



OPEN

Schizophrenia detection from electroencephalogram signals using image encoding and wrapper-based deep feature selection approach

Utathya Aich^{1,6}, Arghyasree Saha^{1,6}, Marcin Woźniak^{3✉}, Muhammad Fazal Ijaz^{4,5✉} & Pawan Kumar Singh^{1,2}

Schizophrenia is a persistent and serious mental illness that leads to distortions in cognition, perception, emotions, speech, self-awareness, and actions. Affecting about 1% of people worldwide, schizophrenia usually emerges in late adolescence or early adulthood. It is characterized by symptoms like hallucinations, delusions, disorganized speech, and cognitive impairments. Despite significant research efforts, the exact cause of schizophrenia is still not fully understood, highlighting the need for continuous investigation into new diagnostic and treatment methods. The electroencephalogram (EEG), which measures brain electrical activity using scalp electrodes, is crucial in schizophrenia research due to its ability to detect subtle brain activity changes due to high temporal information and provide valuable insights into brain function. Many methods have been proposed to identify schizophrenia for diagnosis. Different machine learning and deep learning models have been used to improve the detection of schizophrenia. Through transfer learning using deep learning models, relevant features are selected automatically, outperforming traditional methods in accuracy and speed. Our paper introduces a three-stage framework for detection of schizophrenia from EEG signals. An image encoding method has been used to encode EEG signals to scalogram images to get both spatial and temporal information of the time series data. Using these images in the second step, two pre-trained deep learning models are implemented using transfer learning to extract features for the detection of schizophrenia. In the third step, a newly developed Average subtraction wrapper-based feature selection method has been proposed to lower the number of irrelevant features. The proposed framework has been tested on two datasets. The first (M.S.U) dataset is from M.V. Lomonosov Moscow State University which contains EEG data of 84 individuals where 45 individuals are with schizophrenia symptoms and the rest are 39 individuals are healthy. The second RepOD dataset contains EEG data of 28 individuals where both schizophrenic and healthy individuals are equal in number. Our framework achieved 99.67% and 99.97% accuracy on the first and second dataset, respectively. On both the datasets, our proposed framework outperformed state of the art results.

Keywords Schizophrenia detection, Deep learning model, Electroencephalogram signals, Continuous wavelet transform, Scalogram, Transfer learning, EfficientNet, DenseNet, Average Subtraction based optimization

¹Machine Learning Engineer, CNH Industrial ITC, Greater Noida, India. ²Department of Information Technology, Jadavpur University, Jadavpur University Second Campus, Plot No. 8, Salt Lake Bypass, LB Block, Sector III, Salt Lake City, Kolkata, West Bengal 700106, India. ³Faculty of Applied Mathematics, Silesian University of Technology, Kaszubska 23, Gliwice 44100, Poland. ⁴Torrens University, 196 Flinders St, Melbourne, VIC 3000, Australia. ⁵Centre for Artificial Intelligence Research and Optimisation (AIRO), Torrens University, 46-52 Mountain St, Sydney, NSW 2007, Australia. ⁶Utathya Aich and Arghyasree Saha contributed equally to this work. ✉email: marcin.wozniak@polsl.pl; fazal605@gmail.com

Schizophrenia is a chronic and severe mental condition that affects a person's thoughts, feelings, and behaviour^{1–4}. It is one of the most complex mental health illnesses, including episodes of psychosis that include abnormal thinking and perceptions, hallucinations, and delusions^{5,6}. Approximately 1% of the world's population suffers from it. Despite its prevalence of around 20 million individuals globally, it remains one of the most misunderstood and stigmatised disorders. It affects around 0.3–0.7% of individuals at some point in their lives. In 2017, there were an estimated 1.1 million new cases, with the global total reaching 24 million by 2022. The condition is more prevalent in males, who also tend to develop symptoms earlier than females. The precise origins of schizophrenia remain elusive, but it is widely understood to be influenced by a combination of factors including genetics, brain chemistry, and environmental factors. Individuals with it experience symptoms such as hallucinations, delusions, and various cognitive and motor impairments. Strong evidence suggests a significant genetic component, with several genes, including distributed binding protein 1 (DTNBP1) and neuregulin 1 (NRG1), implicated in its development. Specific mutations in genes involved in "synaptic pruning," the process of eliminating connections between neurons, have also been associated with a higher risk of schizophrenia. Imbalances in brain chemistry, particularly subcortical dopamine deficiency, are believed to be a major contributing factor. Structural abnormalities in specific regions of grey and white matter are observed in individuals with schizophrenia, particularly following the onset of psychosis. These abnormalities are thought to be linked to impairments in executive functioning, attention, and working memory. Additionally, environmental factors such as stressful life experiences can precipitate psychotic episodes. Issues in brain development before birth, as well as disruptions in the connections between different brain regions, are also believed to contribute to the development of schizophrenia.

It is primarily diagnosed through clinical evaluation, relying on the patient's history and symptoms. However, this subjective method can lead to misdiagnosis and delayed treatment, potentially worsening the condition. Therefore, there is a pressing need for objective diagnostic procedures that can accurately and swiftly identify it. Neuroimaging techniques play a crucial role in schizophrenia research, allowing for the visualization of brain structure and function. CT scans utilize X-rays to produce detailed images of the brain, assisting in identifying anatomical abnormalities such as variations in brain volume or the presence of tumours. MRI combines a strong magnetic field with radio waves to create detailed brain images, capable of identifying subtle changes in brain anatomy such as reductions in grey matter volume and white matter integrity. Various types of MRI scans are employed in schizophrenia research. Morphologic MRI provides precise images of brain anatomy to identify abnormalities, while functional MRI (fMRI) detects changes in blood flow to evaluate brain activity, particularly useful in understanding how brain function is altered in individuals with schizophrenia. Diffusion tensor imaging measures the diffusion of water molecules in the brain, revealing the integrity of white matter tracts and connectivity abnormalities related to schizophrenia. Magnetic resonance spectroscopy, a non-invasive technique, detects metabolic changes and specific metabolites in the brain, helping researchers understand the biochemical alterations associated with schizophrenia. Positron emission tomography scans use a radioactive tracer to examine brain metabolism, providing valuable insights into the neurobiological foundations of schizophrenia.

The magnetic fields generated by brain electrical activity is measured by Magnetoencephalography, offering immediate information about the timing and location of brain activity². Meanwhile, Electroencephalogram (EEG) measures brain electrical activity using scalp electrodes, detecting minute changes in brain activity and providing crucial insights into brain function⁷. EEG has emerged as a valuable method for objective analysis in schizophrenia research due to its accessibility, cost-effectiveness, and high temporal resolution^{3,4}. With its ability to reveal underlying neurophysiological abnormalities associated with schizophrenia, EEG has become an essential tool in the diagnosis and monitoring of the disorder. However, interpreting EEG data remains challenging due to its high dimensionality and the subtle nature of disorder-related signal alterations. Nonetheless, EEG's potential for automatic schizophrenia detection using machine learning and deep learning techniques and its ability to access high-quality open databases make it a promising avenue for further research and clinical application^{8–10}.

There has been continuous research work going on with EEG data. EEG signals are characterized by complex temporal patterns across multiple frequency bands, making it difficult to directly process them using traditional machine learning techniques¹¹. Analysing raw EEG data often requires specialized preprocessing steps, such as filtering, artifact removal, and feature extraction, which can be time-consuming and require domain expertise. In EEG data analysis, researchers utilize diverse techniques such as IMF and ERP for temporal and spatial feature extraction. Statistical tests like Kruskal-Wallis aid in identifying significant patterns, while methods like Fuzzy C-Means clustering group similar signals. Complex network analysis reveals connectivity patterns, and advanced approaches like MEMD and Fourier Transformation provide insights into temporal dynamics¹². Additionally, SLBP and time-frequency analysis contribute to understanding EEG data. These techniques enable researchers to select pertinent features, crucial for deciphering cognitive processes and identifying neurological abnormalities. Working with raw EEG data, researchers typically face challenges related to its high dimensionality, temporal dynamics, and noise susceptibility¹⁰.

However, by converting EEG data into images, these challenges can be mitigated to some extent. Transforming EEG signals into visual representations simplifies the data structure and allows for the application of well-established image processing techniques^{13,14}. This conversion enables the utilization of deep learning models, particularly CNNs, which are highly effective in analysing spatial information present in images^{8,15}. One significant advantage of using images derived from EEG data is the compatibility with pre-trained CNN architectures. Researchers can leverage transfer learning techniques by fine-tuning CNN models on these EEG-derived images, benefiting from the rich representations learned from large-scale image datasets. This approach can enhance the robustness and generalization capability of EEG-based classification models, especially when dealing with limited labelled data. This simplification accelerates experimentation and reduces the computational

overhead associated with handling raw EEG signals. Choosing the most pertinent features is consistently difficult. Feature selection methods provide a means to enhance model performance^{16–18}. In the field of feature selection, three primary approaches are recognized: filter, wrapper, and embedded methods. Filter methods rely on statistical properties to select features. These methods are computationally efficient to identify potentially relevant features, but lack model understanding and may miss important feature interactions. Embedded methods take feature selection into the model-building process. These methods incorporate feature selection within the algorithm's training procedure, often through regularization techniques like Lasso regression or tree-based algorithms like Random Forests. Embedded methods automatically select the most relevant features during model training, eliminating the need for a separate feature selection step. While these methods can be efficient and effective, they are limited by the specific algorithm used and may not always capture all relevant features or interactions^{19–22}.

In contrast, wrapper methods evaluate feature subsets by directly measuring the performance of a particular model^{23–25}. These methods typically involve an iterative process where different subsets of features are evaluated using a chosen learning algorithm. By incorporating the model's performance as part of the feature selection process, wrapper methods can identify the most relevant features that contribute to the model's accuracy. This approach can lead to better model performance, especially when complex relationships exist between features and the target variable.

Among these techniques, wrapper-based methods offer a distinct advantage by iteratively evaluating different feature subsets within the context of the model. Wrapper methods can identify features that contribute most significantly to predictive accuracy, potentially leading to improved model generalization and performance^{26–28}. Ghosh et al.²⁹ utilized a wrapper-based approach for human activity recognition, attaining an accuracy of 97.97%. Filter-based methods aid in ranking features, facilitating the selection of top-performing features for improved training outcomes. Yu et al. explored correlation-based feature selection for enhanced results with high-dimensional data. Furthermore, Mondal et al.¹⁶ introduced a tri-stage wrapper filter selection framework for disease classification, attaining a perfect accuracy rate of 100%.

The proposed works contribute to the advancement of disease classification methodologies with a focus on detecting Schizophrenia from EEG signals. The research introduces a unique three-stage framework entered around image encoding and wrapper feature optimization. Initially, EEG signals are transformed into visual representations using an image encoding technique, thereby augmenting the model's ability to identify patterns associated with schizophrenia. Subsequently, a new wrapper-based feature selection method is employed to identify an optimized subset of features. Through this innovative framework, the research aims to significantly reduce computation time also not compromising classification accuracy, offering promising prospects for improved diagnosis and treatment of schizophrenia.

The contributions of the paper are summarized as follows:

1. Two new datasets have been created by encoding EEG signals into scalogram images for the classification of schizophrenia.
2. A three-stage hybrid framework is proposed, consisting of: (i) image encoding using Continuous Wavelet Transform (CWT), (ii) feature extraction using transfer learning with two pre-trained deep learning models, and (iii) feature refinement through an advanced deep wrapper-based feature selection method.
3. A new deep wrapper-based feature selection method has been proposed to identify the most relevant features with the aim of improving classification accuracy.
4. The proposed framework demonstrates high classification accuracy (99.67% and 99.97%) on two independently sourced EEG datasets — from M.V. Lomonosov Moscow State University³⁰ and the RepOD repository³⁴ — validating its generalizability.
5. The framework also outperforms existing methods while using a smaller number of features.

The paper is organized as follows: Section 2 reviews current approaches found in the literature. Section 3 describes the proposed methodology with explanation of the dataset, while Section 4 covers the results and its analysis on using the framework on the dataset and lastly Section 5 concludes the paper and offers suggestions for future research directions.

Literature analysis

In recent years, there has been substantial research into detecting schizophrenia using EEG signals. There are many datasets available publicly. Several research works have been proposed on the M.S.U EEG dataset³⁰. Phang et al.³¹ proposed a study employing Deep Belief Network architecture for automated schizophrenia categorization using EEG effective connectivity. They employed a combination of auto regression-based directed connectivity, graph based network measures as input features. The DNN-DBN outperformed three standard classifiers, leveraging its ability to extract features. These features yielded superior results, achieving a 95% classification accuracy. Rajesh et al.³² developed an automated method for identifying schizophrenia in teenagers using EEG signals, based on Symmetrically Weighted Local Binary Patterns (SLBP). This approach extracts histogram features based on SLBP from each EEG channel. The resulting vector is then fed into a LogitBoost classifier to distinguish between EEG signals of individuals with schizophrenia and those who are healthy. Findings from publicly accessible databases illustrate that SLBP aptly characterizes changes in EEG signals and achieves a classification accuracy of 91.66%. Furthermore, this method surpasses previously suggested techniques for detecting schizophrenia. Sobahi et al.³³ introduced a unique method for schizophrenia detection through EEG recordings, employing a signal-to-image mapping technique. They utilize wavelet transforms to separate EEG channels into rhythms, then encode these signals with 1D local binary patterns (LBP). By merging uniform histograms of LBP-coded beats, they form rows in input images, where channels represent rows and rhythms

represent columns. Data augmentation involves using extreme learning machines (ELM)-based autoencoders (AE). After augmentation, deep transfer learning separates schizophrenic and healthy cases, achieving 97.7% accuracy on M.S.U EEG dataset, outperforming recent methods.

Using the RepOD dataset³⁴ Buettner et al.³⁵ proposed a rapid and highly effective classification method for diagnosing schizophrenia, focusing on EEG data from potentially exposed individuals. The model comprises three main preprocessing steps: Independent Component Analysis, Spectral analysis was conducted using 99-frequency-band method, along with normalization techniques. Classification was performed using Random Forest, achieving a 100% accuracy rate in excluding schizophrenia. This approach has the potential to accelerate diagnosis, enhance accuracy, and lower the cost of ICU treatments. The overall accuracy of the model was 71.43%. On the same dataset Krishnan et al.³⁶ focused on using Multivariate Empirical Mode Decomposition (MEMD) and entropy measures to detect schizophrenia. The study involved both healthy control and schizophrenia subjects. Researchers utilized MEMD to decompose the signals into Intrinsic Mode Functions (IMF), computing entropy measures for checking randomness. Different entropy measures were derived from the IMF signal. Following this various state-of-the-art machine learning classifiers were trained on the entropy values, with SVM-RBF achieving the highest accuracy and F1-score of 93% with 95 features. Sara et al.³⁷ present a novel approach for schizophrenia identification using EEG signals from the RepOD IBIB PAN dataset. Analysing EEG waves from 14 schizophrenia patients and 14 healthy controls, the study employed Transfer Entropy (TE) to derive an effective connection matrix. This matrix was converted into a 19×19 connectivity image and input into five pre-trained CNN models and an LSTM model to extract discriminatory spatiotemporal features. The hybrid CNN-LSTM framework outperformed standalone CNN models, with the EfficientNetB0-LSTM model achieving the highest average accuracy 99.9%, through 10-fold cross-validation. Our proposed work used both M.S.U³⁰ and RepOD³⁴ dataset for evaluation. Timeline of different works with EEG data for Schizophrenia classification have been depicted in Fig. 1.

