Erciyes University Journal of Institue Of Science and Technology

Research Article
10.65520/erciyesfen.1888892

Imprint:
Volume: 42(2)
Year: 2026
Page: 553-580

Ozlem KARABIBER CURA[a*]

[a] Asst. Prof., Izmir Katip Celebi University,
ozlem.karabiber@ikcu.edu.tr

[*] Corresponding Author

Received: 2/13/2026
Accepted: 6/10/2026

Citation:
Ozlem KARABIBER CURA (2026). Automated Schizophrenia Detection via Resnet50-based Ensemble Learning Using Multi-entropy EEG Feature Maps. *Erciyes University Journal of Institue Of Science and Technology*, 42(2), 553-580. https://doi.org/10.65520/erciyesfen.*1888892*

Automated Schizophrenia Detection via Resnet50-based Ensemble Learning Using Multi-entropy EEG Feature Maps

Abstract

Background/Aim: A complicated neuropsychiatric condition, schizophrenia is commonly investigated using one-dimensional EEG signals. In this work, a novel deep learning framework for automated schizophrenia detection is proposed, integrating an ensemble model based on ResNet50 with multi-entropy-based feature maps derived from 10-second EEG segments.

Methods: Six entropy measures—approximate, bubble, fuzzy, permutation, sample, and spectral entropy—are computed to create two-dimensional EEG feature maps (EEG-FMs). A distinct ResNet50 model processes each EEG-FM, and soft voting is utilized for final predictions. The framework's effectiveness is evaluated using a leave-one-group-out cross-validation (LOGO-CV) method on two EEG datasets.

Results: The ensemble framework exhibits high sensitivity and specificity, achieving average accuracies of 81.4% on Dataset 1 and 95.7% on Dataset 2. It evaluates the framework's performance across varied EEG distributions using a combined dataset and the Wilcoxon signed-rank test. The soft voting generally yields more balanced results than individual entropy-based models, although performance declines with increased dataset variability. An independent third EEG dataset (Dataset 3) is used to further assess the framework under differing conditions.

Conclusion: The integration of entropy-based EEG features into the deep learning ensemble improves schizophrenia detection.

Keywords: Schizophrenia, EEG, Entropy-based Feature Maps, Soft Voting, Leave-one-group-out cross-validation

Çoklu Entropi EEG Özellik Haritaları Kullanılarak Resnet50 Tabanlı Topluluk Öğrenimi ile Otomatik Şizofreni Tespiti

Öz

Karmaşık bir nöropsikiyatrik durum olan şizofreni, genellikle tek boyutlu EEG sinyalleri kullanılarak incelenir. Bu çalışmada, 10 saniyelik EEG segmentlerinden üretilen çoklu entropi tabanlı özellik haritalarıyla ResNet50 tabanlı bir topluluk modelini entegre eden otomatik şizofreni tespiti için yeni bir derin öğrenme çerçevesi önerilmiştir. İki boyutlu EEG özellik haritaları (EEG-FM) oluşturmak için altı entropi ölçüsü (yaklaşık, kabarcık, bulanık, permütasyon, örnek ve spektral entropi) hesaplanmıştır. Her EEG-FM ayrı bir ResNet50 modeli tarafından işlenmiş ve nihai tahminler için yumuşak oylama kullanılmıştır. Çerçevenin etkinliği, iki EEG veri kümesi üzerinde bir grup dışarı bırakma çapraz doğrulama (LOGO-CV) yöntemi kullanılarak değerlendirilmiştir. Topluluk çerçevesi, Veri Kümesi 1'de %81,4 ve Veri Kümesi 2'de %95,7 ortalama doğruluk oranlarına ulaşarak yüksek duyarlılık ve özgüllük sergilemiştir. Çerçevenin performansı, birleşik veri kümesi ve

Wilcoxon işaretli sıralama testi kullanılarak çeşitli EEG dağılımları üzerinde değerlendirilmiştir. Performans veri kümesi değişkenliğinin artmasıyla düşse de, yumuşak oylama genel olarak bireysel entropi tabanlı modellerden daha dengeli sonuçlar vermiştir. Çerçeveyi farklı koşullar altında daha ayrıntılı olarak değerlendirmek için bağımsız bir üçüncü EEG veri kümesi (Veri Kümesi 3) kullanılmıştır. Entropi tabanlı EEG özelliklerinin derin öğrenme topluluğuna entegrasyonu, şizofreni tespitini iyileştirmiştir.

Anahtar kelimeler: Şizofreni, EEG, Entropi tabanlı Özellik Haritaları, Yumuşak Oylama, Tek grup dışarıda bırakma çapraz doğrulama

1. INTRODUCTION

Schizophrenia (SCH) is a harmful brain disorder that is frequently linked to sociopsychological issues, anxiety, and sadness. The World Health Organization (WHO) estimates that 24 million people worldwide suffer from schizophrenia, which is equivalent to an overall 0.32% and adults 0.45%. It usually starts in late adolescence or early adulthood and is one of the main causes of disability in the world. Compared to the general population, those with schizophrenia have a two to three times higher chance of dying young. Unfortunately, 69% of people with psychosis do not have access to quality mental health care, especially in low- and middle-income countries. Schizophrenia develops in four stages. Stage I includes children under twelve who are at risk, Stage II includes the beginning of psychosis between the ages of twelve and eighteen, Stage III includes the onset of psychosis between the ages of eighteen and twenty-four, and Stage IV is characterized by chronic impairment [1]. Delusions, positive and negative emotional behavior, and cognitive impairment symptoms are all present in patients with this illness [2]. Families and healthcare systems are strained by long-term, expensive medicine, and mortality is raised by misdiagnosis and treatment delays. It is necessary to have a reliable, reasonably priced, and automated detection system [3, 4]. Innovations in electrophysiological data processing and neuroimaging allow for a network-based investigation of altered brain dynamics, but the neurophysiological mechanisms underlying schizophrenia are still largely unknown. The non-invasive method noted for its temporal accuracy and affordability, electroencephalography (EEG), is very helpful in examining brain activity in schizophrenia and other mental illnesses [5].

There are two main types of research on the distinct complexity of EEG data for SCH diagnosis and detection: those that use deep learning (DL) networks and those that use conventional machine learning (ML). To extract information from EEG recordings and provide new insights for neuroscientific research, ML approaches are essential [6]. A low-complexity bi-orthogonal wavelet filter-bank (LCBWFB) that uses canonical-signed-digit filters to reduce computations is presented in a study. In order to achieve sharp roll-off and provide rational coefficients for near-orthogonality, regularity, and faultless reconstruction, it uses a generalized matrix formulation for filter design. By breaking down EEG signals, employing the Fisher score for channel selection, and utilizing a least square support vector machine improved by the Grey Wolf algorithm, the LCBWFB is used for automatic schizophrenia detection. The findings indicate 95.95% sensitivity, 96.97% specificity, and 96.84% 10-fold accuracy. The leave-one-subject-out scores are 93.92% sensitivity, 92.30% specificity, and 93.77% accuracy for an open-source dataset containing 25 out of 1408 features, showing equivalence to current schizophrenia detection techniques [3]. A hybrid metaheuristic algorithm known as Dwarf Mongoose Optimization and Artificial Ecosystem Optimization (DM-AEO), and EEG subband decomposition are used in another study to choose a channel for SCH detection. The Low Complexity Orthogonal Filter Bank (LCOFB), Wavelet Transform (DWT), Empirical Mode Decomposition (EMD), and Variational Mode Decomposition (VMD) are among the methods it uses to extract biomarkers. Combining the DM-AEO approach with EMD resulted in a 70% reduction in computing costs and a 99.26% accuracy rate [4]. Using a Fourier-based method to split signals into sub-band components, compute statistical characteristics, and use a look-ahead pattern (LAP) for local fluctuations, another work presents an effective methodology for identifying SCH from EEG data. The technique shows promise for real-time detection with accuracy levels of 98.62% and 99.24% [5]. In a research, a novel univariate iterative filtering (IF) method for breaking down multivariate signals into

IMFs is presented. It also describes a method for identifying schizophrenia from multi-channel EEG recordings by applying classifiers like SVM, K-nearest neighbors, and linear discriminant analysis. Using the SVM classifier, this approach produced a classification accuracy of 98.9% [7]. In order to reduce reconstruction error, a study uses a deep stack autoencoder (DSAE) to use the band-limited IMFs (BLIMFs) that are extracted from electroencephalogram signals using parametrized VMD (PVMD). The PVMD-DSAE-RBLS strategy outperforms previous approaches in the classification of neuro-disorders using a robust wide learning system (RBLS). For effective automated diagnosis, a computer-aided diagnosis system for schizophrenia is also devised and built on a reconfigurable FPGA embedded processor. It achieves accuracy rates of 99.98%, 96.91%, and 99.29% across three datasets [8]. In a study, an automated model is created for accurate EEG signal interpretation in SCH circumstances by merging robust VMD (RVMD) with optimized extreme learning machine (OELM) classifiers. Mode duplication and noise are reduced by RVMD, and accuracy is increased without overfitting using the OELM classifier. The model demonstrates good accuracy, precision, specificity, F-1 measure, sensitivity, and Cohen's Kappa by minimizing root mean square error with the use of a whale optimization technique for mode selection. Neuro-experts are assisted in real-time SCH recognition by this RVMD-OELM hybrid decision support system [9]. An alternative approach to detecting schizophrenia uses vector function-derived complexity coefficients to categorize EEG data. Researchers used Random Forest and SVM classifiers in conjunction with model-free feature estimation to reach an accuracy rate of 84.5% in differentiating between teenagers in good health and those suffering from schizophrenia. This demonstrates the efficiency of complexity-based analysis as a practical substitute for conventional model-dependent EEG processing [10]. The histogram of local variance (HLV) and symmetrically weighted-local binary patterns (SLBP) features are used in another research to present an automated technique for identifying schizophrenia in EEG signals. The technique distinguishes between healthy EEG signals and those with SCH using an AdaBoost classifier on a reduced feature vector. The findings show that temporal lobe channel characteristics successfully capture regional differences, resulting in 92.85% and 99.36% classification accuracy [11]. Specifically for schizophrenia, the Wavelet Scattering Transform (WST) is compared to the Continuous Wavelet Transform (CWT) and DWT in another work. Six machine learning techniques are applied after twelve statistical features are collected from the EEG data. The results show that while traditional approaches performed well with WST features, ensemble methods did better with CWT and DWT features. Decision trees produced results of 97.98% accuracy, 98.2% sensitivity, 97.72% specificity, 95.94 Kappa score, and 48.04 seconds execution time [12].

ML has made significant progress in its knowledge of schizophrenia. The rise of DL introduced sophisticated end-to-end EEG analysis techniques. Convolutional neural networks (CNNs), for example, extract temporal information from raw data to improve accuracy in EEG-based tasks [6]. In order to automatically detect schizophrenia from EEG recordings, a paper proposes a convolutional attention-based deep learning system. The model combines an attention mechanism to focus on salient temporal patterns with spatial feature extraction via convolutional layers. The Moscow EEG dataset was used for the experiments, and an accuracy of 73.98% was reported. The short recording time and age fluctuation may have an effect on this performance. On the IBIB PAN dataset, however, the model's accuracy was 98.45%, indicating successful generalisation [1]. The use of deep learning, more especially a three-dimensional CNN (3DCNN), to enhance EEG complexity analysis for early schizophrenia diagnosis is investigated in one research. In order to produce feature-enhanced data, the technique uses the spatiotemporal capabilities of 3DCNN to improve data feature extraction. Knowledge distillation is then used to reintegrate the advanced features into the original channels. In subject-dependent tests on two distinct datasets, an average accuracy of 99.46% and 98.06% are achieved using a 10-fold cross-validation approach. The accuracies on both datasets for subject-independent evaluations are 96.04% and 92.67%, respectively [6]. Another study uses a combination of machine learning classifiers and CNNs to automatically identify schizophrenia. According to the results, a CNN and logistic regression (LR) hybrid with three channels (T4, T3, and Cz) achieves 98% accuracy in non-subject testing and 90% accuracy in subject-based testing [13]. A study presents a three-minute computer-assisted diagnosis of schizophrenia based on EEG analysis of brain micro-states. Using a random walk trajectory and CNN, it reports a 93% success rate in patient classification.

Optimizing electrodes, improving the classifier, and developing new devices are some of the problems that lie ahead [14]. In order to identify transient features in EEG signals for SCH classification, another study presents two complementary pipelines: one that uses spectral scalograms (SS) with the EfficientNet (ENB) architecture, and another that uses a convolutional autoencoder (CAE) with an extreme gradient boosting (XGB) classifier. The SS improves the representation of temporal and spectral data and is derived from the continuous wavelet transform. The approaches' efficacy is demonstrated by the experimental results, which provide mean cross-validation accuracies of 95.3% for CAE-XGB and 97% for SS-ENB [15]. The identification of SCH EEG signals using deep learning and dynamic functional connectivity analysis is investigated in another work. It focuses on the alpha band (8–12 Hz) characteristics of participants and presents a cross-mutual information approach for time-frequency domain functional connectivity research. SCH and healthy control (HC) patients are categorized using a 3D convolutional neural network. The evaluation performed on the publicly available EEG dataset reported an accuracy of 97.74% [16].

