Int J Clin Exp Med 2026;19(8):220-239
www.ijcem.com /ISSN:1940-5901/IJCEM0170762

Original Article

SWPBTN: a hybrid deep learning framework for accurate schizophrenia detection using fMRI analysis

Purnima Shanmugam, Uma Maheswari Oorkavalan

Department of Electronics and Communication Engineering, College of Engineering, Anna University, Chennai, India

Received December 15, 2025; Accepted April 23, 2026; Epub August 15, 2026; Published August 30, 2026

Abstract: Objectives: Schizophrenia is a severe mental condition that is marked by structural and functional abnormalities in the brain, and thus, it is difficult to diagnose early and accurately. Traditional diagnostic tools are dependent on clinical testing that is subjective and might slow down intervention. The proposed research creates a powerful and automated deep learning system of precise schizophrenia identification through the use of functional Magnetic Resonance Imaging (MRI) data. Methods: A novel hybrid architecture is suggested, which is Swin Parrot Barnacle TransdenseConvNet (SWPBTN). The framework combines the advanced preprocessing methods, such as skull stripping with FSL-BET, N4ITK bias field correction, min-max normalization, Contrast Limited Adaptive Histogram Equalization (CLAHE) enhancement, and elastic deformation to augment the data. Swin UNETR is used to anatomically divide major brain regions (hippocampus, thalamus, and prefrontal cortex), and Graph Neural Networks (GNN) are used to model functional brain connectivity. A hybrid Parrot-Gooseneck Barnacle feature selection strategy is used to extract structural, texture (GLCM), and deep learning (DL) features, which are optimized. The last classification is conducted with DenseFormer that is combined with Temporal Convolutional Networks (TCNs). Results: Tests on the OpenNeuro fMRI dataset show a higher accuracy of 99.85%, sensitivity (98.85), and F1-score of 99.9%, which is better than the available state of the art. Conclusions: The suggested framework offers a structural-functional analysis model and has a high potential for an early and valid clinical diagnosis of schizophrenia.

Keywords: Schizophrenia detection, fMRI analysis, Swin UNETR, graph neural network, feature extraction, dense former temporal convolutional networks

Introduction

The diagnosis of schizophrenia is one of the most difficult requirements since it involves severe and chronic mental illness that impacts a person's ability to think, feel, and behave. Despite the frequent difficulty and subjectivity of already available diagnostic techniques, early detection and timely management are crucial for the improvement of patient outcomes [1]. The conventional methods rely to a large extent on self-reports, clinical interviews, and observational evaluations whose quality is liable to human error and variation among physicians. The diversity of symptoms makes the diagnosis of schizophrenia difficult, and consequently, treatment delay worsens the patient's condition [2]. A branch of Artificial Intelligence (AI), DL, gives new opportunities in the solution to such problems while studying large and complex datasets for spotting minute patterns linked with schizophrenia [3]. Unlike typical models of Machine Learning (ML), deep models do not need the procedure of manual feature extraction. They learn feature representations automatically from the raw data itself. This capability enables them to work with a variety of data formats and identify complex patterns within the data [4]. Therefore, adding DL to schizophrenia research has the prospect of improving diagnostic precision and creating individualized treatment regimens for patients [5].

DL applications related to schizophrenia mostly concentrate on data derived from neuroimaging methods. Techniques of EEG and fMRI yield enormous levels of detailed information regarding the state of the human brain, demonstrating the structural and functional abnormalities

https://doi.org/10.62347/ZUWK2755

associated with the disorder [6]. These models, however, are based on CNN algorithms and are therefore capable of more accurately identifying and describing disturbed connectivity, gray matter density loss, and spatial features, among others, in MR images [7]. Similarly, EEG time-series data is analyzed by Recurrent Neural Networks (RNNs), which identify abnormalities in brainwaves associated with schizophrenia [8]. However, apart from the imaging technique, applications were used in the analysis of behavior and speech. Natural Language Processing (NLP) technologies can identify the defining symptom: disordered speech patterns [9]. DL systems provide a holistic view of schizophrenia as they incorporate multimodal data such as behavioral signals, voice, and neuroimaging that eventually provide more accurate diagnostic tools [10].

Despite the many promises of DL in schizophrenia detection, there exist many obstacles ahead. The quality and diversity in datasets are always required for strong models, although the development process is hindered due to privacy and scarcity of data issues [11]. Moreover, the "black-box" nature of a lot of algorithms in DL restricts interpretation and raises clinical decision-making transparency-related questions [12]. Explainable AI (XAI) techniques aim to address this problem, which explains model predictions so that the underlying workings are easy to understand. Other ethical considerations include reducing bias and data fairness, so the results will be fair for all patient populations [13]. These restrictions will have to be overcome by further cooperation among medical practitioners, data scientists, and policymakers [14]. If appropriately applied, DL has the potential to transform mental health diagnostics into better patient care, improved quality of life in schizophrenia patients, and early, accurate detection of the disorder [15].

The key contribution of the research is mentioned as follows:

• In this research, the SWPBTN model has been developed for the accurate detection of schizophrenia, and it is the combination of Swin UNETR, EfficientNet, and DenseFormer.

• Moreover, pre-processing is done based on the stripping, intensity normalization, bias field correction, and augmentation techniques (CLAHE, elastic deformation) to enhance data quality.

• Furthermore, a novel hybrid optimization has been implemented for feature selection, and it is the combination of Parrot and Gooseneck barnacle optimization techniques for enhancing the model performance by selecting the most relevant features.

• To integrate both structural metrics (volumetric, cortical thickness) and texture features (GLCM) for a more detailed analysis of brain alterations associated with schizophrenia.

• To utilize DenseFormer, which incorporates flattened and dense layers, a sigmoid layer, and TCN, to achieve superior performance in terms of accuracy, sensitivity, and specificity.

Methods

The input of the proposed SWPBTN framework is preprocessed raw fMRI volumes as input, which also do not require manual feature engineering. The voxel-level brain images are 3D and divided into patches and run through Swin UNETR, with transformer-based self-attention models that attend to the local anatomical features and global contextual relationships. Simultaneously, the functional connectivity matrices based on the fMRI time-series signals are learned with GNN, where brain regions are the nodes, and the strength of connectivity is described as an edge to comprehend the abnormal interactions between network nodes. EfficientNet is also used to learn deep spatial representations of the fMRI slices, leading to better structural features learning. The structural, functional, and deep features are extracted and optimized as a hybrid Parrot-Gooseneck Barnacle feature selection strategy and then sent to the DenseFormer and TCN layers, where they are finally classified. The model learns hierarchical representations with raw neuroimaging data using this end-to-end design, which is used to diagnose schizophrenia automatically. **Figure 1** presents the schematic description of the proposed SWPBTN framework.

Dataset description

In this study, the diagnosis of schizophrenia is made according to the conventional clinical criteria, which were stipulated in the DSM-5, which requires at least two core symptoms

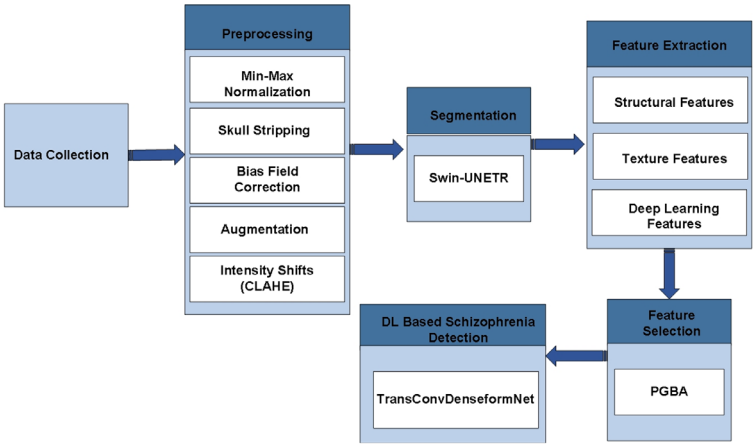


Figure 1. Overview of proposed Framework for schizophrenia detection.

such as delusions, hallucinations, disorganized speech, negative symptoms keep on for a minimum duration of six months and led to a significant impairment of functioning. Each of the subjects in the dataset is clinically diagnosed by qualified psychiatrists based on the available diagnostic criteria. The OpenNeuro fMRI dataset used in this study is a freely available, quality neuroimaging dataset that is intended to be used in the study of brain functionality and mental illnesses. It contains fMRI scans and is organized in the Brain Imaging Data Structure (BIDS) standard, which is consistent and convenient to use.

Pre-processing

The preprocessing stage is the essential stage that is used to enhance the input data for the detection process.

Normalization

Min-Max Scaling [26] is also known as normalization, which is a preprocessing step applied to change the range of feature values. In the case of schizophrenia detection from fMRI, many features used for imaging data (for example, voxel intensities, thickness, and volumes of brain regions) can have varying ranges. Min-Max Scaling makes sure that all these features have been normalized to the same range, often [0, 1], so that the ML algorithm does not favor any of them.

The Min-Max Scaling formula is given in Eq. (1).

$$x' = \frac{X\text{-}X_{min}}{X_{max}\text{-}X_{min}} \quad (1)$$

Where x' represents the scaled value of the feature within the range [0, 1], x illustrates the original value of the feature, x_{max} denoted as the maximum value of the feature in the dataset and x_{min} indicates the minimum value of the feature in the dataset.

Skull stripping

Skull stripping [27] as part of image processing is a refinement in fMRI analysis that erases the additives of the image, such as the skull, scalp, and any other obstruction from the image. This step helps to avoid analyses of the whole body; the brain is the only body part that can reveal structural or functional changes linked to schizophrenia. The most commonly used technique in this case is the Brain Extraction Tool (BET) available in the FMRIB Software Library (FSL).

The BET algorithm employs a deformable spherical surface model to obtain the brain. The deformation is governed by Eq. (2):

$$F_{Total} = F_{int} + F_{ext} + F_{img} \quad (2)$$

Where F_{int} penalizes irregular shapes to maintain a smooth surface, F_{ext} guides the surface towards strong intensity gradients, F_{img} incorporates prior knowledge about brain intensities.

