APSIPA Transactions on Signal and Information Processing, 2024, 13, e33

Original Paper

Riemannian Manifold-based Epileptic Seizure Detection Using Transfer Learning and Artifact Rejection Techniques

Kazi Mahmudul Hassan[1], Xuyang Zhao[1], Hidenori Sugano[2] and Toshihisa Tanaka[1*]

[1] *Graduate School of Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan*
[2] *Department of Neurosurgery, Juntendo University School of Medicine, Tokyo, Japan*

ABSTRACT

Improvement in technology and the availability of electroencephalogram (EEG) data have raised the demand for automated seizure detection in long-term EEG recordings. This study proposes a framework to automate seizure detection from long-term EEG by combining anomaly detection, artifact removal, and seizure detection techniques, along with Riemannian manifold and transfer learning approaches. First, the method identifies potential EEG segments for seizures using Riemannian manifold-based features from covariance matrices. Next, it removes extra-physiological artifacts using power-based features. Finally, it uses Riemannian potato-based features to classify the remaining segments with a LightGBM classifier. The method's performance was evaluated on two datasets–a private dataset (Juntendo) and a public dataset (Siena)–using leave-one-patient-out cross-validation. For the Juntendo dataset, the method achieved an average performance across

*Corresponding author: Toshihisa Tanaka, tanakat@cc.tuat.ac.jp. This work was supported by CREST (JPMJCR1784) of the Japan Science and Technology Agency (JST).

Received 08 June 2024; revised 05 September 2024; accepted 15 October 2024
ISSN 2048-7703; DOI 10.1561/116.20240032

all subjects with a sensitivity of 89.9%, specificity of 96.8%, precision of 33.3%, and an F1–score of 44.5%. On the Siena dataset, the method achieved a sensitivity of 63.8%, specificity of 98.7%, precision of 32.4%, and an F1–score of 40.5%. Processing EEG data in multiple stages helps reduce the class imbalance problem. Therefore, automating the seizure detection process will ease the practitioner's workload.

Keywords: Electroencephalogram, anomaly, seizure, artifact, Riemannian potato

1 Introduction

Epilepsy is a group of neurological disorders that are characterized by unprovoked and recurrent seizures that occur more than 24 hours apart. A seizure can be defined as the generation of excessive and abnormal electrical charges in the entire or parts of the brain, depending on the seizure type. "Unprovoked" seizures mean that they are not happening due to some specific reason, such as alcohol withdrawal, heart problems, hypoglycemia, etc. [37]. Seizures can be the result of a genetic disorder or brain injury, but quite often, the reason is unknown [22]. A person with epilepsy, referred to as an "epileptic patient," might have abnormal physiological and psychological effects due to epilepsy [26]. The patient's overall quality of life may be significantly improved by early, and affordable diagnosis of epilepsy. Electroencephalogram (EEG) contains significant physiological and pathological information that is highly useful in epilepsy diagnosis [10]. EEG is the process of recording the brain electrical activity of a subject through a bunch of electrodes. Scalp EEG-based epilepsy diagnosis is common in the practical domain because of its low cost, non-invasive technique, and easy availability in hospitals and clinics. Detecting seizure events in EEG can help diagnose an epileptic patient. Usually, the task is conducted by practitioners through manual inspection of the long EEG recordings over several hours or days. This process is time-consuming, tedious, and error-prone. Besides, human factors like fatigue and similar experiences also play an important role in the performance of the seizure detection process. In the real world, the ratio between a clinician who specializes in caring for people with epilepsy (aka an *epileptologist*) and patients with epilepsy is significantly high.

In recent years, due to the advancement of technology and low-cost EEG recording equipment, large amounts of EEG data can be recorded in hospitals, clinics, or even in a home with a special setup. A long-term EEG recording is more important to the diagnosis of epileptic patients who do not show

conclusive results or specific biomarkers in the initial standard EEG test. Furthermore, it helps the practitioners gather more information about the patient's condition, like seizure onset location in the brain, frequency of seizure occurrence, and distinguish between epileptic and non-epileptic seizures. Thus, the demand for automatic seizure detection from long-EEG recordings using machine learning has increased in recent years among practitioners. Researchers have developed several methods over the past few decades to automate the seizure detection process from EEG; however, each has its own limitations [69]. The automation process faces several challenges, including heterogeneous recording setups, subject dependencies, imbalanced datasets, and high statistical similarity between artifacts and seizures. Due to the non-invasive recording process, noises and artifacts are quite common in scalp EEG recording compared to invasive processes like stereoelectroencephalography (SEEG) and electrocorticography (ECoG). Artifacts are a group of recorded activities that do not originate from the cerebral cortex and can be categorized as physiological (or internal) and extra-physiological (or external) artifacts based on their origin [39]. Artifacts usually have a high statistical similarity with seizures, which makes the automation of the seizure detection task more challenging (see Figure 1). Several studies have attempted to automate the artifact detection and rejection process. However, to date, it highly depends on manual inspection and a bunch of parameters in the practical applications [34, 39]. Besides, the high similarity between artifact and seizure prevents researchers from applying the state-of-the-art artifact rejection model before the seizure detection task. Therefore, it is necessary to design an effective solution to differentiate between normal, abnormal (including noises and artifacts), and seizure activities in epileptic EEG recordings. Unfortunately, most of the epileptic EEG datasets only contain seizure labels annotated by the epileptologist with very little additional information. For accurate analysis and interpretation of EEG data, specific details about the subjects (e.g., age, gender, medical history), the experimental settings, and the precise circumstances during data collection are valuable. However, such information is sometimes incomplete, missing, or not published due to confidentiality in publicly available datasets. In this study, we introduce a new epileptic EEG dataset that includes comprehensive supplementary information relevant to EEG diagnosis. Additionally, we utilize another publicly available EEG dataset to demonstrate the scalability of our proposed method.

To address the challenges discussed above, this study developed a novel seizure detection method based on Riemannian manifolds, incorporating transfer learning and artifact rejection techniques. In EEG recordings, physiological and extra-physiological artifacts can occur individually or concurrently, depending on the situation. The extra-physiological artifacts include electrode movement, cable movement, power line, and head and body movement, which often results in electrode contact loss in the recording process [39]. The

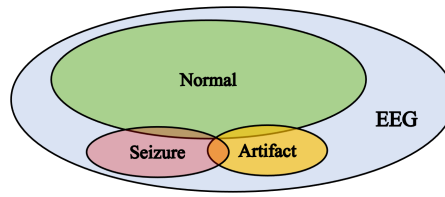


Figure 1: Relationship between normal, artifact, and seizure EEG

physiological artifacts include eye blinking, eye movement, scalp contractions, glossokinetic artifacts, chewing, electrocardiogram (EKG), swallowing, and sweating. [39]. The method presented in this study has three stages, which include multiple artifact detection techniques and a classification process.

The remainder of the paper is organized as follows: Section 2 discusses the related works in EEG classification, seizure detection and artifact handling techniques. Section 3 discusses the datasets, preprocessing, and each step of the proposed method, including detailed feature calculation. The experiment details are explained in Section 4, and Section 5 reports the results and their related discussion with future direction. The concluding remarks are in Section 6.

2 Previous Works

In EEG research a common set of features has already been used to extract valuable information whether or not it is relevant to a specific task [63]. EEG records electrical activities of the brain through electrodes, where features related to complexity, continuity, and connectivity provide significant insights [63]. These features are particularly useful for downstream tasks like postanoxic coma patients, sleep analysis, cardiac arrest studies, epilepsy, attention, perception, and mental illness. The complexity features are used to measure the irregularity, unpredictability, and information content of the EEG. This type of features can be entropy-based (Shannon [65], Tsallis, Sample, Permutation, and Approximate [3]), fractal-based (Higuchi's [29], Katz's fractal dimension [38]), Lyapunov exponents [84], Lempel-Ziv complexity [1], and Hjorth complexity [5] etc. However, these features are highly sensitive to noise and artifacts, and they are also not well generalized across subjects [64]. The continuity features assess the stability, smoothness, and persistence of the recorded EEG over time. The features help to understand the consistency and interruption of the brain's activity as an indicator of different neurological states. The median frequency, band power(δ, θ, α, β, and γ), standard deviation, burst-suppression, diffuse slowing, sharp spike, burst band powers, and the number of suppression, etc., are common continuity features used in the EEG research [72]. Although

they are valuable for understanding the characteristics of EEG signals, those features have certain limitations such as reliance on longer time windows for analysis, and limited temporal resolution. The connectivity features in EEG assess the communication and interaction between the different brain regions. These features are critical for understanding the functional interactions between brain areas and are commonly employed in research on brain networks, neurological diseases, and cognitive processes. The connectivity features like coherence [76], phase locking value [44], mutual information [4], Granger causality [12], directed transfer function [48], and functional connectivity [19] etc., are well-known in many EEG tasks. However, these features may face challenges such as capturing only linear relationships between EEG signals, and complex interpretation of the brain regions connectivity.

