behaviours within nonlinear systems. In summary, chaotic maps can serve as traditional optimisation searches, often leading to superior outcomes compared to random numbers.

The Notion of introducing the chaotic principles in the CAT Swarm Optimisation Model both in seeking and tracing mode is to increase the exploration and search phase so that the proposed chaotic CAT model can give better efficiency. The chaotic variables are used in the seeking and tracing modes of the CAT model and detailed description of chaotic variables. This research incorporates the logistic chaotic maps used for increasing the exploration and search space of the CAT to obtain a better application.

Initially, all the variables in the traditional model are replaced with the chaotic variables which are initialised at the seeking and tracing process. Thus the proposed CTCSSO model based on random behaviour is created by the chaotic concepts thereby increasing the convergence rate by deploying the chaotic maps. The CTCSSO demonstrates a greater ability to evade local traps compared to traditional stochastic techniques. Consequently, the randomness inherent in CTCSSO will substituted which shares characteristics of randomness but offers enhanced attributes. Within a desirable zone, which is anticipated for a brief early duration but behaves stochastically over an extended period. After the tracing period, each position is updated with the chaotic map, to generate the new position. All the random variables in Equation 14 are now replaced with the chaotic variables

$$x(new_chaotic_cat) = (1 + R + SRD) * x(old_chaotic_cat) \quad (14)$$

where $x(new_chaotic_cat)$ represents the new cat logistic chaotic, $x(old_cat)$ is the cat's primary logistic chaotic variable, and R denotes the chaotic duration, derived from the logistic maps. After chaotic positioning of cats, fitness function of the chaotic cats are calculated based on Equation 12.

During the initial cycle, the opening rate is distributed across disordered placements. To explore more search spaces during the tracing process, Equation (14) is updated with the chaotic variables, resulting in the formulation presented in Equation (15).

$$(CAT) = V(CAT) + a * c (x(new_chaotic_cat) - x(old_chaotic_cat)) \quad (15)$$

3.6 Hyperparameter Tuning Process

The Chaotic Cat Optimiser is employed to enhance the weights of GRU's dense networks. Initially, the hyperparameters are chosen randomly and fed into the GRU training network. The novel fitness function which is coined based on the CCO model is given in Equation 16.

$$\text{Fitness Function} = \text{Min}(\text{MSE}(\text{Predicted value} - \text{Actual Value})) \quad (16)$$

The fitness function is calculated relying on the minimum error which is measured by MSE (mean Square Error) among the predicted value and actual value. Once the hyperparameters are optimised using Equation 16, dense training layers classify data into normal and ASD. The complete mode of operation of the recommended model is presented in **Algorithm-2**. Table 2 utilises the specification of the training network obtained after the optimisation.

Steps	Algorithm-2 // Pseudo Code for the Proposed Model
01	Input = Bias weights, Hidden layers, Epochs, Learning Rate
02	Output : Prediction of Normal/ASD
03	Arbitrarily allocate the bias weights, hidden layers, epochs, learning rate
04	Commence the three parameters such as
05	While (true)
06	Measure the output from GRU cells by utilising equation(9)&(10)
07	Measure the Fitness function by utilising the equation(16)
08	For t=1 to N where N= Maximum Iteration
09	Allocate the bias weights and input layers by equation (14)and (15)
10	Measure the fitness function utilising equation(16)
11	If (Fitness function == threshold)
12	Go to Step 17
13	Else
14	Go to Step 08
15	End
16	End
17	If (output value <=1)
18	//Normal is Fixed
19	Else if(output value >1 && output value <=2)
20	// ASD is Fixed
21	Else
22	Go to Step 09
23	End
24	End
225	End

4. Results and Discussion

This section outlines the implementation specifics and performance metric results from two distinct experiments. Comprehensive ablation and statistical analyses have been thoroughly executed and verified. Ultimately, the results are evaluated and contrasted with other established models, highlighting the distinctiveness of the proposed model for ASD prediction.

4.1. Implementation

The suggested foundational schema was implemented utilising Keras, with TensorFlow serving as its backend.

4.2. Effectiveness Review

The effectiveness recommended technique is assessed by utilising various assessment metrics, belongs to accuracy, precision, recall, specificity, and F1-score. These metrics are then compared with those of advanced deep learning (DL) schemas to emphasise the superiority of the proposed approach. All assessment metrics and overhead costs are analysed to depict the scheme as streamlined requiring lesser operational costs. Table 3 illustrates the quantitative equations employed in calculating the assessment metrics. To mitigate issues related to generalisation and overfitting, the early stopping technique is implemented. This method is employed to terminate the training process when the validation performance of the recommended schema ceases to improve over time.