Bias field correction

Bias field correction [28] is a critical preprocessing step in medical image analysis, especially fMRI, to eliminate low-frequency intensity distortions due to inhomogeneities in the magnetic field or coils. These variations can obscure the actual differences in the tissue intensity, which is important for detecting such changes as schizophrenia in the brain. These variations can be corrected using the N4ITK (N4 Bias Field Correction) algorithm, which is one of the most common. It is an updated version of the N3 (Nonparametric Nonuniformity Normalization) algorithm and is meant to increase the algorithm's stability and convergence.

The N4ITK algorithm estimates and removes the bias field (B) from the observed image (I) to

produce a corrected image (J), which has been expressed by Eq. (3):

$$I(x) = J(x).B(x) + e \quad (3)$$

Where $I(x)$ represents as observed fMRI intensity at position x, $J(x)$ denoted as the true underlying tissue intensity at position x, $B(x)$ indicates as a bias field, a smooth multiplicative low-frequency field, e represents noise or random error.

Augmentation

Rotation augmentation is a similar concept to elastic transformation, where the main idea is to artificially increase the data set by applying some realistic distortions [29]. For the detection of schizophrenia from fMRI, it enhances the generalization capability of ML algorithms by creating realistic variations in the training data with respect to brain shape and scanner artifacts.

Elastic deformation brings in spatially smooth and random transformations to the image. It is especially beneficial for such neural networks as CNN, since they are rather picky about spatial characteristics. This augmentation enhances the generalization capability of the model since it draws the model's attention to a diverse range of spatial arrangements.

For an image $I(x)$, where x represents voxel coordinates, the elastically deformed image $I'(x)$ is defined by Eq. (4):

$$I'(x) = I(x + \Delta x) \quad (4)$$

In which x represents the original voxel coordinate, Δx denotes the displacement vector at x, which specifies the elastic deformation, $I(x+\Delta x)$ denoted as interpolated value of the image at the displaced coordinate.

Intensity shifts

For fMRI preprocessing for schizophrenia detection, the enhancement of brightness and contrast is sensitive and fundamental for enhancing the features of images and enhancing the performance of the model. CLAHE [30], for instance, enhances the local contrast in an image by adapting intensity variations autonomously without over-accenting noise.

For an image with intensity values $I(x,y)$, the cumulative distribution function (CDF) is used to map original intensity values to enhanced ones, which are given in Eq. (5):

$$I'(x,y) = CDF(I(x,y)).(L - 1) \quad (5)$$

Where $I(x,y)$ denoted as the original intensity value at position (x,y), $CDF(I(x,y))$ represents a cumulative histogram of intensities, and L indicates the maximum intensity value.

Segmentation

In this segmentation stage, schizophrenia detection combines graph-based connectivity analysis with fMRI data. This strategy tackles two crucial components: The first goal of the segmentation task is to locate Regions Of Interest (ROIs), with an emphasis on the hippocampus, thalamus, and prefrontal cortex regions of the brain that are frequently linked to schizophrenia. This model used a transformer-based architecture called Swin UNETR to achieve high-resolution segmentation by recording local and global context, which allows for accurate anatomical delineation. The extraction of structural features that might indicate the morphological anomalies associated with schizophrenia is made possible by accurate segmentation.

The second part employs graph-based brain connectivity analysis to identify changes in functional brain networks. The connectivity graph has nodes for each segmented brain region and edges for functional connections measured by correlations or statistical dependencies between fMRI time series signals. It examines properties like degree centrality, clustering coefficient, and small-worldness of the network for schizophrenia patients against healthy controls using graph-theory-based measures to look for patterns of connectivity that characterize this set of patients. Such a method carries in itself an integrated diagnostic approach toward schizophrenia by fusing functional connectivity analysis with high-intensity structural segmentation in understanding the morphological changes, in addition to the network abnormalities associated with this disorder. **Figure 2** shows the Swin UNETR architecture, which has a transformer-based encoder and a UNet-like decoder to perform an accurate segmentation.

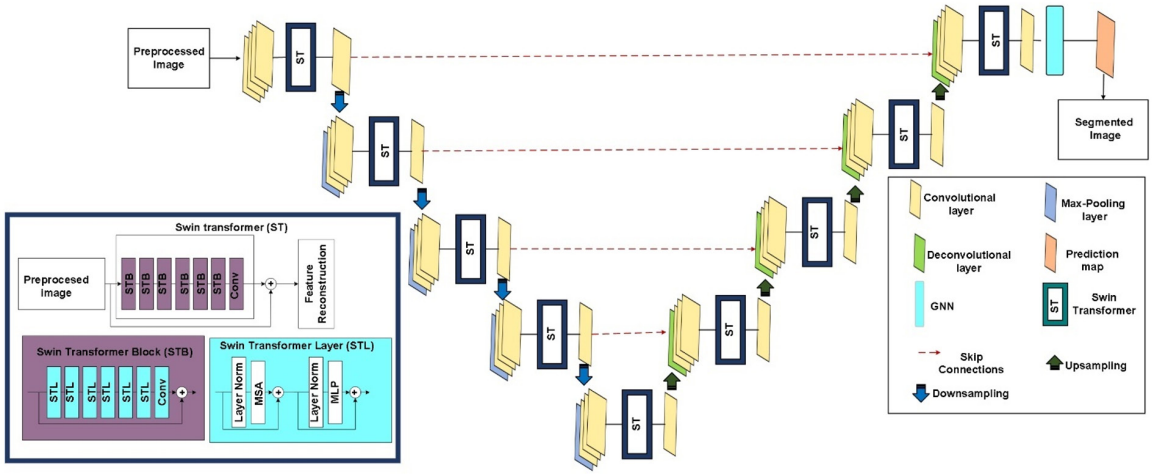


Figure 2. Overview of Swin UNETR.

Segmentation via Swin UNETR

Assuming that the input data denotes the fMRI data as $I_{fMRI} \in \mathbb{R}^{V \times V}$, where V denotes the number of brain regions, and I_{fMRI} is the connectivity matrix representing functional interactions between regions. $Y_{seg} \in \mathbb{R}^{H \times W \times D}$ is the ground truth segmentation map, where each voxel is assigned a class corresponding to the brain region. Furthermore, $Y_{graph} \in \mathbb{R}^{V \times 1}$ is the ground truth label for the brain network analysis task, where each label is either 0 (healthy) or 1 (schizophrenia).

The Swin Transformer [31] backbone of the Swin UNETR architecture learns the spatial features of the fMRI data to do semantic segmentation. Where S_{seg} represents the output segmentation map. The model can be formulated as follows:

i) Encoder (Swin Transformer): Let an encoder based on the Swin Transformer process the input image, I_{MRI}. The Swin Transformer calculates self-attention across the visual patches by using shifted windows:

$$Z^{(l)} = Swin\left(Z^{(l-1)}, W_l\right), l = 1, 2, ..., L \quad (6)$$

Here $Z^{(l)}$ refers to the feature map at the l-th layer of the encoder, which is determined by Eq. (6), W_l denotes the learnable weights at layer l, and the number of transformer layers is represented as L.

ii) Decoder (UNet Decoder): The decoder produces pixel-by-pixel predictions for the segmentation task by refining the features that the encoder retrieved using a U-Net architecture [32] is determined in Eq. (7).

$$S_{seg} = UNetDecoder\left(Z^{(L)}\right) \quad (7)$$

Where, $S_{seg} \in \mathbb{R}^{H \times W \times D}$ is the predicted segmentation map.

iii) Segmentation loss: The segmentation network is trained using a typical Dice loss L_{seg}, which calculates the similarity between the ground truth Y_{seg} and the anticipated segmentation map S_{seg}. The loss function is determined in Eq. (8).

$$L_{seg} = 1 - \frac{2\sum_{i=1}^{N} S_{seg}(i)\, Y_{seg}(i)}{\sum_{i=1}^{N} S_{seg}(i) + \sum_{i=1}^{N} Y_{seg}(i)} \quad (8)$$

Where N is the number of pixels/voxels in the segmentation map.

Graph-based brain connectivity analysis

A graph G is used to simulate the brain, with each node standing for a different part of the brain and each edge denoting the functional connection between those regions. Let $A_{fMRI} \in \mathbb{R}^{V \times V}$ be the adjacency matrix representing the functional connectivity between V brain regions from fMRI data. For every node v_i, the network has a corresponding feature vector h_i, which is first obtained using the functional connectivity matrix A_{fMRI}. The GNN uses message passing to update the node features.

i) Graph Neural Network Message Passing: At the t-th layer of the GNN [33], the node features are updated using Eq. (9).

$$h_i^{(t)} = \sigma\left(\sum_{j\in N(i)} \frac{A_{ij}}{c_{ij}} h_j^{(t-1)} + W_t h_i^{(t-1)}\right) \quad (9)$$

$$h_{graph} = GlobalPooling(h_i^{(T)}) \quad (10)$$

Here, $N(i)$ is the set of neighbors of the node i, A_{ij} is the adjacency matrix element representing the connectivity between regions i and j, the normalized constant is said to be c_{ij}, W_t represents the learnable weight matrix at the t-th layer, and σ is the non-linear activation function. The final node features $h_i^{(T)}$ are aggregated to form the graph-level feature vector h_{graph}, which is used for segmentation, is determined in Eq. (10).

ii) Graph loss: The binary cross-entropy loss L_{graph}, used for the graph segmentation task, which predicts whether the subject is healthy or schizophrenic based on the brain network features, is determined in Eq. (11).

$$L_{graph} = - Y_{graph}\log(p) - (1 - Y_{graph})\log(1 - p) \quad (11)$$

Where p is the predicted probability that the subject has schizophrenia, obtained from a sigmoid activation applied to the final graph feature vector h_{graph}.

iii) Joint loss function: The total loss function combines the segmentation loss and graph-based loss for training this model. The final objective is to reduce both losses simultaneously as determined in Eq. (12).

$$L_{total} = L_{seg} + \lambda L_{graph} \quad (12)$$

Here, λ is the hyperparameter that controls the relative importance of the segmentation stage. This framework used GNNs to analyze brain network connections and Swin UNETR for semantic segmentation. This model captures the structural and functional anomalies linked to schizophrenia by combining the optimization of segmentation with a graph-based model. A combined loss function that pushes the model to concentrate on both tasks at the same time is used to train the final model end-to-end.