The studies mentioned highlight notable advancements in the classification of Schizophrenia, employing various methodologies with EEG data. Despite the diversity in approaches, there have been encouraging results in accurately identifying schizophrenia, with the combination of different features proving pivotal in achieving better outcomes. However, challenges remain, such as the requirement for larger and more diverse datasets, extended processing durations, the creation of reliable and efficient models, and the smooth integration of these models into clinical practices. Tackling these challenges is crucial to aiding medical practitioners in making

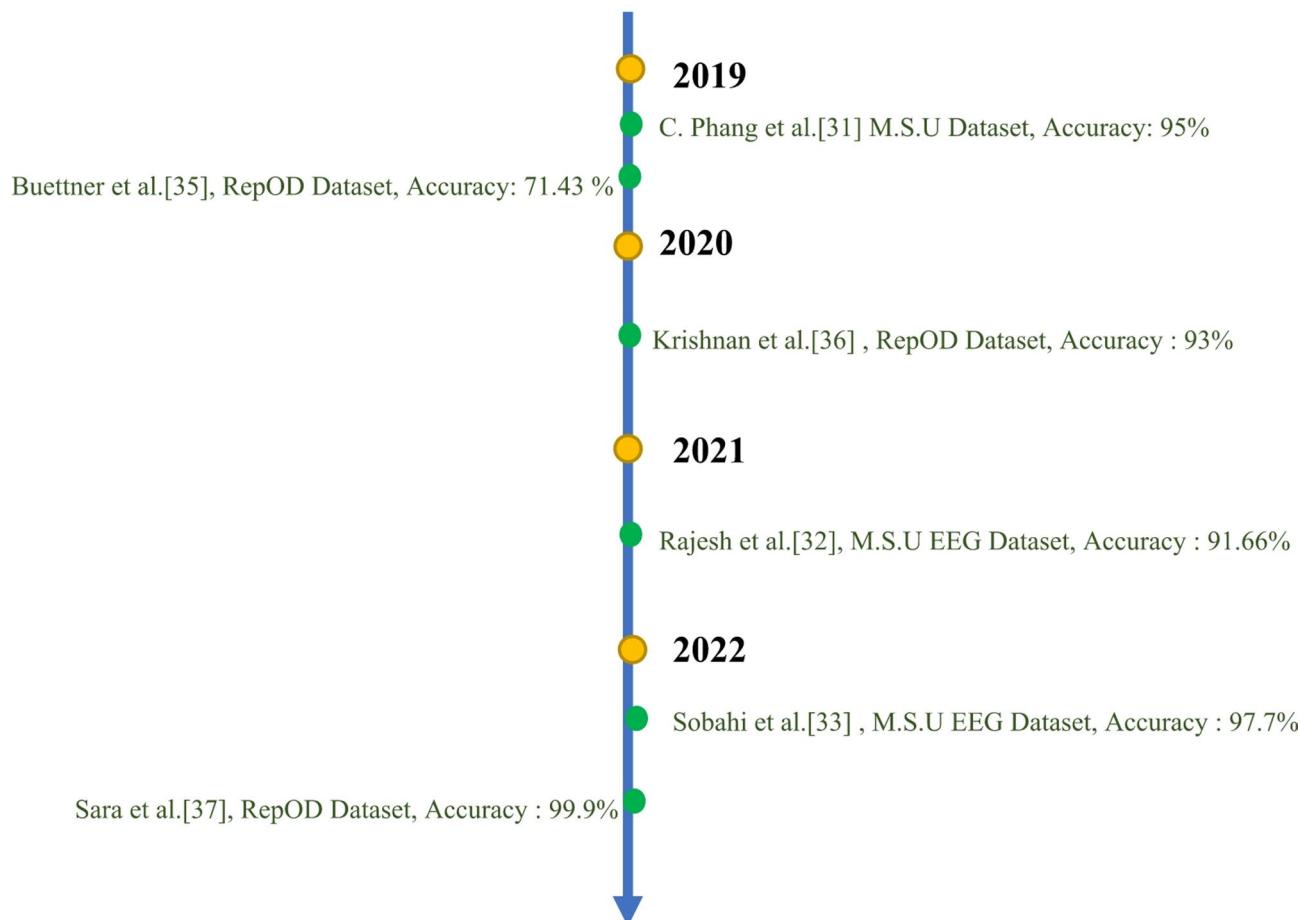


Fig. 1. Research timeline on Schizophrenia classification using on EEG data.

accurate diagnoses and, ultimately, improving patient outcomes. The next section deals with our proposed framework for the detection of schizophrenia.

Motivation of proposed work

Researchers have explored various machine learning and CNN-based models in previous studies on these datasets. However, there is no unified framework applicable to all these datasets. For the first time, raw EEG data has been converted into scalogram images, preserving the spatial and temporal information inherent in the time-series EEG data. This new encoded dataset of scalogram images can be further useful for schizophrenia detection. This transformation redefines the Schizophrenia detection problem as an image classification task. Our proposed three-stage framework first converts the raw EEG data into scalogram images, which are then classified as either Schizophrenia or Normal using optimal features through a subsequent two-stage framework. The detailed description of this framework is provided in the following section.

Materials and framework

CNNs are a specialized type of deep learning model designed to process visual data. These networks consist of several trainable layers of a conventional CNN architecture. The convolutional and pooling layers, in particular, can be optimized by adjusting hyperparameters. The synergy of various CNN architectures with transfer learning techniques has significantly advanced image classification. ImageNet³⁸, especially through the ILSVRC competition since 2012, has played a crucial role in enhancing deep image recognition frameworks, driving the widespread adoption of CNN algorithms in visual recognition tasks.

Figure 2 depicts the comprehensive workflow of a schizophrenia classification framework, spanning from data encoding to image processing and transfer learning, culminating in classification through a proposed Average Subtraction wrapper-based feature selection method. This framework can be delineated into three primary phases:

1. Raw EEG data encoding to Scalogram images.
2. Feature extraction using EfficientNetB3³⁹ and DenseNet169⁴⁰ through transfer learning using ImageNet weights.
3. Average Subtraction Wrapper based feature selection to select optimal features from the concatenated features for schizophrenia classification.

Data preprocessing

Dataset description

M.S.U EEG dataset³⁰ In our research project, we utilized publicly accessible EEG datasets sourced from M.S.U for the purpose of detecting schizophrenia³⁰. These datasets consist of EEG recordings obtained from adolescents who underwent screening by psychiatrists and were subsequently categorized into two distinct groups: 39 healthy individuals and those exhibiting symptoms indicative of schizophrenia are 45 in numbers.

Each EEG file corresponds to a single subject's EEG recording. Within each file, there is a column dedicated to the samples captured from 16 distinct electrode positions. Each number within the column represents the EEG amplitude (in microvolts, μ V) recorded at a specific sample point. The arrangement of the data is such

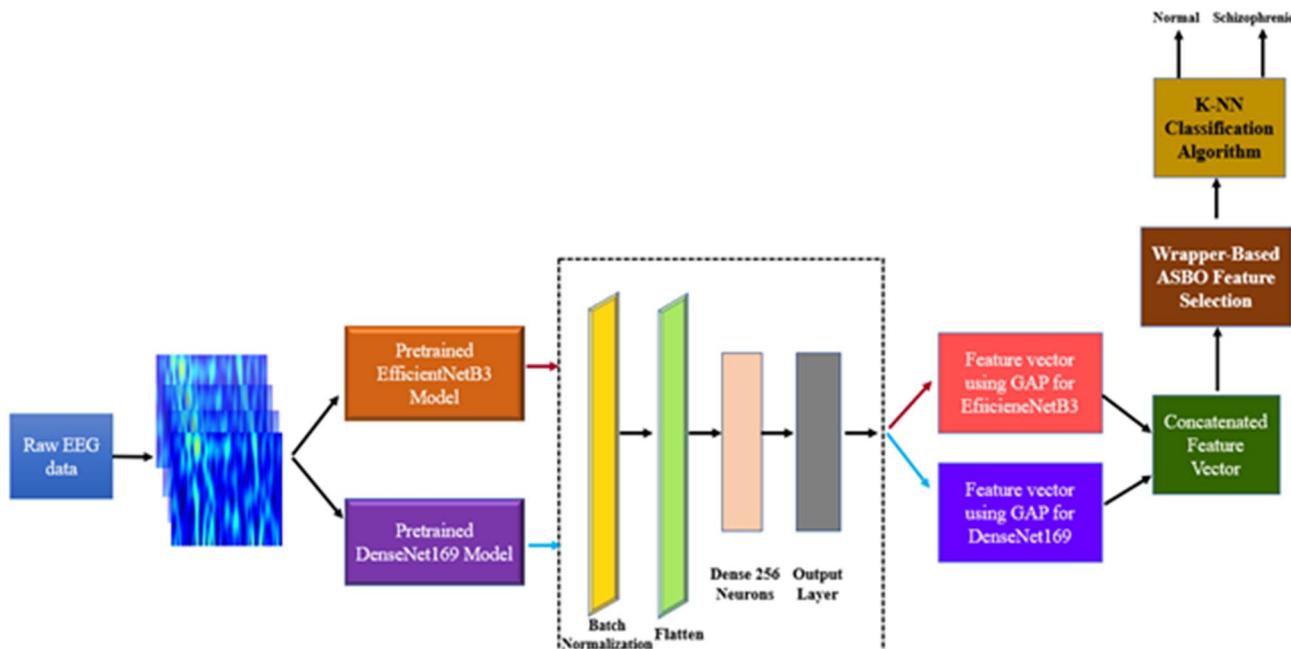


Fig. 2. Workflow of the proposed Average Subtraction wrapper-based feature selection framework.

that the first 7680 samples pertain to the 1st channel, followed by samples for the 2nd channel, and so forth. A duration of 1 min of EEG recording contains 7680 samples, given a sampling rate of 128 Hz. The 16 channels are identified as follows: F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2. The positions of these channel numbers are shown in Fig. 3.

RepOD EEG dataset³⁴ Our proposed framework has been tested on a dataset consisting of EEG recordings from 14 individuals diagnosed with paranoid schizophrenia containing an equal number of 7 males and 7 females. There are also 14 more EEG recordings collected from 7 male and 7 female healthy individuals. Each individual underwent a 15-minute session of EEG recording in a resting state with eyes closed. The EEG data were gathered at a sampling rate of 250 Hz using the standard 10–20 system. 19 electrodes were positioned at Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2, with FCz utilized as the comparison electrode.

Image encoding

Data-to-image encoding is a process of converting raw data, such as numerical values or signals, into visual representations in the form of images. This conversion enables the application of techniques commonly used in image processing and computer vision to analyse and extract features from the original data. In the context of EEG data, which comprises sequences of electrical signals recorded from the brain over time, data-to-image encoding involves transforming these signals into images that encapsulate both temporal and spectral information. In our proposed method, we've implemented a data-to-image encoding technique for EEG data utilizing the CWT. The CWT decomposes the EEG signal into its constituent frequencies and represents the signal's energy distribution across time and frequency dimensions. The resulting image, often referred to as a scalogram, provides a detailed visual representation of the underlying brain activity, with different colours or intensity levels indicating the strength of signal components at different frequencies and time points. This approach enhances the interpretability of deep learning models trained on EEG data and facilitates their integration into clinical decision-making processes, such as diagnosing neurological disorders or monitoring brain health.

In both the MSU³⁰ and RepOD EEG³⁴ datasets, the raw EEG data was initially segmented into 30-second artifact-free intervals, carefully excluding disturbances such as eye movements, cardiac signals, and muscle contractions. These clean segments were then filtered using a second-order Butterworth filter to isolate specific frequency bands: delta (2–4 Hz), theta (4.5–7.5 Hz), alpha (8–12.5 Hz), beta (13–30 Hz), and gamma (30–45 Hz). To enhance signal fidelity and ensure robustness in subsequent analyses, we examined both the original EEG data and data that had undergone re-referencing using three distinct methods: scalp surface Laplacian estimators (also known as current source density or CSD), the AVERAGE method, and the standardized reference electrode technique. Further, to encode the EEG data into images for our proposed framework, we employed scalograms—a form of spectrogram that captures the detailed distribution of signal energy across time and frequency domains. The EEG data was segmented using a 20-second sliding window with a 15-second overlap, resulting in a 5-second shift between consecutive segments to preserve temporal continuity. For each segment, we generated a scalogram using the *ScalogramCWT* function⁴¹, which visualizes dynamic intensity variations in EEG signals across frequencies over time. In Fig. 4, green-coloured dotted lines indicate the current

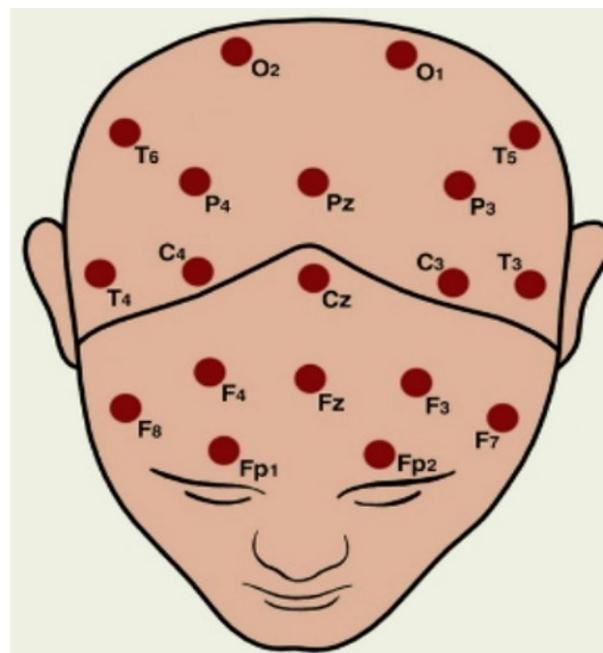


Fig. 3. Topographical positions of channel numbers³⁰.

5-second window for each channel, while red-coloured dotted lines represent the previous overlapping window. These encoded images were subsequently transformed into pixel data for further processing in our model.

In our study, we selected the Morlet wavelet as the basis function for the CWT due to its well-recognized strengths in time-frequency analysis. The Morlet wavelet offers an optimal balance between time and frequency resolution, making it particularly effective for capturing oscillatory patterns across a wide frequency range. Unlike compactly supported wavelets such as Haar or Daubechies—which are better suited for detecting abrupt changes or for denoising—the Morlet wavelet, with its Gaussian-modulated sinusoidal shape, provides superior frequency localization. This property is especially advantageous for identifying rhythmic components and transient features, which are central to our analysis. Although other wavelets, such as the Mexican Hat, may offer better localization for sharp, spike-like features, they generally fall short in frequency resolution compared to the Morlet. Given the characteristics of the signals we are analysing, the Morlet wavelet emerged as a natural and effective choice.⁴²

The scalogram, which represents the squared magnitude of the CWT coefficients, offers a powerful visualization of how signal energy is distributed across both time and frequency. Unlike traditional time-domain or frequency-domain representations, the scalogram provides a joint time-frequency perspective, enabling the detection of non-stationary and transient features in signals—characteristics often missed by standard Fourier analysis. This is particularly useful for analysing EEG-based time-series, where meaningful patterns may vary over time. The high-resolution time-frequency mapping of the scalogram allows for precise localization of oscillatory activity, aiding in the identification of relevant biomarkers or abnormal patterns. When paired with the Morlet wavelet, the scalogram becomes even more effective, as the wavelet's strong frequency localization enhances the clarity and interpretability of these time-varying features.

Feature extraction using CNN models

The images that have been pre-processed in the previous stage are inputted into two transfer learning models known as EfficientNetB3³⁹ and DenseNet169⁴⁰. These models have been trained on extensive datasets of images, enabling them to classify images or similar data with exceptional accuracy.

Feature extraction using EfficientNetB3

EfficientNetB3 is a mobile-friendly convolutional neural network (ConvNet) architecture from the EfficientNet family³⁹. These models provide a novel scaling strategy that employs a compound coefficient to uniformly scale depth, width, and resolution. This strategy strikes an excellent compromise between model size and performance. Mingxing et al. introduced EfficientNetB3³⁹. It was trained on ImageNet³⁸. It is intended for image classification tasks and can be used as a base model or fine-tuned for specific tasks. The architecture consists of depth-wise separable convolutions, squeeze-and-excitation blocks, and fast feature extraction layers.

EfficientNet-B3 extends the capabilities of EfficientNet-B2, aiming to enhance the equilibrium between model efficiency and performance. It presents heightened capacity and accuracy while upholding efficient resource utilization. Employing efficient building blocks like MobileNetV2-like inverted residual blocks and squeeze-and-excitation (SE) blocks, EfficientNet-B3 bolsters its capacity without sacrificing computational

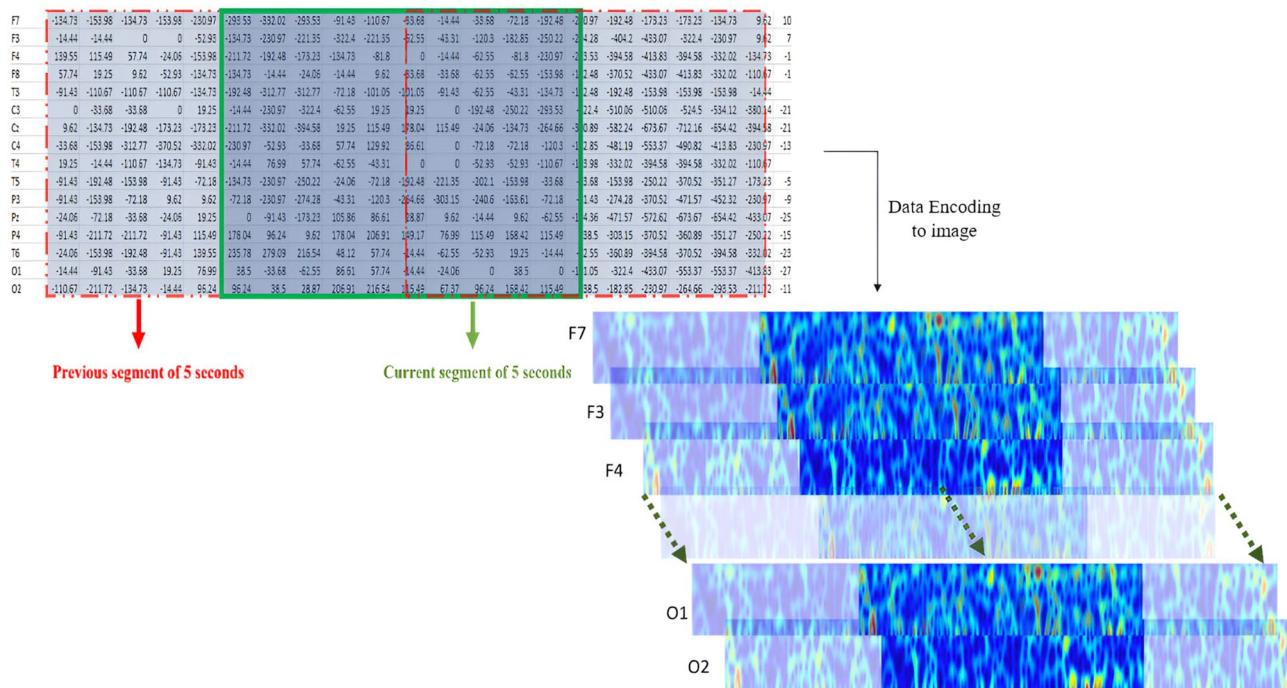


Fig. 4. Raw EEG data encoding to scalogram images.