The EEG is a non-stationary signal that fluctuates with brain activity, exhibiting varying degrees of unpredictability in healthy controls and pathological conditions. Entropy is an example of a complexity metric that is used to diagnose SCH from EEG signals. However, because there aren't enough publicly available datasets, the use of entropy measures for schizophrenia detection is rare. According to earlier research, SCH patients show less randomness in their EEG patterns, which is associated with a decline in cognitive function [2]. Using Multivariate Empirical Mode Decomposition to break down the EEG signal into Intrinsic Mode Functions (IMF), a study employs a multivariate analysis of EEG signals to identify schizophrenia. The study shows a substantial difference between healthy people and those with schizophrenia by calculating the entropy of the IMF signal, suggesting that entropy calculation improves the ability to detect the disease [2]. The method achieved an accuracy of 98.9% using the support vector machines (SVM) classifier. An alternative method is presented by Bagherzadeh et al., in which transfer entropy (TE) from 19 EEG channels is used to create effective connectivity matrices. These matrices created a 19 × 19 connectivity image that represented the spatial and temporal information of the EEG signal. Five pre-trained CNNs use Transfer Learning to process the images. A Long Short-Term Memory (LSTM) model is used in conjunction with the deep features produced by these models to categorise 14 schizophrenia patients against 14 healthy controls. The EfficientNetB0-LSTM model achieved the maximum accuracy of 99.90% and an F1-score of 99.93% through 10-fold cross-validation, outperforming pre-trained CNN models in the hybrid CNN-LSTM framework [17]. Another study produced an Automated Diagnostic Tool (ADT) for categorising EEG signal patterns into normal and schizophrenia groups. EEG series splitting, nonlinear feature extraction, t-test-based feature selection, classification, and validation are all part of the ADT. Entropy-based features, including Rényi, Shannon, Tsallis, bispectrum, and permutation entropy, are among the nonlinear features computed in the research. On the EEG dataset, SVM-RBF achieves the best performance at 92.91% [18]. In order to distinguish between SCH patients and CS, a hybrid deep neural network (DNN) model that combines convolutional neural networks with long short-term memory is suggested. Entropy and Fourier-based RGB pictures with spatial information derived from EEG time series were employed as input for this model. The results show that in brain topography, fuzzy entropy (FuzzyEn) features perform much better than fast Fourier transform (FFT) features. The suggested deep learning approach achieved 99.22% accuracy with FuzzyEn and 96.34% accuracy with FFT [19]. In a different study, 31 traditional EEG features, including entropy-based metrics, and 19 microstate features obtained from resting-state EEG recordings of 14 schizophrenia patients and 14 healthy controls, are examined. EEG microstate features considerably surpassed traditional EEG features in patient versus control classification, according to machine learning multivariate analysis [20]. Another study suggests a deep learning method for detecting schizophrenia based on LSTM. From the EEG recordings of SCH patients and HC individuals, nonlinear features such as approximation entropy and Katz fractal dimension are calculated along with different time domain features. After training on 6000 feature vectors and testing on 790, an LSTM model with four hidden layers of 32 neurons each achieved a detection accuracy of 99.0% [21].

In this work, raw EEG signals are converted into two-dimensional, spatial-temporal entropy-based EEG feature maps (EEG-FM), introducing a unique diagnostic framework for schizophrenia

detection. The method generates 2D inputs for a deep learning model using six entropy measures: approximate, bubble, fuzzy, permutation, sample, and spectral entropy. A ResNet-50-Based ensemble learning strategy is used, focusing on specific entropy-based EEG-FM. Six parallel ResNet50 architectures received each entropy-based EEG-FM as input. To improve the results, a soft voting procedure is used to integrate the architectures' independent predictions.

2. MATERIALS AND METHODS

The entire method, which classifies SCH and control subjects (CS) using deep learning based on entropy-based EEG-FM, is shown in Figure 1. The method starts with preprocessing steps, including filtering and segmentation (e.g., into 10-second segments) to remove artifacts and isolate relevant time periods. Following preprocessing, a set of entropy metrics is calculated for each channel and segment. After organizing these entropy metrics into 9x9 feature matrices, these feature matrices are converted into EEG-FM (224x224) and fed into a pre-trained CNN architecture to detect SCH. MATLAB R2025a, a popular platform for biomedical data analysis and machine learning, is utilized for all EEG preprocessing, entropy-based feature extraction, EEG-FM creation, and soft voting-based ensemble techniques.

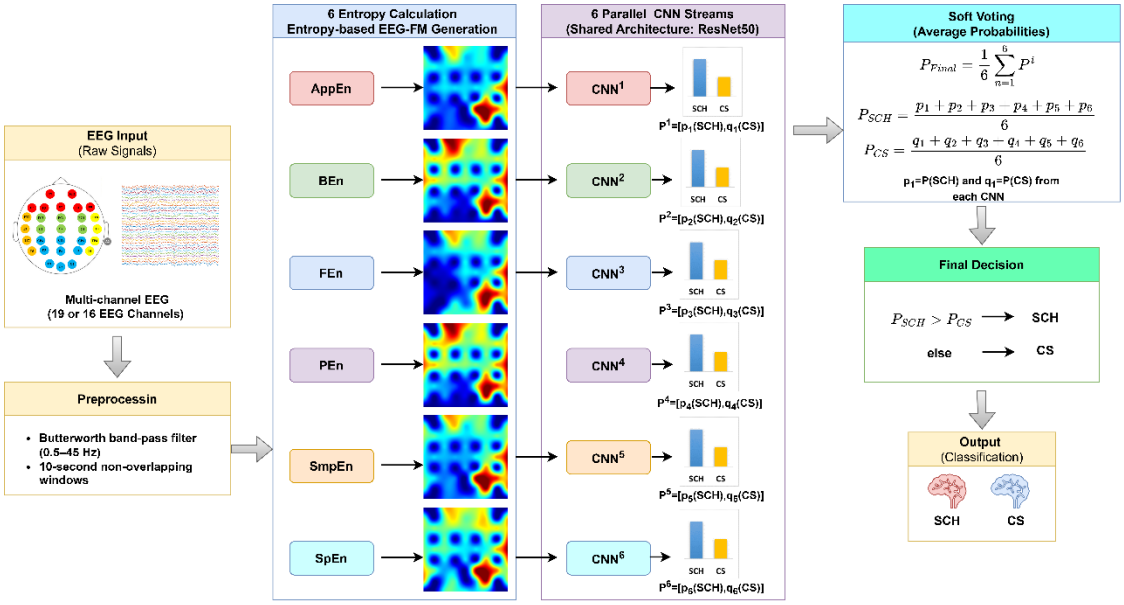

Figure 1. The framework of automatic detection of SCH through EEG-FM and deep learning

2.1. Description of EEG Datasets In this study, two publicly available EEG datasets containing EEG data of schizophrenia patients and healthy controls are used to evaluate the effectiveness of the proposed entropy-based EEG-FM and ensemble soft voting-based deep learning approach. To evaluate the stability of the suggested approach under various circumstances, two separate EEG datasets with different demographic and acquisition characteristics are included. The model's behaviour under greater variability and distributional shifts is also further investigated through a combined dataset evaluation.

2.1.1 Schizophrenia EEG Datasets-1 (Dataset 1)

The study's first open dataset [22], which was gathered at Moscow State University (MSU) and recorded at 128 Hz from 16 electrodes positioned in accordance with the 10–20 channel system. The database is composed of 84 teenagers' resting state EEG data. The participants' ages ranged from 11 to 14. The subjects were split into two groups: 45 individuals with schizophrenia symptoms and 39 individuals without the disorder (healthy age-matched controls). Before the recording, none had received any prior treatment. For one minute, the EEG samples were collected at a sampling rate of 128 Hz, yielding 7680 data points per channel. With predetermined electrode placements, data were collected from 16 electrodes (F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2) located

in the frontal, central, parietal, temporal, and occipital areas of the brain. The data was organized as a 16x7680 matrix for every participant.

2.1.2 Schizophrenia EEG Datasets-2 (Dataset 2)

The second dataset [23] included 14 patients with paranoid schizophrenia and 14 age-and gender-matched CS who were admitted to the Institute of Psychiatry and Neurology in Warsaw, Poland. The patients had finished at least seven days of medication washout before the EEG recordings. The study was authorized by the Institute's Ethics Committee, and each participant gave written informed consent. EEG data were recorded during a 15-minute eyes-closed resting-state session at a sampling frequency of 250 Hz using the standard 10–20 system with 19 electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2). The reference electrode was placed at FCz.

2.2. EEG Pre-processing and Segmentation During the pre-processing stage, we meticulously clean the EEG data to ensure accuracy and remove extraneous signals. Initially, separate 10-second non-overlapping EEG segments are created for Dataset 1 (16 channels, 128 Hz) and Dataset 2 (19 channels, 250 Hz). A second-order Butterworth band-pass filter (0.5–45 Hz) is then applied to each segment to remove low-frequency drifts and high-frequency noise. The processed segments are saved separately for Dataset 1 and Dataset 2 for further analysis after each EEG signal of all subjects has undergone independent filtering and segmentation.

2.3. Calculated EEG Feature Research suggests that aberrant dopamine fluctuations are associated with changes in brain dynamics in schizophrenia. Because of its wide range of symptoms, schizophrenia is a useful topic for complexity studies, where EEG regularity is evaluated using entropy-based methods. Age, medication, and preprocessing all have an impact on EEG complexity in people with schizophrenia; numerous studies have shown a reduction in neural complexity, especially in the frontal lobe. This decrease is consistent with hypofrontality findings from neuroimaging. Additionally, patients show reduced complexity changes in response to stimuli, which suggests that their information processing responsiveness is compromised. All things considered, nonlinear systems theory makes a substantial contribution to our knowledge of the brain abnormalities linked to schizophrenia [6]. This study aimed to create an effective entropy-based EEG-FM by leveraging the complexity of EEG signals using established entropy measures to detect schizophrenia. Approximate Entropy (AppEn), Bubble Entropy (BEn), Fuzzy Entropy (FEn), Permutation Entropy (PEn), Sample Entropy (SmpEn), and Spectral Entropy (SpEn) [24] are the six entropy-based features calculated in this study. Two-dimensional topographical maps, known as EEG-FMs, are produced from these features.

- Approximate Entropy: AppEn evaluates the regularity and predictability of a time series by estimating the likelihood that similar patterns will persist across time. Signals with lower levels are more regular, while those with higher values are more complicated and unpredictable. Higher AppEn in an EEG could be a sign of more neuronal irregularity. Moreover, AppxEn depends on the filtering value and is a biased randomness measure. Here, we take into account the dimension parameter m= 2 and the filter parameter = 0.2 times the standard deviation of the EEG segment [2,24].

- Sample Entropy: By eliminating self-matching, SmpEn—an enhanced form of AppEn—becomes a less biased and more reliable indicator of signal complexity. It calculates the likelihood that comparable signal patterns will remain consistent at the following time step. More complicated and unpredictable neural activity is indicated by greater SampEn values in the EEG segment, while more regular brain dynamics are associated with lower values. SampEn's resilience and decreased reliance on EEG segment length make it a popular choice for biomedical signal processing [2,24].

- Fuzzy Entropy: By substituting a continuous fuzzy membership function for the strict similarity criterion, FEn expands SmpEn. It assigns a degree of similarity to patterns rather than just labelling them as "similar" or "not similar," making comparisons easier and more reliable. Because brain signals are naturally noisy and nonlinear, FEn is especially well-suited for EEG analysis. When working with brief or tainted EEG recordings, FEn offers a more accurate and noise-tolerant measure of signal complexity [2,19,24].

• Permutation Entropy: Instead of using the amplitudes of the values in a time series to determine complexity, PEn uses the order (ranking) of the values. It focuses on the frequency of various patterns of increase or decrease in the signal. PEn is very resilient to noise and amplitude fluctuations because it solely depends on the order of values. It efficiently captures temporal dynamics and pattern variability in EEG analysis, which makes it helpful for detecting alterations in brain activity linked to neurological conditions [2,18,24].

• Spectral Entropy: The distribution of signal strength across various frequency components is measured by SpEn. It shows how uniformly the signal energy is distributed across the frequency spectrum and is based on the normalised power spectral density. High SpEn values in the EEG segment imply a more uniform and disorganised frequency distribution, whereas low values show that power is concentrated in particular frequency bands (more regular activity). Because of this, SpEn is especially helpful for examining abnormalities in brain activity that are connected to frequency [2,24].

• Bubble Entropy: Inspired by bubble sort, bubble entropy measures signal complexity by counting the number of swaps needed to sort values within embedding vectors. Unlike AppEn/SampEn, it doesn't need a tolerance parameter. Robust to parameter selection and noise. Helpful for detecting changes in the dynamic complexity of the EEG segment [24].

By capturing temporal irregularity (AppEn, SampEn), noise-robust similarity (FEn), ordinal pattern dynamics (PEn), frequency-domain organization (SpEn), and structural complexity (BEn), these entropy measures offer complementary depictions of EEG complexity. These many nonlinear features allow the deep learning model to learn richer and more discriminative patterns when converted into entropy-based EEG-FMs, which improves classification performance and strengthens the suggested ensemble framework.