Several intriguing techniques have been researched in epileptic seizure detection task [23, 33, 25, 70]. To identify seizures in long-term EEG recordings, Gotman introduced a straightforward amplitude calculation method in [23]. Decomposition methods like empirical mode decomposition (EMD) and discrete wavelet transform (DWT) are popular due to their effectiveness in feature extraction from the decomposed subband signals of the EEG. Researchers have also used support vector machines (SVM) and wavelet-based features to extract relevant information from single-channel scalp EEG recordings by calculating the DWT's approximation and log variance of the detail-based features [33]. This proposed model selects features based on performance using a patient-specific-based approach. The EMD method can empirically decompose any signal into a set of narrow-band signals. A study [25] used the statistical features calculated from the set of narrow bands and selected significant features for the classification by using the mutual information-based feature selection method. The features extracted from time, frequency, and entropy in each channel were used for seizure detection with a random forest-based model and a leave-one-patient-out cross-validation scheme in [70]. Despite the popularity of the decomposition-based methods, they have several limitations, such as the selection of the number of decomposition levels and the effective features from the feature set. Furthermore, due to the highly subject-dependent nature of epilepsy, those decisions could vary from patient to patient or even recording session after session with the same patient. Additionally, the wavelet-based methods experience some issues when selecting the mother wavelet for the application.

Entropy-based methods are becoming more common in many epilepsy studies to create computer-aided solutions for seizure detection from EEG recordings [60, 43, 24, 71]. Entropy can measure the complexity of any time-series signal, which can be used to differentiate between seizure and non-seizure periods. The authors used an SVM classifier with a permutation entropy (PE)–based feature to categorize normal and epileptic EEG in [60]. According to their study, the PE value of an EEG during an epileptic seizure is lower than

that of a normal EEG. Using wavelet transform, the authors calculated the approximation entropy (ApEn) as a feature to create an automated system for detecting epileptic seizures from EEG signals [24]. In [43], the chaotic dynamics of the EEG signals are measured using the fuzzy approximate entropy (fApEn) of various sub-bands of DWT and used to classify seizure and normal EEG. This study showed that the quantitative value of fApEn decreases throughout the ictal period, indicating that the EEG data from an epileptic patient is more organized than that from a normal subject. Another study [71] showed that the value of the sample entropy decreases suddenly during an epileptic seizure, which is utilized in the proposed system. Besides, several studies used a variation of decomposition procedures to integrate several entropy measurements to enhance the performance of the automatic seizure detection system [32, 59, 58]. The authors proposed an automatic method for identifying epileptic episodes that uses DWT and various forms of entropy and statistical features in [59]. In [58], the statistical and entropy-based features using the details and approximations of DWT are computed, and a graph eigen decomposition-based method is used for feature selection. The authors used higher-ranked features for classification using a feedforward neural network. However, the main drawback of the aforementioned approaches is that parameter selection significantly impacts how well an entropy-based feature will perform on the classification task [85, 57].

Researchers have been using neural networks to enhance the automated system's performance in recent EEG studies [2, 79]. Due to the remarkable performance in other fields like computer vision, several studies have proposed models based on deep learning in recent times [53, 75, 74, 20]. A convolutional neural networks (CNN) and bidirectional long short-term memory (BLSTM)-based model is proposed to detect seizures in [53]. The authors processed multichannel EEG recordings with DWT, and features were extracted with CNN. Later, the features are fed into the BLSTM for classification. The smoothing and collar technique is also used as a post-processing technique to improve the false detection rate (FDR) and sensitivity of the method. In [75], the authors proposed a CNN-based model called "EEGWaveNet" for seizure detection. The model utilizes trainable depth-wise convolutions as discriminative filters, which extract spatial-temporal features from each channel of the EEG for classification. A combination of CNN and patient-specific autoencoder (AE) models called AE-CNN has also been proposed to detect seizures from scalp EEG [74] and improve the performance of the only CNN-based seizure detection model using EEG plot images [20]. Furthermore, using the AE model, the segmented EEG was multi-labeled as non-seizure, non-seizure-but-abnormal, and seizure by importing them as images. Later, the multi-class classification was done using the CNN model. To improve the interpretability of the deep learning-based models, Zhao *et al.* in [87] proposed a visual diagnosis mechanism with the state-of-the-art models such as

LeNet, Visual Geometry Group (VGG), deep residual network (ResNet), and vision transformer (ViT). The method introduced random channel ordering (RCO) data augmentation to create new images from EEG segments by randomize the channel orders. Later, the models decision mechanism are interpreted with Grad-CAM and attention layers. However, deep learning techniques have drawbacks, such as high computational complexity, inadequate interpretation, and the need for large amounts of training data to fine-tune hyperparameters. Additionally, the extreme imbalance in the epileptic dataset makes it challenging to train a deep-learning model effectively.

In machine learning, transfer learning is a technique that uses a pre-trained model on one task to improve performance on a related task. It transfers knowledge from the original task to the target task through the trained model, rather than starting the learning model from the beginning. This learning approach is popular in computer vision and natural language processing tasks. In brain-computer interface (BCI), most methods are domain adaptation based i.e. try to learn structure that is common in all the datasets. As an early attempt, Krauledat *et al.* [41] proposed a generalized filtering technique called "prototypical spatial filters" that need fewer calibration periods. Another study proposed a multi-source manifold metric transfer learning algorithm [67]. The algorithm used the Mahalanobis distance for source selection and manifold mapping to mitigate domain drift. Recently, the authors in [18] demonstrated the performance improvement in EEG categorization through cross-dataset knowledge transfer using data scaling.

Hillyard and Galambos proposed time domain regression techniques for removing ocular activity [30] based on the assumption that each channel is the total cumulative sum of actual EEG data along with some degree of artifact [73]. Because regression approaches require reference channels, they are less effective in removing artifacts from electrooculogram (EOG) and electrocardiogram (ECG) data when using a simplified model [80]. Various filtering techniques were used, including adaptive, wiener, and Bayes filtering, each using a different optimization theory [28]. Unfortunately, EEG artifact removal filtering is limited by its potential to inadvertently remove relevant neural information and difficulties in identifying overlapping frequency spectra of artifacts. The wavelet transform is an effective technique for time-frequency localization and multi-resolution analysis in EEG artifact detection [42]. However, the appropriate selection of Wavelet Basis, parameter tuning, overlapping spectral properties, and selection of the appropriate resolution levels, etc., make practical application difficult.

Blind source separation (BSS) is a term used to refer to a group of unsupervised algorithms for detecting the source or decomposing a mixed signal into a set of component signals based on sources. Standard BSS methods used in EEG signal processing for artifact detection include principal component analysis (PCA), independent component analysis (ICA), and canonical

correlation analysis (CCA), each having its own advantages and limitations [36, 78, 34]. PCA is excellent at identifying dominant signal fluctuations and reducing the data dimensionality [45, 13]. PCA was first applied by Berg and Scherg to eliminate ocular artifacts from EEG [11]. It can effectively capture global artifacts but assumes linearity and a Gaussian distribution. However, its use in EEG may be restricted due to its inability to handle non-Gaussian and non-linear artifacts. ICA separates a multivariate signal into additive, independent components by assuming that the observed signals are linear mixtures of independent sources [31]. It aims to estimate these sources by exploiting their statistical independence. It was first employed by Makeig *et al.* to examine EEG and EPR signals [54]. Using both simulated and real data, Vigaro *et al.* evaluated the effectiveness of the ICA approach in separating signals from their linear mixes and extracting the ocular information present in EOG signals [82, 81]. In recent years, several studies have been conducted on the automatic removal of artifacts based on ICA [21, 62, 55, 35]. However, in some cases, EEG sources may be dependent, and ICA may not accurately separate them, leading to less effective artifact removal. Besides, its performance can be sensitive to initial conditions as well. CCA is a multivariate statistical technique that analyzes the relationships between two sets of variables [47]. It is more suitable for identifying artifacts shared by multiple modalities and performs better than ICA in identifying muscle artifacts [83, 17]. However, it assumes linear relationships and may struggle with non-Gaussian and non-linear artifacts.

Hybrid methods such as EMD-BSS [14], Wavelet-BSS [52], and BSS-SVM [68] are effective in EEG artifact detection; however, each method has its own limitations: EMD-BSS might face challenges related to mode mixing; Wavelet-BSS requires precise parameter selection; and BSS-SVM depends on the assumption of linear separability.

The Riemannian manifold-based BCI study utilizes the geometry of covariance matrices to enhance feature extraction and classification, thereby improving the accuracy and robustness of BCI systems [86]. Recently, this approach has been applied for EEG classification such as sleep stage classification [49, 50], motor imagery [7, 8], artifact removal [6], and seizure detection [66, 61] among others. Most of the Riemannian approaches in BCI typically involve 1) signal preprocessing (including band-pass filtering), 2) source extraction, 3) feature vector generation, and 4) classification. In these approaches, EEG sources can be extracted using conventional PCA, CCA, and common spatial pattern (CSP)-based methods or as covariance matrices. Most state-of-the-art Riemannian approaches have demonstrated improved performance by representing EEG data as covariance matrices for the feature extraction task [86]. Two main categories of features are used in EEG classification: 1) signal energy-based and 2) sample-based. The energy-based features are extracted directly from raw or preprocessed EEG from all or from the subset of all channels

based on the applications. In sample-based features, the covariance matrices are calculated from EEG channels to extract sample or energy information and their inter-channel relationships. The spatial covariance matrices can extract spatio-frequential information of the signal that can be used in many BCI tasks. Tomioka *et al.* [77] was the first to attempt a direct classification of covariance matrices. Later, several other attempts are also made to improve it [7]. However, metric learning plays a crucial role in the Riemannian manifold-based approaches. It is a process of automatically tuning the parameters to measure the similarity or dissimilarity from the data. The common framework is to compute symmetric positive definite (SPD) matrices from EEG signals, then compute the mean (aka "Riemannian Mean") of the SPD matrices of each class. Later, the SPD matrix is calculated from new EEG segments to determine its class based on the distance from the Riemannian mean of each class. Several studies have followed a simple "Minimum Distance to the Mean" classifier for the classification purpose [86].