Fig. 5. Schematic diagram of EfficientNetB3³⁹ Model before sending to the dense layer for classification.

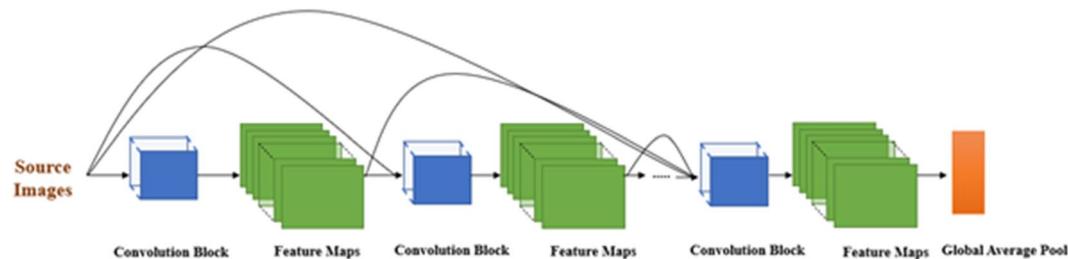


Fig. 6. Schematic Diagram of DenseNet169⁴⁰ Model before sending to dense layer for classification.

efficiency. The compound scaling approach integrated into EfficientNet-B3 facilitates seamless adjustments of the model's size within the EfficientNet³⁹ series, making it adaptable to various scales. This model has displayed impressive performance across a range of tasks, from image classification to segmentation. EfficientNetB3 can be loaded with ImageNet's pre-trained weights. These weights make an excellent starting point for transfer learning problems. EfficientNet³⁹ models demand input images to be float tensors with pixel values ranging from 0 to 255. The model includes input preprocessing (such as rescaling). EfficientNet-B3's benefits, such as its enhanced performance, optimized resource management, transfer learning aptitude, and scalability, render it a valuable selection for a multitude of computer vision applications. The schematic diagram of EfficientNetB3³⁹ is shown in Fig. 5.

Feature extraction using DenseNet169

Densely Connected Convolutional Networks, is a deep learning architecture that has gained prominence for its potential to enhance accuracy and cost-efficiency in many computer vision and text categorization tasks. Huang et al.⁴⁰ introduced DenseNet. The main idea of the architecture is to connect each layer to every other layer in a feed-forward manner, resulting in a network with $L(L+1)/2$ direct connections, where L indicates the number of layers. DenseNet's dense connectivity enhances gradient flow, promotes feature reuse, and minimizes parameter count, making it an effective architecture for a variety of computer vision tasks. DenseNet employs dense connections between layers using Dense Blocks. In the dense blocks, all the layers having feature map sizes matched are directly linked together. Each layer takes additional inputs from the previous layers and sends its own feature maps to the next layers. This dense connectivity structure enables greater feature reuse and gradient flow across the network. DenseNet solves the problem of vanishing gradients in high-level neural networks. The lengthy distance between the input and output layers frequently causes information to evaporate before it reaches its intended destination. Dense connections help to alleviate this issue by ensuring strong gradients throughout the network. The dense connectivity becomes unsustainable as the network goes deeper and deeper. To address this transition layers are introduced. These transition layers use one-by-one convolutions followed by max pooling to reduce feature map size. DenseNet minimizes the number of parameters by reusing characteristics across layers, unlike standard architectures. In our proposed framework we have used DenseNet169. DenseNet169 stands as a notable member within the DenseNet series, boasting 169 layers and serving as a popular choice for various deep learning classification endeavours. Despite its extensive layer count, it notably features a reduced number of trainable parameters when compared with its counterparts possessing fewer layers. This efficiency in parameter usage is accomplished through parameter sharing across layers, thereby leading to a diminished parameter count in contrast to conventional convolutional neural networks. It strikes a balance between model complexity and computational efficiency, making it suitable for a wide range of applications, especially when computational resources are limited. The schematic diagram of DenseNet169 is shown in Fig. 6.

Our initial models, EfficientNetB3³⁹ and DenseNet169⁴⁰, are initialized with weights pretrained on the ImageNet dataset. During deployment, the model is set up with an image size of 256×256 . The final fully connected layer is excluded, and max pooling is applied before the last classification layer. These base models are then integrated into a sequential model—a linear stack of layers. The hyperparameters for both the models are noted in Table 1.

To determine the most effective hyperparameter settings for our model, we conducted extensive and repeated experiments using an empirical approach. These iterative trials allowed us to systematically evaluate a wide range

Parameters	Value
Momentum	0.9
Maximum number of epochs	30
Regularizer	L1 (0.006) & L2 (0.016)
Dropout	0.45
Activation function for hidden layers	ReLU
Activation function output layer	Softmax
Loss	Categorical Crossentropy
Optimizer	Adamax (0.001)

Table 1. Hyperparameter for EfficientNetB3 and DenseNet169 model with pre-trained weights.

of configurations, ultimately identifying parameter values that consistently produced optimal or near-optimal performance. This rigorous trial-and-error methodology enhances the transparency and reproducibility of our study. Looking ahead, we plan to explore more sophisticated hyperparameter optimization strategies—such as grid search, random search, and Bayesian optimization^{43,44}—to automate and further improve the tuning process, thereby increasing both the robustness and generalizability of our model.

Feature optimization module

Optimization involves the selection of the most suitable option from a set of search spaces, with the aim of optimizing or minimizing an objective function while considering constraints. There are various methods used in optimization, including deterministic and stochastic techniques, which can be gradient-based or non-gradient-based. Stochastic optimization techniques can be categorized into four main types: evolutionary, swarm-based, game-based, and physics-based. An example of evolutionary-based methods is the Genetic Algorithm⁴⁵, which imitates natural selection principles to iteratively improve solutions. Swarm-based optimization, demonstrated by Particle Swarm Optimization⁴⁶ and Ant Colony Optimization⁴⁷, relies on cooperation among agents or particles to effectively explore solution spaces.

Optimization algorithms such as Genetic Algorithm⁴⁵, Particle Swarm Optimization⁴⁶, Cuckoo Search Algorithm⁴⁸, Binary Bat Algorithm⁴⁹, and Grey Wolf Optimization Algorithm⁵⁰ require setting parameters to execute effectively, which can be challenging. Conversely, non-parameterized approaches like Average Subtraction based Optimization (ASBO) aim to find nearly optimal solutions by analysing the average and differences between the strongest and weakest solutions within a group. Dehghani et al.⁵¹ proposed Average Subtraction Based Optimization Algorithm (ASBO). ASBO employs three different phases in the process of updating the algorithm population. This improves the solution of the candidates. ASBO identifies promising regions in the search space by steering away from poorly performing areas, without the need for any specific parameters. Interestingly, results obtained using ASBO have been observed to be comparable to those achieved with parameterized optimization algorithms.

Unlike traditional wrapper methods such as Genetic Algorithms⁴⁵ and Particle Swarm Optimization⁴⁶, which rely heavily on biologically inspired operators and require meticulous tuning of control parameters (e.g., mutation rates, crossover probabilities, and learning coefficients), ASBO introduces a deterministic, parameter-free search mechanism. Its core update strategy is based on simple statistical operations—specifically, the average and difference between the strongest and weakest solutions in the population—rather than stochastic transitions or heuristic-driven movements. This results in a more structured and stable search process, with reduced randomness beyond the initial population generation, thereby promoting consistent convergence, improved repeatability, and efficient exploitation of promising regions in the search space.

ASBO has shown the potential to outperform many other optimization algorithms^{42,52}, particularly in high-dimensional or noisy problem settings. Unlike surrogate-model-based approaches, which often become computationally intensive and less accurate as dimensionality increases or when objective functions are irregular, ASBO remains scalable and robust due to its reliance on lightweight statistical computations.

Feature selection plays a vital role in identifying the most impactful feature set from a dataset to attain optimal results for samples. This entails tackling a binary optimization problem where the goal is to use the fewest features possible while maximizing accuracy. Simultaneous optimization of accuracy, to be maximized, and features, to be minimized, is essential. A transfer function such as the sigmoid function is necessary to prioritize features for the selection of the subset.

In our proposed three stage framework, we leverage the wrapper based feature selection using ASBO⁵¹ algorithm in the last stage of the framework for optimal feature selection, ensuring accuracy is maintained without compromise. Wrapper-based approaches consider the interdependencies among features across the entire feature space, leading to enhanced computational performance. Figure 7 illustrates the flowchart of our proposed Average Subtraction wrapper-based feature selection algorithm.

Multiple steps are involved in ASBO⁵¹. Firstly, search space is created. In our proposed framework we choose N to be the total number of candidates in the search space and F to be the Length of the feature vector for each candidate in the search space. In the second step, the fitness function is decided for the optimization problem. Here the fitness function is the accuracy of the classification model. After generating the search space, the fitness value of each candidate is calculated. Step 4 finds the best and worst fitness candidate in the search space for the iteration.

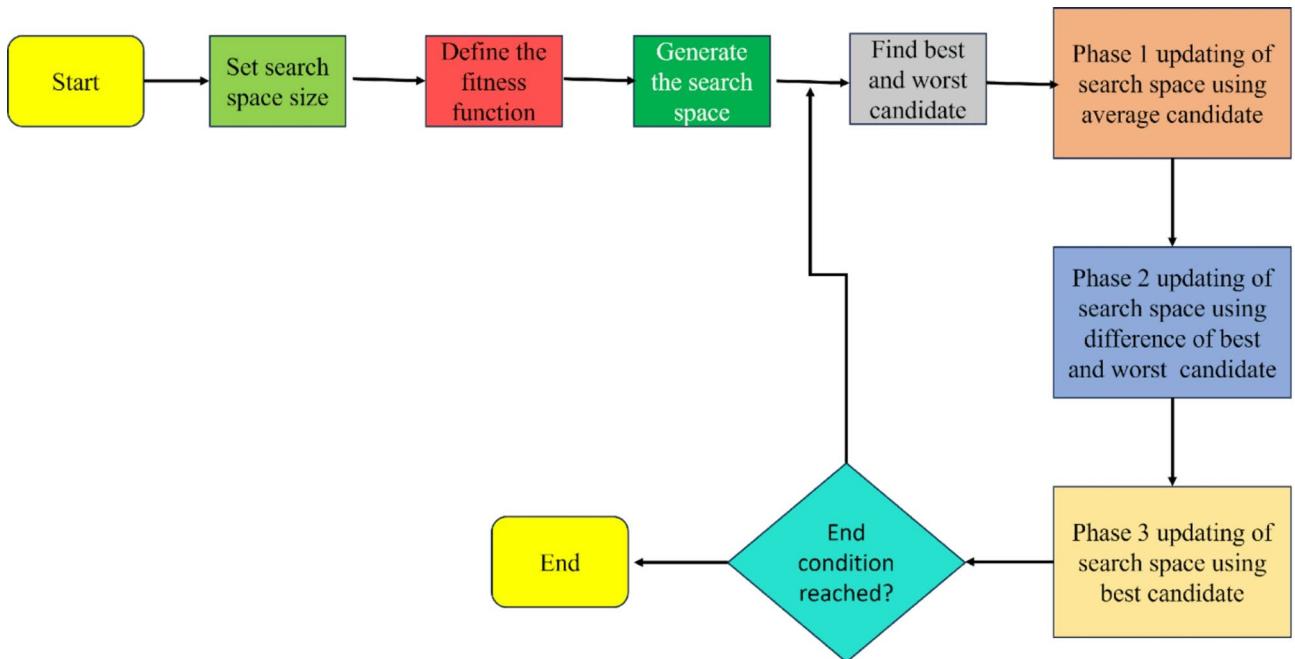


Fig. 7. Flowchart of wrapper based ASBO feature selection in our proposed work.

- Step 1: Set the search space size (N), feature vector length (F) as the parameter.
- Step 2: Specify the classification accuracy as the fitness function.
- Step 3: Randomly generate the search space of N possible candidates.
- Step 4: Calculate fitness value of each member in the search space. Also find the best and worst member for the iteration.
- Step 5: Find and populate the average member in the search space. Update the search space in phase 1 by using the Average member in the equation.
- Step 6: Find the difference between the best and worst member using Step 3. Using the equation in update the search space for phase 2.
- Step 7: Update search space in Phase 3 by using the best member from step 3.
- Step 8: Repeat the steps from 4 to 7 until the termination condition is met.
- Step 9: The optimized subset of features is obtained.

Algorithm 1. ASBO based wrapper feature optimization.

On each new formation of the feature vector, to select or deselect features in the feature vector transformation function is applied over each value in the vector. The values convert to either 0 or 1 using Eqs. 1,2.

$$TF(V_j) = \left\{ \begin{array}{ll} \frac{1}{1 + e^{V_j}} & , V_j > 0, \\ \frac{1}{1 + e^{-V_j}} & , V_j < 0, \end{array} \right. \quad (1)$$

Here V_j is the j^{th} feature in the feature vector V . The value V_j is transformed between 0 and 1 by the sigmoid transfer function expressed in Eq. 1. This means the output will be a high value near to 1 (if the input V_j is positive) or a low value near to 0 (if the input V_j is negative). The output value signifies the feature importance in the feature vector V .

$$V_{newj} = \{1, \quad TF(V_j) > \beta ; 0, \quad \text{else} , \quad (2)$$

Base on the output from the Eq. 1, V_{newj} is determined using Eq. 2. Here β is a random number uniformly distributed between (0,1). The value '1' in V_{newj} indicates the feature is selected whereas '0' indicates that the feature is not selected.

With these values ASBO updates the search space in three phases. In the first phase (P_1) using the following Eqs. 3–5 the search space is updated.

$$L^{P_1} = \frac{X_b + X_w}{2}, \quad (3)$$

Here in Eq. 3, L^{P_1} is the average of the worst (X_w) and best (X_b) candidate in the search space of the iteration.

$$X_{i,d}^{new,P_1} = \{X_{i,d} + r \cdot (L_d^{P_1} - I \cdot X_{i,d}), F_i^{P_1} < F_i; X_{i,d} + r \cdot (X_{i,d} - L_d^{P_1}), else, \quad (4)$$

$F_i^{P_1}$ represents the fitness score for the candidate L^{P_1} calculated using Eq. 3. $L_d^{P_1}$ is the d^{th} dimension of L^{P_1} . In Eq. 4, X_i^{new,P_1} signifies the new status of the i^{th} population member, while F_i^{new,P_1} from Eq. 5 denotes its corresponding fitness function value. X_i^{new,P_1} indicates the d^{th} dimension of X_i^{new,P_1} . I is a random number with a value of either 1 or 2, and is a random number uniformly distributed between 0 and 1. Updation of X_i is done using Eq. 5.

$$X_i = \{X_i^{new,P_1}, F_i^{new,P_1} < F_i; X_i, else, \quad (5)$$

After completion of phase 1, The search space enters for update in phase 2. In this phase the search space is updated from the Eqs. 6–8. Firstly, difference L^{P_2} is calculated between best and worst candidates using Eq. 6.

$$L^{P_2} = X_b - X_w, \quad (6)$$

$$x_{i,d}^{new,P_2} = x_{i,d} + r \cdot L_d^{P_2}, \quad (7)$$

X_i^{new,P_2} represents the updated status of the i^{th} population member, with F_i^{new,P_2} derived from Eq. 8 denoting its respective fitness function value. X_i^{new,P_2} specifies the d^{th} dimension of X_i^{new,P_2} . Lastly, Eq. 8 is utilized to adjust the value of X_i .

$$X_i = \{X_i^{new,P_2}, F_i^{new,P_2} < F_i; X_i, else, \quad (8)$$

Last update of the search space takes place in phase 3 using Eqs. 8–9. In this phase exploitation of the algorithm takes place. The candidates in the search space are updated based on the best solution X_b . Final value of X_i is selected using Eq. 10.

$$x_{i,d}^{new,P_3} = x_{i,d} + r \cdot (X_{i,d} - I \cdot X_{b,d}), \quad (9)$$

$$X_i = \{X_i^{new,P_3}, F_i^{new,P_3} < F_i; X_i, else, \quad (10)$$

Throughout each iteration until termination conditions are met, the entire algorithm updates the search space in three phases. In striving for maximum accuracy, we also prioritize minimizing the number of features. Our proposed wrapper-based feature selection follows the following conditions.