2.4. Development of EEG Feature Maps A 2D EEG-FM is produced to concurrently integrate the spatial, temporal, and frequency characteristics of EEG signals, as seen in Figure 2 [25,26]. First, 10-second epochs are extracted separately from 19-channel and 16-channel EEG recordings. Next, entropy-based feature vectors are computed for each channel, resulting in N-dimensional feature representations (N = 19 and N = 16, respectively) (given in Figure 2 (I)). To maintain the spatial structural information of the electrode site, we then convert the N-dimensional entropy-based feature vector into a compact 2D map (9×9 matrix) based on the electrode placements on the brain (Figure 2 (II-a)). Figure 2 (II-b) displays the 2D map of 19-channel estimated entropy-based features. The matrix's colored regions show the relevant channels' normalized entropy values, while the empty spaces indicate that the signals from those channels are not being used. According to the study [27,28], the empty values of the matrix are found using the following formulation as a function of the point values surrounding them;

$$K_{(i,j)} = \frac{K'_{(i+1,j)} + K'_{(i-1,j)} + K'_{(i,j+1)} + K'_{(i,j-1)}}{M} \quad (1)$$

where j ≤ 9 and 1 ≤ i. The normalized entropy-based feature value K represents the features of the empty spaces, while K^(')indicates the normalized entropy-based feature value of the surrounding point. The number of non-zero components in the numerator is indicated by M's default value of 1.

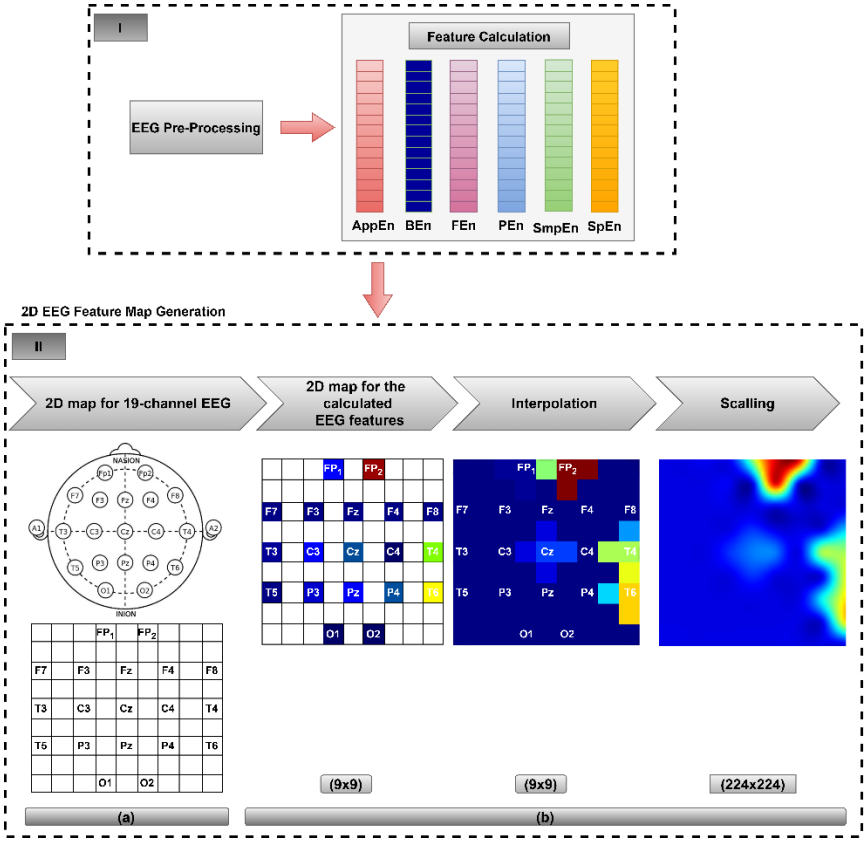

Figure 2. The procedure of representing entropy-based features in a 2D EEG-FM

A "jet" colormap is then used to create a comprehensive EEG feature map of a specific entropy (Figure 2 (II-b-middle)). To improve the resolution of the 2D EEG feature maps, 224×224 EEG feature maps are created using biharmonic spline interpolation (Figure 2 (II-b-right)). After the procedures, two-dimensional EEG FMs are created for each entropy-based feature. In Dataset 1, there are 6 EEG-FMs per participant and an entropy feature, while Dataset 2 has 70 EEG-FMs. Hence, for Dataset 1, 234, and 270 EEG-FMs are generated for each of the six entropy-based features of CS and SCH patients, respectively, by representing all 16 EEG channels in a single matrix. Similarly, for Dataset 2, 980 EEG-FMs are generated for each of the six entropy-based characteristics of both CS and SCH patients by representing all 19 EEG channels in a single matrix. For instance, Figure 2 (II-b-right) displays the EEG feature map of an AppEn-based feature of an EEG epoch of SCH. In reference to the "jet" colormap, a dark blue color indicates a completely inactive electrode, whereas red indicates an electrode that is active. In particular, each N-channel EEG segment in Datasets 1 and 2 undergoes an independent and iterative EEG-FM generation process, where entropy-based feature maps are generated separately for each EEG segment using segment-wise min–max normalization. In order to avoid inter-segment information leakage and maintain subject-level independence, this guarantees that entropy values are scaled within each N-Channel EEG segment independently. As a result, segment-specific features are reflected in the created EEG-FMs, allowing the deep learning model to acquire reliable and objective representations.

2.5. Convolutional Neural Network Model and Soft Voting for Classification A ResNet50 architecture is employed to classify the generated entropy-based EEG feature maps because its residual learning structure facilitates stable deep feature extraction from complex nonlinear EEG representations while mitigating the vanishing gradient problem. In this study, ResNet50 is selected as a representative deep CNN backbone not primarily to identify the most computationally efficient architecture, but rather to investigate the effectiveness of combining complementary entropy-based EEG-FM representations through decision-level ensemble learning. Through the use of six parallel ResNet50 streams, each of which is devoted to a particular entropy-based feature, the model is able to

capture a variety of nonlinear dynamics. These are then combined via a soft voting layer to get the final diagnostic prediction. A 50-layer deep version of the Residual Network (ResNet) family is called ResNet50. The development of "identity shortcut connections" or "skip connections" that get around one or more layers is its main innovation. Accuracy has historically tended to saturate and deteriorate as neural networks get deeper (the vanishing gradient problem). By letting the gradient pass through the shortcut connections, ResNet50 resolves this issue and enables efficient training of extremely deep architectures. Its building blocks are designed in a "bottleneck" fashion. The three convolutional layers that make up each block are 1x1, 3x3, and 1x1 convolutions. The 3x3 layer has reduced input/output dimensions to increase computing efficiency, whereas the 1x1 layers are in charge of first decreasing and then restoring (raising) dimensions [29].

ResNet50's 50 layers can extract both high-level abstract patterns and low-level edges from entropy-based EEG-FMs (such as AppEn, BEn, FEn, PEn, SmpEn, and SpEn) because they depict the nonlinear intricacies of the brain. Typically, the ImageNet dataset is used to pre-train ResNet50. Our model starts with a deep understanding of visual forms thanks to these pre-trained weights, and it needs less EEG data to achieve high accuracy. For a particular entropy type, each ResNet50 functions as an independent expert, as shown in Figure 1. Stable and accurately calibrated probability outputs (p_1, q_1) are required for a successful Soft Voting mechanism to arrive at the ultimate prediction (SCH vs. CS).

A Soft Voting ensemble layer, which combines the class probabilities produced by the six simultaneous ResNet50 streams, is used for the final classification. The model makes use of the combined intelligence of several non-linear complexity metrics by determining the average projected probability across all entropy-based feature maps. This method improves the detection system's overall stability and accuracy by ensuring that the ultimate diagnostic decision (SCH vs. CS) is not only based on a single feature type. Soft Voting is an ensemble learning technique that forecasts the class labels based on the predicted probabilities (p) instead of merely the final discrete labels (0 or 1). It is sometimes referred to as weighted or average probability voting. Unlike Hard Voting, which is a "winner-takes-all" method, Soft Voting considers the individual confidence level of each ResNet50 model. A probability distribution for the classes of Schizophrenia (SCH) and Healthy Controls (CS) is generated for each CNN stream. The final probability for each class is determined by taking the arithmetic mean of the probabilities provided by each of the six entropy-based Feature maps (p^1 to p^6), as your diagram shows (Figure 1). The class that attains the highest average probability makes the final forecast. In mathematics, the final probability P_{final} for class "j" for a binary classification is determined as follows:

$$P_{final}(j) = \frac{1}{N}\sum_{i=1}^{N} p_j^i \tag{2}$$

where p_j^i is the probability of class j predicted by the i-th CNN, and N=6 is the number of entropy-based feature maps [30].

2.6. Model Training and Performance Evaluation In our study, Leave-One-Group-Out cross-validation (LOGO-CV) is applied for both datasets. The principal aim of group-independent testing is to prevent identical data from being included in both the training and testing sets at the same time. The 39 CS and 45 SCH participants in Dataset 1 are divided into 13 groups (Fold1–Fold12: 3 CS and 3 SCH subjects per fold; Fold13: remaining 3 CS and 9 SCH subjects). In each iteration, a 13-fold LOGO-CV-based evaluation is performed, where one group is used for testing, and the remaining groups are used for training. For Dataset 2, through the use of the LOGO-CV approach, each participant (14 SCH, 14 CS) is split up into 14 groups (Fold1 to Fold14) according to their individual differences, with one SCH and one CS participant constituting each group. One group is used for testing, and thirteen groups are employed for training among these participants. Every group goes through both the training and

testing phases thanks to this procedure's 14 iterations [6]. Additionally, to prevent overfitting during training for both datasets, the training dataset is divided into 70% training and 30% validation sets in each fold, and the training process is conducted accordingly. MATLAB R2025a with GPU-supported deep learning training processes is used to implement the suggested framework. The core CNN model is a pretrained ResNet50 architecture initialized with ImageNet weights. During training, every layer of the network is adjusted; no layer freezing technique is used. Using augmentedImageDatastore preprocessing, EEG feature maps are scaled to the default ResNet50 input size (224 × 224 × 3). The Adam optimizer is used for training, using a cross-entropy loss function, 10 epochs, a mini-batch size of 20, and an initial learning rate of 0.001. In terms of the multi-entropy framework, a different ResNet50 backbone is used to process each entropy-based EEG-FM. The final classification is obtained by combining the output probabilities of several independently trained streams at the decision level using a soft voting technique. To ensure consistency and fair comparison, all hyperparameters are experimentally chosen based on initial trials and maintained constant throughout all datasets.

Several evaluation metrics, including accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and false discovery rate (FDR), are also used to assess the performance of the proposed approach.

$$
\begin{aligned}
ACC &= \frac{TP + TN}{TP + FN + FP + TN} \\
SEN &= \frac{TP}{TP + FN} \\
SPE &= \frac{TN}{FP + TN} \\
PPV &= \frac{TP}{TP + FP} \\
FDR &= \frac{FP}{TP + FP}
\end{aligned}
\tag{3}
$$

The number of correctly identified EEG segments in SCH and CS is denoted by TP and TN, respectively. Additionally, FP and FN stand for the number of incorrectly diagnosed EEG segments in CS and SCH, respectively [28].

2.7. Ethical Approval Ethical approval is not required for this study.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

The ResNet50-based model and entropy-based EEG-FM, which detect EEG segments of SCH patients via soft voting, are presented in this paper. The suggested method begins by computing six different entropy-based features (AppEn, BEn, FEn, PEn, SmpEn, and SpEn) for 10-s EEG segments utilizing 16-channel EEG signals from Dataset 1 and 19-channel EEG signals from Dataset 2. EEG-FMs are generated using the estimated entropy-based features and input into the ResNet50 architecture. A conclusive decision (0: CS or 1: SCH) is derived from the six distinct ResNet50 outputs via a soft voting methodology. Initially, to analyze the statistical significance of the computed entropy features for each EEG channel, the one-way analysis of variance (ANOVA) test at a 95% confidence level (α = 0.05) is employed. Small p-values ($p \leq 0.05$) from the ANOVA test indicate a substantial difference between the groups and provide credence to the null hypothesis' rejection [28]. Using an EEG feature map presentation, the p-values for Datasets 1 and 2 are displayed in Figures 3 and 4, respectively. Gray in these graphs means that the channels in those parts are not used to calculate features. P-values of 0-0.01 are represented by blue, 0.01-0.05 by green, 0.05-0.08 by yellow, and >0.08 by red. Notably, only a few channels' computed entropy features for Dataset 1 showed statistically significant differences between groups (Figure 3).