In [66], the authors proposed a Riemannian geometry-based automatic seizure detection method. They extracted features from the symmetric and positive definite covariance matrices and made classifications with SVM. In [61], the authors proposed a Riemannian geometry-based method with SVM where training data selection processes are patient-specific, which improves the performance to some extent. Besides, Riemannian geometry has also been used for artifact removal in recent times. A multivariate automatic and adaptive technique called "Riemannian potato" was proposed by Barachant *et al.* [6] for identifying artifacts in continuous EEG. Later, the concept is extended, called the "Riemannian potato field," to capture different types of artifacts in continuous EEG [9].

3 Materials and Methods

3.1 Datasets

3.1.1 Juntendo Epileptic Dataset

A private EEG database recorded at Juntendo University Hospital, Japan, is used in this study. The research is approved by the Ethics Committee of the Juntendo University Hospital and the Ethics Committee of the Tokyo University of Agriculture and Technology (TUAT). Hereafter, the dataset will be denoted as "Juntendo." The EEG dataset is recorded from the scalps of epileptic patients (subjects); each patient has several hours of recordings. 22 patients were chosen for this study, 14 males and 8 females with an age range 1 to 42 years. The International 10-20 standard follows the electrode positions, where the sampling frequency (F_s) is 500 Hz. Most of the recordings contain 21 common channels, which are Fp1, Fp2, F7, F8, F3, F4, T3, T4, T5, T6, C3,

C4, P3, P4, Fz, Cz, Pz, O1, O2, A1, and A2. We have used all the channels except A1 and A2. Table 1 gives more information about the dataset.

Table 1: Description of the "Juntendo" EEG dataset

Pt.	Sex	Age (Yr.)	Age (M.)	Seizure types	Channels with Seizure	Record Length(hr.)	Seizure time (sec.)	Seizure ratio(%)
1	F	19	10	Generalized, Motor, Tonic-clonic	All Channels except Fp1,F7	4	217	1.507
2	M	16	5	Generalized, Motor, Epileptic spasms	C3, F3, T3	4	247	1.715
3	M	26	11	Focal, Nonmotor, Behavior arrest	F7, T3	8	371	1.288
4	M	1	4	Generalized, Motor, Epileptic spasms	All Channels	2	17	0.236
5	M	8	5	Focal, Nonmotor, Behavior arrest	C3, F3	6	441	2.042
6	M	30	9	Focal, Nonmotor, Behavior arrest	F7, T3	6	416	1.926
7	M	27	0	Focal, Nonmotor, Behavior arrest	F8, T4	4	161	1.118
8	M	26	4	Focal, Motor, Automatisms	F7, T3	4	229	1.59
9	F	19	4	Generalized, Motor, Tonic-clonic	All Channels except Fp1, F7	2	1150	15.972
10	M	24	7	Focal, Focal to bilateral tonic-clonic	F7, T3	10	218	0.606
11	F	27	8	Focal, Motor, Automatisms	T5	4	343	2.382
12	M	11	7	Focal, Motor, Tonic	F7,F8, T3,T4,T5,T6	6.03	1581	7.282
13	F	27	9	Focal, Motor, Automatisms	All Channels except Fz,Cz,Pz, C1,C2	2	135	1.875
14	M	35	5	Focal, Motor, Automatisms	F7, F8	12	388	0.898
15	F	5	2	Focal, Motor, Epileptic spasms	C4, F4	4	125	0.868
16	M	23	10	Generalized, Motor, Tonic-clonic	All Channels	2	94	1.306
17	M	42	0	Generalized, Motor, Tonic-clonic	All Channels	2	23	0.319
18	F	12	11	Unknown, Motor, Epileptic spasms	C3, F3, F7, T3	4	59	0.41
19	F	11	8	Generalized, Nonmotor, Typical absence	All Channels	2	144	2.0
20	F	11	8	Generalized, Nonmotor, Typical absence	All Channels	8	226	0.785
21	F	35	0	Focal, Motor, Automatisms	T3, T5	10	495	1.375
22	M	31	6	Focal, Motor, Tonic	T3, T4, T5, T6	2	27	0.375

3.1.2 Siena Epileptic Dataset

We also utilized a publicly available EEG dataset in this study that was recorded at the Unit of Neurology and Neurophysiology of the University of Siena. Hereafter, the dataset will be referred to as "Siena." It contains recordings of 14 patients, including 9 males (ages 25-71) and 5 females (ages 20-58). These recordings adhere to the International 10-20 standard for the electrode positions, where the sampling frequency (F_s) is 512 Hz. For the purposes of

this experiment, we downsampled the data to 500 Hz. Each patient's clinical and electrophysiological data were diagnosed and labeled (seizure events) by an expert clinician according to the International League Against Epilepsy criteria. The dataset includes three types of seizures: focal onset impaired awareness (IAS), focal onset without impaired awareness (WIAS), and focal to bilateral tonic-clonic (FBTC) present in the dataset. More detailed information about the dataset is summarized in Table 2. The dataset is publicly available here: https://physionet.org/content/siena-scalp-eeg/1.0.0/.

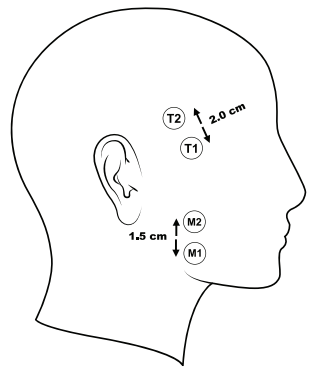
Table 2: Description of the "Siena" EEG dataset. (Pt.ID: Patient ID; Gender: Male(M), Female(F); IAS: focal onset impaired awareness, WIAS: focal onset without impaired awareness, FBTC: focal to bilateral tonic-clonic; Loc.: Localization, T: temporal lobe, F: frontal lobe; Lat.: Lateralization, R: right, L: left, Bilateral: right and left simultaneously.)

Pt.ID	Age	Gender	Seizure types	Loc.	Lat.	Record Length(hr.)	Num. of Seizures	Seizure time (sec.)	Seizure ratio (%)
PN00	55	M	IAS	T	R	3.20	5	325	2.821
PN01	46	M	IAS	T	L	13.48	2	128	0.264
PN03	54	M	IAS	T	R	24.22	2	244	0.280
PN05	51	F	IAS	T	L	6.02	3	104	0.480
PN06	36	M	IAS	T	L	12.05	5	282	0.650
PN07	20	F	IAS	T	L	8.73	1	62	0.197
PN09	27	F	IAS	T	L	6.83	3	203	0.825
PN10	25	M	FBTC	F	Bilateral	18.68	10	365	0.543
PN11	58	F	IAS	T	R	2.40	1	55	0.637
PN12	71	M	IAS	T	L	6.07	4	290	1.328
PN13	34	F	IAS	T	L	8.63	3	264	0.849
PN14	49	M	WIAS	T	L	20.43	4	163	0.222
PN16	41	F	IAS	T	L	4.85	2	230	1.317
PN17	42	M	IAS	T	R	5.10	2	153	0.833

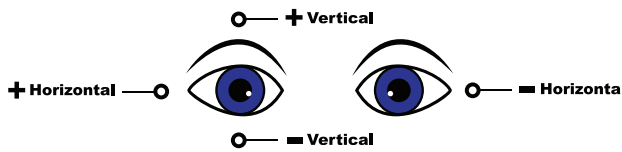
3.1.3 Artifact Dataset

For artifact analysis, we used another private scalp EEG dataset, which was recorded at the Tokyo University of Agriculture and Technology (TUAT), Tokyo, Japan. Hereafter, the dataset will be denoted as "TUAT." The research is approved by the Ethics Committee of the Tokyo University of Agriculture and Technology (TUAT). This scalp EEG dataset was recorded using Polybench software (Twente Medical Systems International B.V., TMSi, Oldenzaal, Netherlands) with a sampling frequency of 2,048 Hz, which was further downsampled to 500 Hz. Five healthy subjects were used to record the EEG with a 64-electrode EEG gel head cap, based on the 10-10 international system. For the experiment, we selected only 19 channels, like the epileptic dataset discussed in the previous section. The horizontal and vertical eye movements (blink, left-right, and up-down) are recorded with two bipolar (vertical and horizontal) EOG electrodes (refer to Figure 2b). The vertical electrodes were placed on the sides of both eyes, and the horizontal electrodes were placed

above the eyebrow and below the bottom eyelid. We used two bipolar electrodes on the anterior temporal and masseter muscles (refer to Figure 2a) to capture the muscle activity and detect the "ground truth" of EOG and EMG artifacts. A brief description of the dataset is given in Table 3.