Feature extraction

In this stage, three features are employed, such as structural features, texture features, and DL features, for the extraction purpose.

Structural features

Structural features [34] play a significant role in analyzing segmented ROIs in medical imaging, especially in neurodegenerative disorders. The two principal structural features are measured mainly through the measurements of hippocampal atrophy and cortical thickness and surface area. The extraction of these features was performed using volumetric and shape-based analysis.

Hippocampal atrophy measurements: Hippocampal atrophy is the most significant biomarker for neurodegenerative diseases such as Alzheimer's. It is measured using volumetric analysis.

i) Volume of the Hippocampus: Eq. (13) computes the volume of a segmented hippocampal region:

$$V = \sum_{i=1}^{N} V_i \quad (13)$$

where V refers to the total hippocampal volume, N denotes the number of voxels in the hippocampal region, and v_i represents the volume of the i-th voxel.

ii) Atrophy rate: Eq. (14) gives the rate of hippocampal atrophy over time t:

$$A = \frac{V_t - V_{t+\Delta t}}{V_t} \times 100\% \quad (14)$$

where A represents the percentage atrophy rate, V_t denotes the hippocampal volume at initial time t, $V_{t+\Delta t}$ signifies the hippocampal volume after time Δt. A progression of neurodegenerative disorders is indicated by high atrophy.

Cortical thickness and surface area: Cortical thickness and surface area are two important structural features that are derived from the cortical surface of the brain. These measurements are based on mesh-based representations of fMRI scans.

i) Cortical thickness measurement: The distance between the pial surface, which is the outer boundary of gray matter, and the white matter boundary is termed cortical thickness. It is computed in Eq. (15):

$$T(p) = \left\| C_p^{pial} - C_p^{wm} \right\| \quad (15)$$

where $T(p)$ denotes the cortical thickness at point p, C_p^{pial} represents the coordinate of the pial surface at p, C_p^{wm} refers to the coordinate of the white matter surface at p, $\|\cdot\|$ denotes the Euclidean norm. The average cortical thickness over a ROI is determined in Eq. (16):

$$\bar{T} = \frac{1}{N}\sum_{i=1}^{N} T(p_i) \tag{16}$$

where N refers to the total number of sampled points in the ROI.

ii) Cortical surface area measurement: This estimate for the surface area of the cortex is obtained by adding up the areas of small triangular patches forming the cortical mesh using Eq. (17).

$$S = \sum_{i=1}^{M} A_i \tag{17}$$

where S represents the total surface area of the cortical, M indicates the quantity of triangular elements in the mesh, A_i stands for the area of the i-th triangle calculated according to the following Eq. (18):

$$A_i = \frac{1}{2} \| (v_2 - v_1) \times (v_3 - v_1) \| \tag{18}$$

where v_1, v_2, v_3 signifies the vertices of the triangle, and × represents the cross-product.

These structural features provide critical biomarkers for evaluating brain morphology and neurodegeneration. Hippocampal volume and atrophy rate measure brain shrinkage, and cortical thickness and surface area indicate integrity within the cortex. These features have increasingly been used for early diagnosis and disease progression monitoring within medical imaging analyses of neurodegenerative diseases.

Texture features

Texture analysis [35] in medical imaging helps identify patterns in brain tissue that suggest abnormalities, such as those examined in neurodegenerative diseases. A Gray Level GLCM is a well-known method to extract second-order statistical texture features from the image. GLCM characterizes the spatial relationship between pixel intensities by counting how often pairs of pixel intensities (gray levels) appear for a given distance and direction. The GLCM, $P(i,j,d,\theta)$, is defined by Eq. (19) for an image $I(x,y)$, where gray levels range from 0 to G-1.

$$P(i,j,d,\theta) = \sum_{x=1}^{M}\sum_{y=1}^{N} \begin{cases} 1, \text{ if } I(x,y) = i \text{ and } I(x',y') = j \\ 0, \quad \text{otherwise} \end{cases} \tag{19}$$

where (x,y) and (x',y') represents the pixel pairs separated by a distance d in a given direction θ, i,j denotes intensity values in the image, and M,N characterize the dimensions of the image. GLCM-based texture analysis is a strong tool in medical imaging. It gives great information about the structural organization of the brain tissue that allows subtle textural changes due to disease.

DI features

DL models, specifically CNNs, are very efficient in extracting high-dimensional feature sets of raw fMRI slices. Because of its scalability and optimal performance, the state-of-the-art CNN architecture that is quite handy is EfficientNet. EfficientNet: this is a CNN model learned through an optimal accuracy vs. the computational cost trade-off. As opposed to regular CNNs, where depth or width scaling is the case, this compound scaling, in the case of the proposed EfficientNet approach [36], modifies depth, width, and resolution simultaneously. The EfficientNet model is adopted in Eq. (20):

$$\text{compound scaling: } d = \alpha^{\phi}, w = \beta^{\phi}, r = \gamma^{\phi} \tag{20}$$

where d, w and r are the representations for network depth, network width, and input resolution, respectively. α,β,γ denote the scaling coefficients, and ϕ is a user-defined scaling parameter. EfficientNet extracts features with high dimensionalities from DL from the raw fMRI slices. These represent very powerful expressions for brain structure representations that facilitate detection and further analysis of the diseases.

Feature selection

In the feature selection stage, the Parrot Gooseneck Barnacle Algorithm is employed for the selection of features for the schizophrenia detection framework. This algorithm is combined with parrot optimizer (PO) [37] and the Gooseneck Barnacle Optimization Algorithm (GBA) [38].

Parrot optimizers

The popular parrot species Pyrrhura Molinae is a favorite among pet owners because of its appealing appearance, strong attachment to its owners, and simplicity of training. Four unique behavioral characteristics of Pyrrhura molinae have been identified by prior research and breeding initiatives: foraging, staying, communicating, and the fear of strangers. When food is plentiful, domesticated Pyrrhura Molinae prefer to forage in small groups, which makes their foraging behavior intriguing. This stage used the group's presence and their owner's position to find the meal by moving in its direction. Also, it used visual indicators and smell to improve its hunt. As part of their staying behavior, Pyrrhura Molinae haphazardly perch on different parts of their owner's body. These outgoing birds use unique sounds to communicate among themselves, which facilitates social interaction and the dissemination of knowledge. Because of their innate fear of strangers, which is a characteristic shared by all birds, Pyrrhura Molinae avoid unexpected people and seek refuge with their owners. Additionally, the motivation for this optimization is emphasized by the unpredictable behavior of Pyrrhura Molinae, as each individual exhibits these four behaviors at random throughout each cycle within domesticated flocks.

i) Population initialization: The PO's initialization formulation is considered as follows in Eq. (21): swarm size of N, maximum iterations of Max_{iter}, and search space limits of lb (lower bound) and ub (upper bound).

$$X_i^0 = lb + rand(0,1) \cdot (ub - lb) \tag{21}$$

Here, $rand(0,1)$ represents the random number in the range of [0, 1], X_i^0 is the i^{th} Pyrrhura Molinae's position in the initial phase.

ii) Foraging behavior: When PO forages, birds mostly use their observations of the food's location or the owner's position to estimate the approximate location of the food before flying in the direction of the site. Consequently, the positional movement adheres to Eq. (22).

$$X_i^{t+1} = (X_i^t - X_{best}) \cdot Levy(dim) + rand(0,1) \cdot \left(1 - \frac{t}{Max_{\text{iter}}}\right)^{\frac{2t}{Max_{\text{iter}}}} \cdot X_{mean}^t \tag{22}$$

Eq. (22) used X_i^t represent the current location and X_i^{t+1} represent the location of the next update. The average location within the current population is represented by X_{mean}^t, and the Levy distribution is indicated by $Levy(D)$, which is used to characterize parrot flying. In addition to representing the host's current position, X_{best} indicates the best position that has been found from initiation to the present. The current number of iterations is indicated by t. $(X_i^t - X_{best}) \cdot Levy(dim)$ denotes movement according to one's position with respect to the owner, and $rand(0,1) \cdot (1 - \frac{t}{Max_{\text{iter}}})^{\frac{2t}{Max_{\text{iter}}}}$. To better focus the food's orientation, X_{mean}^t reflects the observation of the population's overall position. The average location of the current swarm is shown by X_{mean}^t is obtained using Eq. (23).

$$X_{mean}^t = \frac{1}{N}\sum_{k=1}^{N} X_k^t \tag{23}$$

$$Levy(dim) = \frac{\mu - \sigma}{|v|^{\frac{1}{\gamma}}} \tag{24}$$

Where, $\mu \sim N(0,\text{dim})$, $v \sim (0,\text{dim})$ and $\sigma = \left(\frac{\lambda(1+\gamma)\cdot\sin\left(\frac{\pi\gamma}{2}\right)}{\lambda\left(\frac{1+\gamma}{2}\right)\cdot\gamma\cdot 2^{\frac{1+\gamma}{2}}}\right)^{\gamma+1}$. The Levy distribution is attained on rule in Eq. (24), where γ is assigned the value of 1.5.

iii) Staying behavior: The staying behavior of Pyrrhura Molinae, a very gregarious creature, mostly consists of abruptly flying to any area of its owner's body and then remaining still for a while. The staying behavior process is determined in Eq. (25).

$$X_i^{t+1} = X_i^t + X_{best} \cdot Levy(dim) + rand(0,1) \cdot ones(1,dim) \tag{25}$$

Here, $ones(1,dim)$ represented as the all-1 vector of dimension dim.X_{best}·$Levy(dim)$ is the flight to the host, the process of randomly stopping at a part of the host's body is represented as $rand(0,1)$·$ones(1,dim)$.

iv) Communicating behavior: The social nature of Pyrrhura Molinae parrots is typified by their close communication within their groups. Both speaking without flying to the flock, and flying to the flock are included in this communication behavior. The mean position of the current population is used to represent the flock's center in the PO, where both behaviors are considered to

occur with equal probability. The communication behavior process is determined in Eq. (26).