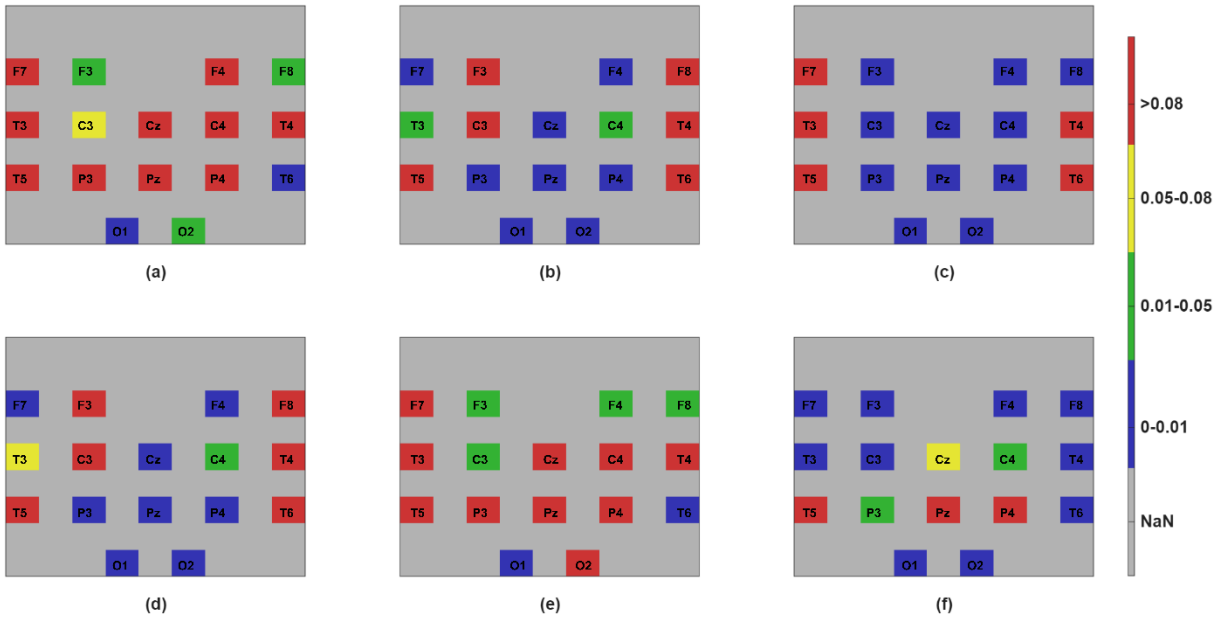

Figure 3. ANOVA test results of various entropy features (a) AppEn, (b) BEn, (c) FEn, (d) PEn, (e) SmpEn, and (f) SpEn for Dataset 1 are expressed in terms of the *p*-value. Note: In these graphs, gray indicates that the channels are not utilized for feature computation. The colors blue, green, and yellow indicate p-values of 0-0.01, 0.01-0.05, 0.05-0.08, and >0.08, respectively.

On the other hand, Dataset 2 showed statistically significant differences between groups with p-values less than 0.05 for nearly all channels and all entropy features (Figure 4). Visualizations of statistical feature maps show distinct patterns of significance among EEG channels, with smaller p-values signifying more pronounced group differences. Consistently low p-values observed in particular EEG channels suggest that certain entropy-based EEG features may provide useful discriminatory information for computer-aided schizophrenia-related EEG classification. Reduced p-values in the graphic representation highlight the fact that, overall, significant impacts are focused in specific channel-feature pairings. As is evident, formulating a general conclusion based on the findings of several entropy-based EEG feature maps requires a common conclusion.

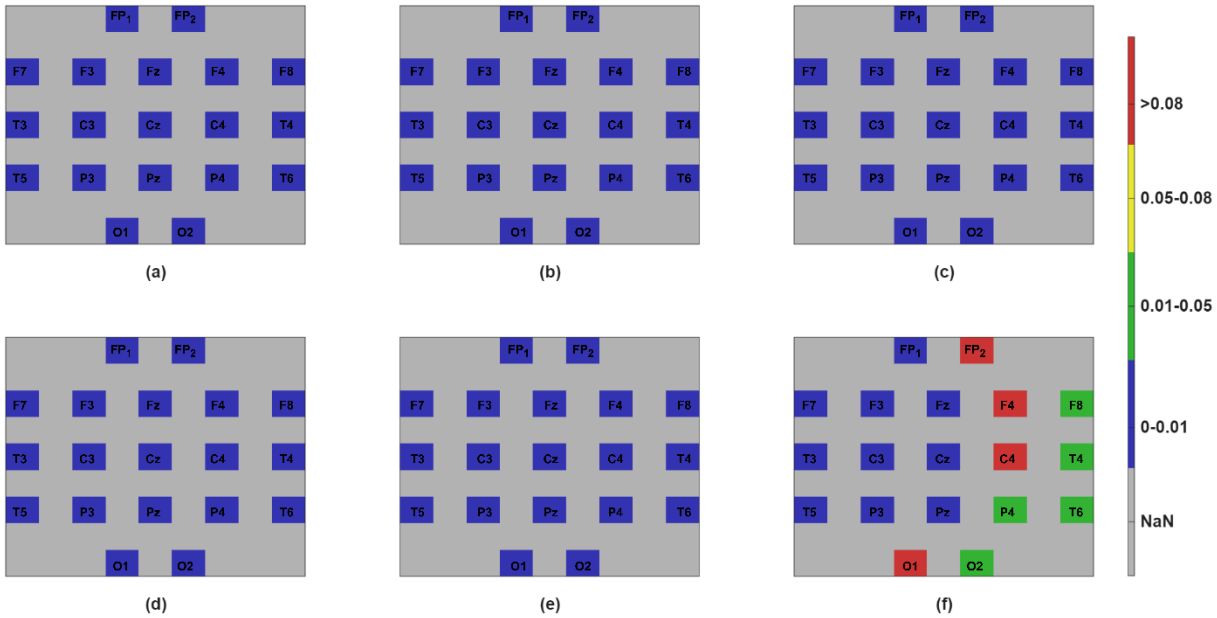

Figure 4. ANOVA test results of various entropy features (a) AppEn, (b) BEn, (c) FEn, (d) PEn, (e) SmpEn, and (f) SpEn for Dataset 2 are expressed in terms of the *p*-value. Note: In these graphs, gray indicates that the channels are not utilized for feature computation. The colors blue, green, and yellow indicate p-values of 0-0.01, 0.01-0.05, 0.05-0.08, and >0.08, respectively.

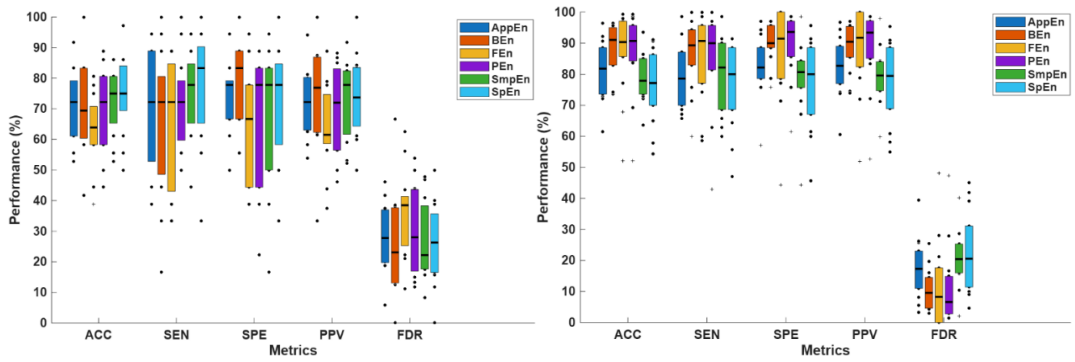

Figure 5. SCH detection performance across individual entropy-based feature maps for (a) Dataset 1 and (b) Dataset 2 is analyzed using boxplots.

Firstly, the LOGO-CV evaluation is used to assess the suggested model's performance, and Figures 5(a) and (b) for Datasets 1 and 2, respectively, display the results of an independent performance evaluation of an entropy-based EEG-FMs. The model is assessed using the following common performance evaluation metrics: FDR, PPV, SPE, SEN, and ACC. Different entropy types perform differently in Dataset 1 (Figure 5 (a)), suggesting that some non-linear features are more discriminative than others. The median ACC of SpEn and SmpEn is approximately 75%, but the majority of entropy maps attain 70% to 80% ACC. SEN varies greatly; features like BEn and FEn are less reliable in identifying SCH, whereas SpEn peaks at about 83%. In contrast to FEn and PEn, BEn has the highest median SPE, at about 85%, and successfully identifies CS. The high False Discovery Rate (20–30%) emphasizes the necessity of a Soft Voting ensemble in order to reduce errors. Across the majority of entropy types, Dataset 2 shows a noticeable performance improvement (Figure 5 (b)). The accuracy is high, over 90% for both FEn and PEn in particular. Sensitivity demonstrates that FEn and PEn successfully detect the SCH class with minimal outliers and medians above 90%. Both FEn and

PEn are close to 95%, indicating good SPE. Additionally, FEn and PEn achieve medians below 10% in Dataset 2, indicating a considerable reduction in the FDR.

Table 1. Comparative analysis of the suggested soft voting ensemble model's classification performance on two separate datasets.

	Dataset 1 (%)						**Dataset 2 (%)**					
Fold	**ACC**	**SEN**	**SPE**	**PPV**	**FDR**	**AUC**	**ACC**	**SEN**	**SPE**	**PPV**	**FDR**	**AUC**
Fold1	97.2	100.0	94.4	94.7	5.3	99.69	97.9	95.7	100.0	100.0	0.0	99.71
Fold2	75.0	66.7	83.3	80.0	20.0	79.94	100.0	100.0	100.0	100.0	0.0	100.00
Fold3	86.1	100.0	72.2	78.3	21.7	98.77	99.3	98.6	100.0	100.0	0.0	99.96
Fold4	83.3	77.8	88.9	87.5	12.5	90.43	97.9	97.1	98.6	98.6	1.4	99.86
Fold5	69.4	83.3	55.6	65.2	34.8	74.07	97.1	97.1	97.1	97.1	2.9	99.99
Fold6	58.3	50.0	66.7	60.0	40.0	71.91	96.4	95.7	97.1	97.1	2.9	99.20
Fold7	94.4	94.4	94.4	94.4	5.6	99.38	90.7	88.6	92.9	92.5	7.5	96.45
Fold8	94.4	88.9	100.0	100.0	0.0	97.84	100.0	100.0	100.0	100.0	0.0	100.00
Fold9	86.1	94.4	77.8	81.0	19.0	93.52	85.0	77.1	92.9	91.5	8.5	94.86
Fold10	63.9	61.1	66.7	64.7	35.3	73.77	98.6	100.0	97.1	97.2	2.8	100.00
Fold11	83.3	88.9	77.8	80.0	20.0	93.52	89.3	92.9	85.7	86.7	13.3	95.96
Fold12	94.4	88.9	100.0	100.0	0.00	95.06	92.9	85.7	100.0	100.0	0.0	99.78
Fold13	72.2	61.1	83.3	78.6	21.4	80.86	97.9	97.1	98.6	98.6	1.4	99.80
Fold14							97.1	100.0	94.3	94.6	5.4	99.33
Avg	***81.4***	***81.2***	***81.6***	***81.9***	***18.1***	***88.37***	***95.7***	***94.7***	***96.7***	***96.7***	***3.3***	***98.92***

Soft Voting determines the final diagnostic result by combining the projected probabilities from each entropy-based model (e.g., AppEn, FEn, PEn) according to the analytical framework explained. By using the confidence of each sub-model, this approach usually produces a more reliable performance than individual feature maps. Table 1 presents the findings of soft voting. Across individual entropy features, Dataset 1 provides a baseline with poor performance (Figure 5 (a)). The ensemble model outperforms the best individual feature SpEn (average ACC 74.6%) by almost 7%, achieving an average accuracy of 81.4% with a high area under the curve (AUC) of 88.37%. The best average results shown in Figure 5 (a) are about 5% lower than the average SPE for CS and the average SEN for SCH, which are roughly 82%. The average false discovery rate is also decreased to about 18% as a result of the agreement between multiple entropy measures. It is noteworthy that, in addition to average decisions, soft voting-based results achieved an ACC of over 85% for SCH detection, with a few exceptions (Fold5, Fold6, Fold10, and Fold13). Due to improved signal-to-noise ratios or more distinct physiological patterns, Dataset 2 has noticeably increased diagnosis performance with 95.7%ACC and a high AUC of 98.92%. With accuracy rates above 95% for the majority of folds and typically approaching 90%, the soft voting-based method is close to state-of-the-art. It successfully lowers false positives by an average 94.7% SEN for the identification of SCH and an average 96.7% SPE for CS. It is also noteworthy that the FDR rate has fallen to an average of 3.3%. The recording period, age variables, and demographic differences are thought to have an impact on the performance difference between the two datasets.

Table 2. Statistical significance analysis of the proposed soft voting ensemble on the Dataset 1 and Dataset 2 using Wilcoxon signed-rank testing and Benjamini–Hochberg FDR correction.