1. Number of iterations = 100
2. Size of search space = 50

The parameter settings for the ASBO algorithm were determined through an extensive series of repeated experimental evaluations. Specifically, we systematically explored a wide range of values for the search space size (N) and the number of iterations to assess their impact on the algorithm's performance. These trials led us to a configuration that consistently delivered strong and reliable results across multiple runs. Although the selection process was empirical, the consistency observed throughout the experiments provides strong evidence of the robustness and effectiveness of the chosen values. Moving forward, we plan to conduct sensitivity analyses and employ automated hyperparameter tuning strategies to further explore how variations in parameter settings influence performance.

For several compelling reasons, the Average Subtraction Based feature optimization algorithm is incorporated in our proposed wrapper-based feature selection framework. This algorithm is highly accurate and efficient when it comes to solving complex problems. ASBO's approach to navigating and leveraging the search space holds the potential for achieving optimal solutions. ASBO algorithm⁵¹ eliminates the need for extensive parameter tuning, streamlining the optimization process.

Classification

Our framework incorporates the K-Nearest Neighbours (KNN)⁵³ algorithm in the wrapper-based feature selection. The algorithm was originally devised by Evelyn Fix and Joseph Hodges. KNN is a cornerstone in supervised machine learning, renowned for its versatility in both classification and regression tasks. The fundamental principle underlying KNN is its reliance on the proximity of data points in a feature space. This principle is demonstrated in Fig. 8.

When applied to classification tasks, KNN determines the class label of a new data point by considering the class labels of its nearest neighbours. It does so by calculating distances, typically using metrics like Euclidean, Minkowski, Manhattan distance, between the new data point and all points in the training set. The 'K' nearest neighbours with the shortest distances are then selected, and the class label for the new data point is determined by majority voting among these neighbours.

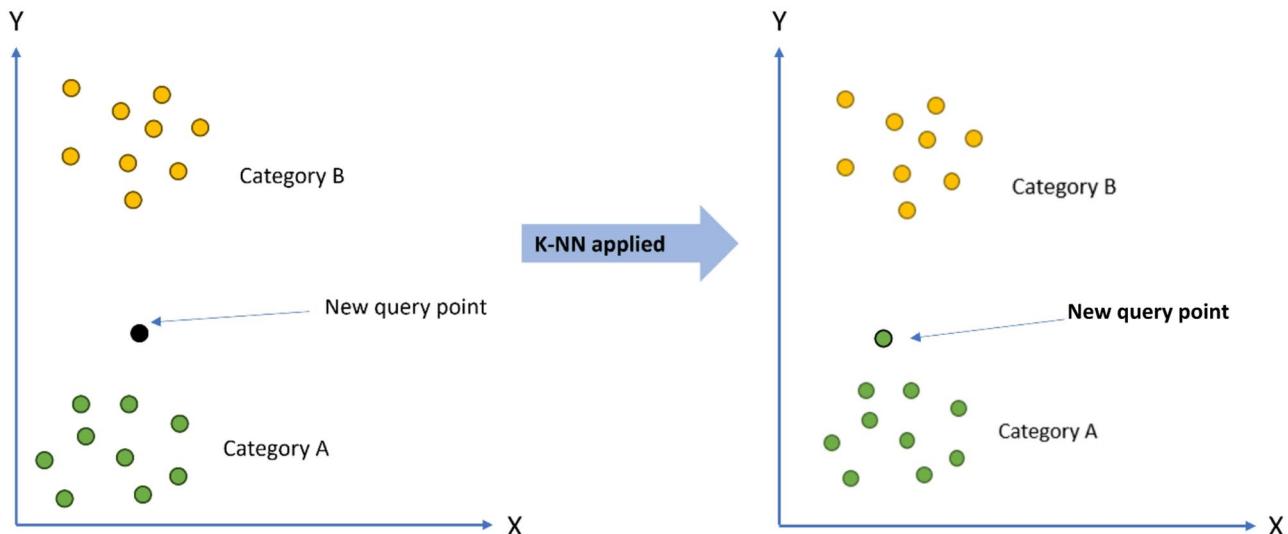


Fig. 8. K-Nearest Neighbour classifier⁵³.

One notable aspect of KNN is its non-parametric nature, meaning it doesn't make assumptions about the underlying data distribution. Additionally, it belongs to the family of "lazy learning" models, which means it doesn't require a separate training stage. Instead, it retains the entire training dataset and performs computations at the time of prediction. Nevertheless, KNN remains widely used across various domains due to its interpretability, ease of implementation, and ability to provide competitive performance in many scenarios. In Fig. 8, we have shown a new query point has been moved to category A by the concept of KNN.

In our proposed framework, the default parameters of KNN are used which are available in Scikit-learn. With respect to original feature set accuracy, our framework chooses the minimum features required to get similar or higher accuracy.

Results

We're utilizing the Kaggle platform along with Python programming for implementing preprocessing techniques. Specifically, we're employing the NVIDIA GPU T4×2-15 GB on the Kaggle platform for our computational needs.

The performance of our proposed three-stage framework was assessed by calculating various outcome-based evaluation metrics, such as true positives (T_i), true negatives (T_j), false positives (F_i), and false negatives (F_j). These metrics together constitute the Confusion Matrix and carry the following interpretation within our context. Formula to calculate Accuracy (A), Precision (P), Recall (R), F_1 Score (F_1) are discussed below.

Accuracy (A) : This determines the frequency of accurate predictions made by the model.

$$A = \frac{T_i + T_j}{T_i + T_j + F_i + F_j}$$

Precision (P) : This indicates the model's reliability in predicting schizophrenia

$$P = \frac{T_i}{T_i + F_i}.$$

Recall (R) : This evaluates the model's capability to detect actual schizophrenia classes.

$$R = \frac{T_i}{T_i + F_j}$$

F_1 – Score (F_1): This metric provides a balanced result by considering both precision (P) and recall (R) .

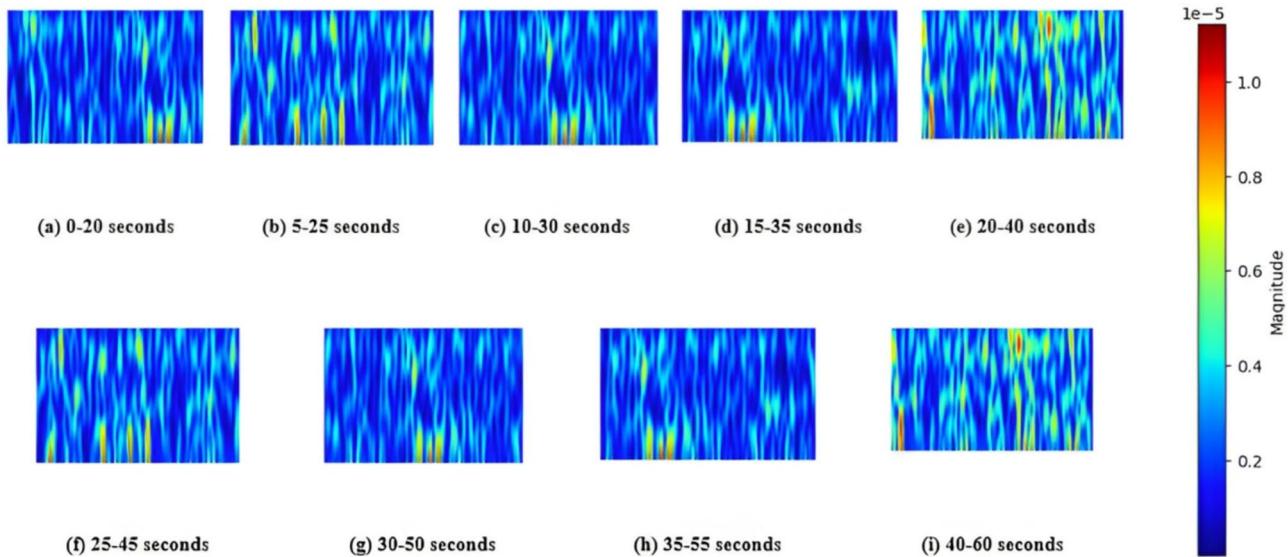
$$F_1 = 2 \times \frac{P \times R}{P + R}$$

Encoded scalogram dataset

*M.S.U EEG dataset*³⁰

In our research, we've used the encoded images derived from raw EEG data through the preprocessing method outlined in Sect. 3.1.2. Using a 20 s window with a 15 s overlap, a total of 12,096 scalogram images were generated from a collection of recorded raw EEG data of 84 individuals. With each image file name, a label type indicating normal or schizophrenic is present. The detailed dataset statistics are represented in Table 2.

M.S.U EEG dataset ³⁰		Encoded scalogram images
Normal	39	5616
Schizophrenic	45	6480
Total	84	12,096

Table 2. Distribution of classes in M.S.U EEG dataset³⁰.**Fig. 9.** Sample encoded scalogram images of the C3 channel for a normal adolescent, generated using a 20-second sliding window with a 15-second overlap.

There are a total of 12,096 scalogram images encoded from the raw data. Normal scalogram images are 5616 in number which is 46.4% of the total images, whereas there are 6480 Schizophrenic images which covers 53.6% of the total dataset.

Sample scalogram images of C3 channels for both Normal and Schizophrenic Adolescents are shown in Figs. 9 and 10, respectively.

We trained deep learning models using pre-trained weights from ImageNet for 30 epochs to extract features. The dataset was divided into training, validation, and testing sets. Table 3 contains distribution of scalogram images for our proposed framework. The training set consists of 9676 scalogram images, with 5197 images categorized as schizophrenic and 4479 as normal. The testing set comprises 2420 scalogram images, with 1137 normal and 1283 schizophrenic images.

RepOD EEG dataset³⁴

For the RepOD dataset³⁴, the scalogram images are generated from 15 min raw EEG data. Using a 20 s segment with 15 s overlap, a total of 89,830 images are encoded from the raw EEG data of 28 individuals. The statistics of the dataset are depicted in Table 4. There are 46,854 encoded normal scalogram images and 42,976 encoded schizophrenic scalogram images in total.

Sample scalogram images of the Fp1 channel for both Normal and Schizophrenic Adolescents are shown in Figs. 11 and 12, respectively.

Using the encoded scalogram images we trained deep learning models using transfer learning. We used the pretrained weights from ImageNet and trained the model for 30 epochs to extract the features. 20% of the images are kept for testing and the rest of the images are kept for training. The validation dataset consists of 20% of the training images. It is noted from Table 5 that the Training set contained 71,864 images. Schizophrenic class contains 34,403 scalogram images in the training dataset whereas the rest 37,461 images belong to Normal class. The testing dataset comprises a total 17,966 images. In the test dataset, the Normal class contains 9385 images and Schizophrenic class contains 8581 images.

Results and discussions using our proposed framework

Our proposed framework is trained over two datasets. In our proposed framework, we used both EfficientNetB3³⁹ and DenseNet169⁴⁰ models to train the scalogram images. Due to the significant computational resources and time typically required for training deep learning models from scratch, we chose to employ transfer learning. Using ImageNet weights through transfer learning, we trained both the models for 30 epochs on the datasets.

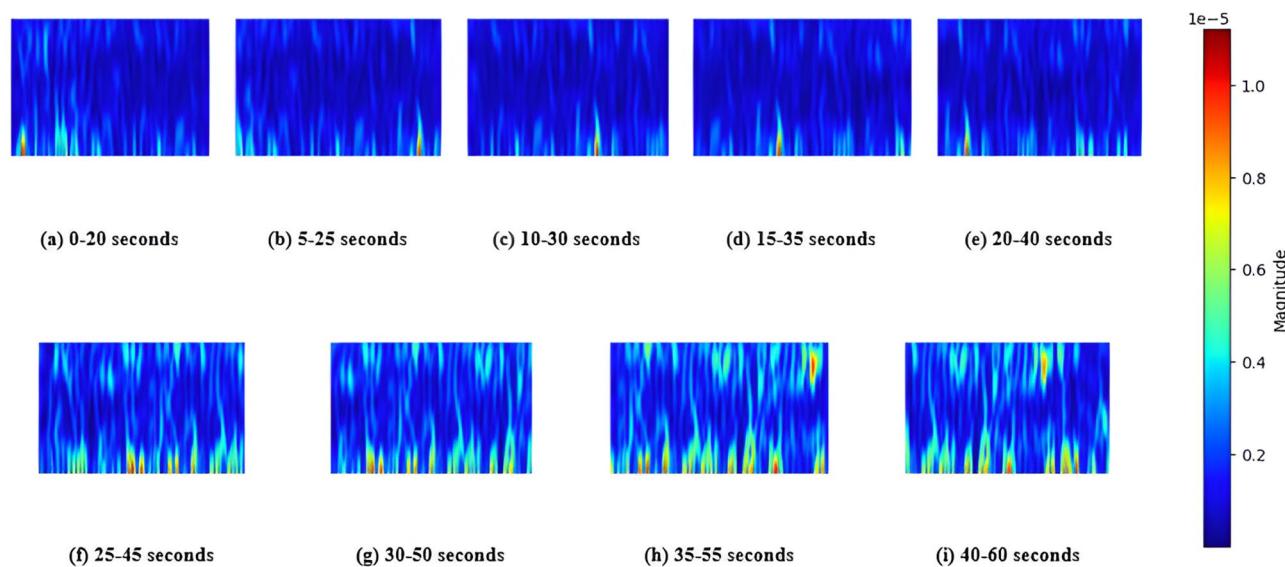


Fig. 10. Sample encoded scalogram images of the C3 channel for a schizophrenic adolescent, generated using a 20-second sliding window with a 15-second overlap.

Encoded scalogram images		Class distribution	
		Schizophrenia (Label 0)	Normal (Label 1)
Training set	9676	5197	4479
Testing set	2420	1283	1137
Total	12,096	6480	5616

Table 3. Data distribution of encoded image dataset.

RepOD EEG dataset ³⁴		Encoded scalogram images
Normal	14	46,854
Schizophrenic	14	42,976
Total	28	89,830

Table 4. Distribution of classes in RepOD EEG dataset³⁴.

Transfer learning leverages the existing knowledge encoded in models trained on the extensive ImageNet dataset, thereby enhancing feature selection for our specific task. Transfer learning frequently leads to enhanced performance, particularly in scenarios involving relatively small datasets like ours, where the risk of overfitting is heightened.

Further features of the images are extracted using the models through Global average pooling. These features are further used for classification using KNN. The results are explained in the further sections.

Results on M.S.U dataset

Our proposed framework underwent a training regimen utilizing two deep learning CNN-based models: EfficientNetB3³⁹ and DenseNet169⁴⁰. The training effectively captures a range of high-level features present in the scalogram images of the M.S.U EEG Dataset³⁰.

Training with EfficientNetB3 A consistent reduction in the loss function values was observed during training. The initial loss at the 1st epoch was recorded at 5.3765, progressively decreasing to a minimum of 0.05125 by the 30th epoch. This indicates effective optimization of the model's parameters through gradient descent. Figure 13. demonstrates the model's loss on training dataset over time. EfficientNetB3's validation performance saw consistent improvement over 30 epochs. Validation loss decreased from 3.65457 to 0.1135. Figure 13 confirms the model's ability to learn from training data and generalize well to unseen validation data, highlighting its strong performance and robust generalization capabilities.

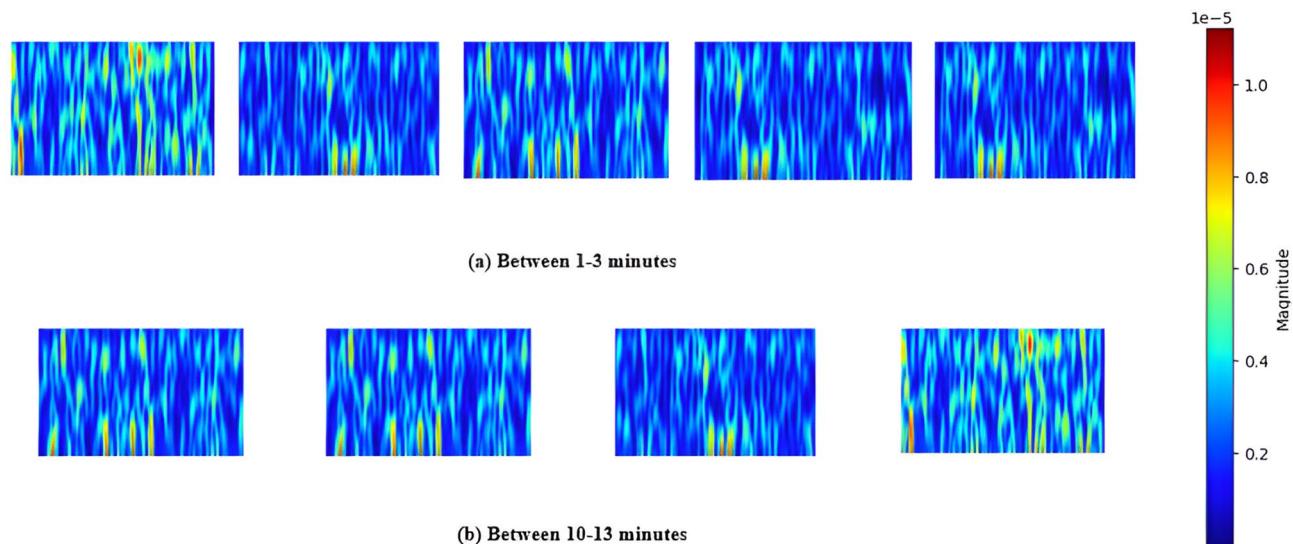


Fig. 11. Sample encoded scalogram images of the C3 channel for a normal adolescent, generated using a 20-s sliding window with a 15-s overlap: (a) Between 1 and 3 min (b) Between 10 and 13 min.