$$X_i^{t+1} = \begin{cases} 0.2 \cdot rand(0,1) \cdot \left(1 - \frac{t}{Max_{iter}}\right) \cdot (X_i^t - X_{mean}^t), P \le 0.5 \\ 0.2 \cdot rand(0,1) \cdot exp\left(-\frac{t}{rand(0,1).Max_{iter}}\right), p > 0.5 \end{cases} \quad (26)$$

where $0.2 \cdot rand(0,1) \cdot \left(1 - \frac{t}{Max_{iter}}\right) \cdot (X_i^t - X_{mean}^t)$ indicates how an individual joins a parrot's group to communicate, and $0.2 \cdot rand(0,1) \cdot exp\left(-\frac{t}{rand(0,1).Max_{iter}}\right)$ indicates the process of an individual taking off right away after communicating. Since both behaviors are possible, a randomly generated *P* between 0 and 1 is used to implement them.

v) Fear of strangers' behavior: Typically, birds, like Pyrrhura Molinae parrots, have an innate mistrust of strangers. In pursuit of a safe environment, they have been known to avoid new people and seek refuge with their owners. The Fear of strangers' behavior process is determined using Eq. (27).

$$X_i^{t+1} = X_i^t + rand(0,1) \cdot \cos\left(0.5\pi \cdot \frac{t}{Max_{iter}}\right) \cdot (X_{best} - X_i^t) - \cos(rand(0,1) \cdot \pi) \cdot \left(\frac{t}{Max_{iter}}\right)^{\frac{2t}{Max_{iter}}} \cdot (X_i^t - X_{best}) \quad (27)$$

Here, $rand(0,1) \cdot \cos\left(0.5\pi \cdot \frac{t}{Max_{iter}}\right) \cdot (X_{best} - X_i^t)$ represents the reorienting process to fly towards the owner and $\cos(rand(0,1) \cdot \pi) \cdot \left(\frac{t}{Max_{iter}}\right)^{\frac{2t}{Max_{iter}}} \cdot (X_i^t - X_{best})$ represents the process of moving away from the strangers.

Gooseneck barnacle optimization algorithm

The most crucial aspects of gooseneck are its mating behavior, which involves casting sperm, and how wind and wave motion impact reproduction. The mathematical model of the GBA algorithm is determined using Eq. (28).

$$(X + I)_{i+1} = (X + I)_{i+1} + WD_i + T_{dim} + S((X + I)_i (Sp_{water})_j) + Hs.(X + I)_i \quad (28)$$

Whereas (*X*+*I*) specifies the goose barnacle's position in the i-th iteration, *WD* is the wind's direction in the i-th iteration, T_{dim} is the target dimension to move towards the target or to the optimal solution, $S((X+I)_i(Sp_{water})_j)$. Between the i-th $(X+I)_i$ and the jth sperm area in water, *S* is the logarithmic spiral, and *Hs* is the significant wave height. In the section that follows, each stage of the mathematical model has been thoroughly explained.

i) Initialization: The GBA algorithm assumes that gooseneck barnacles are the candidate solution, and the position of the gooseneck is one of the problem space's variables. The gooseneck's penis extended seven or eight times to reach their body, allowing them to move through any dimensional space. The population of gooseneck barnacles, a type of stalked barnacle, is the candidate of solution X for initialization, as GBA is a population-based method. This population is expressed as in Eq. (29).

$$X = \begin{bmatrix} (x + I)_{1,1} & (x + I)_{1,2} & \cdots & (x + I)_{1,d} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ (x + I)_{n,1} & (x + I)_{n,2} & \cdots & (x + I)_{n,d} \end{bmatrix} \quad (29)$$

where *n* and *d* stand for the number of dimensions or variables that need to be optimized and the total populations, respectively. Each gooseneck has a different length, *I*, which will be chosen at random because of its edible structure. There is a matching area in the water with sperm for mating for every gooseneck. This is an additional crucial element of the suggested algorithm called "sperm_region", which shares the following matrix with *X*, as determined in Eq. (30).

$$Sperm\_Region = \begin{bmatrix} (sr)_{1,1} & (sr)_{1,2} & \cdots & (sr)_{1,d} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ (sr)_{n,1} & (sr)_{n,2} & \cdots & (sr)_{n,d} \end{bmatrix} \quad (30)$$

where *d* is the number of variables or dimensions as *X* and *n* is the number of goosenecks. Although sperm areas and gooseneck barnacle locations are also potential solutions, the issue domain will treat them differently in each iteration. The sperm area in the water depicts the ideal mating location, whereas real search agents, such as those present in the gooseneck barnacles' penis, traverse the search space. The sperm region could be considered the new gooseneck in the search space for the sake of mating and reproduction. The goosenecks thus perform a search radius around the new location and update it in the event that they find a better option. This technique guarantees that there will always be the best gooseneck option accessible.

Gooseneck barnacles are found in the upper and intermediate intertidal zones. Barnacles predominate in waves that are significantly below 0.8 to 1.5-3 meters above mean low water. Using this equation, $Hs = \sqrt{H \hat{} 2T}$ is typically used to calculate significant wave height. In this version of GBA, however, compute *Hs* using the following Eq. (31) (taking into account the tolerance of barnacles in their life cycle):

Significant Wave Height,

$$Hs = 1.5 - \left(\frac{Iteration(1.5\text{-}0.2)}{Maximum\ iteration}\right) \quad (31)$$

Since it has been discovered that barnacles withstand wave intensities between 0.2 and 1.5 or 3.0 m, *Hs* decreases with each iteration to investigate the wave intensity range and choose the ideal mating location. For sperm casting, a neighboring logarithmic spiral region is defined in Eq. (32).

$$S((X + I)_i(Sp_{water})_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + (Sp_{water})_j \quad (32)$$

where *t* is a random number in [-1, 1], b is a constant that defines the shape of the logarithmic spiral, and D_i denotes the distance between the i-th barnacle and the j-th sperm area for mating. The following formula is used to determine distance: The equation is $D_i=(X+I)_i-(Sp_{water})_j$. where D_i denotes the distance of the i-th barnacles for the j-th sperm casting region, $(Sp_{water})_j$ denotes the j-th sperm casting region, $(X+I)$ denotes the i-th barnacles.

ii) Offspring generation: It is projected that sperm-cast mating behavior combines a few of the patterns covered in this model. A movement vector, $\Delta(X+I)_i$, defined as follows in Eq. (33-35), has been taken into consideration to update the position of fresh gooseneck barnacles in a search space and simulate the migration towards new offspring production.

$$\triangle(X + I)_{i+1} = WD_i + T_{dim} + Hs \cdot \triangle(X + I)_i \quad (33)$$

$$(X + I)_i = (X + I)_i + \triangle(X + I)_{i+1} \quad (34)$$

$$(X + I)_{i+1} = (X + I)_i + Levy * (X + I)_i \quad (35)$$

Using the degree range [0, 359] on the search space radius, WD_i *i* specifies the wind direction, which is then included in the best solution together with the target dimension T_{dim}. By assuming that the wind direction is always the goal, the dimension toward the target, T_{dim} is the value of the dimension in the target. The Levy flight is determined in Eq. (36).

$$Levy(N) = 0.01 \times \frac{r_{1\times\sigma}}{|r_2|^{\frac{1}{\beta}}} \quad (36)$$

Here, β is a constant set to 1.5, r_1 and r_2 are random numbers [0, 1], and Eq. (37) is as follows:

$$\sigma = \left(\frac{\tau(1+\beta)\times\sin\left(\frac{\pi\beta}{2}\right)}{\tau\left(\frac{1+\beta}{2}\right)\times\beta\times 2^{\left(\frac{\beta\text{-}1}{2}\right)}}\right)^{\frac{1}{\beta}} \quad (37)$$

Were, $\tau(y)=(y\text{-}1)!$. According to the previous description, the mathematical model requires the gooseneck barnacle to move over several iterations in the logarithmic spiral sperm region in the direction of a target mating zone for sperm throwing. However, since this model needs to determine the location of the global optimum, there are no directions in the actual search space. Therefore, select a mating target for every optimization cycle. Other goosenecks are compelled to approach the mating zone in GBA because the gooseneck with the highest objective value at the moment of optimization is regarded as the target. Nevertheless, it also aids GBA in preserving the most promising target in the search space for every iteration. A random pool of potential solutions is created by the GBA to start the optimization process. The search agents then adjust their positions in relation to the target based on Eq. (34), taking into account the direction of the wind and any notable wave movement or action. Sometimes, because of the wind and wave intensity, there may only be one remaining sperm zone to replicate the actual situation. If so, Eq. (35) will be used to update their location. Lastly, the position is updated repeatedly until the last condition is satisfied. The position and quality of the best target are provided at the end, along with the best global estimate. The following section provides the pseudocode for the PGBA algorithm.