	Model	p_value	FDR_adjusted_p	Cohens_d	Significant
Dataset 1	**AppEn**	0.00098	0.00200	1.6282	true
	Ben	0.00098	0.00200	1.2025	true
	FEn	0.00098	0.00200	1.4883	true
	PEn	0.00630	0.00630	0.9258	true
	SmpEn	0.00340	0.00510	1.0339	true
	SpEn	0.00490	0.00590	0.9744	true
Dataset 2	**AppEn**	0.00012	0.00024	1.4634	true
	BEn	0.00024	0.00037	1.0074	true
	FEn	0.01040	0.01040	0.7064	true
	PEn	0.00012	0.00024	1.0627	true
	SmpEn	0.00012	0.00024	1.4946	true
	SpEn	0.00061	0.00073	1.0898	true

The suggested soft voting ensemble model is statistically compared with individual entropy-based EEG-FM models using the Wilcoxon signed-rank test with Benjamini–Hochberg false discovery rate (BH-FDR) correction. This non-parametric statistical test controls the false discovery rate brought on by multiple comparisons and assesses whether the performance differences between paired models across folds are statistically significant. Cohen's d effect sizes are computed in addition to corrected p-values to measure the extent of performance gains made possible by the suggested ensemble framework. A significance criterion of $\alpha = 0.05$ is used to interpret corrected p-values. Table 2 shows that, following BH-FDR correction (adjusted $p < 0.05$ for all comparisons), the suggested soft voting ensemble considerably outperformed all individual entropy-based EEG-FM models in terms of classification accuracy for both Dataset 1 and Dataset 2. The greatest accuracy gains for Dataset 1 are seen against AppEn (d = 1.63), FEn (d = 1.49), and BEn (d = 1.20), suggesting significant effect sizes. According to Cohen's d interpretation, significant accuracy improvements are also seen in Dataset 2, especially for SmpEn (d = 1.49), AppEn (d = 1.46), and SpEn (d = 1.09), indicating large impact sizes once more. Overall, the statistical analysis verifies that, when compared to single entropy-based representations, the suggested soft voting technique regularly yields noticeably better classification accuracy. The observed increases in EEG-based schizophrenia classification performance are both statistically significant and practically useful, according to the substantial effect sizes.

3.1. Combined Dataset Evaluation Results To further investigate the behavior of the proposed framework under heterogeneous EEG distributions, a combined dataset evaluation strategy is additionally employed alongside the individual LOGO-CV analyses performed separately for each dataset. This method combined the corresponding training and test groups from Datasets 1 and 2 at each iteration. In particular, the test groups from both datasets are merged to create a unified test set for a given fold, while the remaining groups from both datasets are joined to create a joint training set. To manage the learning process and minimize overfitting, the combined training data are then divided into 70% training and 30% validation subsets. After training, the combined test set for that fold is used to assess the model. This process allows the model to learn from a more varied distribution of EEG signals while ensuring that subject independence is maintained within each dataset. Additionally, the unified evaluation methodology makes it possible to evaluate the model's resilience across a variety of datasets without combining subject-specific data from the training and testing stages.

Table 3 summarises the proposed multi-entropy-based EEG-FM and CNN-based framework's classification performance on the combined dataset. The presented values show the average performance over all folds. PEn and BEn, two of the individual entropy-based models, performed comparatively well, with accuracies of 68.1% ± 11.4 and 68.1% ± 12.6, respectively. FEn, on the other hand, performed the worst with an accuracy of 51.1% ± 13.0 and a lower specificity of 29.5% ± 23.3, showing a limited capacity for SCH detection. Across all evaluation metrics, the suggested soft voting ensemble strategy consistently performed better than any individual entropy model. In particular, it

had the lowest FDR of 29.8% ± 10.0 and the greatest ACC of 70.3% ± 11.4, SEN of 76.2% ± 13.7, SPE of 64.4% ± 22.4, and PPV of 70.2% ± 12.6. These results imply that combining the performance of entropy-based EEG-FM through ensemble learning enhances the model's ability to capture different nonlinear EEG properties, leading to better and more balanced classification performance when compared to single entropy EEG-FM-based techniques.

Table 3. Comparative analysis of the classification performance of the suggested soft voting ensemble model on the combined dataset.

	Performance Metrics (%)				
	ACC	**SEN**	**SPE**	**PPV**	**FDR**
AppEn	65.6±10.9	62.7±19.3	68.5±12.8	66.5±12.1	33.6±14.7
BEn	68.1±12.6	70.2±16.6	66.1±20.9	68.7±12.3	31.3±13.1
FEn	51.1±13.0	72.6±14.0	29.5±23.3	51.7±10.4	48.3±9.8
PEn	68.1±11.4	69.3±19.3	66.9±19.7	69.1±11.9	30.9±13.3
SmpEn	57.1±13.5	53.2±21.9	60.9±19.3	58.6±16.8	41.4±17.4
SpEn	59.7±11.8	67.3±14.9	52.11±23.0	60.9±14.4	39.1±10.2
Soft-Voting	***70.3±11.4***	***76.2±13.7***	***64.4±22.4***	***70.2±12.6***	***29.8±10.0***

The Wilcoxon signed-rank test with BH-FDR is used to assess the statistical significance of performance differences between the suggested soft voting-based ensemble and individual entropy-based EEG-FM models. Figure 6 provides a visual summary of the associated results obtained for the combined dataset. Large effect sizes are indicated by statistically significant improvements after FDR correction, especially for FEn (adjusted $p < 0.001$, $d = 1.74$), SmpEn (adjusted $p < 0.01$, $d = 0.92$), and SpEn (adjusted $p < 0.05$, $d = 0.80$). Conversely, following correction, AppEn, BEn, and PEn is not exhibit statistically significant differences (adjusted $p > 0.05$), while moderate effect sizes point to possible improvements.

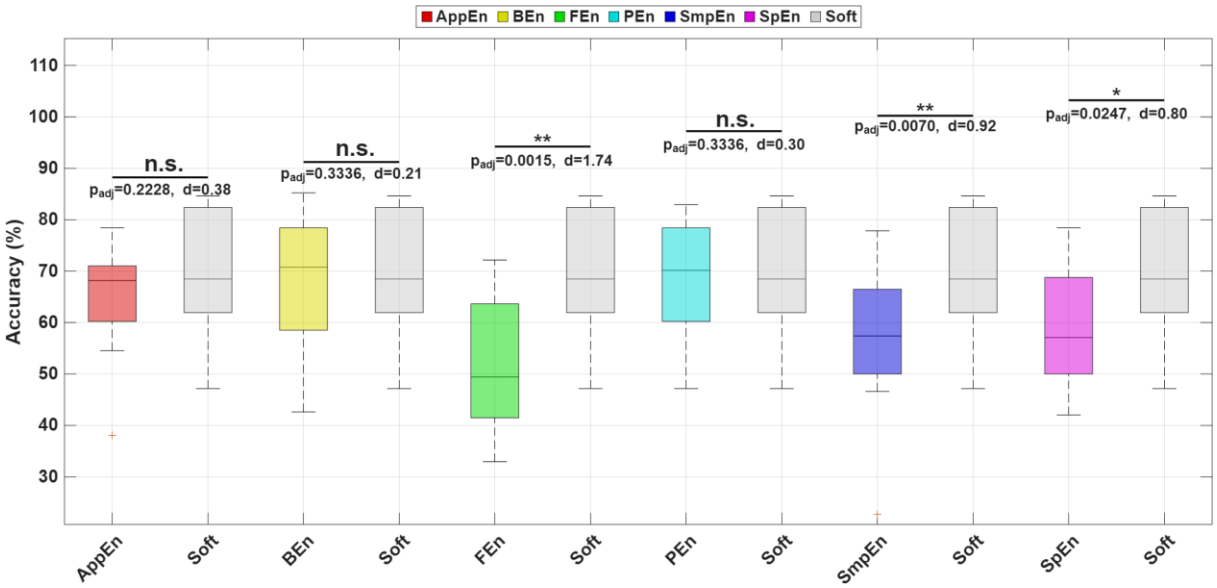

Figure 6. Pairwise boxplot comparison of classification accuracy between individual entropy-based EEG-FM models and the proposed soft voting ensemble approach for the combined dataset. Statistical significance is evaluated using the Wilcoxon signed-rank test with BH-FDR correction (* adjusted $p < 0.05$, ** adjusted $p < 0.01$, *** adjusted $p < 0.001$, n.s.: not significant).

3.2. Subject Level Evaluation Results To provide a more clinically significant evaluation, the performance of the suggested technique is also assessed at the subject level rather than the segment level. Because each subject has numerous EEG segments, segment-wise predictions are first made separately, then the dominant predicted class across all segments is used to create the final subject

label. By treating correlated segments from the same person as independent samples, this majority-based decision approach minimizes artificially inflated performance and reduces segment-level bias. In order to provide a more transparent interpretation of statistical uncertainty and performance variability, especially for datasets with relatively small numbers of subjects, 95% confidence intervals (CIs) calculated using the standard error of the mean (SEM) are reported for each evaluation metric. SEM is calculated as $p(1-p)/n$ for a given metric p, where n is the number of subjects relevant to the metric. Next, $p \pm 1.96 \times SEM$ is used to generate the 95% confidence interval. In this case, n represents the total number of subjects for ACC (n=TP+TN+FP+FN), the number of positive subjects for SEN (n=TP+FN), the number of negative subjects for SPE (n=TN+FP), and the number of anticipated positive subjects for PPV (n=TP+FP).

Table 4. Subject-level classification performance with 95% confidence intervals of Dataset 1.

	Performance Metrics							
	ACC (%)		**SEN (%)**		**SPE (%)**		**PPV (%)**	
	Value	**95% CI**	**Value**	**95% CI**	**Value**	**95% CI**	**Value**	**95% CI**
AppEn	81.0	72.6-89.3	80.0	68.3-91.7	82.1	70.0-94.1	83.7	72.7-94.8
BEn	70.2	60.5-80.0	60.0	45.7-74.3	82.1	70.0-94.1	79.4	65.8-93.0
FEn	72.6	63.1-82.2	80.0	68.3-91.7	64.1	49.0-79.2	72.0	59.6-84.4
PEn	73.8	64.4-83.2	73.3	60.4-86.3	74.4	60.7-88.1	76.7	64.1-89.4
SmpEn	78.6	69.8-87.3	77.8	65.6-89.9	79.5	66.8-92.2	81.4	69.8-93.0
SpEn	82.1	74.0-90.3	80.0	68.3-91.7	84.6	73.3-95.9	85.7	75.1-96.3
Soft-Voting	***90.5***	***84.2-96.8***	***88.9***	***79.7-98.1***	***92.3***	***83.9-100.0***	***93.0***	***85.4-100.0***

Table 5. Subject-level classification performance with 95% confidence intervals of Dataset 2

	Performance Metrics							
	ACC (%)		**SEN (%)**		**SPE (%)**		**PPV (%)**	
	Value	**95% CI**	**Value**	**95% CI**	**Value**	**95% CI**	**Value**	**95% CI**
AppEn	92.9	83.3-100.0	85.7	67.4-100.0	100.0	100.0-100.0	100.0	100.0-100.0
BEn	92.9	83.3-100.0	92.9	79.4-100.0	92.9	79.4-100.0	92.9	79.4-100.0
FEn	89.3	77.8-100.0	92.9	79.4-100.0	85.7	67.4-100.0	86.7	69.5-100.0
PEn	96.4	89.6-100.0	100.0	100.0-100.0	92.9	79.4-100.0	93.3	80.7-100.0
SmpEn	96.4	89.6-100.0	100.0	100.0-100.0	92.9	79.4-100.0	93.3	80.7-100.0
SpEn	78.6	63.4-93.8	71.4	47.8-95.1	85.7	67.4-100.0	83.3	62.2-100.0
Soft-Voting	***100.0***	***100.0-100.0***	***100.0***	***100.0-100.0***	***100.0***	***100.0-100.0***	***100.0***	***100.0-100.0***

The performance measures at the subject level and their associated 95% confidence intervals are presented in Tables 4 and 5, for Dataset 1 and Dataset 2, respectively. The suggested soft voting ensemble performs better than individual entropy-based feature maps across every evaluation metric, as shown by the findings in both tables. The model performs nearly perfectly in Dataset 2, with Soft Voting achieving 100% accuracy, sensitivity, specificity, and precision, along with extremely tight confidence intervals, indicating extremely consistent and dependable subject-level predictions. Although individual entropy approaches like PEn and SmpEn also exhibit good performance (ACC = 96.4%), their performance is less consistent across measures, and their confidence intervals are slightly broader than those of the ensemble. Dataset 1, on the other hand, offers a more challenging circumstance, with individual entropy approaches performing moderately (ACC ranging from ~70%

to 82%), indicating higher variability and poorer discriminative power. By reaching 90.5% accuracy with balanced sensitivity (88.9%) and specificity (92.3%) and comparatively lower confidence intervals as compared to individual methods, the soft voting strategy, however, greatly enhances overall performance. These results imply that more balanced and consistent classification performance is typically achieved by integrating complementary nonlinear entropy features. However, Dataset 1's larger confidence ranges suggest greater uncertainty, which is probably related to subject variability and dataset heterogeneity.