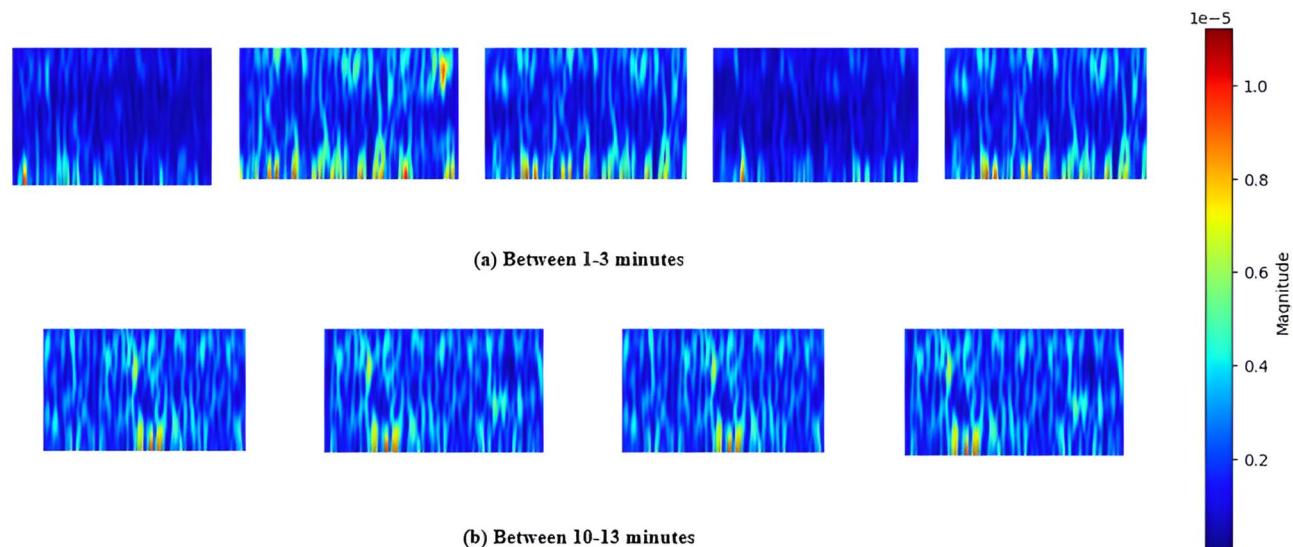


Fig. 12. Sample encoded scalogram images of the C3 channel for a schizophrenic adolescent, generated using a 20-s sliding window with a 15-second overlap: (a) Between 1 and 3 min (b) Between 10 and 13 min.

Encoded scalogram images	Class distribution	
	Schizophrenia (Label 0)	Normal (Label 1)
Training Set	71,864	33,591
Testing Set	17,966	9385
Total	89,830	42,976
		46,854

Table 5. Data distribution of encoded image dataset.

We employed the pre-trained model without its fully-connected top layer to extract comprehensive features from every image in both the training and testing datasets. These features were then processed through a Global Average Pooling 2D layer, yielding a single scalar value per feature map. The resulting feature vector, comprising 1536 features, encapsulates the vital characteristics of each input image and acts as input for our subsequent machine learning tasks.

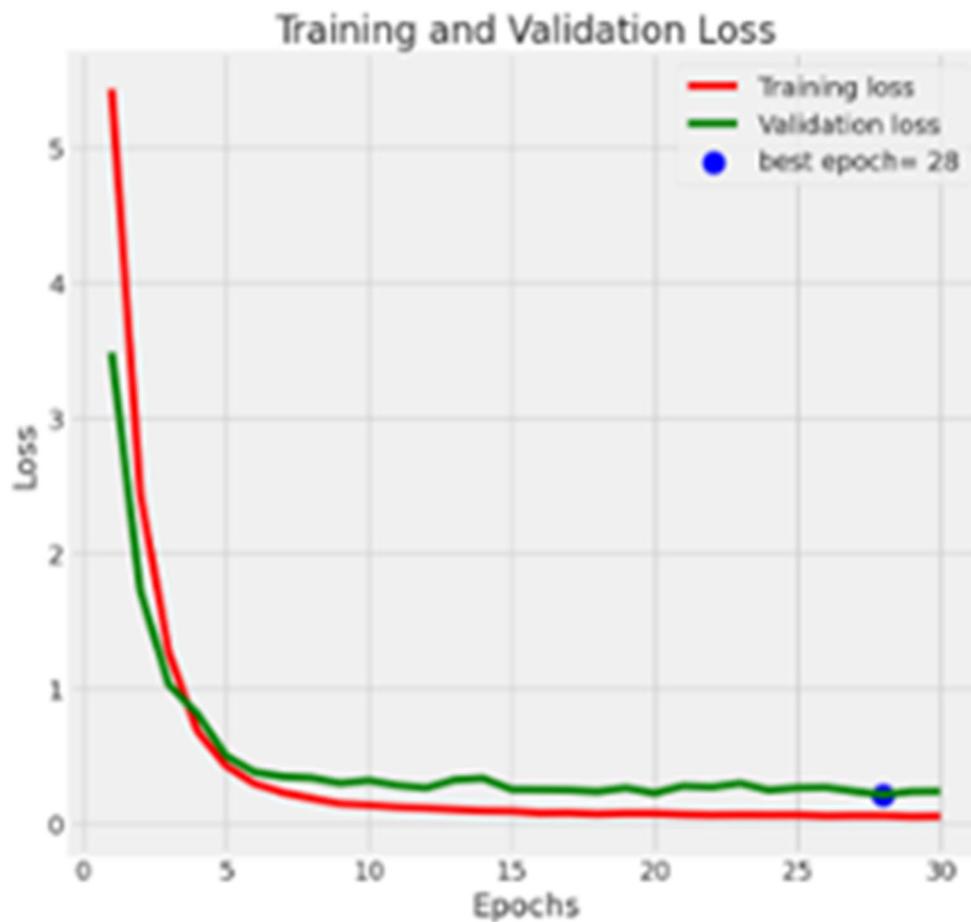


Fig. 13. Graphs plotted for training and validation loss of EfficientNetB3 Model³⁹ training on M.S.U EEG Dataset³⁰.

Training with DenseNet169 During the training of the second deep learning CNN model, DenseNet169⁴⁰, a continuous decrease in the loss function values was observed in Figure. Starting at 3.3245 during the 1st epoch, the loss gradually decreased to a minimum of 0.0424 by the 30th epoch, indicating effective parameter optimization through gradient descent. This demonstrates the model's enhanced predictive precision on the training dataset over time. Additionally, DenseNet169's validation performance exhibited consistent improvement over 30 epochs. Figure 14 depicts that the validation loss decreased from 2.00145 to 0.0924. This tells that the model is robust and has generalization capabilities.

We employed a pre-trained model to extract features from images in both training and testing datasets. These features were processed through a Global Average Pooling 2D layer, resulting in a feature vector of 1664 dimensions, which encapsulates essential image characteristics for further machine learning and deep learning tasks. The performance of the CNN models - EfficientNetB3³⁹ and DenseNet169⁴⁰ using scalogram images was evaluated on the utilized datasets, as summarized in Table 6.

The extracted features from both the models EfficientNetB3³⁹ and DenseNet169⁴⁰, respectively, have been merged into a single feature vector of size 3200. This consolidated feature vector captures diverse information from both models. Further in our proposed framework, we have proposed a new wrapper based feature selection technique to identify the most effective subset of features from the dataset.

Application of ASBO algorithm In our framework, we leverage the KNN⁵³ classifier to select optimal features through wrapper-based feature selection to classify whether a patient is normal or schizophrenic. The K-Nearest Neighbours (KNN) classifier is renowned for its simplicity and effectiveness in classification tasks, particularly in scenarios where data distributions are non-linear or complex. KNN makes predictions based on the majority class among its nearest neighbours.

In our proposed wrapper-based feature selection, we select the optimal features using ASBO algorithm. After global average pooling in EfficientNetB3³⁹, we obtained 1536 features for each scalogram image. Using K nearest neighbour, the accuracy turned out to be 95.1%. From Table 7 it has been observed that our proposed ASBO based wrapper feature selection improves the accuracy by 1 by reducing 28% features. With only 1105 features, the accuracy reached 96.25%.

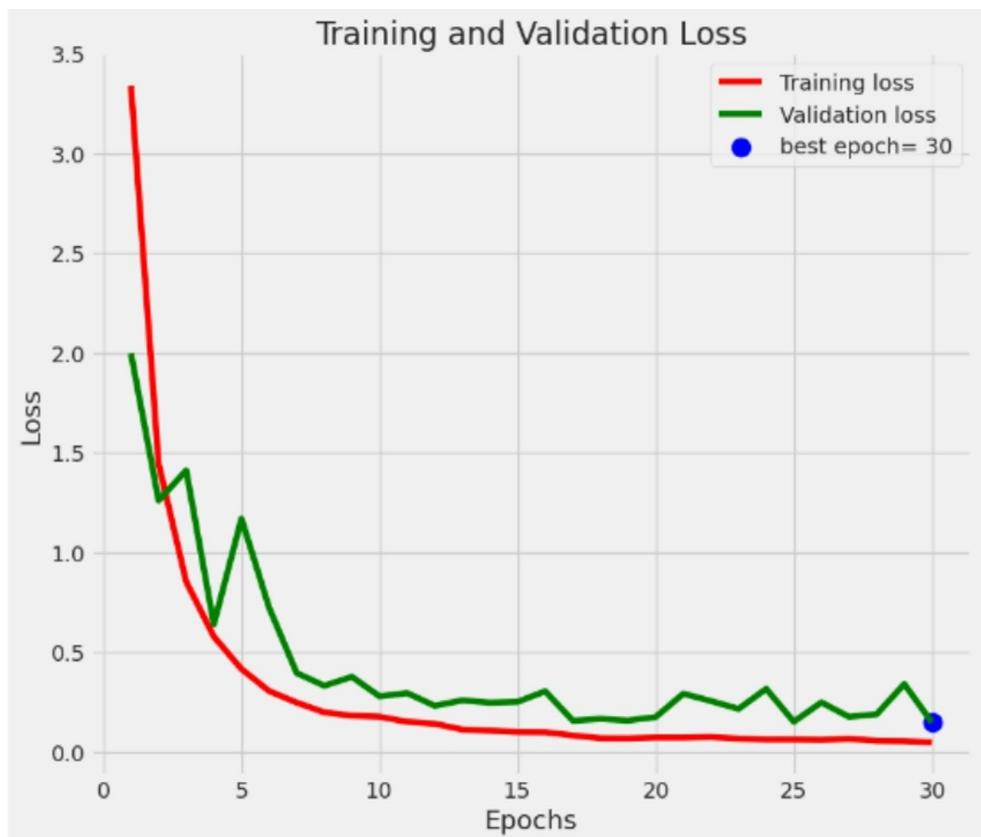


Fig. 14. Graphs plotted for training and validation loss.

Dataset	Evaluation using EfficientNetB3 (%)	Evaluation using DenseNet169 (%)
M.S.U ³⁰	96.38	95.33
RepOD ³⁴	94.26	94.10

Table 6. Evaluation of CNN models for schizophrenia detection using scalogram images, on the utilized datasets.

Dataset	Techniques used	Number of extracted features	Classification using K-NN (%)
			Accuracy
M.S.U EEG dataset ³⁰	Global Average Pooling	1536	95.1
	Proposed ASBO algorithm	1105	96.25

Table 7. Results using proposed ASBO feature selection using EfficientNetB3 model and classification using KNN classifier⁵³ on M.S.U dataset³⁰.

Similar steps have been followed after training with DenseNet169⁴⁰. The results in Table 8 showed that using proposed Average Subtraction wrapper-based feature selection, the result improved by almost 1.5 with reduction of 37% features leading to 1055 features from 1664 features.

Concatenating the features from both the models showed promising results. Using KNN classifiers on the concatenated features, it shows an improvement from using individual models. The result improved by 3.71% and 3.09% from individual EfficientNetB3³⁹ and DenseNet169⁴⁰ models, respectively. Our proposed method evaluated the influence of the K-NN classifier's accuracy based on the number of features chosen using Average Subtraction Wrapper-based feature selection. Our experimental findings revealed that our proposed feature selection algorithm achieved nearly better performance with a classification accuracy of 99.96% by selecting 892 features. This represents a reduction of approximately 72% features from the original concatenated feature set. The convergence graph depicted in Fig. 15 highlights the iterative refinement process of our feature selection algorithm. Across 100 iterations, the search space continuously updates, guiding all candidates toward the best

Dataset	Techniques used	Number of extracted features	Classification using K-NN(%)	
			Accuracy	
M.S.U EEG dataset ³⁰	Global average pooling	1664	95.56	
	Proposed ASBO algorithm	1055	96.87	

Table 8. Results using proposed average Subtraction wrapper-based feature selection using DenseNet169⁴⁰ model and classification using KNN classifier⁵³ on M.S.U dataset³⁰.

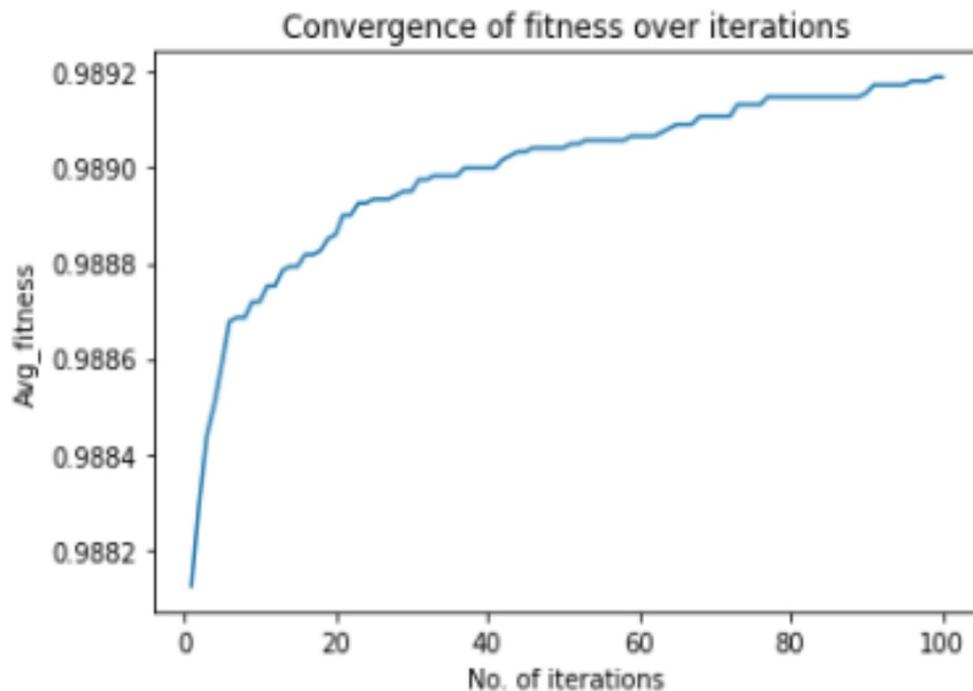


Fig. 15. Convergence graph for proposed Average Subtraction based Wrapper feature selection.

Dataset	Techniques used	Number of extracted features	Classification using K-NN (%)			
			Accuracy	Precision	Recall	F1-score
M.S.U EEG dataset ³⁰	Global average pooling	3200	98.76	98.72	98.85	98.77
	Our proposed framework	892	99.67	99.46	99.92	99.69

Table 9. Results after proposed average Subtraction based wrapper feature selection on concatenated features on M.S.U dataset³⁰.

possible solution. This emphasizes the importance of meticulous feature selection in optimizing the model's performance.

Our proposed Average Subtraction-based wrapper feature selection method has resulted in a 1.2% enhancement in classification accuracy while reducing the number of features by 72%, from 3200 to 892. Table 9 shows that the average precision and recall are 99.67% and 99.46%, respectively, when using the K-NN algorithm. The F1-score achieved with our framework is 99.96%.

After thoroughly analysing the results from the proposed framework on the test data, our results are promising and demonstrate the framework's resilience. Notably, in the Schizophrenic class, it attained a precision of 99.46%, signifying that nearly all Schizophrenic patients were accurately diagnosed. This achievement is remarkable as it minimizes the likelihood of unnecessary treatments for healthy individuals and provides reassurance to those undergoing testing. Based on the data presented in Fig. 16, it is evident that our proposed framework achieved impressive metrics for both the Normal and Schizophrenic classes. Specifically, for the Normal class precision, recall, accuracy, and F1-Score were 99.91%, 99.38%, 99.69%, and 99.64%, respectively. In the Schizophrenic class, these metrics were 99.46%, 99.92%, 99.75%, and 99.69%, respectively. These findings collectively demonstrate the high reliability of our framework in schizophrenia detection, potentially enabling earlier diagnosis and treatment, thereby enhancing patient outcomes.