Algorithm 1: Parrot GBA

Step 1: Initialize population

Initialize population size *N*, max iterations Max_{iter}, bounds *lb*, *ub*

Initialize positions X_i^0 randomly using:

$X_i^0 = lb + rand(0,1) \cdot (ub - lb)$

Initialize GBA-specific variables

Initialize sperm_region matrix corresponding to search space dimensions

Initialize wind direction *WD*, significant wave height *Hs* and target dimension T_{dim}

Initialize random parameters (b,μ,σ,γ) for Levy flights

Step 2: Main optimization loop

for t = 1 to Max_{iter}:

#Calculate the Mean Position (X_{mean}) for PO

$X_{mean}^t = \frac{1}{N}\sum_{k=1}^{N} X_k^t$

Track the Best Position Found so far (X_{best})

Identify X_{best} as the best position found so far

Update the significant wave height Hs for GBA

$Hs = 1.5 - \left(\frac{Itertaion(1.5\text{-}0.2)}{Maximum\ iteration}\right)$

Step 3: Behavior updates for each parrot (PO)

for each parrot *i* in population:

Foraging Behavior (PO)

Levy = *calculate_Levy*(*dim*)

$X_i^{t+1} = (X_i^t - X_{best}) \cdot Levy(dim) + rand(0,1) \cdot \left(1 - \frac{t}{Max_{iter}}\right)^{\frac{2t}{Max_{iter}}} \cdot X_{mean}^t$

Staying Behavior (PO)

$X_i^{t+1} = X_i^t + X_{best} \cdot Levy(dim) + rand(0,1) \cdot ones(1,dim)$

Communicating Behavior (PO)

$P = rand(0,1)$

if $P \leq 0.5$:

$$X_i^{t+1} = \begin{cases} 0.2 \cdot rand(0,1) \cdot \left(1 - \frac{t}{Max_{iter}}\right) \cdot (X_i^t - X_{mean}^t), P \leq 0.5 \\ 0.2 \cdot rand(0,1) \cdot exp\left(-\frac{t}{rand(0,1).Max_{iter}}\right), p > 0.5 \end{cases}$$

Fear of Stranger Behavior (PO)

$X_i^{t+1} = X_i^t + rand(0,1) \cdot \cos\left(0.5\pi \cdot \frac{t}{Max_{iter}}\right) \cdot (X_{best} - X_i^t) - \cos(rand(0,1) \cdot \pi) \cdot \left(\frac{t}{Max_{iter}}\right)^{\frac{2t}{Max_{iter}}} \cdot (X_i^t - X_{best})$

Step 4: Behavior Updates for Each GBA

for each gooseneck *i* in population:

for each sperm region *j* in sperm_region:

Calculate distance D_i between the barnacle and sperm region

$D_i = (X + l)_i - (Sp_{water})_j$

Logarithmic Spiral Movement for sperm casting

$S((X + l)_i (Sp_{water})_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + (Sp_{water})_j$

Offspring Generation (GBA)

$\triangle (X + l)_{i+1} = WD_i + T_{dim} + Hs \cdot \triangle (X + l)_i$

$(X + l)_i = (X + l)_i + \triangle (X + l)_{i+1}$

$(X + l)_{i+1} = (X + l)_i + Levy * (X + l)_i$

Step 5: Update target and best position

Identify the best position as the new target for both PO and GBA

Update X_{best} if a better solution is found

Return best solution found so far after all iterations

return X_{best} best fitness

Helper function: Levy Flight Calculation

function *calculate_Levy*(*dim*)

μ~N(0,*dim*)

v~(0,*dim*)

gamma = 1.5

$$\sigma = \left(\frac{\lambda(1+\gamma) \cdot \sin\left(\frac{\pi\gamma}{2}\right)}{\lambda\left(\frac{1+\gamma}{2}\right) \cdot \gamma \cdot 2^{\frac{1+\gamma}{2}}}\right)^{\gamma+1}$$

return

$$Levy(dim) = \frac{\mu - \sigma}{|v|^{\frac{1}{\gamma}}}$$

DL based detection

In this detection stage, the TransConvDenseformNet model is employed for the detection of schizophrenia disease. **Figure 3** shows the architecture of the TransConvDenseformNet model.

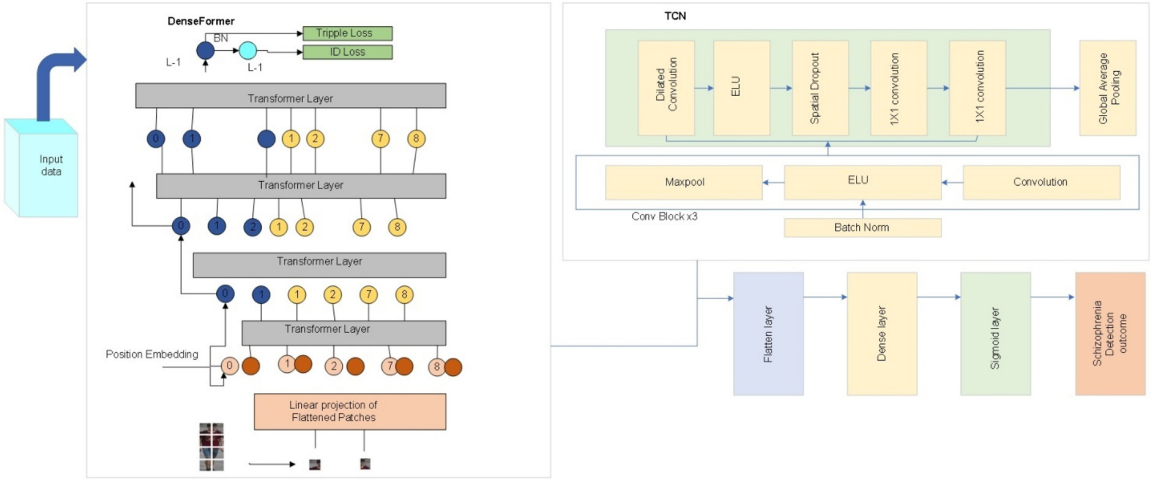


Figure 3. TransConvDenseformNet architecture.

Denseformer

Vision transformer and TransReID's architecture is based on the standard transformer encoder [39], comprising multihead self-attention and multilayer perceptron blocks. Each block uses Layernorm (LN) for normalization. A transformer block is given in Eqs. (38) and (39):

$$E_l^{'} = MSA(LN(E_{l-1})) + E_{l-1} \tag{38}$$

$$E_l = MLP(LN(E_l^{'})) + E_l^{'} \tag{39}$$

Every hidden feature E_l that comes from each transformer layer could be represented as in Eq. (40):

$$E_l = \left[z_l^0, z_l^1, z_l^2, \ldots, z_l^N\right] \tag{40}$$

where z_l^0 represents the class token after layers of transformers, indicating the global feature of the image. The length of the feature sequence E_l keep unchanged via transformer encoder layers. Although this kind of method could already attain high performance in ReID tasks, the features of each transformer layer were still not fully utilized, thus limiting the expressive ability of the final model. To solve this problem, inspired by DenseNet, a densely connected framework has been proposed to enhance information flow through transformer layers. **Figure 4** depicts the framework of Denseformer.

Additionally, the first-class token, which was, z_l^0 of each hidden feature E_l is connected towards input E_{l+1} at the front. The figure represents how the features are densely connected. Accordingly, the hidden feature of each layer E_{l+1} could be expressed as in Eq. (41):

$$E_{l+1} = \left[z_l^0, z_{l+1}^0, z_{l+1}^1, z_{l+1}^2, \ldots z_{l+1}^{N+l}\right] \tag{41}$$

It is rather easy to deduce that each z_l^0 is equal to the class token z_0^0 in the first input sequence E_0, and each z_{l+1}^0 is the first transformation of z_l^0, z_{l+2}^0 is the second transformation of z_{l-1}^0 and so on by the transformer layer. In other words, each of the global representations of the former layers might be reproduced in the back layers. This operation could well merge more detailed features of the shallow layers into the deep layers, which could not have been done before.

In contrast to DenseNet, where all previous features are directly concatenated into the back layers, our method does the feature fusion implicitly: all the class tokens connected to the back layers are transformed by several transformer layers except z_l^0. Last but not least, we obtain the L class tokens if we use L transformer encoder layers. The most transformed class token carries the most feature information, so we use z_L^{L-l} as our final global representation.

TCN

TCN is mainly composed of causal dilated convolution and a residual block. By using this method, vanishing and exploding gradients could be well prevented, and time series feature extraction could be well conducted.

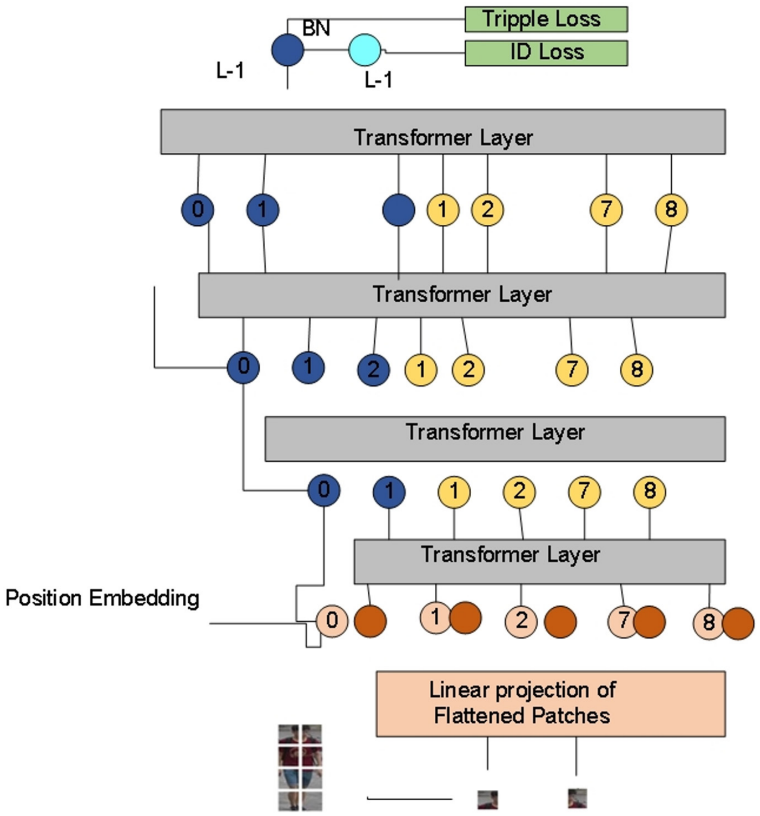


Figure 4. Framework of denseformer.

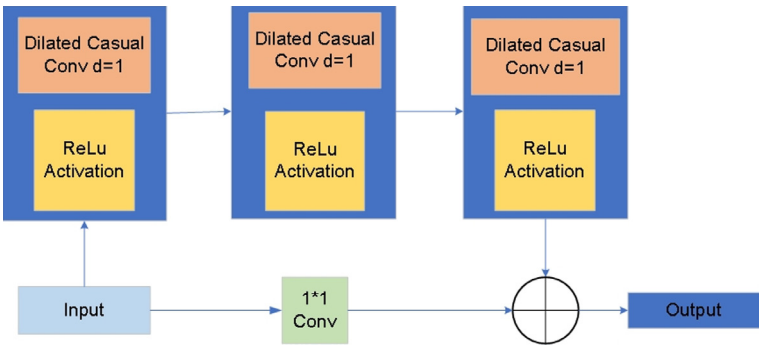


Figure 5. Residual block structure.

i) Causal dilated convolution: The causal convolution in the TCN could be defined as the value at moment t of the layer *h* network depends only on the value at moment t and the immediate previous moment in layer *h*-1. The convolution structure is different from the traditional convolution in that the convolution area is limited strictly, and it is expressed in time constraint in the time series; that is, the value of the future moment is not taken into account in feature extraction.