(a) EMG electrode placement



(b) EOG electrode placement

Figure 2: The (a) EMG, (b) EOG electrode placement

Table 3: Description of the "Artifact" EEG dataset

Subject	Sex	Age(Yr.)	Age(M.)	Record Length
1	F	30	0	12 min.
2	M	30	10	12 min.
3	M	22	8	12 min.
4	M	21	11	12 min.
5	M	22	5	12 min.

3.2 Riemannian Manifold in EEG

A "Riemannian manifold" is a smooth, curved space where we can measure distances and angles between points in a special way. This concept is useful for studying data with complex structures, such as brain signals on curved surfaces. This makes it particularly valuable for analyzing data like EEG, which has distinct patterns and structures.

In seizure detection, Riemannian manifolds are used to analyze EEG data to extract suitable features from complex patterns. Initially, EEG signals are represented as covariance matrices, capturing the relationships between different brain regions. Each covariance matrix can be viewed as a point on the Riemannian manifold [66, 61]. In BCI and medical imaging, Symmetric Positive Definite (SPD) matrices play a crucial role in the Riemannian manifold framework. They are square, symmetric, and positive definite (i.e., all eigenvalues are positive). A set of all SPD matrices can form a curved space called a Riemannian manifold. In EEG processing, SPD matrices can represent the functional connectivity of the brain more efficiently than other techniques [15]. The calculation of SPD matrices and the construction of Riemannian manifolds using them are discussed in Section 3.5.1.

The Riemannian mean is a method to calculate the average of data points (like covariance matrices) on the Riemannian manifold (discussed in Section 3.5.2). Unlike the regular arithmetic mean, the Riemannian mean gives a more accurate idea of the "middle" point among the data by considering the curved space of the manifold. To do so, measuring the distance between the points in a curved space like Riemannian manifolds can be done using techniques like Euclidean, log-Euclidean distance, and Affine Invariant Riemannian Metric (AIRM), etc. [66]. The purpose of that Riemannian metric on a manifold is to minimize the distance between intra-class samples and maximize the separation of inter-classes samples. Here, by calculating distances between covariance matrices on the Riemannian manifold, we can extract features to differentiate the different brain states (like normal, abnormal, or seizure).

However, SPD matrices cannot be used directly in most classifiers as features because of their existence on a curved space like a Riemannian manifold. Specialized metrics and transformations are required to project them into a linear space suitable for classification tasks. Here, the tangent space mapping (TSM) technique helps to handle complex curved data on a Riemannian manifold by converting it into a linear space. The detailed calculation of the TSM process is given in Section 3.5.3.

Finally, the extracted feature vectors can be fed into a classifier for the classification task. As we discussed in Section 2, the long recordings of EEG have normal and abnormal (i.e., noise, artifacts, seizures, etc.) segments with most segments reflecting normal activities. So, using the Riemannian manifold-based techniques, we can calculate the Riemannian mean from the whole recordings and extract features to detect abnormal EEG segments through the classification process. Typically, the distribution of abnormal EEG segments will be significantly different from that of normal EEG activities. Later, those abnormal EEG segments can be further processed for seizure detection tasks.

3.3 Preprocessing

In practice, the recording procedure introduces many noise and distortions into scalp EEG recordings. A FIR filter with a passband between 4 and 80 Hz was applied for every EEG recording. The power line artifacts are usually rejected with a notch filtering process, which is effective in most cases. Here, power noise from alternative current power lines was eliminated using a 50 Hz notch filter. Each patient's EEG was split every 5s with non-overlapping EEG to select possible seizure candidates and every 1s and 500ms with 80% overlapping EEG to remove artifacts and noisy EEG and detect seizures.

3.4 Proposed Method

The flowchart of the proposed method is shown in Figure 3. The proposed method can be divided into three stages.

I **Anomaly detection**: Pre-screening of EEG using Riemannian manifolds

II **Artifact rejection**: Extra-physiological and general artifacts rejection

III **Seizure detection**: Physiological artifacts (including unknown general artifacts) and seizure classification based on Riemannian potato and power-based features

The following sections contain the details of the proposed method.

3.5 Pre-screening of EEG Using Riemannian Manifolds

From the viewpoint of epileptic seizure detection, the segments of an epileptic EEG recording can be classified into three categories: normal (or regular), abnormal (it might contain artifacts, undefined abnormal EEG but not seizure), and seizure. The main purpose of the pre-screening stage is to separate "normal" EEG segments from "abnormal" and "seizure" segments. Here, anything apart from the "normal" EEG is called "anomaly." To separate "normal" segments, we have used Riemannian manifold-based features [61] to calculate the anomaly score corresponding to each EEG segment, and categorize them as "normal" and "anomaly" based on the calculated threshold. Figure 4 shows three types of EEG segments: normal (Figure 4a), artifact (Figure 4b), and seizure (Figure 4c). The entire process of calculating the anomaly score using the Riemannian manifold from EEG is summarized in Figure 5. The Riemannian manifold-based features and calculation of the anomaly score are discussed in the following section.

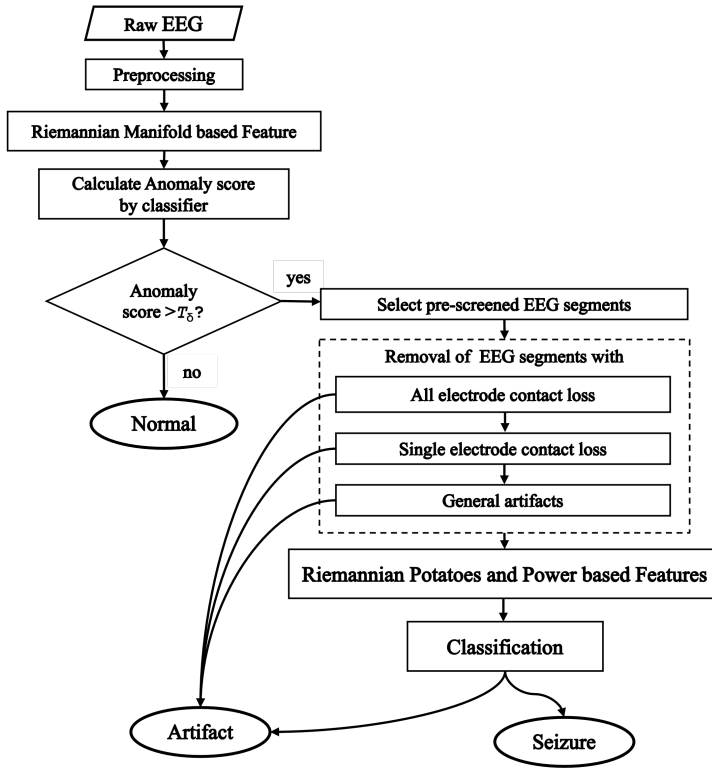


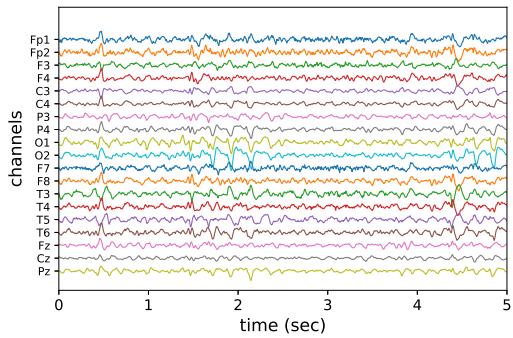
Figure 3: Flow chart of the proposed method

3.5.1 Spatial and Temporal Covariance Matrix

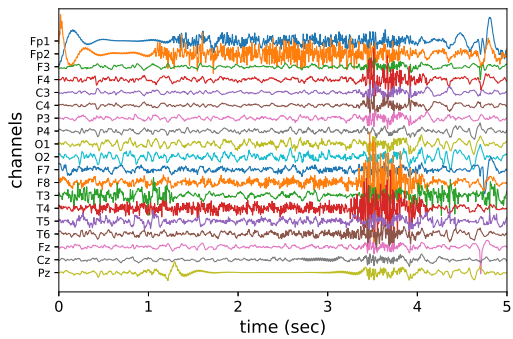
Important information is provided by the temporal and spatial patterns during epileptic seizures to differentiate seizures from other brain activity. Here, spatial patterns mean the EEG patterns between electrodes in a fixed time frame or segment. Similarly, temporal patterns mean the EEG patterns of an electrode in between consecutive EEG segments within the same electrodes.