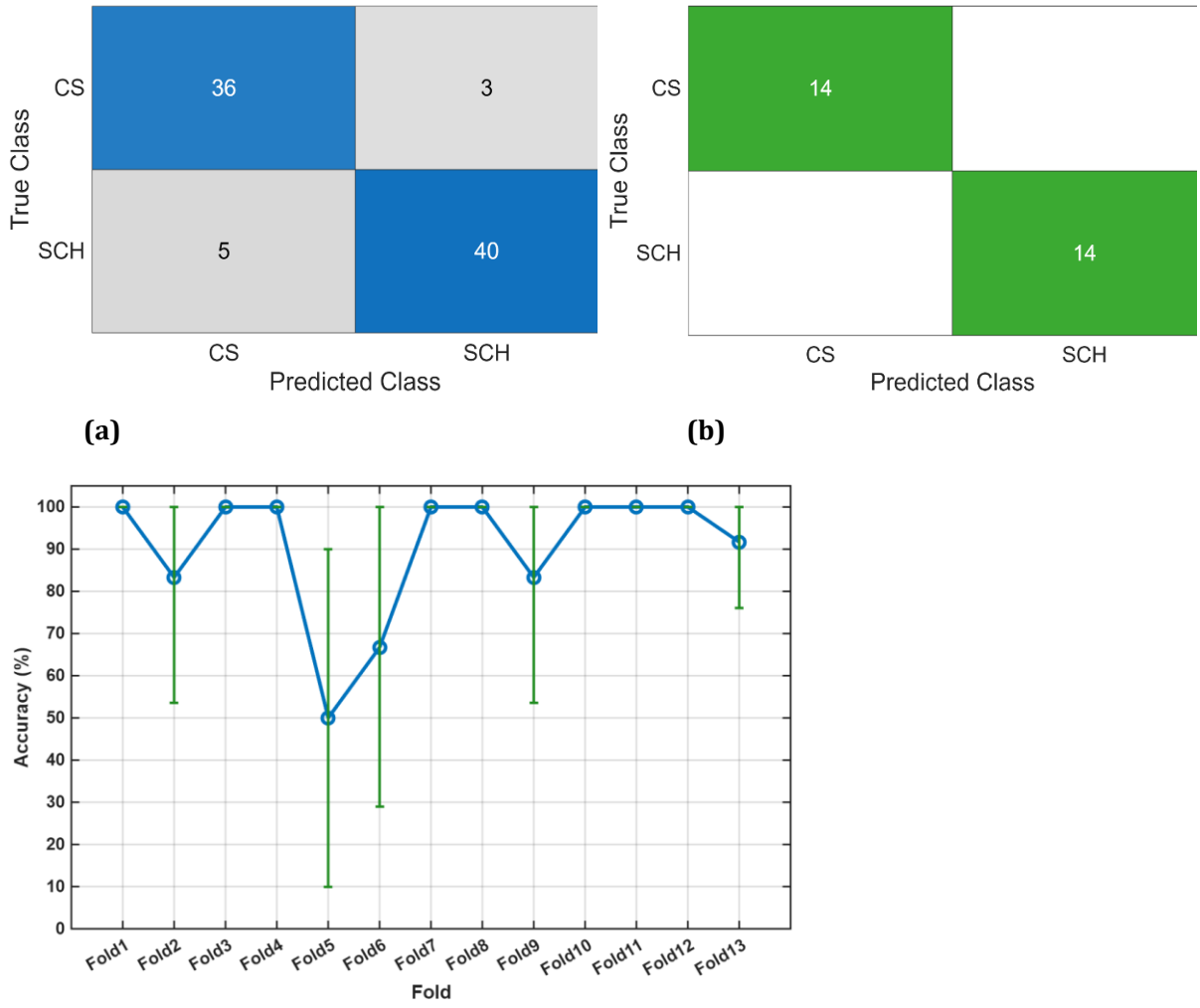

Figure 7. Subject-level confusion matrices of the proposed soft voting ensemble model for (a) Dataset 1 and (b) Dataset 2; (c) Fold-wise subject-level accuracy with 95% confidence intervals of the proposed soft voting ensemble model for Dataset 1.

Figure 7 displays the subject-level confusion matrices derived from the proposed soft voting ensemble method for both datasets. 36 CS and 40 SCH participants are accurately categorised by the model in Dataset 1 (Figure 7 (a)), but 3 CS are incorrectly labelled as SCH and 5 SCH as CS. This indicates a reasonable level of classification performance with some degree of overlap between classes. Dataset 2 (Figure 7 (b)), on the other hand, shows flawless classification performance, with all 14 CS and 14 SCH patients accurately identified and no misclassifications. This outcome shows great model performance under LOGO-CV evaluation and highly discriminative EEG patterns within Dataset 2. Overall, the confusion matrices show how well the soft voting approach improves subject-level classification while also highlighting variations in dataset separability and complexity. The fold-wise

classification accuracy and 95% confidence intervals for Dataset 1 utilising the soft voting ensemble framework are shown in Figure 7 (c). The findings emphasise the challenges of LOGO-CV-based schizophrenia classification by demonstrating variability between folds. While some folds (Fold1, Fold3, Fold4, Fold7, Fold8, Fold10, Fold11, and Fold12) demonstrated near-perfect accuracy, others (Fold5 and Fold6) showed susceptibility to subject variability with poorer accuracy and broader confidence ranges. Because Dataset 1 is heterogeneous, the performance varies according to the subject distribution in each fold. Furthermore, Dataset 2 produced consistent confidence intervals with 100% accuracy across all folds; a comparable fold-wise confidence interval analysis is not offered. Because of the small sample size and bootstrap-based confidence interval estimate process, some subject-level evaluations may result in incredibly narrow or perfect confidence intervals. As a result, it is important to evaluate these intervals carefully.

3.3. External Validation Using an Independent Dataset An additional independent dataset (Dataset 3) is included to further assess the behaviour of the suggested framework under various recording situations. 153 people (76 SCH, Age: 40 ± 12; 77 CS, Age: 38 ± 13) provided 19-channel EEG recordings for this dataset (Note: Although the reference article states there are 16 channels, the dataset contains 19 channels of EEG data). The EEG recordings are acquired using two different systems: Contec KT-2400 devices with a sampling frequency of 200 Hz and BrainMaster Discovery24-E systems operating at 256 Hz under harmonized recording protocols [31]. To maintain methodological consistency, the same preprocessing pipeline, entropy-based feature map creation, and ResNet50-based soft voting technique are used without any changes. Similarly, LOGO-CV is applied for that dataset. In Dataset 3, the 77 CS and 76 SCH participants are divided into 15 groups for LOGO-CV-based evaluation (Fold1–Fold14: 5 CS and 5 SCH subjects per fold; Fold15: remaining 7 CS and 6 SCH subjects). During each iteration, one group is used for testing and the remaining groups are used for training. In each iteration, a 15-fold cross-validation is performed, utilising one group for testing and the remaining groups for training.

The suggested soft voting ensemble technique consistently outperformed the individual entropy-based EEG-FM models on Dataset 3 at the segment level, as shown by the results in Table 6. The classification performance of individual entropy metrics is moderate, with ACC values typically falling between roughly 69% and 75%. The discrimination capability of PEn (74.8%) and SpEn (71.3%) is the strongest among the individual entropy models, whereas the other entropy measures performed similarly but marginally worse. The suggested soft voting ensemble, on the other hand, performed the best overall, with an ACC of 83.6% ± 12.6, SEN of 77.5% ± 24.4, SPE of 82.7% ± 18.8, and PPV of 84.0% ± 13.6. Furthermore, the ensemble model yielded the lowest FDR (16.0% ± 13.6), suggesting a significant decrease in false positive predictions. Compared to single-feature entropy models, these results imply that incorporating complementary nonlinear entropy representations increases classification stability and lowers prediction errors.

Table 6. The suggested soft voting ensemble model's classification performance on Dataset 3.

	Performance Metrics (%)				
	ACC	**SEN**	**SPE**	**PPV**	**FDR**
AppEn	69.2±14.1	66.6±19.2	67.8±22.1	67.4±20.8	32.6±20.8
BEn	69.7±8.6	68.8±20.9	66.9±15.5	64.1±15.6	35.9±15.6
FEn	69.8±13.6	65.1±18.7	69.1±23.8	70.3±17.9	29.8±17.9
PEn	74.8±10.1	74.8±21.4	70.9±16.7	69.7±15.4	30.3±15.4
SmpEn	68.8±12.8	61.3±27.1	70.0±18.9	63.6±19.9	36.4±19.9
SpEn	71.3±12.2	70.9±24.4	66.9±16.6	64.1±18.7	35.9±18.7
Soft-Voting	***83.6±12.6***	***77.5±24.4***	***82.7±18.8***	***84.0±13.6***	***16.0±13.6***

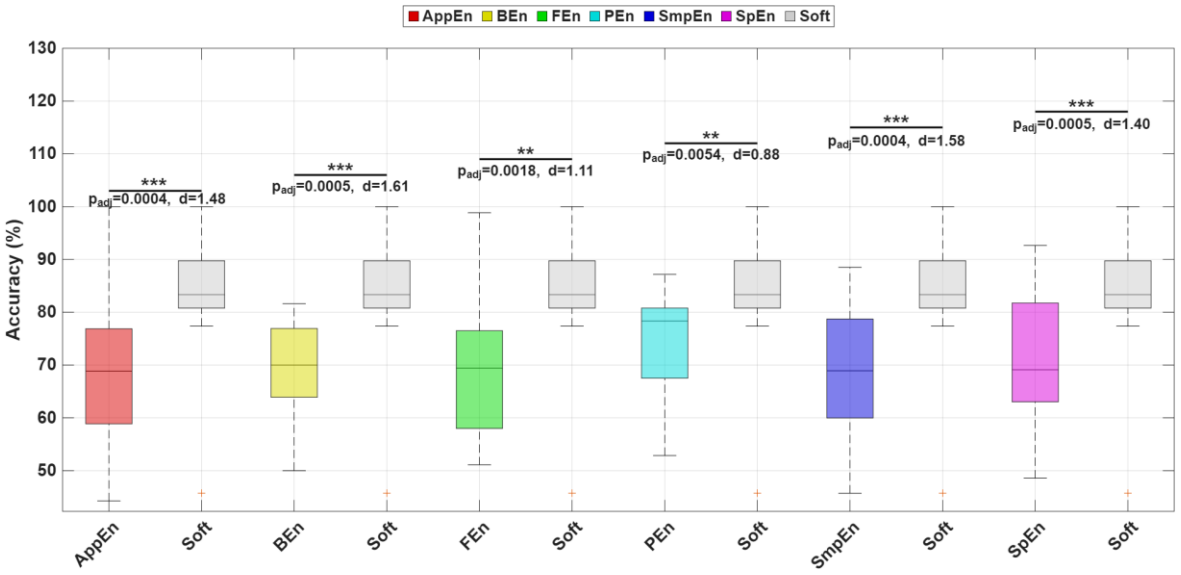

Figure 8. Pairwise boxplot comparison of classification accuracy between individual entropy-based EEG-FM models and the proposed soft voting ensemble approach for the Dataset 3. Statistical significance is evaluated using the Wilcoxon signed-rank test with BH-FDR correction (* adjusted p < 0.05, ** adjusted p < 0.01, *** adjusted p < 0.001, n.s.: not significant).

Figure 8 shows a pairwise boxplot comparison of classification accuracy between the individual entropy-based EEG-FM models and the suggested soft voting ensemble framework using the Wilcoxon signed-rank test with BH-FDR correction in order to further examine the statistical significance of these improvements. After FDR correction, the soft voting technique outperformed all individual entropy models in a statistically significant way. Specifically, AppEn (adjusted p = 0.0004, d = 1.48), BEn (adjusted p = 0.0005, d = 1.61), SmpEn (adjusted p = 0.0004, d = 1.58), and SpEn (adjusted p = 0.0005, d = 1.40) showed significant effect sizes. While PEn also showed statistically significant improvement (adjusted p = 0.0054), the corresponding effect size is much less (d = 0.88), suggesting a more moderate gain in comparison to the other entropy metrics.

The results summarised in Table 7 further support the efficacy of the suggested ensemble framework at the subject level, which offers a more clinically useful evaluation. Together with SEN of 77.6%, SPE of 87.0%, and PPV of 85.5%, the soft voting model obtained an ACC of 82.4% with a 95% confidence interval of 76.3–88.4. The ensemble framework yielded more balanced performance across all evaluation metrics when compared to the individual entropy-based approaches. Additionally, compared to the individual entropy measures, the soft voting model's confidence intervals are typically shorter and more consistent, suggesting lower variability and increased subject-level dependability.