Table 3. Assessment measures in identifying the adopted in the evaluation.

Performance Measures	Expression
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Recall	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Precision	$\frac{TP}{TP + FP}$
F1-Score	$2 \cdot \frac{Precision * Recall}{Precision + Recall}$

A predictive case comprises four components: the first is the true positive (TP), where values are recognised as true and are indeed true. Then the false positive (FP), where values are incorrectly signified as true despite being false. Then the false negative (FN), in which the true is chosen as opposite to positive. Lastly, the fourth is the true negative (TN), where a negative value is accurately recognised as negative.

4.3. Experimental Outcomes

To substantiate the excellence of the recommended model, performance evaluation tests (PET) and Statistical Validation tests (SVT) are conducted. In the PET analysis, existing models such as CNN's variants such as MobileNets, Hybrid MobileNets, MobileNets+Squeezenets, GRU networks, Bi-GRU Networks, and HCLA_BiGRU are considered, while their performance was calculated based on the collected datasets.

4.3.1. PET Analysis

The PET experimentation involves a ten-fold cross-validation of the recommended scheme, utilising various optimisers as previously noted. The PET process is conducted by evaluating assessment metrics by partitioning the dataset into seventy percent for training, twenty percent for testing, and ten percent for validation, respectively.

4.3.1.1. PET Study Findings

Figure 6 illustrates the confusion matrix for the proposed model designed for differentiating between ASD persons and mild average persons. It is evident from the matrix that the recommended scheme demonstrates exceptional performance in telling apart ASD patients and

non-ASD individuals. Additionally, Figure 7 presents the validation loss during the training process of the recommended scheme.

The validation curves signify that the discrepancy between the actual and predicted values remains remarkably low, at just 0.001, which underscores the potency of the suggested scheme in identifying both ASD and non-ASD patients. Figure 8 shows the performance of validation Curves of the Recommended Schema in identifying the health of the complete datasets. Table 4 assessment metrics related to the network and film stricture frameworks in identifying ASD patients when $k=5$, where k represents the Number of subsets in the dataset. Table 5 assessment metrics associated with the deep learning frameworks for the Healthy when $k=5$. Table 6 assessment metrics of the deep learning frameworks in identifying the ASD patients when $k=10$. Table 7 result metrics of the deep learning frameworks in detecting healthy patients when $k=10$.

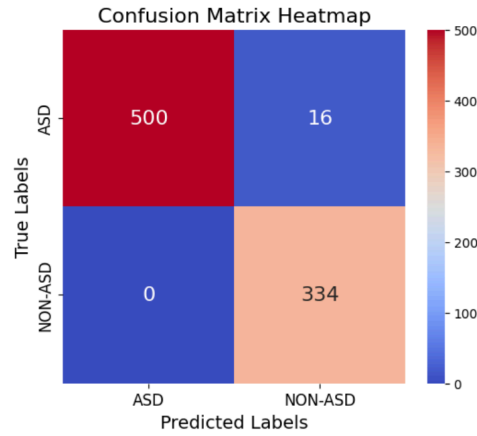


Figure 6. Confusion Matrix of the Recommended Schema Model in identifying the ASD and Non-ASD patients.

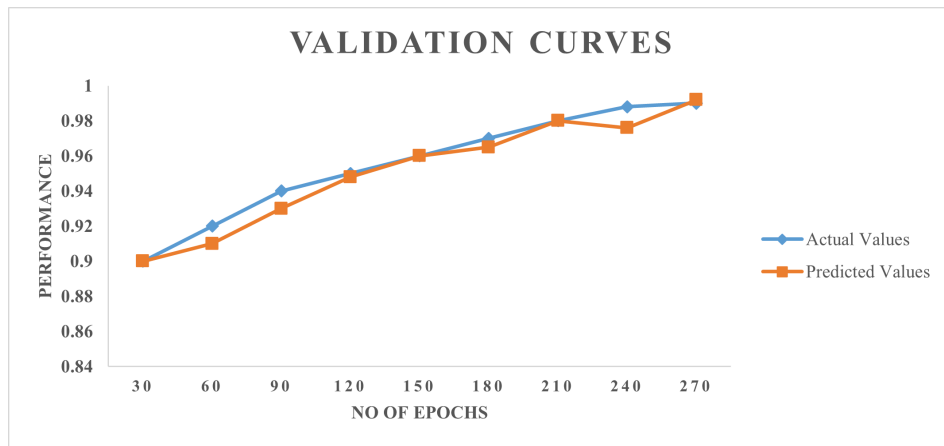


Figure 7. Validation Curves of the Recommended Schema in identifying the ASD in the complete datasets.