Let $x_i^m \in \mathbb{R}^{N\times 1}$ represent the m-th channel of the i-th indexed EEG segment; then, it can be defined as $\boldsymbol{X}_i = [{x_i}^1, {x_i}^2, \ldots, {x_i}^M]^T \in \mathbb{R}^{M\times N}$, where $\cdot^T$ indicates a transpose operation, N is the number of time samples, and M is the number of channels in an EEG segment. Suppose a signal at each row or channel has zero mean. Then, the spatial covariance matrix (SCM) of the i-th indexed EEG segment, $\boldsymbol{C}_i \in \mathbb{R}^{M\times M}$ of $\boldsymbol{X}_i$ can be defined as [61]:

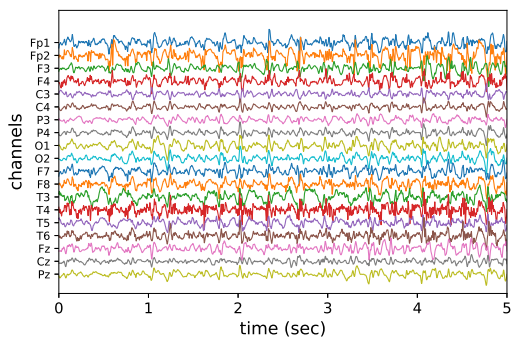
$$\boldsymbol{C}_i = \frac{1}{N-1}\boldsymbol{X}_i\boldsymbol{X}_i^T. \tag{1}$$



(a) Normal EEG Segment



(b) Artifact EEG Segment



(c) Seizure EEG Segment

Figure 4: Five-seconds long (a) normal, (b) artifact, (c) seizure EEG segments

Furthermore, temporal variations in an EEG also provide important details on the development of seizures [16]. Let the temporal variation for consecutive L EEG segments with respect to the i-th index is $\boldsymbol{Y}_i^m =$

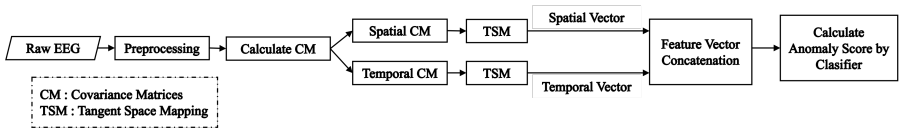


Figure 5: Flow chart of the Riemannian Manifold-based anomaly detection from Epileptic EEG recordings

$[x_i^m, x_{i+1}^m, \ldots, x_{i+(L-1)}^m]^T \in \mathbb{R}^{L\times N}$. For the m-th channel of the i-th indexed EEG, the temporal covariance matrix (TCM), $\boldsymbol{Q}_i^m$ can be defined as [61]:

$$\boldsymbol{Q}_i^m = \frac{1}{N-1}\boldsymbol{Y}_i^m(\boldsymbol{Y}_i^m)^T \in \mathbb{R}^{L\times L}. \tag{2}$$

3.5.2 Riemannian Mean from a Set of EEG Covariance Matrices

Using Equations (1) and (2), we represent an EEG recording with a set of covariance matrices. Here, each covariance matrix can be treated as a point on a Riemannian manifold, which is a set of symmetric positive definite (SPD) matrices [61]. To transform covariance matrices to Riemannian manifold ($\mathcal{M}$), we need Riemannian mean $\boldsymbol{C}_{\text{ref}}$, which can be calculated from the covariance matrices set as:

$$\boldsymbol{C}_{\text{ref}} = \underset{\boldsymbol{C}\in\mathcal{M}}{\operatorname{argmin}} \sum_i \delta_R^2(\boldsymbol{C}, \boldsymbol{C}_i) \tag{3}$$

where $\delta_R(\boldsymbol{C}, \boldsymbol{C}_i)$ is the Riemannian distance between $\boldsymbol{C}$ and $\boldsymbol{C}_i$. If $\boldsymbol{C}$ and $\boldsymbol{C}_i$ are SPD matrices with an of dimension $M \times M$ in a Riemannian manifold $\mathcal{M}$, then $\delta_R(\boldsymbol{C}, \boldsymbol{C}_i)$ can be calculated as [61]:

$$\delta_R(\boldsymbol{C}, \boldsymbol{C}_i) = \left\| \log\left(\boldsymbol{C}^{-\frac{1}{2}}\boldsymbol{C}_i\boldsymbol{C}^{-\frac{1}{2}}\right)\right\|_F, \tag{4}$$

where $\log(\cdot)$ is the matrix logarithm, and $\|\cdot\|_F$ is the Frobenius norm.

3.5.3 Feature calculation through tangent space mapping

The nonlinearity of the Riemannian manifold makes it challenging to apply most of the state-of-the-art classifiers to classify the feature vectors. Tangent space mapping (TSM) is a mapping technique that can map matrices from nonlinear to linear space, like Euclidean, and vectorize them for convenience [7]. The mapping process can be defined as:

$$\boldsymbol{p} = \text{upper}\left(\log\left(\boldsymbol{C}_{\text{ref}}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{C}_{\text{ref}}^{-\frac{1}{2}}\right)\right) \tag{5}$$

where the upper(·) function considers the upper triangular part of a matrix that is multiplied by the $\sqrt{2}$ coefficient to the non-diagonal elements [66].

Now, using Equations (1), (3), (4), and (5), we calculate a feature vector set $\boldsymbol{s}_i$ from the set of $\boldsymbol{C}_i$. Similarly, using Equations (2), (3), (4), and (5), we can calculate feature vector sets $\boldsymbol{t}_i^m$ from the set of $\boldsymbol{Q}_i^m$ where m is the index of electrodes $m = 1, 2, \ldots, M$. Finally, for the i-th indexed EEG segment, we concatenate the spatial vector $\boldsymbol{s}_i$ and temporal vector $\boldsymbol{t}_i^m$ to produce the final feature vector:

$$\boldsymbol{v}_i = [\boldsymbol{s}_i, \boldsymbol{t}_i^1, \boldsymbol{t}_i^2, \ldots, \boldsymbol{t}_i^M]^T. \tag{6}$$

If K is the total number of EEG segments in a recording, the feature vector set $\boldsymbol{V}$ will be:

$$\boldsymbol{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_K\}. \tag{7}$$

Later, $\boldsymbol{V}$ is fed to the anomaly detection classifier to calculate the anomaly score (a_i) set. Here, threshold T_δ is calculated from the set of a_i as:

$$T_\delta = \mu_a + \delta\sigma_a, \tag{8}$$

where μ_a and σ_a are the mean and absolute standard deviation of the anomaly-scores a_i of the corresponding EEG recording. Now, we calculated the moving average of anomaly scores a_i, denoted as A_i for the decision making purpose. Finally, a segment with index i can be labeled as anomaly or not as:

$$\mathcal{D}_i = \begin{cases} 1, & \text{Anomaly, if } A_i > T_\delta \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

3.6 Extra-physiological and General Artifacts Rejection

In this stage, we will consider only those segments labeled as anomalies by the anomaly detection algorithm.

3.6.1 All electrode contact loss

An EEG recording discontinuity is typically caused by a transient or prolonged loss of contact between all the electrodes. These discontinuities are usually followed by a typical oscillatory period, which eventually goes back to normal EEG [9]. During this artifact, a similar oscillatory pattern can be seen in a group of electrodes situated far from each other [27]. Here, we observed that the Bipolar montage (taken from bilateral positions) exhibits important properties for this artifact type. The montages used in this study are Fp1-Fp2,

F7-F8, T3-T4, T5-T6, F3-F4, C3-C4, P3-P4, Fz-Cz, Cz-Pz, and O1-O2. The absolute power-rate of segment i, denoted as B_i, is calculated as:

$$B_i = \frac{1}{MN} \sum_{c=1}^{M} \sum_{t=1}^{N} |b_{i,c}(t)|, \tag{10}$$

where c, i, t, M, and N are the channel index, segment index, time index, number of bipolar channels, and length of bipolar EEG segment b_i, respectively, where $c = 1, 2, \ldots, M$; and $M = 10$. From an EEG recording with K segments, the B_μ can be calculated as:

$$B_\mu = \frac{1}{K} \sum_{i=1}^{K} B_i. \tag{11}$$

Now, we considered the anomaly segments, which are consecutive to each other as a group ($\mathcal{G}$). Here, if any j, such that $\min(B_j) < B_\mu/\rho$, we reject all the anomaly segments i as artifacts that are in the same group of segment j, i.e., $i, j \in \mathcal{G}$.

3.6.2 Single electrode contact loss

Similar to contact loss across all electrodes, the disruption of a specific electrode connection is a frequent issue during EEG recordings. Typically, an electrode experiencing connection loss exhibits a distinctive oscillation with higher power than its surrounding electrodes [27]. In ordinary situations, given the volume-conduction characteristics of scalp EEG recordings, it is uncommon for a specific electrode to display a substantial increase in power without affecting its neighboring electrodes. If we define the sum of absolute power E_i for EEG segment i as:

$$E_i = \frac{1}{MN} \sum_{c=1}^{M} \sum_{t=1}^{N} |x_{i,c}(t)|, \tag{12}$$

and the absolute power ratio $R_{i,c}$, of channel c of segment i can be calculated as:

$$R_{i,c} = \frac{\sum_{t=1}^{N} |x_{i,c}(t)|}{MNE_i} \tag{13}$$

Now, we rejected an EEG segment as an artifact if any c, such that $R_{i,c} > \eta$, where $c = 1, 2, \ldots, M$ (for an example, see Figure 6).