Table 7. Subject-level classification performance with 95% confidence intervals of Dataset 3

	Performance Metrics							
	ACC (%)		**SEN (%)**		**SPE (%)**		**PPV (%)**	
	Value	**95% CI**	**Value**	**95% CI**	**Value**	**95% CI**	**Value**	**95% CI**
AppEn	73.9	66.9-80.8	67.1	56.5-77.7	80.5	71.7-89.4	77.3	67.2-87.4
BEn	71.9	64.8-79.0	63.2	52.3-74.0	80.5	71.7-89.4	76.2	65.7-86.7
FEn	67.3	59.9-74.8	60.5	49.5-71.5	74.0	64.2-83.8	69.7	58.6-80.8
PEn	75.2	68.3-82.0	69.7	59.4-80.1	80.5	71.7-89.4	77.9	68.1-87.8
SmpEn	68.6	61.3-76.0	53.9	42.7-65.2	83.1	74.7-91.5	75.9	64.5-87.3
SpEn	75.8	69.0-82.6	69.7	59.4-80.1	81.8	73.2-90.4	79.1	69.4-88.8
Soft-Voting	***82.4***	***76.3-88.4***	***77.6***	***68.3-87.0***	***87.0***	***79.5-94.5***	***85.5***	***77.2-93.8***

Figure 9 (a) displays the subject-level confusion matrix of the suggested soft voting model for Dataset 3. The soft voting model's subject-level confusion matrix on Dataset 3 (TP = 59, TN = 67, FP = 10, FN = 17) shows that both classes' classification performance is fairly balanced. While misclassifying 10 CS as SCH and 17 SCH as CS, the model properly recognised 67 CS subjects and 59 SCH subjects. Overall, the confusion matrix indicates that the proposed soft voting ensemble provides relatively balanced subject-level classification performance, with a slightly higher capability for identifying control subjects compared with schizophrenia cases. The comparatively low number of false positives and moderate number of false negatives suggest that the framework maintains a reasonable balance between sensitivity and specificity under Dataset 3 conditions. However, the findings should be interpreted cautiously because EEG classification performance may be influenced by subject variability, dataset heterogeneity, and recording differences. The incorporation of Dataset 3 provides an additional evaluation of the proposed framework under different subject distributions and EEG acquisition conditions, rather than definitive evidence of clinical generalizability. The fold-wise subject-level classification accuracy and matching 95% confidence intervals for Dataset 3 utilising the suggested soft voting ensemble architecture are shown in Figure 9 (b). Because LOGO-CV-based EEG classification under heterogeneous recording conditions is challenging, the results demonstrate noticeable variability across folds. With accuracies close to 100% and comparatively narrow confidence intervals, several folds (Fold3, Fold7, Fold8, Fold9, Fold13, and Fold14) demonstrated near-perfect classification performance, suggesting persistent differentiation between schizophrenia and control participants for those subject groups. On the other hand, some folds, especially Fold5, Fold6, and Fold11, showed broader confidence ranges and worse accuracy, indicating greater sensitivity to inter-subject EEG variability and subject composition as well as increased uncertainty. These variations are to be expected in LOGO-CV assessments, particularly when individual folds contain a small number of subjects. The fold-wise analysis indicates that the suggested soft voting technique offers comparatively balanced and constant subject-level performance across Dataset 3, despite the observed heterogeneity between folds. Confidence intervals highlight the impact of subject composition and diverse EEG recording settings on classification performance and provide a more explicit presentation of statistical uncertainty.

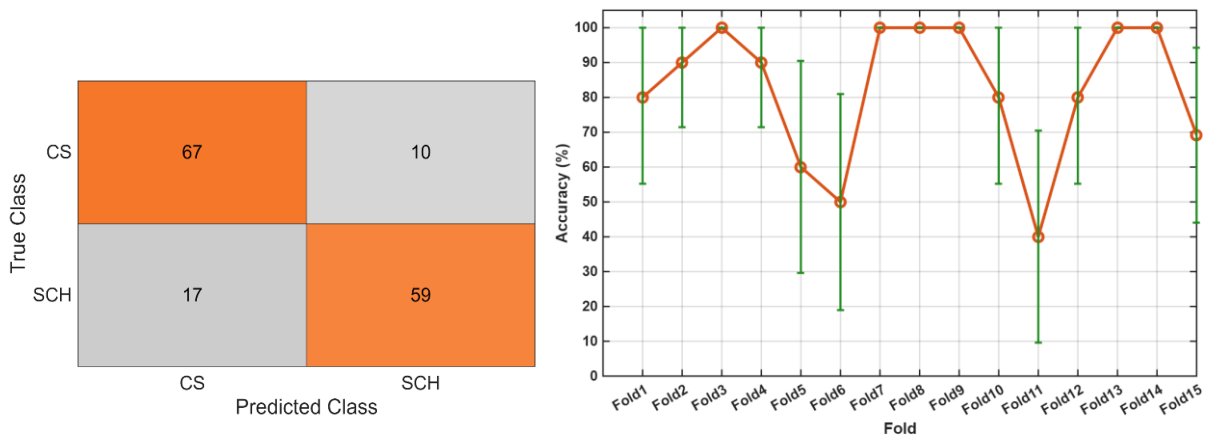

Figure 9. (a) Subject-level confusion matrix, (b) fold-wise subject-level accuracy with 95% confidence intervals of the proposed soft voting ensemble model for Dataset 3.

3.4. Ablation Study and Conventional Machine Learning Baselines An additional ablation study is carried out using traditional machine learning classifiers under the same fold configurations to further explore whether the observed classification improvements primarily stem from the entropy-based EEG-FM representations or from the suggested ensemble learning strategy. In this investigation, KNN-based classification is carried out independently, utilising both the suggested soft voting technique and individual handcrafted entropy features. To give a more thorough review of fold-wise variability and classification stability across Datasets 1, 2, and 3, segment-level and subject-level

assessments with 95% confidence intervals are also reported.

Table 8. Segment-level ablation analysis of entropy-based features and soft voting strategy using the KNN classifier across Dataset 1, Dataset 2, and Dataset 3 (mean ± standard deviation and 95% confidence intervals).

	Dataset 1		**Dataset 2**		**Dataset 3**	
Feature	**ACC**	**CI (% 95)**	**ACC**	**CI (% 95)**	**ACC**	**CI (% 95)**
AppEn	51.8 ± 3.6	47.8 - 55.7	57.82 ± 11.6	55.96 - 59.7	50.6 ± 7.5	47.8 - 53.4
BEn	49.9 ± 4.2	45.9 - 53.9	59.77 ± 15.6	58.01 - 61.5	55.5 ± 13.4	52.7 - 58.2
FEn	52.4 ± 5.1	48.4 - 56.4	60.12 ± 15.1	58.34 - 61.9	52.1 ± 11.0	49.3 - 54.9
PEn	50.0 ± 3.2	46.1 - 53.9	58.25 ± 15.2	56.47 - 60.0	52.6 ± 10.7	49.9 - 55.4
SmpEn	52.8 ± 2.9	48.8 - 56.7	57.37 ± 11.0	55.52 - 59.2	51.9 ± 8.8	49.1 - 54.7
SpEn	50.2 ± 3.4	46.2 - 54.2	50.73 ± 5.7	48.83 - 52.7	56.4 ± 13.3	53.8 - 59.1
Soft-Voting	52.7 ± 3.9	48.7 - 56.7	61.76 ± 17.7	60.04 - 63.4	55.6 ± 11.9	52.8 - 58.3

Table 9. Subject-level ablation analysis of entropy-based features and soft voting strategy using the KNN classifier across Dataset 1, Dataset 2, and Dataset 3 (mean ± standard deviation and 95% confidence intervals).

	Dataset 1		**Dataset 2**		**Dataset 3**	
Feature	**ACC**	**CI (% 95)**	**ACC**	**CI (% 95)**	**ACC**	**CI (% 95)**
AppEn	55.1 ± 15.8	22.4 - 87.8	67.9 ± 24.9	35.71 - 100.0	50.9 ± 8.0	21.4 - 80.5
BEn	45.5 ± 20.6	16.7 - 78.5	67.9 ± 24.9	35.71 - 100.0	58.9± 19.6	31.1 - 84.7
FEn	60.9 ± 17.8	28.2 - 89.1	60.7 ± 34.9	35.71 - 85.7	55.6 ± 12.9	26.4 - 83.3
PEn	45.5 ± 12.1	14.1 - 78.2	64.3 ± 23.4	28.57 - 100.0	53.4 ± 14.5	24.9 - 81.6
SmpEn	53.2 ± 13.4	20.5 - 84.6	64.3 ± 23.4	28.57 - 100.0	55.4 ± 13.1	26.1 - 82.9
SpEn	44.9 ± 14.3	13.5 - 78.2	64.3 ± 36.3	42.86 - 85.7	61.3 ± 18.8	33.1 - 86.8
Soft-Voting	60.3 ± 17.4	30.1 - 89.4	67.9 ± 24.9	35.71 - 100.0	56.9 ± 12.2	27.5 - 85.1

Segment-level classification results employing individual handcrafted entropy features are comparatively mild across all datasets, as Table 8 illustrates. The suggested soft voting approach typically offers consistent classification performance, even though some entropy features show relatively greater discriminating skills in particular datasets. Specifically, Dataset 2 shows significantly better segment-level accuracies than Datasets 1 and 3, while Datasets 1 and 3 show more difficult classification conditions. Additionally, fold-wise variability and dataset heterogeneity—particularly in Datasets 2 and 3—are indicated by the provided confidence intervals. In comparison to segment-level analysis, Table 9 shows that subject-level evaluation typically increases classification performance, indicating that majority voting across EEG segments improves prediction consistency at the patient level. In several datasets, the proposed soft voting strategy either achieves the highest or among the highest subject-level accuracies while also providing comparatively balanced confidence intervals. The ensemble-based soft voting strategy typically yields more stable and consistent behaviour across different subject distributions and recording situations, even when some individual entropy features occasionally achieve comparable performance levels.

Overall, the ablation analysis indicates that using ensemble learning to combine complementary entropy-based handcrafted features improves classification stability; nevertheless, the suggested EEG-FM + ResNet50 framework continues to perform better than traditional handcrafted-feature-based machine learning techniques.

Table 10. Summary of relevant studies using the same dataset.

Study	Dataset	Method	Results
[2]	Dataset 2	MEMD	ACC: 93%
[3]	49 SCH, 32 CS 64-channel EEG (Dataset 4)	LCBWFB	CV; ACC: 96.84%, SEN: 95.95%, SPE: 96.97%, LOSO; ACC: 93.77%, SEN: 93.92%, SPE: 92.30%,
[4]	Dataset 1	DM-AEO, EMD, VMD, DWT, and LCOFB	ACC: 99.26%
[5]	Dataset 2 Dataset 4	LAP	ACC: 98.62% for Dataset 2 ACC: 99.24% for Dataset 4
[6]	Dataset 1 Dataset 2	3DCNN	CV; 98.06% ACC for Dataset 1 99.46% ACC for Dataset 2 LOSO; 92.67% ACC for Dataset 1 96.04% ACC for Dataset 2
[7]	Dataset 2	IF and SVM	ACC: 98.9%
[8]	Dataset 1 Dataset 2 Dataset 4	PVMD and BLIMFs	ACC: 96.91% for Dataset 1 ACC: 99.98% for Dataset 2 ACC: 99.29% for Dataset 4
[10]	Dataset 1	Complexity of Continuous Functions	ACC: 84.5%
[11]	Dataset 1 Dataset 2	HLV and SLBP	ACC: 92.85% for Dataset 1 ACC: 99.36% for Dataset 2
[12]	Dataset 2	WST, CWT, and DWT	ACC: 97.98%, SEN: 98.2%, SPE: 97.72%
[13]	Dataset 2	CNN	CV: 98% ACC LOSO: 90% ACC
[14]	Dataset 2	CNN	ACC: 93%
[15]	Dataset 1 Dataset 2	CAE-XGB SS-ENB	ACC: 95.3% for CAE-XGB ACC: 97% for SS-ENB
[16]	Dataset 1	Functional Connectivity	ACC: 97.74%
Proposed Method	***Dataset 1*** ***Dataset 2*** ***Dataset 3***	***Entropy-based EEG feature maps and CNN***	***LOGO;*** ***ACC: 81.4%, SEN: 81.2%, SPE: 81.6% for Dataset 1*** ***ACC: 95.7%, SEN: 94.7%, SPE: 96.7% for Dataset 2*** ***ACC: 83.6%, SEN: 77.5%, SPE: 82.7% for Dataset 3***

The findings of this work show that using a ResNet50-based CNN architecture in conjunction with a soft voting ensemble technique for the automated identification of schizophrenia from entropy-based EEG-FMs is highly effective. In order to capture the non-linear dynamics of brain activity linked to the disease, the suggested model converts one-dimensional EEG signals into multi-dimensional spatial-temporal representations (EEG-FM). A crucial discovery of this study is the notable enhancement of the performance of the soft voting mechanism in comparison to individual entropy-based results. Individual entropy-based EEG-FMs, such as AppEn, BEn, and SmpEn, showed lower median accuracies and significant variance, especially in Dataset 1 (60% to 75%). In comparison, the suggested ensemble approach combined these measurements and produced consistent average accuracies of 95.72% for Dataset 2 and 81.38% for Dataset 1.