As illustrated in **Figure 5**, interval sampling is done with reference to the causal convolution to widen the receptive field. In contrast to the multilayer convolution and pooling for the expansion of the receptive field, this method could avoid information loss and does not need unit convolution kernels; in addition, the input and output time steps are the same. Eq. (42) illustrates the operation of causal dilated convolution.

$$y_{h,t} = \sum_{i=0}^{k} f_i \cdot y_{h-1}, t - 2id \quad (42)$$

Where $y_{h,t}$ represents the series value of layer *h* in the network at time *t*, f_i represents the filter, *k* represents the size of the convolution kernel, and *d* represents the convolutional dilation rate.

ii) Residual block: As described in **Figure 5**, the TCN set up in this paper has three causal dilated convolution layers, and the dilation rates are 1, 2, and 4. The second is to employ the unit convolution kernel to process the input in order to make the structure of the two paths a series for the summation operation.

Flatten layer

Flatten is another layer in neural networks that transforms the incoming multidimensional data into a vector of one dimension. It is normal to apply a DL model as an activation function after a convolution or pooling layer before feeding the data to fully connected layers. For instance, if the input has dimensions of (*h*,*w*,*d*) where *h* - height, *w* - width, *d* - depth, the Flatten layer changes it to a vector of size *h*.*w*.*d*. This transformation guarantees compatibility with dense layers, which take a one-dimensional input. Flattening does not remove any of the data values, but only reshuffles their form. It is useful in areas such as fMRI-based schizophrenia detec-

tion, integrating feature extraction and classification phases.

Dense layer

A Dense layer is another name for a fully connected layer in which each neuron is connected with every neuron in the previous layer. It performs a linear transformation of the input, followed by an optional activation function:

$$y = \sigma(Wx + b) \tag{43}$$

Here, W represents the weight matrix, b indicates the bias, x represents the input, and σ indicates the activation function.

Full layers take the features extracted and combine them to make predictions or classifications. In schizophrenia detection using fMRI, they assist in making the features, such as the volumes of the brain regions or the texture of the images, contribute to the diagnostic results. The dense layers are usually positioned at the output of the neural network.

Sigmoid layer

The second layer is the Sigmoid layer, which is normally the final layer in a neural network when binary classification is being undertaken. It converts the raw outputs, or logits, of the last layer into a probability value between 0 and 1, the probability of the positive class (schizophrenia). The Sigmoid function is defined as in Eq. (44):

$$P(y) = \frac{1}{1 + e^{-z}} \tag{44}$$

Where z refers to the logit in the previous layer and $P(y)$ denotes the probability of the positive class that is predicted. The output is simply interpreted as a probability in this layer, thus making it easy to perform threshold-based classification. A predicted probability of 0.5 with a threshold of 0.5. $P(y) \geq 0.5$ represents the schizophrenia category. $P(y) < 0.5$ denotes the healthy control group. The sigmoid layer plays a crucial role in fMRI-based schizophrenia detection to make the network give only one probability score, enabling the network to draw a correct binary diagnosis.

Statistical analysis

In order to assess the strength and the generalization of the proposed SWPBTN framework, the 5-fold cross-validation is used. There are five equal subsets of the dataset. Training and testing are performed in four folds in every iteration. This is done 5 times, such that each fold is a test set at one time. The last performance measures are given as the average standard deviation of all the folds. The accuracy, precision, recall (sensitivity), specificity, F1-score, false positive rate (FPR), and false negative rate (FNR) were used to measure the classification performance. To compare the proposed model with the baseline models, paired t-tests are conducted on the values of fold-wise accuracy to find out whether the differences in performance were statistically significant. The p-value that is deemed statistically significant is less than 0.05. All the statistical procedures are done with Python libraries such as NumPy, SciPy, and scikit-learn.

Results

In this section, the performance attained by the proposed model has been discussed in detail. Moreover, the performance of the suggested model has been compared with the existing models to validate the performance of the suggested model. Apart from these, this section provides a discussion about the dataset and performance metrics.

Performance metrics evaluation

The metrics that are used for measuring the performance of the suggested model have been discussed under this section. **Table 1** gives the Evaluation of performance metrics.

Comparison analysis

This section provides a comparison of the proposed model along with the existing models for validating the performance of the suggested model. For comparison, the existing techniques include Deep ResNet [18], CNN [23], HKNN-FL [25], and SchizoGoogLeNet [16]. Moreover, the comparison scores show that the proposed model provides a better detection score than the existing models.

Table 2 shows a comparison of the performances of different models of schizophrenia detection. The presented SWPBTN model is superior to others as it has the best accuracy (99.85%), F1-score (99.9%), and sensitivity (99.95%), which means the best prediction abil-

Table 1. Evaluation of performance metrics

Metric	Description	Formula
Accuracy	Percentage of correctly predicted cases (both positive and negative).	$A = \frac{tp+tn}{tp+fp+tn+fn}$
Precision	Proportion of predicted positives that are actual positives.	$P = \frac{tp}{tp+fp}$
Sensitivity (Recall)	Ability to correctly identify true positive cases (schizophrenia cases).	$s_n = \frac{tp}{tp+fn}$
Specificity	Ability to correctly identify true negative cases (non-schizophrenia cases).	$s_p = \frac{tn}{tn+fp}$
F1-Score	Harmonic mean of Precision and Sensitivity, balancing false positives/negatives.	$F1\text{-}score = 2.\frac{P.s_n}{P+s_n}$
NPV (Negative Predictive Value)	Proportion of predicted negatives that are actual negatives.	$NPV = \frac{tn}{tn+fn}$
MCC (Matthews Correlation Coefficient)	Balanced measure considering all confusion matrix elements.	$MCC = \frac{(tp.tn)-(fp.fn)}{\sqrt{(tp+fp)(tn+fn)(tp+fn)(tn+fp)}}$
FPR	Proportion of non-schizophrenia cases incorrectly classified as positive.	$FPR = \frac{fp}{tn+fp}$
FNR	Proportion of schizophrenia cases incorrectly classified as negative.	$FNR = \frac{fn}{tp+fn}$

Table 2. Performance comparison

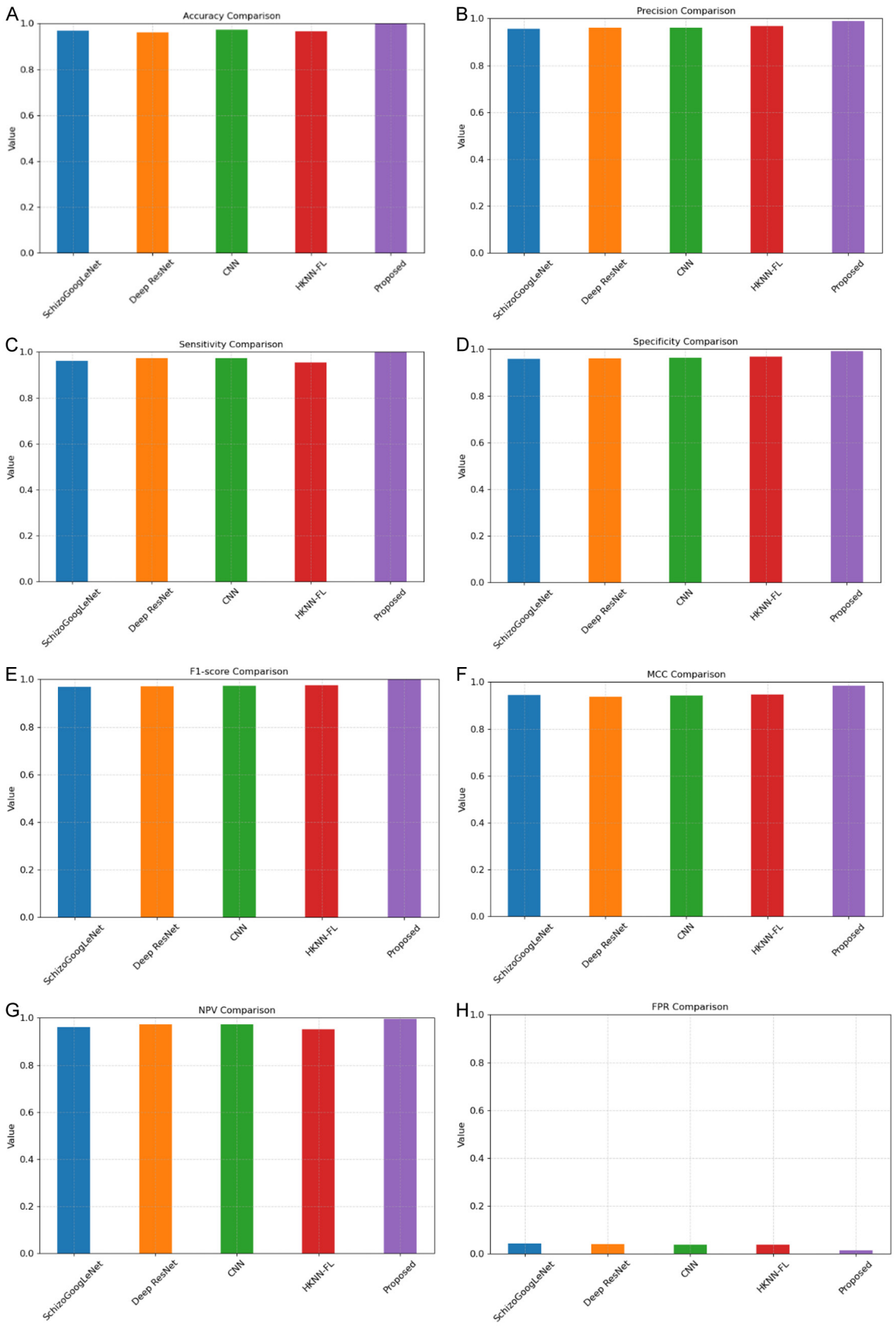
Model	Accuracy	Precision	F1-score	Specificity	Sensitivity	NPV	MCC	FPR	FNR
SchizoGoogLeNet [16]	0.968	0.955	0.969	0.9575	0.961	0.961	0.944	0.0425	0.029
Deep ResNet [18]	0.96	0.96	0.97	0.96	0.972	0.9715	0.9385	0.04	0.028
CNN [23]	0.972	0.962	0.9725	0.9625	0.973	0.9719	0.9415	0.038	0.027
HKNN-FL [25]	0.965	0.9675	0.975	0.968	0.9535	0.9525	0.9465	0.037	0.02
Proposed	0.9985	0.9885	0.999	0.991	0.9995	0.9955	0.9835	0.015	0.0075

ity. It also indicates the best false positive (1.5%) and false negative (0.75%) rates, which indicates high diagnostic reliability. On the whole, the suggested model proves to be extremely efficient in comparison with the current DL strategies.