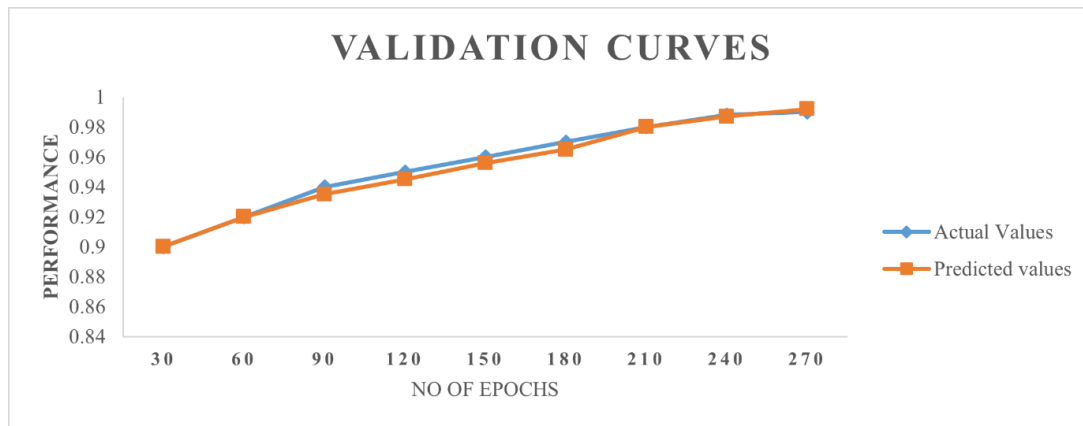


Figure 8. Validation Curves of the Recommended Schema recognising grow healthy in the complete datasets.

Table 4. Assessment Metrics of the deep learning frameworks in identifying the Healthy when $k=5$.

Algorithms	assessment metrics				
MobileNets	0.83	0.82	0.814	0.83	0.822
SqueezeNets	0.823	0.812	0.810	0.82	0.820
MobileNets+Squeezenets	0.840	0.837	0.832	0.83	0.84
GRU Networks	0.843	0.834	0.831	0.831	0.823
Bi-GRU Networks	0.85	0.843	0.841	0.839	0.84
HCLA-Bi-GRU	0.912	0.903	0.902	0.90	0.92
Proposed Model	0.982	0.981	0.972	0.965	0.982

Table 5. Assessment Metrics of the deep learning frameworks in identifying the Healthy when $k=5$.

Algorithms	Assessment Metrics				
	Accuracy	Precision	Recall	Specificity	F1-score
MobileNets	0.825	0.819	0.820	0.820	0.831
SqueezeNets	0.829	0.812	0.810	0.82	0.820
MobileNets+Squeezenets	0.845	0.837	0.832	0.83	0.84
GRU Networks	0.842	0.834	0.831	0.831	0.823
Bi-GRU Networks	0.854	0.843	0.841	0.839	0.84
HCLA-Bi-GRU	0.923	0.903	0.902	0.90	0.92
Proposed Model	0.982	0.981	0.972	0.965	0.982

Table 6. Assessment Metrics frameworks in identifying ASD patients when $k=10$.

Algorithms	Assessment Metrics				
	Accuracy	Precision	Recall	Specificity	F1-score
MobileNets	0.833	0.819	0.820	0.820	0.829
SqueezeNets	0.842	0.812	0.810	0.82	0.819
MobileNets+Squeezenets	0.856	0.837	0.832	0.83	0.838
GRU Networks	0.854	0.834	0.831	0.831	0.832
Bi-GRU Networks	0.873	0.843	0.841	0.839	0.842
HCLA-Bi-GRU	0.931	0.903	0.902	0.90	0.90
Proposed Model	0.984	0.981	0.969	0.966	0.983

Table 7. Assessment Metrics in detecting the healthy patients when $k=10$.