On the other hand, the findings from the study of the combined datasets (given in Table 3) offer significant new information about how entropy-based EEG-FMs behave in varied environments. The suggested framework is subjected to a wider range of EEG features, including differences in age, recording techniques, and illness stages, by combining the training and test folds from both datasets. Given the increasing variability and distributional disparities between the datasets, the overall classification performance under this context showed a moderate decline in comparison to evaluations of individual datasets. Despite these challenges, the soft-voting-based ensemble method generally produced more balanced classification performance than individual entropy-based models by integrating complementary nonlinear EEG representations (ACC of 70.3% ± 11.4, SEN of 76.2% ± 13.7, SPE of 64.4% ± 22.4, and PPV of 70.2% ± 12.6). However, these findings should be interpreted cautiously because substantial dataset heterogeneity and domain differences may influence overall classification performance. In particular, the ensemble approach greatly improved entropy metrics like FEn, SmpEn, and SpEn, which performed poorly when employed separately. Rather than implying biological equivalence between datasets, the combined dataset analysis serves as an exploratory assessment of the model's behavior under heterogeneous data distributions. The observed decrease in performance highlights the sensitivity of EEG-based models to dataset-specific factors and domain variability. The encouraging outcomes of the experiment demonstrate that using entropy-based EEG-FMs to analyse the spatial and temporal features of the EEG in people with schizophrenia may assist in detection. Additionally, the experimental findings from further investigations that used the same dataset are presented in Table 10. This table summarises that numerous prior studies demonstrating elevated accuracies on Dataset 1 utilised traditional CV methodologies, wherein EEG segments from the same subject may be present in both training and testing sets. These configurations may result in too optimistic performance predictions and inadvertently create subject-specific information leaks. The suggested approach, on the other hand, mainly uses a LOGO-CV structure to provide rigorous subject independence across the training and testing stages. The suggested approach produced less accuracy than Study [6] (92.67%), but it reached balanced performance on Dataset 1 (ACC: 81.4%, SEN: 81.2%, SPE: 81.6%) under this more clinically realistic evaluation technique. However, under LOGO-CV evaluation conditions, the suggested framework showed resilient and competitive behaviour. Additionally, the suggested approach outperformed the LOSO-based results given in Study [6] (96.04% ACC) and in Study [13] (90% ACC), with an accuracy of 95.7% on Dataset 2. The suggested soft voting ensemble demonstrated enhanced dependability and balanced classification performance in addition to good accuracy and a low false discovery rate (FDR = 3.29%). The primary contribution of this work is the combination of complementary entropy-based EEG representations using decision-level ensemble learning, even though the suggested framework does not consistently reach the maximum accuracy when compared with all previous studies. Additionally, the proposed methodology is evaluated under LOGO-CV-based conditions using three independent EEG datasets with distinct demographic and recording characteristics, providing a broader assessment of entropy-based EEG classification behavior under heterogeneous data conditions.

4. CONCLUSION

This study presents a ResNet50-based ensemble learning model that uses an entropy-based EEG feature map as input for the automatic detection of SCH using resting-state EEG signals. The suggested method efficiently combines complementary information that cannot be captured by a single feature representation by converting several entropy measures into spatially structured inputs and combining their individual performance through a soft voting procedure. The experimental findings showed that on both datasets, the ensemble technique consistently performed better than individual entropy-based models. Specifically, the suggested framework maintained a balanced trade-off between sensitivity and specificity, achieving average accuracies of 81.4% and 95.7% for Datasets 1 and 2, respectively. Furthermore, statistical investigations verified that the ensemble model significantly outperforms a number of separate entropy-based EEG-FM models, especially for features with low standalone discriminative ability. The combined dataset evaluation demonstrated that, although classification performance decreased under heterogeneous EEG distributions and domain-shift conditions, the proposed soft voting ensemble generally produced more balanced results than individual entropy-based EEG-FM models. Furthermore, the additional evaluation conducted on Dataset 3 suggests that

the proposed entropy-based EEG-FM and ensemble learning framework can provide relatively stable classification performance under different recording conditions and subject distributions. Overall, the findings indicate that the proposed framework may serve as a supportive computer-aided EEG classification approach for schizophrenia-related EEG analysis rather than a definitive clinical diagnostic system.

5. LIMITATIONS AND FUTURE WORK

The suggested soft voting-based framework integrating entropy-based EEG feature maps (EEG-FMs) and ResNet50 showed competitive performance for schizophrenia-related EEG classification, despite several drawbacks. Three publicly accessible EEG datasets related to schizophrenia are used to assess the suggested methodology; however, the datasets range significantly in terms of population size, demographic composition, recording settings, and clinical features. Age, gender, medication status, disease stage, length of recording, and the number of EEG channels are some of the factors that might significantly impact entropy characteristics and, in turn, model performance.

Additionally, although the proposed framework demonstrated comparatively balanced performance under heterogeneous conditions, the combined dataset analysis should not be interpreted as evidence of biological equivalency between datasets or direct clinical generalizability. Similarly, while Dataset 2 achieved outstanding subject-level classification performance, this result should be interpreted cautiously due to the relatively limited sample size of the dataset. Therefore, the reported findings should be considered experimental observations obtained under controlled dataset conditions rather than definitive evidence of clinical applicability or generalizability.

Several potential future research directions can be taken into consideration based on the study's findings:

- Increasing the size and diversity of subject populations in EEG datasets related to schizophrenia may enhance the stability and wider applicability of entropy-based deep learning techniques.

- A more thorough evaluation framework might be produced by adding more clinical and demographic factors, such as medication status, disease progression, and symptom severity.

- The use of six independent ResNet50 models increases computational complexity and training cost. Therefore, future studies may investigate lighter CNN architectures and feature-level fusion techniques for more computationally efficient implementations.

- To better evaluate partition sensitivity and statistical stability, different validation procedures like LOSO or leave-pair-out assessment, repeated subject-independent partitioning strategies, and repeated LOGO-CV designs might be investigated.

- Future research may also include treatment-stage analysis, risk prediction applications, and longitudinal monitoring, as the current study primarily focused on EEG classification related to schizophrenia.

- Lastly, to effectively control inter-dataset variability and distributional differences, future research should concentrate on validating the suggested framework using bigger, more standardised, multi-center EEG datasets and investigating sophisticated domain adaptation strategies.

Peer-review: External, Independent.

Acknowledgements:

The author would like to thank all institutions and publicly available data sources that contributed to the development of the dataset used in this study. No additional acknowledgements are declared.

Declarations:

1. Statement of Originality:

This work is original.

2. Author Contributions:

Concept: OKC; ***Conceptualization:*** OKC; ***Literature Search:*** OKC; ***Data Collection:*** OKC; ***Data Processing:*** OKC; ***Analysis:*** OKC; ***Writing – original draft:*** OKC; ***Writing – review & editing:*** OKC.

3. Ethics approval:

Not applicable.

4. Funding/Support:

This work has not received any funding or support.

5. Competing Interests:

The author declares no competing interests.

6. GenAI Usage Statement:

No GenAI tools were used at any stage of the study.

7. Sustainable Development Goals:

REFERENCES

[1] Jangde, A. S., & Verma, G. K. 2026. EEG-Based Schizophrenia Classification Using Attention-Integrated Deep Convolutional Networks. Psychiatry Research: Neuroimaging, 112138.

[2] Krishnan, P. T., Raj, A. N. J., Balasubramanian, P., Chen, Y. 2020. Schizophrenia detection using Multivariate Empirical Mode Decomposition and entropy measures from multichannel EEG signal. Biocybernetics and Biomedical Engineering, 40(3), 1124-1139.

[3] Puri, D. V., Kachare, P. H., Al–Shourbaji, I., Jabbari, A., Kirner, R., Alameen, A. 2025. Design of efficient bi-orthogonal wavelets for EEG-based detection of Schizophrenia. Engineering Science and Technology, an International Journal, 68, 102090.

[4] Puri, D. V., Kachare, P. H., Sangle, S. B., Al-Shourbaji, I., Jabbari, A., Kirner, R. 2026. A hybrid dwarf mongoose and artificial ecosystem-based EEG channel optimization for Schizophrenia detection. Biomedical Signal Processing and Control, 112, 108513.

[5] Agarwal, M., Singhal, A. 2023. Fusion of pattern-based and statistical features for Schizophrenia detection from EEG signals. Medical Engineering & Physics, 112, 103949.

[6] Guo, Z., Wang, J., Jing, T., & Fu, L. 2024. Investigating the interpretability of schizophrenia EEG mechanism through a 3DCNN-based hidden layer features aggregation framework. Computer Methods and Programs in Biomedicine, 247, 108105.

[7] Das, K., Pachori, R. B. 2021. Schizophrenia detection technique using multivariate iterative filtering and multichannel EEG signals. Biomedical Signal Processing and Control, 67, 102525.

[8] Parija, S., Sahani, M., Rout, S. K. 2025. Robust broad learning system with parametrized variational mode decomposition for schizophrenia diagnosis. Engineering Applications of Artificial Intelligence, 158, 111294.

[9] Khare, S. K., Bajaj, V. 2022. A hybrid decision support system for automatic detection of Schizophrenia using EEG signals. Computers in biology and medicine, 141, 105028.

[10] Piryatinska, A., Darkhovsky, B., Kaplan, A. 2017. Binary classification of multichannel-EEG records based on the ϵ-complexity of continuous vector functions. Computer methods and programs in biomedicine, 152, 131-139.

[11] Kumar, T. S., Rajesh, K. N., Maheswari, S., Kanhangad, V., Acharya, U. R. 2023. Automated Schizophrenia detection using local descriptors with EEG signals. Engineering Applications of Artificial Intelligence, 117, 105602.

[12] Gosala, B., Kapgate, P. D., Jain, P., Chaurasia, R. N., Gupta, M. 2023. Wavelet transforms for feature engineering in EEG data processing: An application on Schizophrenia. Biomedical Signal Processing and Control, 85, 104811.

[13] Hassan, F., Hussain, S. F., Qaisar, S. M. 2023. Fusion of multivariate EEG signals for schizophrenia detection using CNN and machine learning techniques. Information Fusion, 92, 466-478.

[14] Lillo, E., Mora, M., Lucero, B. 2022. Automated diagnosis of schizophrenia using EEG microstates and Deep Convolutional Neural Network. Expert Systems with Applications, 209, 118236.

[15] Mudavath, U. K. N., Ahamed, S. R., Wong-Lin, K. 2025. Toward Interpretable Schizophrenia Detection from EEG Using Autoencoder and EfficientNet. medRxiv, 2025-05.

[16] Shen, M., Wen, P., Song, B., Li, Y. 2023. Automatic identification of schizophrenia based on EEG signals using dynamic functional connectivity analysis and 3D convolutional neural network. Computers in Biology and Medicine, 160, 107022.

[17] Bagherzadeh, S., Shahabi, M. S., & Shalbaf, A. 2022. Detection of schizophrenia using hybrid of deep

learning and brain effective connectivity image from electroencephalogram signal. Computers in Biology and Medicine, 146, 105570.

[18] Jahmunah, V., Oh, S. L., Rajinikanth, V., Ciaccio, E. J., Cheong, K. H., Arunkumar, N., & Acharya, U. R. 2019. Automated detection of schizophrenia using nonlinear signal processing methods. Artificial intelligence in medicine, 100, 101698.

[19] Sun, J., Cao, R., Zhou, M., Hussain, W., Wang, B., Xue, J., & Xiang, J. 2021. A hybrid deep neural network for the classification of schizophrenia using EEG Data. Scientific Reports, 11(1), 4706.

[20] Kim, K., Duc, N. T., Choi, M., & Lee, B. 2021. EEG microstate features for schizophrenia classification. PloS one, 16(5), e0251842.

[21] Nikhil Chandran, A., Sreekumar, K., & Subha, D. P. (2020). EEG-based automated detection of schizophrenia using long short-term memory (LSTM) network. In Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019 (pp. 229-236). Singapore: Springer Singapore.

[22] M.V. Lomonosov Moscow State University. EEG of healthy adolescents and adolescents with symptoms of schizophrenia http://brain.bio.msu.ru/eeg_schizophrenia.htm (Erişim Tarihi: 15.06.2025)

[23] Olejarczyk, E., & Jernajczyk, W. 2017. Graph-based analysis of brain connectivity in schizophrenia. PloS one, 12(11), e0188629.

[24] Fontaine, S. A. S., Amigó, J. M. 2025. Applications of entropy in data analysis and machine learning: A review. arXiv preprint arXiv:2503.02921.

[25] Shen, F., Dai, G., Lin, G., Zhang, J., Kong, W., & Zeng, H. 2020. EEG-based emotion recognition using 4D convolutional recurrent neural network. Cognitive Neurodynamics, 14(6), 815-828.

[26] Kim, S., Kim, T. S., & Lee, W. H. 2022. Accelerating 3D convolutional neural network with channel bottleneck module for EEG-based emotion recognition. Sensors, 22(18), 6813.

[27] Li, Y., Huang, J., Zhou, H., & Zhong, N. 2017. Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. Applied Sciences, 7(10), 1060.

[28] Cura, O. K., Akan, A., & Atli, S. K. 2024. Detection of Attention Deficit Hyperactivity Disorder based on EEG feature maps and deep learning. Biocybernetics and Biomedical Engineering, 44(3), 450-460.

[29] Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. 2023. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability, 15(7), 5930.

[30] Reddy, J., Mundra, S., Mundra, A. 2024. Ensembling Deep Learning Models for Fake News Classification. Procedia Computer Science, 235, 2766-2774.

[31] Mosaku, S. K., Olateju, E. O., Ayodele, K. P., Akinsulore, A., Ajiboye, P. O., Oloniniyi, O. I., ... & Oyekunle, W. A. 2025. An open-access EEG dataset from indigenous African populations for

schizophrenia research. Data in Brief, 111934.