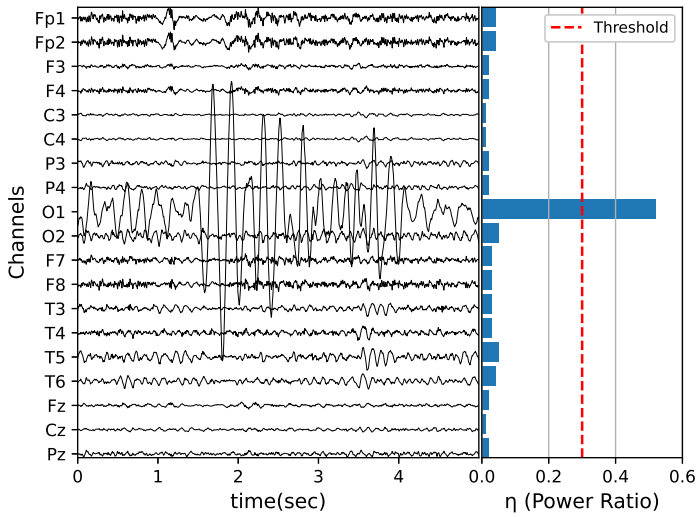


Figure 6: Detection of the single electrode contact loss artifact using power ratio based feature

3.6.3 General artifacts rejection

Typically, during a seizure event, a cluster of neurons simultaneously becomes active within a specific region or throughout the entire brain, resulting in an abnormal EEG pattern characterized by a substantial amount of energy [27]. In practical terms, the average power of the segments associated with a seizure event is significantly greater than that across all EEG segments recorded. The mean absolute power μ_{K} of an EEG recording is given as:

$$\mu_{\mathrm{K}} = \frac{1}{K}\sum_{i=1}^{K} E_i. \tag{14}$$

Suppose j represents the index and G represents the total number of segments in a group of anomalies $\mathcal{G}$. In that case, the mean absolute power (μ_{G}) corresponding to $\mathcal{G}$ can be computed as:

$$\mu_{\mathrm{G}} = \frac{1}{G}\sum_{j=1}^{G} E_j. \tag{15}$$

To separate artifacts and seizures, we calculated threshold E_{th}, which can be derived as:

$$E_{\mathrm{th}} = \gamma\mu_{\mathrm{K}}. \tag{16}$$

Finally, if $\mu_G < E_{th}$, all EEG segments within the anomaly group are dismissed as common artifacts, and the remaining EEG segments are identified as segments of seizure activity.

3.7 Artifact and Seizure Classification Based on Riemannian Potato and Power-based Features

In this final stage, we calculated eight features from the EEG segments to classify seizure and anomaly EEG segments. Those calculated features can be divided into two categories: "Riemannian potato" based features and "power"-based features. The details of the feature calculation are discussed in the following section.

3.7.1 Riemannian potato-based features

The Riemannian potato (RP) is a novel multivariate approach designed to detect artifacts from EEG data [6]. It utilizes covariance matrices processed through Riemannian geometry to measure dispersion, represented as a standardized z-score. A reference of artifact-free EEG is used to compute this z-score, together with the distribution of distances between covariance matrices. The RP method defines a region of acceptability based on a z-score threshold, referred to as the "potato." The RP method rejects epochs with covariance matrices that fall outside that RP threshold. In most cases, the z-score serves as a quality index (SQI) for identifying and rejecting artifacts [9]. From a mathematical viewpoint, this z-score denotes how close or far the distribution of the EEG data is from the clean one. Here, one major problem with the RP-based method is the "reference clean EEG data" and good "initial EEG data" [9]. Additionally, as discussed in Section 1, some seizure events have very high statistical similarity with artifacts.

We utilized a transfer learning-based approach to overcome those challenges by generating reference Riemannian potatoes from known artifacts instead of clean EEG. By adopting this approach, we circumvented the need for "reference clean EEG data" and transferred the artifact's knowledge to the seizure detection domain. Seizures are complex neural events that often involve artifacts due to muscular activities. These artifacts can also contain valuable information related to seizure events. In particular, ocular and muscle activities are major sources of artifacts observed in EEG recordings, even in healthy subjects [39, 34]. Cardiac activity (EKG) is also commonly recorded in EEG, and most EEG recording devices capture this as a reference signal, later removing EKG from EEG during data processing [34]. In this study, we considered artifact sources that are commonly observed and likely to be present in both normal and seizure activities. Other sources of artifacts depend on the subject's condition, and we categorize them under "general artifacts."

Additionally, we aimed to preserve as much EEG information as possible. This study considered chewing activity and three significant types of eye movement artifacts: (a) blink, (b) left-right movements, and (c) up-down movements. To detect artifacts, we trained four Riemannian potatoes using the "TUAT" dataset (discussed in Section 3.1.3). In each recording session, the participants were asked to imitate an artifact. Later, the onset of the events was automatically detected from the vertical, horizontal, and masseter bipolar electrodes. Figure 7 demonstrates the process of extracting the automatic ground truth for the "blink" event. We used a similar approach to detect the ground truth for left-right, up-down eye movements and chewing using horizontal, vertical, and masseter bipolar electrodes, respectively.

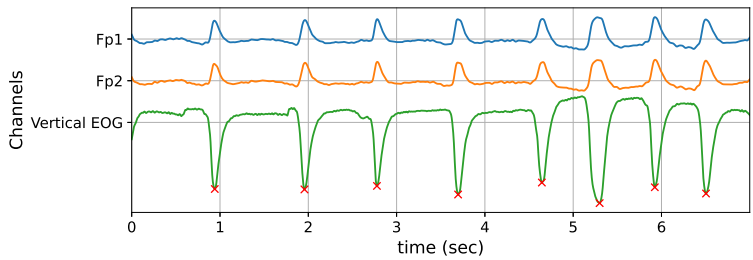


Figure 7: Blink event detection in "TUAT" Dataset

The preprocessing and operational parameter list for potatoes is given in Table 4. Generally, the Riemannian potato is calculated using selected channels more affected by that artifact. Then, the appropriate frequency band was selected with bandpass filtering, which shows more dominant features than the others. Later, spatial covariance matrices were generated to calculate the Riemannian potato-based z-score. Let $\boldsymbol{X}_i$ be the preprocessed i-th EEG segment with selected channels and bandpass filtering; its corresponding spatial covariance matrix is $\boldsymbol{C}_i$. In this study, we employed the "Oracle Approximating Shrinkage (OAS)" estimator to process $\boldsymbol{C}_i$ addressing the issue of ill-conditioned input matrices that could affect the performance of Riemannian potatoes. The z-score calculation process from the spatial covariance matrix set is described as:

$$d_i = \delta_R \left(\boldsymbol{C}_i, \boldsymbol{C}_{\text{ref}} \right), \tag{17}$$

$$\mu_d = \frac{1}{S} \sum_{i=1}^{S} d_i \tag{18}$$

$$\sigma_d = \sqrt{\frac{1}{S} \sum_{i=1}^{S} (d_i - \mu_d)^2} \tag{19}$$

Table 4: Parameters with corresponding values to generate different Riemannian potatoes from the "Artifact" and "Epileptic" datasets for the seizure detection task

Potato Type	Dataset	Passband [Hz]	Ground Truth	Considered Channels
Eyeblink	Artifact	1–20	Vertical EOG	Fp1, Fp2
Eye Left-Right Movement	Artifact	1–20	Horizontal EOG	Fp1, Fp2, F7, F8
Eye Up-Down Movement	Artifact	1–20	Vertical EOG	Fp1, Fp2, F3, F4
Chewing	Artifact	1–35	Masseter Electrodes	F7, F8, F3, F4, C3, C4, T3, T4, T5, T6
EMG	Epileptic	55–95	-	F3, F4, C3, C4, P3, P4
General Artifact	Epileptic	32–45	-	All Channels
Seizure	Epileptic	1–60	-	All Channels

where d_i, S, $\boldsymbol{C}_{\text{ref}}$, μ_d, and σ_d represent the Riemannian distance of index i, considered number of segments, the Riemannian mean, the mean and standard deviation of the distances, respectively. Here, $\boldsymbol{C}_{\text{ref}}$ can be calculated from the spatial covariance matrices using Equation (3). By using $\boldsymbol{C}_{\text{ref}}$, μ_d, and σ_d, we can define a Riemannian potato, denoted as $\boldsymbol{P}$, where $\boldsymbol{P} = \{\boldsymbol{C}_{\text{ref}}, \mu_p, \sigma_p\}$.

Now, using d_i, μ_d, and σ_d, we calculated the z-score of the i-th segment (z_i) as:

$$z_i = \frac{d_i - \mu_d}{\sigma_d}. \tag{20}$$

Algorithm 1 shows the process of calculating the z-scores from a set of covariance matrices ($\boldsymbol{C}^{\text{set}}$) using the Riemannian potato ($\boldsymbol{P}$). Figure 8 shows the overall process of calculating the z-score from the EEG datasets using the transfer learning technique. Here, the "source EEG" dataset is used to generate the Riemannian potatoes and applied to the "target EEG" to calculate z-scores for each recording. Using the transfer learning strategy, we can calculate the Riemannian potato parameters from the "source EEG" and apply them to the "target EEG" segments. In this case, the source EEG is "TUAT," and the target EEG is "Juntendo." The final z-score is calculated by taking the average of z-scores from all the subjects of Riemannian potatoes in "TUAT."