Algorithms	Assessment Metrics				
	Accuracy	Precision	Recall	Specificity	F1-score
MobileNets	0.835	0.819	0.820	0.820	0.829
SqueezeNets	0.840	0.812	0.810	0.82	0.819
MobileNets+Squeezenets	0.866	0.837	0.832	0.83	0.838
GRU Networks	0.862	0.834	0.831	0.831	0.832
Bi-GRU Networks	0.875	0.843	0.841	0.839	0.842
HCLA-Bi-GRU	0.929	0.903	0.902	0.90	0.90
Proposed Model	0.987	0.981	0.969	0.966	0.983

4.3.2. Discussions

Figure 6 represents the confusion matrix of the recommended schema in predicting ASD and non-ASD. Figure 7 demonstrates the recommended schema's validation performance and it shows the mean square error is almost zero among the actual and predicted values. In Figure 7, it is evident that the recommended schema shows its stable performance in detecting ASD and NON-ASD from the EEG signals. As the next step, the evaluation involves the K-fold validation process. This approach ensures the trustworthy potency of a Schema's effectiveness performance by utilising various training and testing datasets and calculating the average outcomes. It also aids in reducing the impact of specific biases present in the data by guaranteeing that every subset is employed for training and testing purposes. Different proportions of K values are utilised to examine the suggested model. The potency of the schema at the different values of K is presented in Tables 4-7. The proposed model shows the most promising performance since it combines the strength of the pre-trained model, Residual GRU, and Optimised Learning networks. It is evident from the table (K=5), in the detection of ASD and healthy subjects, that the accuracy of the model is 0.987, precision is 0.981, recall of 0.969, and F1-score is 0.983, optimal and suboptimal values, average, central point, measures of spread such as deviation and fluctuation, and variability indicators. As the K increases to 10, the performance of the recommended scheme sails to a higher performance by outperforming the other models in detecting ASD and healthy subjects respectively. The entries in Table s indicate the MobileNets and Squeezenets produced the lowest performance (0.835 and 0.840) whereas the hybrid models such as HCLA-BIGRU (accuracy is 0.91) and the proposed model (accuracy is 0.987) produced the best performance of prediction. The integration of chaotic cat swarm and modified GRU networks has achieved the best performance in classifying normal and ASD patients.

4.4. Statistical Analysis

Following the PET analysis in binary classification results, the findings of the proposed schema have been statistically verified. A thorough examination of varied schemas has been conducted, highlighting their unique features and limitations. The table portrays the statistical validation of numerous chaotic optimiser models, with meta-heuristic algorithms infused with CNN+GRU being trained, tested, and evaluated. The results have been cultivated and analysed with recommended schema. The meta-heuristic algorithms employed in the experiments include Chaotic+Particle Swarm Optimisation (CPSO), Chaotic Genetic Algorithm (CGA), Chaotic Whale Optimisation Algorithm (CWOA), Chaotic+Spider Optimisation Algorithm (CSOA), Chaotic Sand Cat Optimisation (CSCO), and Chaotic White Shark Optimisers (CWSO). High-performing optimisers are infused with recommended deep learning designs, further, validation will be performed by comparing them with the Chaotic Cat Swarm Optimisation in the model. Additionally, statistical results are computed on the fitness function evaluation over 100 iterations, like optimal, inferior, average, median, dispersions, and variance, respectively. Furthermore, the results of the indicator functions are also measured. Table 8 Findings shaped on Fitness Function Assessment for Multiple Optimisers. Table 9 Outcomes from Indicator Function applications across diverse optimisers

Table 8. Findings shaped on Fitness Function Assessment for Multiple modifiers.

Algorithm	Best	Worst	Mean	Median	SD	Variance
CPSO	0.7590	0.6475	0.72510	0.022919	0.06550	7.4×10^{-6}
CGA	0.730330	0.63525	0.69034	0.020202	0.07033	6.390×10^{-6}
CWOA	0.7523	0.6763	0.65372	0.027820	0.068903	5.892×10^{-5}
CSOA	0.77435	0.6902	0.71239	0.02039	0.054637	4.1290×10^{-4}
CSCO	0.80234	0.64389	0.73402	0.0302-2	0.059034	3.903×10^{-4}
CWSO	0.85634	0.69034	0.74022	0.039303	0.045890	3.6790×10^{-4}
Proposed Model	0.99763	0.81202	0.85640	0.06344	0.037630	2.28930×10^{-4}

Table 9. Outcomes from Indicator Function applications across diverse modifiers.