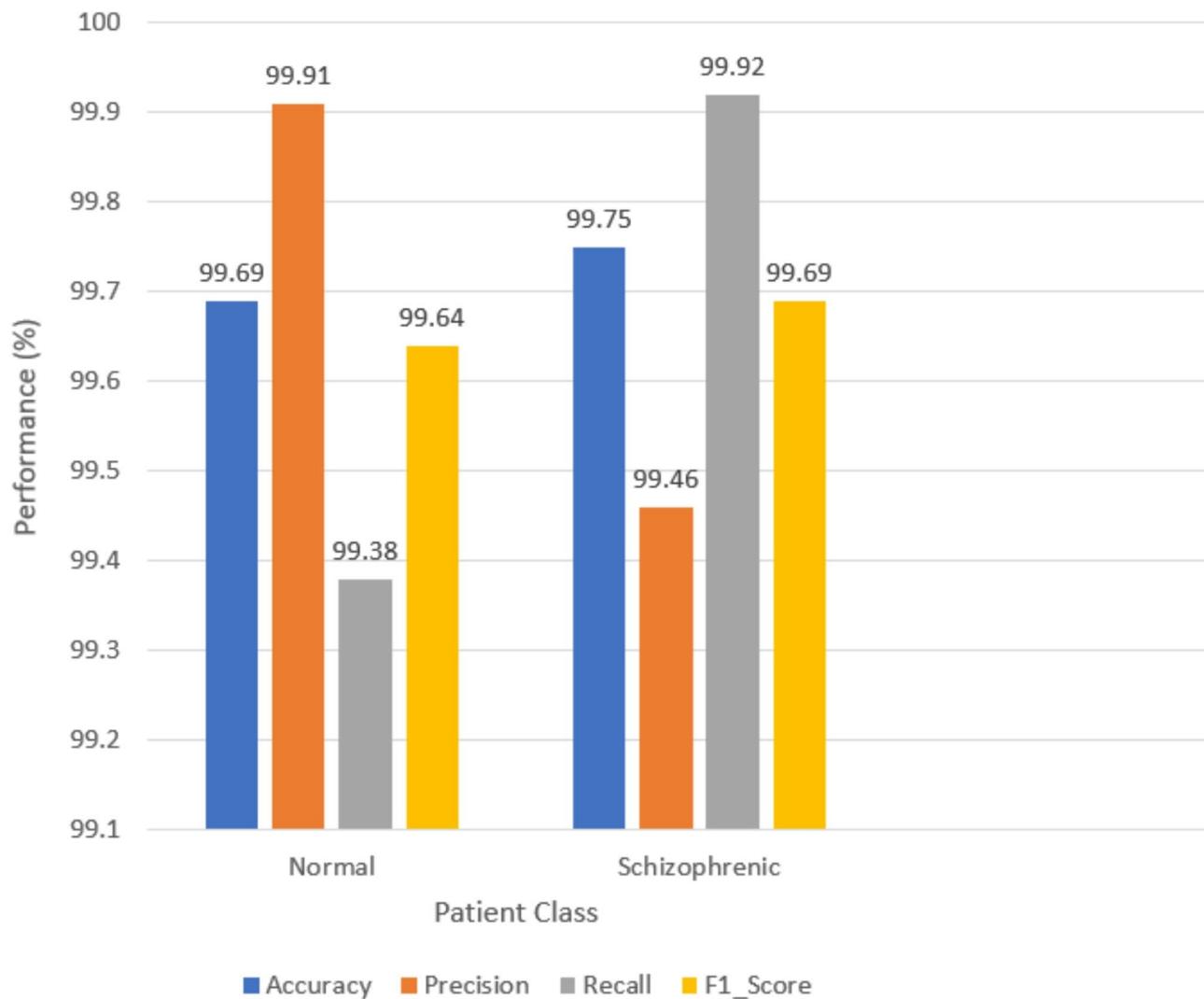


Fig. 16. Performance of our wrapper ASBO based feature selection.

Even with a significant 72% reduction in extracted features, our framework achieved impressive accuracies of 99.75% for correctly identifying the Schizophrenia Class and 99.69% for accurately detecting the Normal class. Figure 17 illustrates the confusion matrix for the classification task, where the colour gradient highlights the distribution of categorizations. It has been observed from the confusion matrix that only 1 schizophrenic patient has been misclassified as normal, whereas 7 normal patients have been misclassified as schizophrenic.

Performance comparison with existing feature selection techniques In Table 10, it is evident that for the M.S.U EEG dataset³⁰, the proposed Average Subtraction wrapper based feature selection achieved the highest classification accuracy while selecting the optimal number of features. It is evident from Table 10 that ASBO beats other optimization algorithms with respect to selecting optimum features with highest accuracy. It can be concluded that ASBO stands out as one of the most suitable wrapper algorithms for the M.S.U EEG dataset³⁰. Our proposed framework outperformed Binary Bat, Whale optimization by almost 1% while reducing nearly 100 features with respect to those.

Our proposed framework is also evaluated with RepOD dataset³⁴. Similar training took place and results are discussed in the following sections.

Results on RepOD dataset

Training with EfficientNetB3 It is observed from Fig. 18 that the loss has reached to almost 0 while training and validation. The model showed the lowest loss in the 29th epoch. The training losses began with 1.72 and reached to almost 0 at the end of 29th epoch. It has been observed from Fig. 18 that there has been a sharp reduction of training loss in the first four epochs. Though validation loss started with 0.48, it has shown similar sharp reduction in the first few epochs and ended reaching almost 0 at the end of 30 epochs.

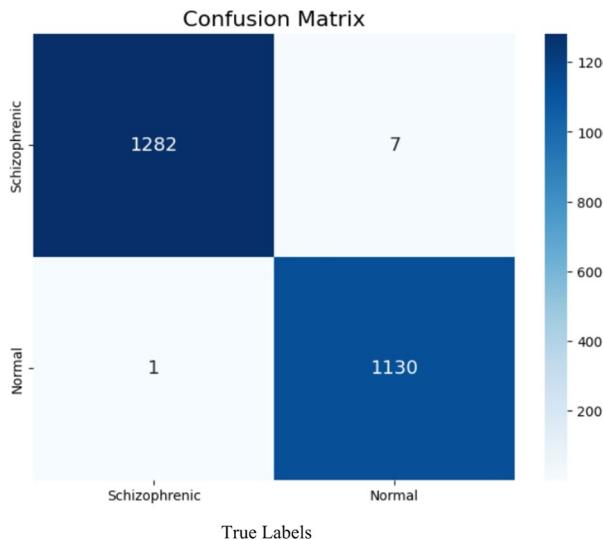


Fig. 17. Confusion matrix given by the proposed framework.

Dataset	Techniques used	Number of extracted features	Classification accuracy using K-NN(%)
M.S.U EEG dataset ³⁰	Original feature set	3200	98.76
	Binary bat algorithm ⁴⁶	966	98.77
	Whale optimization algorithm ⁵⁰	995	98.75
	Gravitational search algorithm ⁵¹	1499	98.85
	Our proposed ASBO algorithm	892	99.67

Table 10. Comparison results with existing feature selection techniques.

The features are extracted before the fully connected layer using Global Average Pooling. The resultant feature vector comprises 1536 features, which is further used for classification using KNN⁵³ classifier to detect schizophrenia.

Training with DenseNet169 The scalogram images are again trained with another deep learning model DenseNet169⁴⁰ to extract the features. A continuous decrease in the loss function values has been observed in Fig. 19. Starting at 1.32 during the 1st epoch, the loss gradually decreased to a minimum of 0.09 by the 30th epoch, indicating effective parameter optimization through gradient descent. This demonstrates the model's enhanced predictive precision on the training dataset over time. Additionally, DenseNet169 validation performance exhibited consistent improvement over 30 epochs. Figure 19 depicts that the validation loss decreased from 0.7 to 0.04 which confirms the model's ability to learn from the training data and generalize effectively to unseen validation data, highlighting its robust performance and generalization capabilities.

We employed a pre-trained model to extract features from images in both training and testing datasets. These features were processed through a Global Average Pooling 2D layer, resulting in a feature vector of 1664 dimensions, which encapsulates essential image characteristics for further machine learning and deep learning tasks. The extracted features from both the models EfficientNetB3³⁹ and DenseNet169⁴⁰, respectively, have been merged into a single feature vector of size 3200. This consolidated feature vector captures diverse information from both models.

Further in our proposed framework, we have proposed an Average Subtraction wrapper-based feature selection technique to identify the most effective subset of features from the dataset.

Average Subtraction wrapper based feature optimization We evaluated our proposed Average Subtraction-based wrapper feature selection technique on different feature sets derived from two models. Initially, we applied the Average Subtraction-based optimization to select optimal features from the feature set obtained using EfficientNetB3³⁹, which provided 1536 features per scalogram image. Utilizing KNN⁵³, we achieved an accuracy of 96.02%. According to Table 11, our proposed wrapper-based feature selection improved the accuracy by 2% while reducing the features by 22%. Consequently, the accuracy increased to 98.25% with this optimal set of features.

We followed a similar approach after training with DenseNet169⁴⁰. The results in Table 12 showed that using proposed wrapper-based feature selection, the result improved by nearly 1% with reduction of 33% features.

Combining features from both models resulted in promising outcomes, with the KNN classifier⁵³ showing a 1% improvement compared to using features from either the EfficientNetB3³⁹ or DenseNet169⁴⁰ models alone.

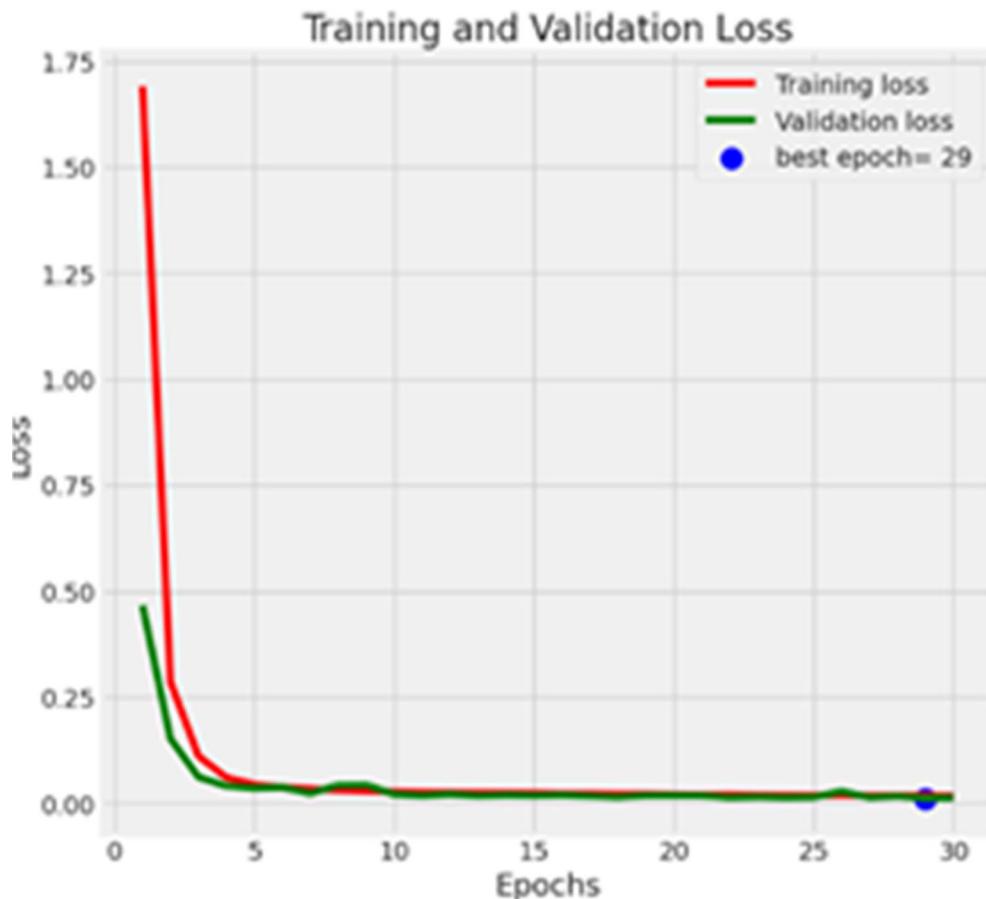


Fig. 18. Graphs plotted for training and validation loss of EfficientNetB3 Model³⁹ training on RepOD Dataset³⁴.

Employing the proposed Average Subtraction Wrapper-based feature selection, our method nearly reached optimal performance, achieving a classification accuracy of 99.97% by selecting 1,016 features, a 69% reduction from the original feature set. The convergence graph in Fig. 20 illustrates the iterative refinement of our feature selection algorithm over 100 iterations, showing continuous updates to the search space and guiding candidates toward the optimal solution. This underscores the vital importance of careful feature selection in optimizing model performance.

As a result, our proposed Average Subtraction wrapper based feature selection method has led to an approximate 1% enhancement in classification accuracy, while reducing the number of features by 69%, from 3200 to 1016. Table 13 demonstrates that the average precision and recall stand at 99.95% and 99.96%, respectively, following classification using the K-NN algorithm⁵³. There has been a 1% increase in precision and recall after reducing irrelevant features through our proposed framework.

After thoroughly evaluating the performance of the proposed framework on the test data, our results are highly promising and demonstrate the framework's robustness. Notably, for the Schizophrenic class, the framework achieved a precision of 99.97%, ensuring that almost all Schizophrenic patients were correctly identified. This exceptional accuracy reduces the risk of unnecessary treatments for healthy individuals and provides reassurance to those undergoing testing. As shown in Fig. 21, our framework delivered impressive metrics for both the Normal and Schizophrenic classes. Specifically, for the Normal class, the precision, recall, accuracy, and F1-score were 99.95%, 99.97%, 99.94%, and 99.95%, respectively. For the Schizophrenic class, these metrics were 99.97%, 99.95%, 99.95%, and 99.96%, respectively. These results collectively highlight the high reliability of our framework in detecting schizophrenia, potentially enabling earlier diagnosis and treatment, thereby improving patient outcomes.

Despite a substantial 69% reduction in extracted features, our framework achieved impressive accuracies: 99.95% for correctly identifying the Schizophrenia class and 99.94% for accurately detecting the Normal class. Figure 22 presents the confusion matrix for the classification task, where the colour gradient represents the distribution of categorizations. A strong diagonal signifies correct classifications, while deviations from this diagonal indicate potential misclassifications. As shown in Fig. 22, the confusion matrix reveals that 5 schizophrenic and 3 normal patients are misclassified by the framework after reducing the original feature set by 2,184 features.

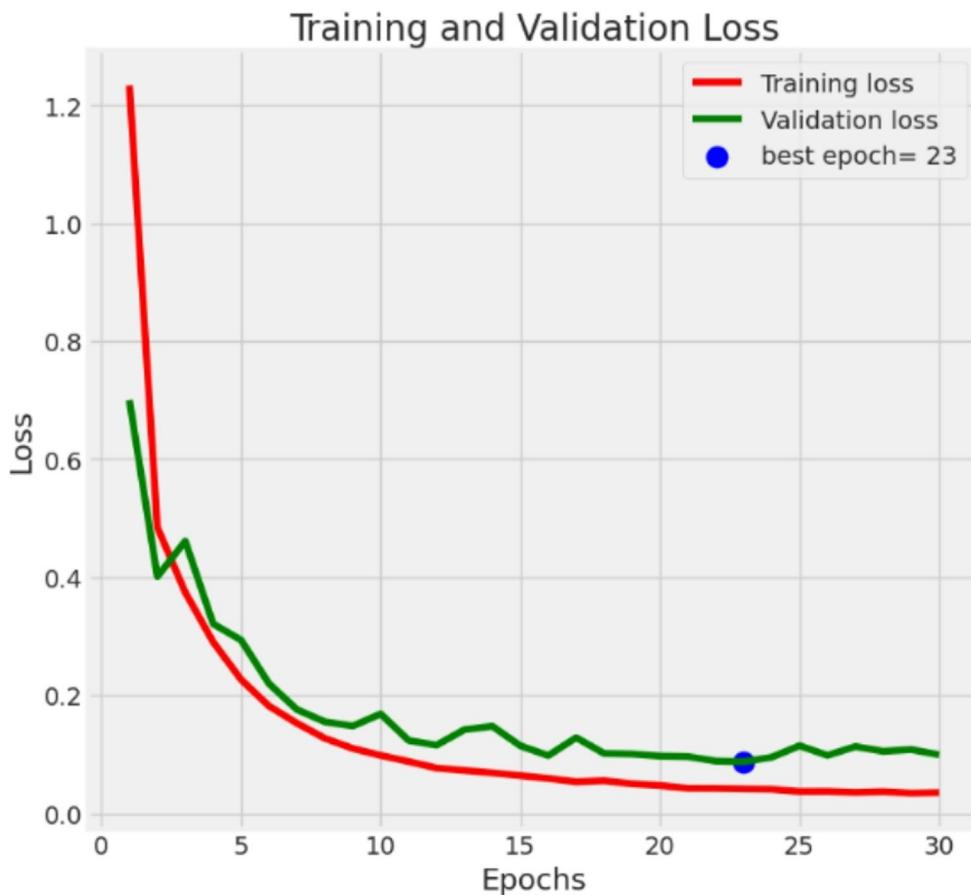


Fig. 19. Graphs plotted for training and validation loss of DenseNet169 model⁴⁰ training on RepOD Dataset³⁴.