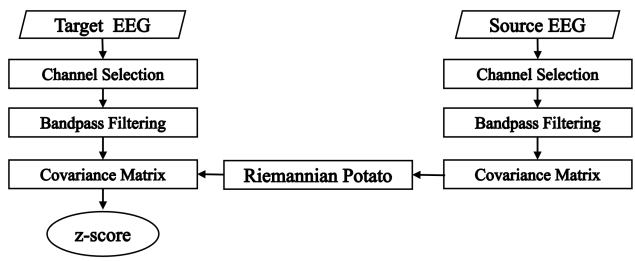


Figure 8: Riemannian potato-based feature calculation using transfer learning approach

We also calculated Riemannian potatoes for "EMG" and "General Artifacts" from each EEG recording and their z-score. Here, $\boldsymbol{C}_{\text{ref}}$, d_i, μ_d, and σ_d are

Algorithm 1 An algorithm to calculate z-scores using the Rimannian potato

Require: Riemannian potato $\boldsymbol{P}$, where $\boldsymbol{P} = \{\boldsymbol{C}_{\text{ref}}, \mu_p, \sigma_p\}$; Covariance matrices set $\boldsymbol{C}^{\text{set}} = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots, \boldsymbol{C}_S\}$;
Ensure: $\boldsymbol{Z}$ = RIEMANNIAN_POTATO_ZSCORE($\boldsymbol{C}^{\text{set}}$, $\boldsymbol{P}$)

$\boldsymbol{Z} \leftarrow \emptyset$ ▷ Set of z-scores
for $i \leftarrow 1$ to S **do**
  $d_i \leftarrow \delta_R(\boldsymbol{C}_i, \boldsymbol{C}_{\text{ref}})$ ▷ Riemannian distance function
  $z_i \leftarrow \dfrac{d_i - \mu_p}{\sigma_p}$ ▷ z-score of a segment
  $\boldsymbol{Z} \leftarrow \boldsymbol{Z} \cup z_i$
end for
return $\boldsymbol{Z}$

calculated from the same EEG recording that is tested for seizure detection, i.e., the source and target EEG recordings are the same.

To include the seizure properties, we generated Riemannian potatoes from each seizure event and calculated their z-score individually. The flowchart of the operation is shown in Figure 9. We considered the seizure event from all the patients except the test patient to generate Riemannian potatoes. By using Equations (3) and (17)–(20), we can calculate the Riemannian potato ($\boldsymbol{P}^{(j)}$) of each seizure event with corresponding parameters $\boldsymbol{C}_{\text{ref}}^{(j)}$, $\mu_p^{(j)}$, and $\sigma_p^{(j)}$, where $j = 1, 2, \ldots, J$, and the value of J depends on the training set. Here, $\boldsymbol{C}_{\text{ref}}^{(j)}$, $\mu_p^{(j)}$, and $\sigma_p^{(j)}$ represent the Riemannian mean, mean, and standard deviation of the Riemannian distances with respect to the $\boldsymbol{C}_{\text{ref}}^{(j)}$, respectively. By using Algorithms 1 and 2, we calculated the set of z-scores of the EEG segments from the test patients, denoted by $\boldsymbol{Z}_{\text{SZ}}$. Algorithm 2 shows the z-scores ($\boldsymbol{Z}_{\text{SZ}}$) calculation process for a single EEG recording by using all the Riemannian potatoes ($\boldsymbol{P}^{\text{train}}$) generated from the training set. Initially, the z-scores are calculated with respect to each Riemannian potato ($\boldsymbol{P}^{(j)}$) from the seizures of the training set. Then, the z-scores were standardized, and moving averaged for each $\boldsymbol{P}^{(j)}$ was determined. Finally, the average over the z-scores was considered to get the final z-scores for a single EEG recording.

3.7.2 Power based feature

Power is an important feature in seizure events. By using Equations (12) and (14), we get E_i, and μ_{K}, respectively. Then, the feature F_i can be calculated as:

$$F_i = E_i - \mu_{\text{K}} \tag{21}$$

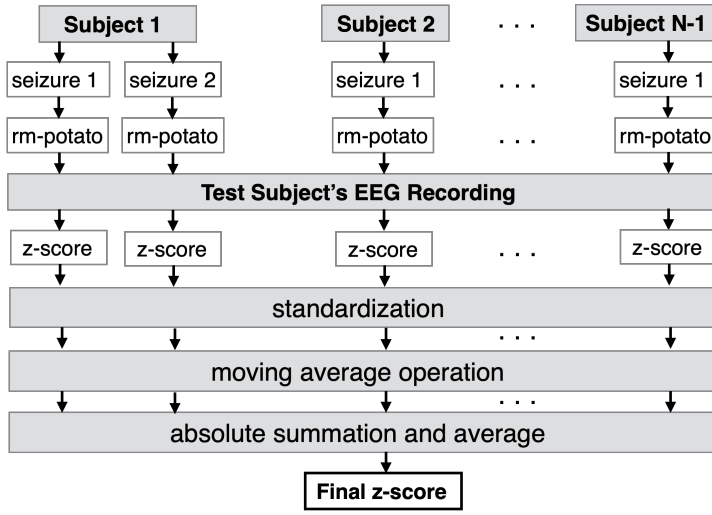


Figure 9: z-score calculation by Riemannian potatoes of seizures

Algorithm 2 An algorithm to calculate z–scores using Riemannian potatoes from the seizure events

Require: The Riemannian potatoes set from the seizure events of the training dataset, $\boldsymbol{P}^{\text{train}} = \{\boldsymbol{P}^{(1)}, \boldsymbol{P}^{(2)}, \ldots, \boldsymbol{P}^{(J)}\}$; Moving averaging window, I; Test covariance matrices, $\boldsymbol{C}^{\text{set}} = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots, \boldsymbol{C}_K\}$;

Ensure:

$\boldsymbol{Z}_{\text{SZ}} \leftarrow \emptyset$ ▷ Set of z-scores
for $j \leftarrow 1$ to J **do**
  $\boldsymbol{Z} \leftarrow \emptyset$
  for $i \leftarrow 1$ to K **do**
    $z_i \leftarrow$RIEMANNIAN_POTATO_ZSCORE$(\boldsymbol{C}_i, \boldsymbol{P}^{(j)})$ ▷ from Algorithm 1
    $\boldsymbol{Z} \leftarrow \boldsymbol{Z} \cup z_i$
  end for
  $\boldsymbol{Z}_{\text{std}} \leftarrow$STANDARDIZATION$(\boldsymbol{Z})$
  $\boldsymbol{Z}_{\text{mavg}} \leftarrow$MOVING_AVERAGE$(\boldsymbol{Z}_{\text{std}}, I)$
  $\boldsymbol{Z}_{\text{SZ}} \leftarrow \boldsymbol{Z}_{\text{SZ}} + \dfrac{|\boldsymbol{Z}_{\text{mavg}}|}{K}$ ▷ element-wise summation
end for
return $\boldsymbol{Z}_{\text{SZ}}$

4 Experiment

In this study, we used data from 22 subjects in the "Juntendo" dataset and all 14 subjects in the "Siena" dataset to evaluate the proposed method. Both datasets contain one or more recordings of the EEG for each subject. The length of each EEG recording varies, ranging from a few minutes to several hours. The experimental details for each stage are described in the following subsections.

4.1 Experiment of Pre-screening of EEG

In the pre-screening stage, the EEG recordings are preprocessed with bandpass (4–80 Hz) filtering and 50 Hz notch filtering. Then, each recording is segmented with a window of 5s (non-overlapping) duration with 19 selected channels. If an EEG recording has K segments of duration Ds and M number of channels; then, for each segment i, the dimension of the corresponding spatial covariance matrix $\boldsymbol{C}_i$ will be $M \times M$ (using Equation (1)). Similarly, for channel index m, the dimension of the corresponding temporal covariance matrix Q_i^m will be $L \times L$, where L is the number of consecutive segments considered. Here, $D = 5$, $M = 19$, and $L = 2$ are used in the pre-screening stage; the number of time samples in a segment is $N = DF_{\mathrm{s}}$. For the seizure detection task, we considered seizure events whose duration was equal to or greater than 10s. Now, from each EEG segment i, we have a total of $1+M$ number of covariance matrices. After the tangent space mapping operation (see Equation (5)), the dimension of the "Spatial Vector" will be $\mathcal{T}_{\boldsymbol{C}} = M(M+1)/2$. Similarly, for each channel, the dimension of the "Temporal Vector" will be $\mathcal{T}_{\boldsymbol{Q}} = L(L+1)/2$. So, after concatenating all the vectors for segment i, the final dimension of the feature vector for the anomaly detection task will be $\mathcal{T}_v = \mathcal{T}_{\boldsymbol{C}} + M\mathcal{T}_{\boldsymbol{Q}}$.

The anomaly detection task has experimented with three anomaly detection methods: (a) one-class SVM (OCSVM), (b) isolation forest (IForest), and (c) empirical cumulative distribution functions for unsupervised outlier detection (ECOD). The OCSVM is an anomaly detection algorithm using support vector machines [56], while the IForest uses random forests to isolate anomalies from datasets efficiently. Besides, ECOD estimates the cumulative distribution of a dataset, helping identify anomalies by assessing deviations from an expected distribution [51]. Here, all three methods are experimented with a default contamination percentage of 0.1 to calculate the anomaly scores of each segment. Later, a moving averaged filtering is applied, where $I = 7$ is used for smoothing the anomaly scores. Here, $I = 7$ is equivalent to a 35s window for filtering.