Algorithm	Best	Worst	Mean	Median	SD	Variance
CPSO	0.74910	0.6455	0.72510	0.023819	0.06540	7.4×10^{-6}
CGA	0.730330	0.63525	0.69034	0.020202	0.07033	6.390×10^{-6}
CWOA	0.7523	0.6763	0.65372	0.027820	0.068903	5.892×10^{-5}
CSOA	0.77435	0.6902	0.71239	0.02039	0.054637	4.1290×10^{-4}
CSCO	0.80234	0.64389	0.73402	0.0302-2	0.059034	3.903×10^{-4}
CSWO	0.85634	0.69034	0.74022	0.039303	0.045890	3.6790×10^{-4}
Proposed Model	0.99763	0.81202	0.85640	0.06344	0.037630	2.28930×10^{-4}

The data in these tabulations clearly show that the recommended schema has achieved optimal results compared to the varied optimisation algorithms. The reliability of the schema is portrayed in Figure 11. The recommended schema has brought the optimal findings while assessing the varied residing schema. The convergence analysis of the recommended approach in comparison with the previously outlined plans for varying loops is illustrated in Figure 9.

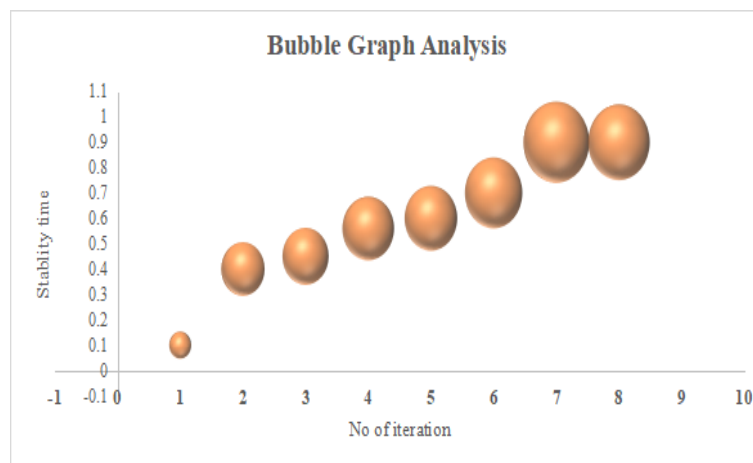


Figure 9. Bubble Graph Assessment of the Proposed Optimisation techniques incorporated with the DL model.

The recommended optimisation utilising the chaotic principle has improved marginally over the varied optimisation schemes. While the CSO has shown the optimal findings among traditional algorithms, the incorporation of logistic chaotic maps into CSO has resulted in superior performance, making it an optimal choice for adjusting the hyperparameters of dense training networks.

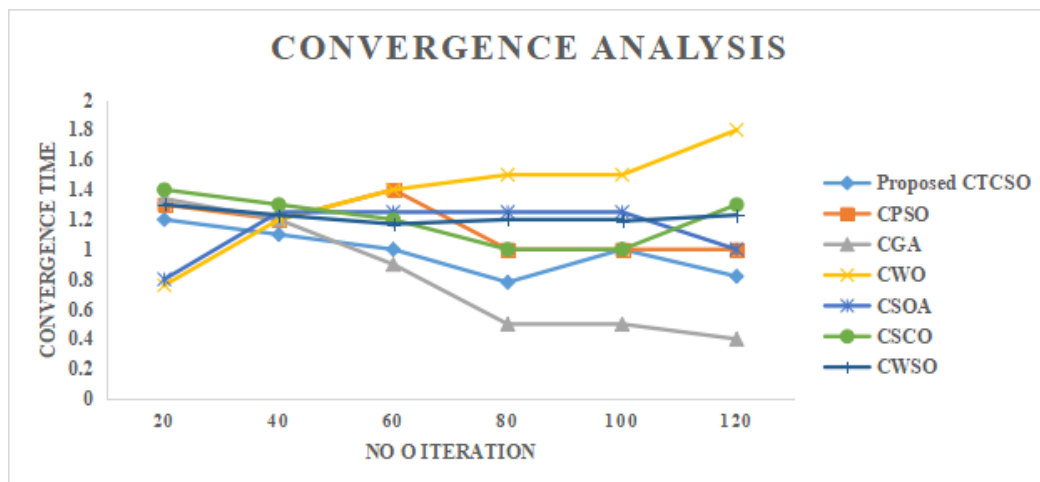


Figure 10. Convergence Assessment for the Varied Optimisation Schemes.