Dataset	Techniques used	Number of extracted features	Classification using K-NN (%)	
			Accuracy	
RepOD ³⁴	Global average pooling	1536	96.02	
	Proposed wrapper based feature selection	1209	98.25	

Table 11. Results using proposed average Subtraction wrapper-based feature selection using EfficientNetB3 model³⁹ and classification using KNN⁵³ classifier on RepOD dataset³⁴.

Dataset	Techniques used	Number of extracted features	Classification using K-NN (%)	
			Accuracy	
RepOD ³⁴	Global average pooling	1664	97.46	
	Proposed wrapper based feature selection	1120	98.32	

Table 12. Results using proposed average Subtraction wrapper-based feature selection using DenseNet169 model⁴⁰ and classification using KNN⁵³ classifier on RepOD dataset³⁴.

The experimental findings indicate that although the individual models perform competitively when evaluated in isolation, their standalone capabilities are further enhanced through a synergistic approach involving feature concatenation followed by feature selection. By integrating the feature representations learned from both models, the framework captures a more comprehensive and complementary set of information relevant to schizophrenia classification. The application of a feature selection module plays a crucial role in refining this combined feature space by filtering out redundant or less informative features, thus retaining only the most discriminative patterns essential for accurate classification. This refined feature representation contributes to a significant boost in overall model performance, as reflected by notable improvements across various evaluation

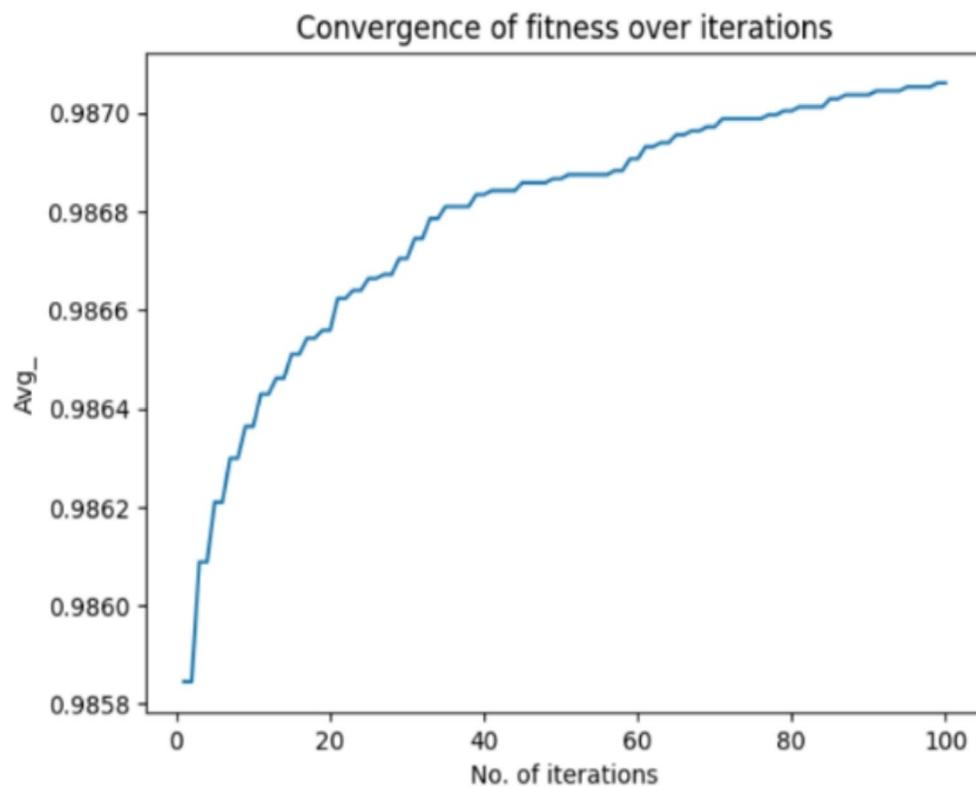


Fig. 20. Convergence graph for wrapper ASBO based feature selection.

Dataset	Techniques used	Number of extracted features	Classification using K-NN (%)			
			Accuracy	Precision	Recall	F1-score
RepOD ³⁴	Global average pooling	3200	98.76	98.72	98.85	98.77
	Our proposed framework	1016	99.97	99.95	99.96	99.955

Table 13. Results after average Subtraction wrapper based feature selection feature selection encoded scalogram images of dataset.

metrics such as accuracy, precision, recall, and F1-score. Moreover, the reduction in both false positive and false negative rates highlights the robustness and reliability of the proposed method in identifying schizophrenic cases, suggesting its potential utility in real-world clinical diagnostic scenarios.

Performance comparison with existing feature selection techniques Table 14 demonstrates that for the EEG dataset, the proposed Average subtraction wrapper-based method achieved the highest classification accuracy while selecting the optimal features. The binary bat algorithm reduced the feature space to 1,046 features, attaining 99.22% accuracy. In contrast, our proposed framework achieved a superior accuracy of 99.97% by retaining only 1,016 features. This highlights an advantageous balance between classification accuracy and the number of selected features, with our method maintaining minimal feature count while achieving exceptional accuracy.

Comparison with existing schizophrenia classification models

The results produced by our proposed framework have been extensively compared with the existing Schizophrenic models for EEG dataset in this section. Most of the approaches tend to develop the model. Our framework outperformed existing models through deep feature selection which not only improved the accuracy but also reduced total processing time.

Comparison results for M.S.U dataset³⁰

On the M.S.U dataset³⁰, C. Phang et al.³¹ achieved a 95% accuracy by integrating Direct Connectivity and Complex Network (DC-CN) features with DNN-DBN, underscoring the significance of complex network structures and advanced machine learning for classification. Rajesh et al.³² reached a 91.66% accuracy using SLBP based histogram features with a LogitBoost Classifier, focusing on texture-based features and leveraging weak classifiers to enhance performance. Sobahi et al.³³ attained a 97.7% accuracy by employing Time-Frequency features with an Extreme Learning Machine (ELM) based Autoencoder (AE), effectively capturing both temporal and spectral

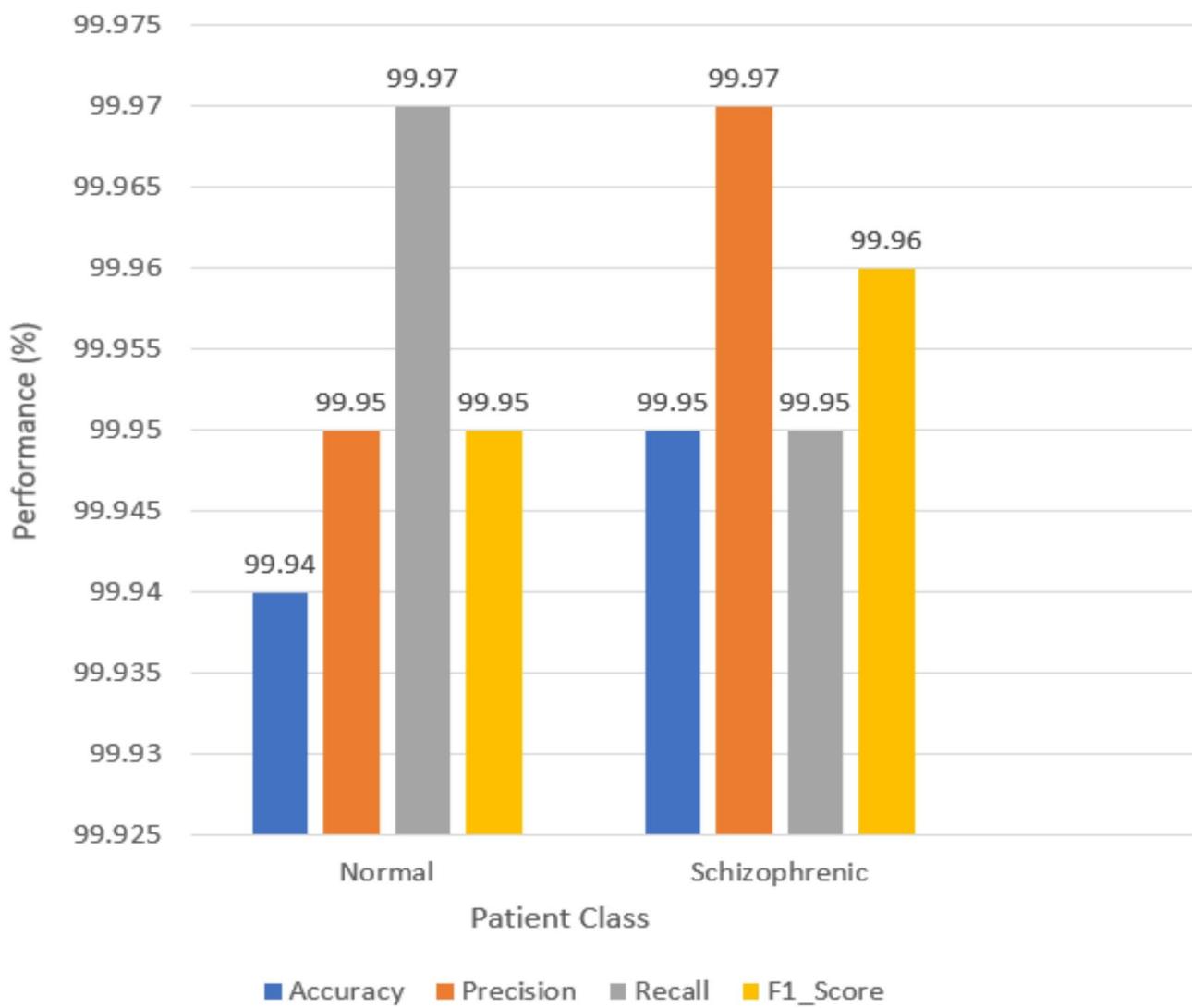


Fig. 21. Performance of our proposed Average Subtraction wrapper-based feature selection.

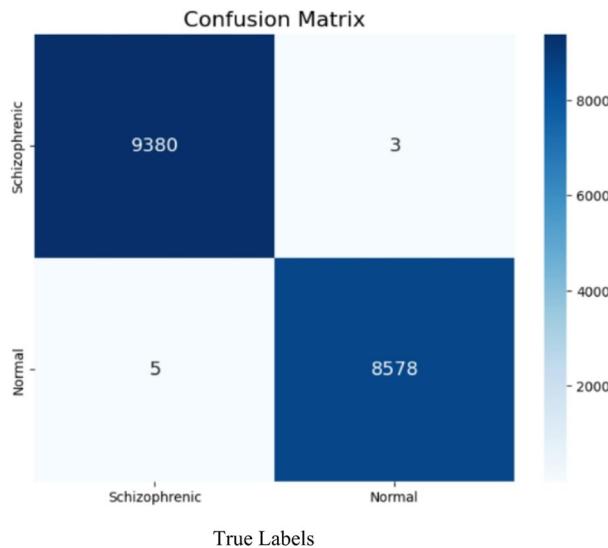


Fig. 22. Confusion matrix given by the proposed framework.

Dataset	Techniques used	Number of selected features	Classification accuracy (%) using KNN
RepOD ³⁴	Concatenated feature set	3200	98.96
	Binary bat algorithm ⁴⁶	1046	99.22
	Whale optimization algorithm ⁵⁰	1495	99.45
	Gravitational search algorithm ⁵⁴	1299	98.92
	Our proposed framework through ASBO	1016	99.97

Table 14. Comparison results with existing feature selection techniques.

References	Method used	Year	Accuracy (%)
C. Phang et al. ³¹	DNN-DBN	2019	95
Rajesh et al. ³²	SLBP	2021	91.66
Sobahi et al. ³³	ELM based AE	2022	97.7
Bagherzadeh et al. ⁵⁵	PDC, dDTF, TE	2024	98.46
Our proposed framework		2025	99.67

Table 15. Comparison of the proposed framework methodology with existing models on the M.S.U. EEG dataset³⁰.

data for class distinction. Compared to the fused effective connectivity approach by Bagherzadeh et al.⁵⁵, which achieved 98.46% accuracy with MSU dataset³⁰, using an ensemble of CNNs on connectivity images derived from Transfer Entropy (TE), Partial Directed Coherence (PDC), and Directed Transfer Function (dDTF), our proposed method offers a simpler yet more effective alternative. By encoding EEG signals as scalogram images via CWT and employing optimized CNN architectures with a novel Average Subtraction wrapper-based feature selection, we achieve a superior accuracy of 99.67%. While their method emphasizes inter-channel interactions and benefits from connectivity fusion and ensemble learning, our approach captures rich time-frequency patterns with lower complexity. This results in a more scalable, interpretable, and generalizable model, demonstrating the power of spectral encoding and targeted feature reduction over complex connectivity-based strategies.

In comparison, our framework demonstrates superior performance of 99.67%, achieving nearly a 2% increase in accuracy. By employing a transfer learning approach, it facilitates efficient feature extraction with minimal training time. Unlike existing methods that primarily aim to refine models for accuracy improvement, our framework emphasizes both maximizing accuracy and reducing features. This dual focus not only boosts classification accuracy but also decreases overall computation time during testing. The comprehensive comparison of our framework with existing approaches on the dataset is detailed in Table 15.

Comparison results for RepOD dataset³⁴

We evaluated our framework on the RepOD dataset³⁴ and recorded outstanding results, achieving state-of-the-art performance in less time with optimal features. Buettner et al.³⁵ proposed a model incorporating ICA, spectral analysis with 99 frequency bands, normalization, and Random Forest classification. While their model achieved 100% accuracy in excluding schizophrenia, the overall accuracy was 71.43%. Krishnan et al.³⁶ utilized Multivariate Empirical Mode Decomposition (MEMD) and entropy measures on EEG data from the RepOD dataset, achieving the highest accuracy of 93% with an SVM-RBF based approach.

Bagherzadeh et al.⁵⁶ employed effective connectivity through TE and used pre-trained CNN-LSTM models for schizophrenia detection, achieving an accuracy of 99.90% by generating 19×19 connectivity matrices from EEG signals. In contrast, our proposed framework adopts a different data encoding strategy by transforming EEG signals into scalograms using CWT, capturing rich temporal and spectral information. This enables more effective feature extraction, leading to a higher classification accuracy of 99.97%. Additionally, our novel wrapper-based feature selection method eliminates redundant features, resulting in a lightweight and efficient model that outperforms existing approaches.

Shoeibi et al.⁵⁷ used Discrete Wavelet Transform (DWT) and dDTF to extract 2D feature maps from EEG signals, which were then classified using pretrained CNN and Transformer models. The ConvNeXt-Tiny model achieved the highest accuracy of 96% in the beta sub-band. In comparison, our framework achieves a significantly higher accuracy of 99.97% on the RepOD dataset³⁴ by transforming EEG signals into scalogram images using CWT. Deep features extracted via EfficientNetB3³⁹ and DenseNet169⁴⁰, combined with our novel Average Subtraction wrapper-based feature selection (reducing dimensionality by 72%), enhance both accuracy and efficiency. Unlike connectivity-based methods, our image-based approach is more scalable, computationally efficient, and clinically interpretable.

Also, Bagherzadeh et al.⁵⁵ achieved 99.51% accuracy with RepOD dataset³⁴, using an ensemble of CNNs on connectivity images derived from TE, PDC, and dDTF. In contrast, our method converts EEG signals into scalograms using CWT, enabling rich time-frequency representation. With optimized CNN architectures and a novel Average Subtraction wrapper-based feature selection, we achieve a higher accuracy of 99.97%. Our approach avoids complex connectivity fusion, offering better scalability, interpretability, and efficiency.

References	Method used	Year	Accuracy (%)
Buettner et al. ³⁵	ICA-random forest	2019	71.43
Krishnan et al. ³⁶	MEMD + SVM RBF	2020	93
Bagherzadeh et al. ⁵⁶	TE, Inception	2022	99.90
Shoeibi et al. ⁵⁷	DWT, dDTF, ConvNeXt-Tiny	2024	96
Bagherzadeh et al. ⁵⁵	PDC, dDTF, TE	2024	99.51
Sara et al. ³⁷	EfficientNetB0-LSTM	2022	99.90
Our Proposed Framework		2025	99.97

Table 16. Comparison of the proposed framework methodology with existing models on the RepOD EEG dataset³⁴.

Sara et al.³⁷ identified schizophrenia using an EfficientNetB0-LSTM model, attaining an average accuracy of 99.90%. In comparison, our framework scored 28% and 6% higher than Buettner et al.³⁵ and Krishnan et al.'s³⁶ approach, respectively, as detailed in Table 15. While Sara et al.'s³⁷ model achieved 99.9% accuracy, it is computationally heavy due to the combination of EfficientNetB0 and LSTM, which involves a large number of parameters. Our proposed framework, on the other hand, is lightweight and less parameter-intensive, resulting in faster processing times. These comparisons illustrate that our framework achieves similar or superior accuracy to state-of-the-art models while maintaining computational efficiency. Detailed comparisons with existing approaches on the dataset³⁴ are presented in Table 16.