Figure 6 shows the comparison chart of performance measures between models, which indicates the advantage of the proposed model in diagnostic measures of accuracy and reliability. The suggested SWPBTN model is significantly better than the current approaches in all assessment indicators. It has the best accuracy, precision, and F1-score, which indicates its efficiency in the accurate detection of schizophrenia cases. The model has also demonstrated high sensitivity and specificity, and this implies that it is highly reliable in terms of the diagnosis. On the whole, it greatly decreases false predictions, and it is very appropriate to use in clinical practice.

Figure 7 demonstrates the consecutive results of the preprocessing and segmenting processes that are performed on fMRI scans. Preprocessing involves intensity normalization to scale the intensity, skull stripping to extract brain tissue, and bias correction to provide uniform contrast. CLAHE, flipping, and rotation are other augmentations to enhance the quality of the images and the size of the dataset to train a robust model. Segmentation using the Swin UNETR model is precise in the definition of the major parts of the brain, such as the hippocampus, thalamus, and prefrontal cortex. The accurate segmentation helps in the targeted feature extraction to enhance the detection of schizophrenia.

Table 3 shows the fold-wise performance of the proposed SWPBTN model, which is tested with 5-fold cross-validation. The model has a high level of performance with a consistent high performance in all five folds, which implies high levels of robustness and generalization. The


A
Accuracy Comparison
B
Precision Comparison
C
Sensitivity Comparison
D
Specificity Comparison
E
F1-score Comparison
F
MCC Comparison
G
NPV Comparison
H
FPR Comparison
Value
0.0
0.2
0.4
0.6
0.8
1.0
SchizoGoogLeNet
Deep ResNet
CNN
HKNN-FL
Proposed

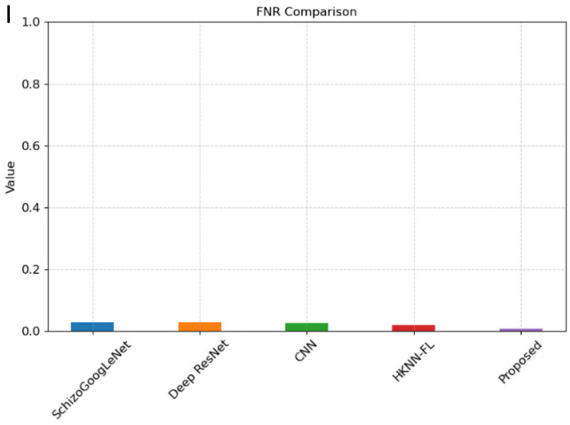


Figure 6. Analysis of performance metrics, (A) accuracy, (B) precision, (C) sensitivity, (D) specificity, (E) F1-score, (F) MCC, (G) NPV, (H) FPR, (I) FNR.

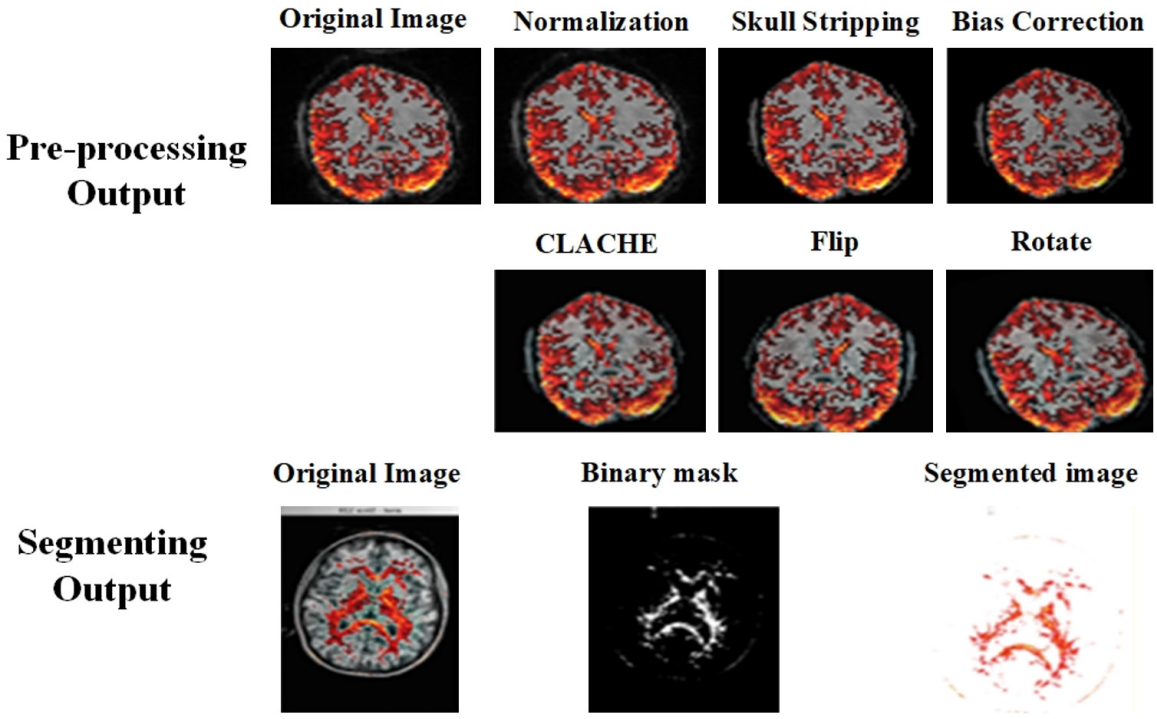


Figure 7. Pre-processing and segmentation outputs of brain MRI.

Table 3. Five-fold cross-validation performance of the suggested SWPBTN model

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Fold 1	99.72	99.85	99.8	99.82	0.998
Fold 2	99.88	99.92	99.9	99.91	0.999
Fold 3	99.83	99.87	99.85	99.86	0.998
Fold 4	99.91	99.95	99.93	99.94	0.999
Fold 5	99.91	99.96	99.95	99.95	0.999
Mean ± SD	99.85 ± 0.07	99.91 ± 0.04	99.89 ± 0.06	99.90 ± 0.05	0.999 ± 0.001

Table 4. Paired t-test comparison between SWPBTN and baseline models

Comparison	Mean Accuracy Difference (%)	t-value	*p*-value
SWPBTN vs. CNN	3.12	4.85	0.003
SWPBTN vs. ResNet	2.87	3.96	0.006
SWPBTN vs. HKNN-FL	2.45	3.21	0.012

difference between folds is not as high as indicated by low values of standard deviation. The average accuracy, precision, recall, F1-score, and AUC scores prove the existence of stable and trustworthy classification performance. These findings indicate that the proposed framework has a high predictive performance with varying data partitions, which minimizes the chances of overfitting.

The findings of the paired t-test that is used to compare the performance of the proposed SWPBTN model and the baseline models in the five cross-validation folds are presented in **Table 4**. The test is carried out on the accuracy values on a fold-wise basis to find out whether the differences in the performance were significant. The t-values and the *p*-values reported show that the obtained improvements by the proposed model are statistically significant ($P < 0.05$). The results of these findings validate the fact that the high performance of the SWPBTN framework is not a random variation but a steady improvement compared to the current methods.

Discussion

The suggested SWPBTN model has shown excellent results in the classification of schizophrenia based on fMRI data with an accuracy of 99.85, sensitivity of 98.85, and F1-score of 99.9. These findings are better than the previous ones and show that the combination of advanced preprocessing, structural-functional brain analysis, and hybrid feature extraction with DL classifiers is effective. EEG and retinal imaging have been used in previous studies to identify schizophrenia with a high level of precision. As an example, SchizoGoogLeNet [16] with a combination of GoogLeNet features and SVM had 99.02% accuracy, and VGGNet-based models [17] and ResNet-based models [18] had over 99% accuracy. CNN models that are based on time-frequency, like SchizoNET [20, 21], proved to be effective in differentiating several datasets [19]. On the same note, retinal imaging and fMRI-based techniques [24, 25] established that multi-modal analysis has the potential to offer robust biomarkers of schizophrenia.

In contrast to these previous methods, SWPBTN combines both structural and functional data by anatomically dividing important brain areas (hippocampus, thalamus, and prefrontal cortex) with Swin UNETR and simulates functional connectivity with GNNs [23]. The Parrot-Goose-neck Barnacle hybrid feature selection strategy allows the derivation of structural, texture, and DL features, which are optimized towards classification. This structural-functional combination, along with DenseFormer and TCNs, offers high-precision and improved interpretability, overcoming the drawbacks of single-modality approaches or purely CNN-based ones.

The preprocessing techniques, such as skull stripping, bias field correction, CLAHE enhancement, and elastic deformation, lead to high-quality data and better generalization, comparable to augmentation techniques, such as WGAN-GP and VAE, which have been employed in previous EEG studies [22]. The high performance of the framework in various measures indicates that it can be used in clinical diagnosis at an earlier and more accurate stage, which is an improvement over the conventional subjective diagnosis techniques.

Although such encouraging outcomes are achieved, there are certain limitations. The performance might differ according to the various fMRI acquisition protocols/populations, and extensive multi-center validation is needed to completely determine generalizability. Future studies can consider incorporating more neuroimaging modalities, more efficient feature selection, and uncertainty estimation to improve clinical reliability.