5. Conclusion and Future Scope

Scientific real-time work that recommended the model revealed impressive findings in classifying the ASD and normal data, achieving an accuracy of 0.982, precision of 0.97, recall of 0.98, and F1-score of 0.98. Attainment related to this model relies on the hybrid combination of modified GRU and chaotic cat-optimised architectures underscoring its effectiveness in classifying ASD and non-ASD patients. The proposed model's exceptional performance validates its potential to reliably classify both positive and negative cases. Furthermore, precision and recall portray the capacity of the recommended schema to differentiate the non-ASD and ASD from the EEG signals.

The application of Modified GRU networks, pre-trained layers, and optimised classification components has a major impact in effectively captivating the complex temporal and spatial characteristics of electroencephalogram signals with the accurate identification of ASD data. This innovative research represents a major leap forward in the ASD domain, providing a key asset for healthcare professionals and stakeholders aiming regarding the elevation of the diagnosis management appropriately ASD. Although additional studies are essential to confirm the schema's effectiveness on larger datasets related to practical medical environments, initial results underscore the potency of complexity-aware deep learning methods to enhance healthcare results.

Acknowledgement

The authors wish to thank the Research Management Centre (RMC) of Universiti Tun Hussein Onn Malaysia (UTHM) for supporting this work through the W029 international research grant. The authors would also like to thank and appreciate the management, and administrative officials of UTHM, Malaysia, and M/s. AK Akan Mining and Infrastructures from India for their support in carrying out the research work successfully.

References

- Abbas, H., Garberson, F., Liu-Mayo, S., Glover, E., & Wall, D.P. (2020). Multi-modular AI approach to streamline autism diagnosis in young children. *Scientific Reports*, 10, 5014. <https://doi.org/10.1038/s41598-020-61213-w>.
- Abdolzadegan, D., Moattar, M.H., & Ghoshuni, M. (2020). A robust method for early diagnosis of autism spectrum disorder from EEG signals based on feature selection and DBSCAN method. *Biocybernetics and Biomedical Engineering*, 40(1), 482–493. <https://doi.org/10.1016/j.bbe.2020.01.008>.
- Ali, N.A., Syafeeza, A.R., Jaafar, A.S., & Alif, M.K.M.F. (2020). Autism spectrum disorder classification on electroencephalogram signal using a deep learning algorithm. *IAES International Journal of Artificial Intelligence*, 9(1), 91–99. <http://dx.doi.org/10.11591/ijai.v9.i1.pp91-99>
- Amrani, G., Adadi, A., Berrada, M., Souirti, Z., & Boujraf, S. (2021). EEG signal analysis using deep learning: A systematic literature review. *2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS)*, 1–8. <https://doi.org/10.1109/ICDS53782.2021.9626707>.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1412.3555>
- Ghembaza, F., & Djebbari, A. (2023). Novel quadratic time-frequency features in EEG signals for robust detection of epileptic seizure. *Research on Biomedical Engineering*, 1–23. <https://doi.org/10.1007/S42600-022-00256-6>.
- Goel, N., Grover, B., Anuj, Gupta, D., Khanna, A., & Sharma, M. (2020). Modified Grasshopper Optimization Algorithm for detection of Autism Spectrum Disorder. *Physical Communication*, 41, 101115. <https://doi.org/10.1016/j.phycom.2020.101115>.

- Patel, M., Bhatt, H., Munshi, M., Pandya, S., Jain, S., Thakkar, P., & Yoon, S. (2024). CNN-FEBAC: A framework for attention measurement of autistic individuals. *Biomedical Signal Processing and Control*, 88, 105018. <https://doi.org/10.1016/j.bspc.2023.105018>.
- Siddiqui, F., Mohammad, A., Alam, M.A., Naaz, S., Agarwal, P., Sohail, S.S., & Madsen D.Ø. (2023) Deep Neural Network for EEG Signal-Based Subject-Independent Imaginary Mental Task Classification. *Diagnostics (Basel)*, 13(4), 640. <https://doi.org/10.3390/diagnostics13040640>.
- Xu, L., Liu, Y., Yu, J., Li, X., Yu, X., Cheng, H., Li, J. (2020). Characterizing autism spectrum disorder by deep learning spontaneous brain activity from functional near-infrared spectroscopy. *Journal of Neuroscience Methods*, 331, 108538. <https://doi.org/10.1016/j.jneumeth.2019.108538>