Our framework achieves comparable accuracy to existing models on the mentioned datasets by optimizing the feature space with minimal training time. Existing models often rely on deep learning to enhance accuracy, which requires significant training time and large datasets to prevent overfitting⁵⁸. In contrast, our approach leverages transfer learning to reduce training time and extract important features. Additionally, we proposed an average subtraction wrapper-based feature selection method to eliminate irrelevant features, further improving accuracy. By incorporating a wrapper-based feature optimization step after merging features from both models, our proposed framework not only enhances classification accuracy but also reduces irrelevant features and computational time. Unlike other methods that primarily focus on accuracy improvement, this framework balances maximizing accuracy and reducing feature count, thus boosting classification performance while cutting down on computation time.

As a future direction, we recognize the importance of visualizing the feature space after applying feature selection techniques to better understand the data distribution and the separability between classes. Techniques such as t-SNE and UMAP⁴² can offer valuable insights into how the selected features contribute to the clustering of schizophrenic and healthy subjects in the reduced-dimensional space. Furthermore, interpretability analysis is essential for enhancing the clinical relevance of the proposed model. Future work will focus on identifying significant brain regions and frequency bands that contribute most to classification decisions, using techniques such as saliency maps, Gradient-weighted Class Activation Mapping (Grad-CAM), or SHapley Additive exPlanations (SHAP). This will help in understanding the neurobiological underpinnings of schizophrenia, in line with the approaches demonstrated in recent studies^{59,60}. Integrating such interpretability frameworks will provide transparency and support the adoption of AI-based tools in real-world clinical settings.

Conclusion and future scope

Our research underscores the potential of transfer learning models in effectively classifying schizophrenia from raw EEG data. Our findings demonstrate that pre-trained models, when fine-tuned appropriately, can successfully identify complex patterns within encoded scalogram images, accurately distinguishing between healthy individuals and those with schizophrenia. Although further improvement of the model is necessary for optimal results, our approach shows that eliminating redundant features can also enhance accuracy. Our proposed approach has been implemented and tested on two standard benchmark datasets namely, M.S.U. EEG dataset and RepOD dataset. The classification accuracies of 99.67% and 99.97%, respectively, have been achieved while utilizing just 1/3 of the original feature set.

Previous methods mainly focused on applying classical machine learning and deep learning models to raw EEG data. The models were heavy due to the large number of parameters. Our proposed approach is lightweight and can be deployed in lightweight devices for inferencing results. For the first time, the scalogram approach is used for schizophrenia detection. Our approach encodes raw EEG data into scalogram images, preserving both spatial and temporal information. Transforming EEG data into images for deep learning presents a promising avenue for advancing brain signal analysis. This technique bridges the fields of neuroscience and computer vision, offering great potential for improving classification accuracy, enabling multimodal integration, and streamlining analysis pipelines in various applications, including healthcare and neuroscience research.

These images are then used to extract relevant features through deep learning models. Our proposed feature selection framework has shown that transfer learning, combined with training the model for a few epochs to extract features, can significantly reduce training time. Moreover, reducing less important features after this step enhances the framework's efficiency compared to using the model alone. Our study indicates that the proposed feature-based reduction techniques significantly outperform previous approaches. This novel approach accelerates classification accuracy by optimizing both efficiency and time. Most of the previous models were

specific to the dataset. Our approach showed overwhelming results for multiple datasets. This shows that our framework has the generalization ability.

Further our framework can be explored to solve similar aligned problems. Future research will focus on extending this study to create an end-to-end framework for various objectives such as segmentation and detection. The framework can be further explored with different models like a combination of CNN and RNN based. Additionally, we plan to explore filter-based and embedded feature reduction techniques to further improve results. Also, more extensive comparisons of the ASBO-based wrapper method will be conducted with a broader range of wrapper methods, including biologically and nature-inspired algorithms. Additionally, a deeper analysis will focus on its convergence behaviour across diverse objective landscapes, as well as explore potential biological analogues, such as group-based decision-making processes observed in nature.

The implications of this research are substantial, potentially leading to more accurate and timely diagnosis, a deeper understanding of the neurobiological underpinnings of schizophrenia, and the development of personalized treatment strategies, ultimately improving the quality of life for individuals affected by schizophrenia.

Data availability

Publicly available datasets are used in this study can be downloaded from below links:1. M.S.U. EEG dataset³⁰: http://brain.bio.msu.ru/eeg_schizophrenia.htm2. RepOD dataset [34]: <https://www.kaggle.com/datasets/srinivasapanchavedi/eeg-data-for-schizophrenia>.

Received: 30 March 2025; Accepted: 6 June 2025

Published online: 01 July 2025

References

1. Luvsannyam, E. et al. Neurobiology of schizophrenia: a comprehensive review. *Cureus*, **14**(4). (2022).
2. Dabiri, M. et al. Neuroimaging in schizophrenia: A review Article. *Front. NeuroSci.* **16**, 1042814 (2022).
3. Aslan, Z. & Akin, M. A deep learning approach in automated detection of schizophrenia using scalogram images of EEG signals. *Phys. Eng. Sci. Med.* **45** (1), 83–96 (2022).
4. Teixeira, F. L. et al. A narrative review of speech and EEG features for schizophrenia detection: progress and challenges. *Bioengineering* **10** (4), 493 (2023).
5. Schizophrenia, [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/schizophrenia>
6. Schizophrenia [Online]. Available: <https://www.nimh.nih.gov/health/statistics/schizophrenia>
7. EEG Test (Electroencephalogram). Purpose, Procedure, And Risks, [Online]. Available: <https://www.simplypsychology.org/what-is-an-eeg.html>
8. Ranjan, R., Sahana, B. C. & Bhandari, A. K. Deep learning models for diagnosis of schizophrenia using EEG signals: emerging trends, challenges, and prospects. *Arch. Comput. Methods Eng.* **31** (4), 2345–2384 (2024).
9. Shoeibi, A. et al. Automatic diagnosis of schizophrenia in EEG signals using CNN-LSTM models. *Front. Neuroinformatics*. **15**, 777977 (2021).
10. Shameedha Begum, B., Faruk Hossain, M., Jose, J. & Krishnapriya, B. EEG-Based Identification of Schizophrenia Using Deep Learning Techniques. In *International Conference on Computation Intelligence and Network Systems* (pp. 26–37). Cham: Springer Nature Switzerland. (2023), October.
11. Gashkarimov, V., Sultanova, R., Efremov, I. & Asadullin, A. Machine learning techniques in diagnostics and prediction of the clinical features of schizophrenia: A narrative review. *Consortium Psychiatricum*, **4** (3 (eng)), 43–53 (2023).
12. Medalia, A., Saperstein, A. M., Hansen, M. C. & Lee, S. Personalised treatment for cognitive dysfunction in individuals with schizophrenia spectrum disorders. *Neuropsychological Rehabilitation*. **28** (4), 602–613 (2018).
13. Sharma, G. & Joshi, A. M. SzHNN: A novel and scalable deep Convolution hybrid neural network framework for schizophrenia detection using multichannel EEG. *IEEE Trans. Instrum. Meas.* **71**, 1–9 (2022).
14. Verma, S. et al. Machine learning techniques for the schizophrenia diagnosis: A comprehensive review and future research directions. *J. Ambient Intell. Humaniz. Comput.* **14** (5), 4795–4807 (2023).
15. Sun, J. et al. A hybrid deep neural network for classification of schizophrenia using EEG data. *Sci. Rep.* **11** (1), 4706 (2021).
16. Mandal, M., Singh, P. K., Ijaz, M. F., Shafi, J. & Sarkar, R. A tri-stage wrapper-filter feature selection framework for disease classification. *Sensors* **21** (16), 5571 (2021).
17. Bhattacharyya, T. et al. Mayfly in harmony: A new hybrid meta-heuristic feature selection algorithm. *IEEE Access*. **8**, 195929–195945 (2020).
18. Ghosh, K. K., Ahmed, S., Singh, P. K., Geem, Z. W. & Sarkar, R. Improved binary sailfish optimizer based on adaptive β -hill climbing for feature selection. *IEEE Access*. **8**, 83548–83560 (2020).
19. Theng, D. & Bhoyer, K. K. Feature selection techniques for machine learning: a survey of more than two decades of research. *Knowl. Inf. Syst.* **66** (3), 1575–1637 (2024).
20. Pudjhartono, N., Fadason, T., Kempa-Liehr, A. W. & O'Sullivan, J. M. A review of feature selection methods for machine learning-based disease risk prediction. *Front. Bioinf.* **2**, 927312 (2022).
21. Solorio-Fernández, S., Carrasco-Ochoa, J. A. & Martínez-Trinidad, J. F. A survey on feature selection methods for mixed data. *Artificial Intell. Review*, 1–26. (2022).
22. Tadist, K., Najah, S., Nikolov, N. S., Mrabti, F. & Zahri, A. Feature selection methods and genomic big data: a systematic review. *J. Big Data* **6** (1), 1–24 (2019).
23. Chattopadhyay, S., Singh, P. K., Ijaz, M. F., Kim, S. & Sarkar, R. SnapEnsemFS: a snapshot ensembling-based deep feature selection model for colorectal cancer histological analysis. *Sci. Rep.* **13** (1), 9937 (2023).
24. Sahoo, K. K. et al. Wrapper-based deep feature optimization for activity recognition in the wearable sensor networks of healthcare systems. *Sci. Rep.* **13** (1), 965 (2023).
25. Marik, A., Chattopadhyay, S. & Singh, P. K. A hybrid deep feature selection framework for emotion recognition from human speeches. *Multimedia Tools Appl.* **82** (8), 11461–11487 (2023).
26. Bhattacharya, D., Sharma, D., Kim, W., Ijaz, M. F. & Singh, P. K. Ensem-HAR: an ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring. *Biosensors* **12** (6), 393 (2022).
27. Das, A. et al. A hybrid meta-heuristic feature selection method for identification of Indian spoken languages from audio signals. *IEEE Access*. **8**, 181432–181449 (2020).
28. Sheikh, K. H. et al. EHHM: electrical harmony based hybrid meta-heuristic for feature selection. *IEEE Access*. **8**, 158125–158141 (2020).
29. SahSahoo, K. K. et al. Wrapper-based deep feature optimization for activity recognition in the wearable sensor networks of healthcare systems. *Sci. Rep.* **13** (1), 965 (2023).

30. Gorbachevskaya, N. N. & Borisov, S. V. EEG of healthy adolescents and adolescents with symptoms of schizophrenia, M.V. Lomonosov Moscow State University, [Online]. Available: http://brain.bio.msu.ru/eeg_schizophrenia.htm
31. Phang, C. R., Ting, C. M., Samdin, S. B. & Ombao, H. Classification of EEG-based effective brain connectivity in schizophrenia using deep neural networks. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 401–406). IEEE. (2019), March.
32. Rajesh, K. N. & Kumar, T. S. Schizophrenia detection in adolescents from EEG signals using symmetrically weighted local binary patterns. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 963–966). IEEE. (2021), November.
33. Sobahi, N., Ari, B., Cakar, H., Alcin, O. F. & Sengur, A. A new signal to image mapping procedure and convolutional neural networks for efficient schizophrenia detection in EEG recordings. *IEEE Sens. J.* **22** (8), 7913–7919 (2022).
34. RepOD Dataset. <https://repod.icm.edu.pl/dataset.xhtml?persistentId=doi:10.18150/repod.0107441>, [Online].
35. Buettner, R. et al. High-performance exclusion of schizophrenia using a novel machine learning method on EEG data. In *2019 IEEE International conference on E-health networking, application & services (HealthCom)* (pp. 1–6). IEEE. (2019), October.
36. Krishnan, P. T., Raj, A. N. J., Balasubramanian, P. & Chen, Y. Schizophrenia detection using multivariateempirical mode decomposition and entropy measures from multichannel EEG signal. *Biocybernetics Biomedical Eng.* **40** (3), 1124–1139 (2020).
37. Bagherzadeh, S., Shahabi, M. S. & Shalbaf, A. Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal. *Comput. Biol. Med.* **146**, 105570 (2022).
38. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, **115**, 211–252.
39. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105–6114). PMLR. (2019), May.
40. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708). (2017).
41. Spkit <https://pypi.org/project/spkit/0.0.9.3/>, [Online].
42. Shahriari, B., Swersky, K., Wang, Z., Adams, R. & De Freitas, N. Taking the human out of the loop: A review of bayesian optimization. *Proc. IEEE.* **104** (1), 148–175 (2015).
43. Ke, H. et al. Cloud-aided online EEG classification system for brain healthcare: A case study of depression evaluation with a lightweight CNN. *Software: Pract. Experience.* **50** (5), 596–610 (2020).
44. Ke, H. et al. Improving brain E-health services via high-performance EEG classification with grouping bayesian optimization. *IEEE Trans. Serv. Comput.* **13** (4), 696–708 (2019).
45. Man, K. F., Tang, K. S. & Kwong, S. Genetic algorithms: concepts and applications [in engineering design]. *IEEE Trans. Industr. Electron.* **43** (5), 519–534 (1996).
46. Kennedy, J. & Eberhart, R. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4, pp. 1942–1948). ieee. (1995), November.
47. Dorigo, M., Birattari, M. & Stutzle, T. Ant colony optimization. *IEEE Comput. Intell. Mag.* **1** (4), 28–39 (2007).
48. Gandomi, A. H., Yang, X. S. & Alavi, A. H. Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. *Eng. Comput.* **29**, 17–35 (2013).
49. Nakamura, R. Y. et al. BBA: a binary bat algorithm for feature selection. In *2012 25th SIBGRAPI conference on graphics, patterns and images* (pp. 291–297). IEEE. (2012), August.
50. Mirjalili, S. & Lewis, A. The Whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016).
51. Dehghani, M., Hubálovský, Š. & Trojovský, P. A new optimization algorithm based on average and Subtraction of the best and worst members of the population for solving various optimization problems. *PeerJ Comput. Sci.* **8**, e910 (2022).
52. Wang, F., Ke, H. & Tang, Y. Fusion of generative adversarial networks and non-negative tensor decomposition for depression fMRI data analysis. *Inf. Process. Manag.* **62** (2), 103961 (2025).
53. K nearest Neighbour. <https://www.jstor.org/stable/1403796>, [Online].
54. Rashedi, E., Nezamabadi-Pour, H. & Saryazdi, S. GSA: a gravitational search algorithm. *Inf. Sci.* **179** (13), 2232–2248 (2009).
55. Bagherzadeh, S. & Shalbaf, A. EEG-based schizophrenia detection using fusion of effective connectivity maps and convolutional neural networks with transfer learning. *Cognitive Neurodynamics*, 1–12. (2024).
56. Bagherzadeh, S., Maghsoudi, A. & Shalbaf, A. Detection of schizophrenia patients using convolutional neural networks from brain effective connectivity maps of electroencephalogram signals. *Tehran Univ. Med. Sci. J.* **79** (10), 754–763 (2022).
57. Shoeibi, A. et al. Diagnosis of Schizophrenia in EEG Signals Using dDTF Effective Connectivity and New PreTrained CNN and Transformer Models. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 150–160). Cham: Springer Nature Switzerland. (2024), May.
58. Saha, A., Park, S., Geem, Z. W. & Singh, P. K. Schizophrenia detection and classification: A systematic review of the last decade. *Diagnostics* **14** (23), 2698 (2024).
59. Ke, H., Wang, F., Ma, H. & He, Z. ADHD identification and its interpretation of functional connectivity using deep self-attention factorization. *Knowl. Based Syst.* **250**, 109082 (2022).
60. Ke, H., Chen, D., Yao, Q., Tang, Y., Wu, J., Monaghan, J., ... McAlpine, D. (2023). Deep factor learning for accurate brain neuroimaging data analysis on discrimination for structural MRI and functional MRI. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Author contributions

Conceptualization, U.A., A.S., M.W., M.F.I., P.W.K.,; methodology, M.W. M.F.I.; software, U.A., A.S., P.K.S.,; validation, A.H., A.D., S.G., M.W., M.W., M.F.I., , P.K.S., ; formal analysis, U.A., A.S., M.W., P.K.S., ; investigation, M.W., M.F.I.; resources, M.W., M.F.I., data curation, U.A., A.S.,,P.K.S., writing—original draft preparation, U.A., A.S.,P.K.S., ; writing—review and editing, , M.W., M.F.I.; visualization, U.A., A.S.,, supervision M.W., and M.F.I., P.K.S ; project administration, P.K.S.,; M.W.,; and M.F.I ; funding acquisition, M.W., M.F.I.; All authors have read and agreed to the published version of the manuscript.

Funding

Authors would like to acknowledge contribution to this research from the Rector of the Silesian University of Technology, Gliwice, Poland under proquality grant no. 09/020/RGJ25/0041.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.W. or M.F.I.

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

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

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

© The Author(s) 2025