Conclusion

This paper suggested a DL model, which is automated to detect schizophrenia with fMRI

data accurately. The experimental findings prove that the proposed method is superior to the current models in all the essential evaluation measures, and it has significant gains in accuracy, precision, and F1-score. The framework is comprehensive in the way it incorporates structural and functional brain analysis to give a comprehensive picture of abnormalities in schizophrenia. The results point to the opportunities for the development of sophisticated DL approaches to objective and reliable diagnosis of neuropsychiatric disorders. The findings are encouraging, but it needs additional confirmation in larger and multicenter studies to be sure of generalizability. The future will be concerned with multi-modal integration and clinical testing in the real world to increase the predictive strength and the practical use.

Acknowledgements

The authors would like to thank the Deanship of Anna University for supporting this work.

Disclosure of conflict of interest

None.

Address correspondence to: Purnima Shanmugam, Department of Electronics and Communication Engineering, College of Engineering, Anna University, Chennai, India. Tel: +91 9944119994; Fax: +91 9944119994; E-mail: purnimasres25@gmail.com

References

[1] Tyagi A, Singh VP and Gore MM. Towards artificial intelligence in mental health: a comprehensive survey on the detection of schizophrenia. Multimedia Tools Appl 2023; 82: 20343-20405.

[2] Verma S, Goel T, Tanveer M, Ding W, Sharma R and Murugan R. Machine learning techniques for the schizophrenia diagnosis: a comprehensive review and future research directions. J Ambient Intell Humaniz Comput 2023; 14: 4795-4807.

[3] Cortes-Briones JA, Tapia-Rivas NI, D'Souza DC and Estevez PA. Going deep into schizophrenia with artificial intelligence. Schizophr Res 2022; 245: 122-140.

[4] Sadeghi D, Shoeibi A, Ghassemi N, Moridian P, Khadem A, Alizadehsani R, Teshnehlab M, Gorriz JM, Khozeimeh F, Zhang YD and Nahavandi S. An overview of artificial intelligence techniques for diagnosis of Schizophrenia based on magnetic resonance imaging modalities: Methods, challenges, and future works. Comput Biol Med 2022; 146: 105554.

[5] Zhang J, Rao VM, Tian Y, Yang Y, Acosta N, Wan Z, Lee PY, Zhang C, Kegeles LS, Small SA and Guo J. Detecting schizophrenia with 3D structural brain MRI using deep learning. Sci Rep 2023; 13: 14433.

[6] Wanderley Espinola C, Gomes JC, Mônica Silva Pereira J and dos Santos WP. Detection of major depressive disorder, bipolar disorder, schizophrenia and generalized anxiety disorder using vocal acoustic analysis and machine learning: an exploratory study. Res Biomed Eng 2022; 38: 813-829.

[7] Santos Febles E, Ontivero Ortega M, Valdés Sosa M and Sahli H. Machine learning techniques for the diagnosis of schizophrenia based on event-related potentials. Front Neuroinform 2022; 16: 893788.

[8] Feng Y, Shen J, He J and Lu M. Schizophrenia and cell senescence candidate genes screening, machine learning, diagnostic models, and drug prediction. Front Psychiatry 2023; 14: 1105987.

[9] Tanveer M, Jangir J, Ganaie MA, Beheshti I, Tabish M and Chhabra N. Diagnosis of schizophrenia: a comprehensive evaluation. IEEE J Biomed Health Inform 2023; 27: 1185-1192.

[10] Jiang Z, Luskus M, Seyedi S, Griner EL, Rad AB, Clifford GD, Boazak M and Cotes RO. Utilizing computer vision for facial behavior analysis in schizophrenia studies: a systematic review. PLoS One 2022; 17: e0266828.

[11] Yang Q, Xing Q, Yang Q and Gong Y. Classification for psychiatric disorders including schizophrenia, bipolar disorder, and major depressive disorder using machine learning. Comput Struct Biotechnol J 2022; 20: 5054-5064.

[12] Keihani A, Sajadi SS, Hasani M and Ferrarelli F. Bayesian optimization of machine learning classification of resting-state EEG microstates in schizophrenia: a proof-of-concept preliminary study based on secondary analysis. Brain Sci 2022; 12: 1497.

[13] Kalinich M, Ebrahim S, Hays R, Melcher J, Vaidyam A and Torous J. Applying machine learning to smartphone-based cognitive and sleep assessments in schizophrenia. Schizophr Res Cogn 2022; 27: 100216.

[14] Hofmann LA, Lau S and Kirchebner J. Advantages of machine learning in forensic psychiatric research - uncovering the complexities of aggressive behavior in schizophrenia. Appl Sci 2022; 12: 819.

[15] Jafari M, Sadeghi D, Shoeibi A, Alinejad-Rokny H, Beheshti A, García DL, Chen Z, Acharya UR and Gorriz JM. Empowering precision medicine: AI-driven schizophrenia diagnosis via EEG signals: a comprehensive review from 2002-2023. Appl Intell 2024; 54: 35-79.

[16] Siuly S, Li Y, Wen P and Alcin OF. SchizoGoogLeNet: the GoogLeNet-based deep feature extraction design for automatic detection of schizophrenia. Comput Intell Neurosci 2022; 2022: 1992596.
[17] Ko DW and Yang JJ. EEG-based schizophrenia diagnosis through time series image conversion and deep learning. Electronics 2022; 11: 2265.
[18] Siuly S, Guo Y, Alcin OF, Li Y, Wen P and Wang H. Exploring deep residual network-based features for automatic schizophrenia detection from EEG. Phys Eng Sci Med 2023; 46: 561-574.
[19] Appaji A, Harish V, Korann V, Devi P, Jacob A, Padmanabha A, Kumar V, Varambally S, Venkatasubramanian G, Rao SV and Suma HN. Deep learning model using retinal vascular images for classifying schizophrenia. Schizophr Res 2022; 241: 238-43.
[20] Khare SK, Bajaj V and Acharya UR. SchizoNET: a robust and accurate Margenau-Hill time-frequency distribution-based deep neural network model for schizophrenia detection using EEG signals. Physiol Meas 2023; 44.
[21] Aydemir E, Dogan S, Baygin M, Ooi CP, Barua PD, Tuncer T and Acharya UR. CGP17Pat: automated schizophrenia detection based on a cyclic group of prime order patterns using EEG signals. Healthcare 2022; 10: 643.
[22] Saadatinia M and Salimi-Badr A. An explainable deep learning-based method for schizophrenia diagnosis using generative data augmentation. IEEE Access 2024; 12.
[23] Shen M, Wen P, Song B and Li Y. Automatic identification of schizophrenia based on EEG signals using dynamic functional connectivity analysis and 3D convolutional neural network. Comput Biol Med 2023; 160: 107022.
[24] Rani KEE and Baulkani S. Construction of deep learning model using ResNet-50 for schizophrenia prediction from rsfMRI images. Research Square 2022.
[25] Salam MA, Badr E, Monier E and Mohamed A. Schizophrenia diagnosis using optimized federated learning models. IJCSNS 2022; 22: 829.
[26] Ahammed MS, Niu S, Ahmed MR, Dong J, Gao X and Chen Y. DarkASDNet: classification of ASD on functional MRI using deep neural network. Front Neuroinform 2021; 15: 635657.
[27] Ruan G, Liu J, An Z, Wu K, Tong C, Liu Q, Liang P, Liang Z, Chen W, Zhang X and Feng Y. Automated skull stripping in mouse functional magnetic resonance imaging analysis using 3D U-Net. Front Neurosci 2022; 16: 801769.
[28] Huang P, Correia MM, Rua C, Rodgers CT, Henson RN and Carlin JD. Correcting for superficial bias in 7T gradient echo fMRI. Front Neurosci 2021; 15: 715549.
[29] Jönemo J, Abramian D and Eklund A. Evaluation of augmentation methods in classifying autism spectrum disorders from fMRI data with 3D convolutional neural networks. Diagnostics 2023; 13: 2773.
[30] Alayba AM, Senan EM and Alshudukhi JS. Enhancing early detection of Alzheimer's disease through hybrid models based on feature fusion of multi-CNN and handcrafted features. Sci Rep 2024; 14: 31203.
[31] Zhou J, Wei Y, Li X, Zhou W, Tao R, Hua Y and Liu H. A deep learning model for early diagnosis of Alzheimer's disease combined with 3D CNN and video Swin transformer. Sci Rep 2025; 15: 23311.
[32] Soltanpour S, Chang A, Madularu D, Kulkarni P, Ferris C and Joslin C. 3D Wasserstein generative adversarial network with dense U-Net-based discriminator for preclinical fMRI denoising. J Imaging Inform Med 2025; 1: 1-20.
[33] Gao J, Liu J, Xu Y, Peng D and Wang Z. Brain age prediction using graph neural network based on resting-state functional MRI in Alzheimer's disease. Front Neurosci 2023; 17: 1222751.
[34] Khatri U and Kwon GR. Alzheimer's disease diagnosis and biomarker analysis using resting-state functional MRI functional brain network with multi-measures features and hippocampal subfield and amygdala volume of structural MRI. Front Aging Neurosci 2022; 14: 818871.
[35] Chelladurai A, Narayan DL, Divakarachari PB and Loganathan U. fMRI-based Alzheimer's disease detection using the SAS method with multilayer perceptron network. Brain Sci 2023; 13: 893.
[36] Preetha R, Priyadarsini MJ and Nisha JS. Automated brain tumor detection from magnetic resonance images using fine-tuned EfficientNet-B4 convolutional neural network. IEEE Access 2024; 12: 112181-95.
[37] Lian J, Hui G, Ma L, Zhu T, Wu X, Heidari AA, Chen Y and Chen H. Parrot optimizer: algorithm and applications to medical problems. Comput Biol Med 2024; 172: 108064.
[38] Ahmed M, Sulaiman MH, Mohamad AJ and Rahman M. Gooseneck barnacle optimization algorithm: a novel nature-inspired optimization theory and application. Math Comput Simul 2024; 218: 248-265.
[39] Alharthi AG and Alzahrani SM. Multi-slice generation sMRI and fMRI for autism spectrum disorder diagnosis using 3D-CNN and vision transformers. Brain Sci 2023; 13: 1578.