When calculating the anomaly threshold, T_δ, the parameter δ plays an important role (see Equation (8)). This study experimented with the anomaly detection methods with values of δ from 0 to 2.0 with almost equal intervals.

The Python library of the anomaly detection methods and their details are publicly available here: https://github.com/yzhao062/pyod [88].

4.2 Experiment of Extra-physiological and General Artifacts

In this stage, for the "all electrode contact loss" artifact, we used 5s (non-overlapping) EEG segments to detect the event. For the "single electrode contact loss," we used 5s (non-overlapping), 1s (overlapped 80%), and 0.5s (overlapped 80%) EEG segments sequentially. In the case of the "general artifacts rejection," only the 0.5s (overlapped 80%) EEG segments were used to detect the events. Here, the number of channels M was 10, 19, and 19 used for "all electrode contact loss," "single electrode contact loss" and "general artifacts rejection," respectively, where the number of time samples N depends on D. The values of the parameters ρ, η, and γ were selected using the grid-search method, where the optimal performance was achieved with $\rho = 3.0$, $\eta = 0.3$, and $\gamma = 1.25$. Here, the general artifacts are more frequently observed in the recordings, which also indicate the "general artifacts rejection" has a higher contribution to the performance compared to the other two. Besides, the presence of "all electrode contact loss," and "single electrode contact loss" highly depends on the recording environment, whereas general artifacts are a common part of the EEG recordings.

4.3 Experiment of Artifact and Seizure Classification

In the final stage of our proposed method, we used the 0.5s (overlapped 80%) EEG segments for the artifact and seizure classification task. Due to the different parameter requirements of the Riemannian potatoes, we preprocessed the raw EEG with 50 Hz notch filtering, with bandpass filtering parameters depending on the potato types (see Table 4). For the power-based feature (see Equation (21)), we used an FIR filter with a passband between 1 and 60 Hz. The preprocessed EEG was then used to calculate the z-score for each Riemannian potato. Due to the Riemannian mean and moving average operation, we must calculate the features from all the EEG segments in a recording. In total, 8 features (7 based on Riemannian potatoes + 1 based on power) are calculated for a single 0.5s segment. The final feature vector was taken over 2.5s windows (equivalent to 21 segments concatenated together) to elevate the performance; that is, in total, 168 features fed into the classifier for a single segment i for the classification purpose. We also considered the 2.5s window ($I = 21$) as a parameter for the moving average filtering to calculate the seizure potato-based feature. The feature set was standardized before the training process. The whole feature vector generation process from EEG is shown in Figure 10.

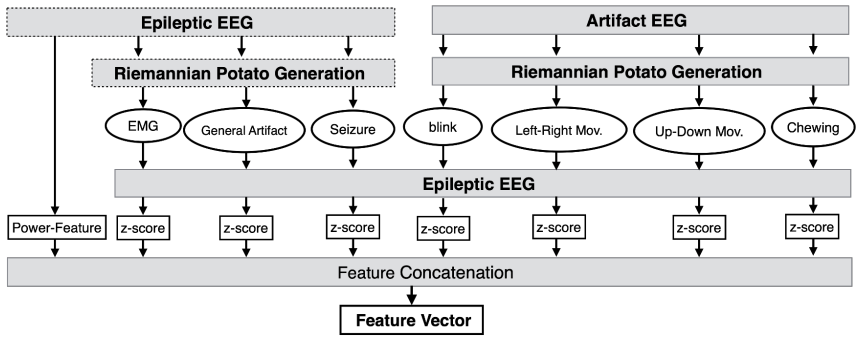


Figure 10: Feature vector generation for seizure classification

The LightGBM is a state-of-the-art method in machine learning. It is a distributed gradient-boosting framework that is free and open-source [40]. It is utilized for several purposes, such as ranking, classification, and other machine learning applications, and is based on decision tree algorithms. It is memory-efficient, supports categorical features, and exhibits high performance. It is suitable for the classification or regression of large-scale and high-dimensional datasets. The leave-one-patient-out cross-validation approach evaluated each patient's performance in each stage. For the binary classification, the LightGBM classifier gives us the predicted probability $\mathcal{P}$ in the $[0-1]$ range. Let $\mathcal{P}_i$ be the predicted probability of i-th segment, and the predicted probability threshold for a positive class is $\mathcal{P}_{\text{th}}$. Now, if $\mathcal{P}_i > \mathcal{P}_{\text{th}}$, then the corresponding segment will be considered as a positive class element (i.e., seizure segment). In a regular case, $\mathcal{P}_{\text{th}} = 0.5$ is used, but, tuning this $\mathcal{P}_{\text{th}}$ can improve the performance. In this study, we selected $\mathcal{P}_{\text{th}} = 0.1$ using the grid search method. For training purposes of the LightGBM classifier, we used the parameters "num_leaves": 31, and "num_round": 30. Here, "num_leaves" controls the complexity of the tree model, and "num_round" represents the number of iterations in the training process.

4.4 Experiment of End-to-end Approaches

In recent years, deep learning-based end-to-end approaches have gained popularity among researchers, driven by advances in computational power and the availability of datasets. We also conducted end-to-end experiments using our private dataset to compare the performance with our proposed method. For this comparison, we selected two state-of-the-art end-to-end models: "EEGNet-8,2" and "EEGWaveNet."

4.4.1 EEGNet-8,2 Model

EEGNet is a compact CNN architecture designed for EEG signal processing [46]. The EEGNet-8,2 is a variant of the EEGNet model with defined parameters. It has 8 filters in the first layer and a depth multiplier of 2, which aims for efficiency and performance for tasks with specific EEG data characteristics. The architecture is optimized for faster inference and lower computational complexity due to its fixed configuration. The network consists of several layers, each with a specific role in the classification process. The input to the network is typically shaped as $(1, M, N)$, where M represents the number of EEG channels, and N represents the number of time points. We used preprocessed EEG segments (5-s window, non-overlapping) as input to the model. The architecture applied three types of convolution sequentially: temporal convolution, depthwise convolution, and separable convolution followed by average pooling, flattening, and dense layer. First, temporal convolution with 8 filters is used to learn temporal features from each EEG channel separately. Then, a depthwise convolution is applied across all channels. It allows the model to capture spatial features while preserving channel-specific information. Afterward, a separable convolutional layer combines the depthwise convolutions with pointwise convolutions to further extract both temporal and spatial features. Later, the dimensionality of the feature map is reduced with the average pooling layer, then, the features are flattened and passed through a dense layer with softmax activation for classification. Finally, the output layer provides the probability of each class for an individual feature vector, which is used for epilepsy detection. The detailed architecture is given in Appendix A. The model was trained with the following parameters: a batch size of 1024, 200 training epochs, and a learning rate of 10^{-5}. The experiment was conducted using standardized EEG data as it demonstrated superior performance with the EEGNet-8,2 architecture.

4.4.2 EEGWaveNet Model

"EEGWaveNet" is a deep learning model for EEG signal processing, proposed in [75]. It is designed to capture long-range temporal dependencies in EEG data. This architecture is intended to capture both temporal and channel-wise features, making it a more robust framework for EEG signal classification compared to the other models. EEGWaveNet has three major components: A) a multi-scale convolution module, B) a spatial-temporal convolution feature module, and C) a predictor module classifier. The EEGWaveNet architecture uses a series of convolutional layers and pooling operations to process EEG signals. The architecture begins with six temporal convolutional layers, which utilize group-wise convolutions to reduce the sequence length while keeping channel information. After that, to extract features across channels, the

network applies pooling operations through multiple sequential blocks. Each block contains convolutional layers, batch normalization, and LeakyReLU activation. These pooled features from these blocks are then concatenated into a single vector. In the final predictor module, a fully connected classifier is implemented that maps the concatenated features to the output classes with the log-softmax function used for class probabilities. The input in the model is the same as the EEGNet-8,2 model discussed in the previous section. The detailed architecture is given in the Appendix B. The training batch size is 1024, the number of training epochs is 200, and the learning rate is 10^{-4}. The more detailed information and implementation code can be found here: https://github.com/IoBT-VISTEC/EEGWaveNet.

5 Results, Discussion and Future Direction

5.1 Performance Metrics

To evaluate the proposed method, we used some well-known performance metrics: accuracy (Acc.), precision (Pre.), sensitivity (Sen.) or recall, specificity (Spe.), and F1-score (F1). The definitions of the metrics are given below:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{22}$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \tag{23}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \tag{24}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100 \tag{25}$$

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \times 100 \tag{26}$$

where, TP, TN, FP, and FN represent the total number of correctly detected positive, correctly detected negative, incorrectly detected positive, and incorrectly detected negative EEG segments, respectively.

5.2 Anomaly Detection

Figure 11 shows the pre-screening process for a 2-hour-long EEG recording (from the Juntendo dataset) by the OCSVM classifier. The x-axis represents time in second, while the y-axis indicates the anomaly score assigned by the OCSVM classifier. The green line in the figure indicates the actual anomaly score, whereas the green line is the outcome score after moving average filtering. We marked the seizure event location with a faded red color. The red dashed

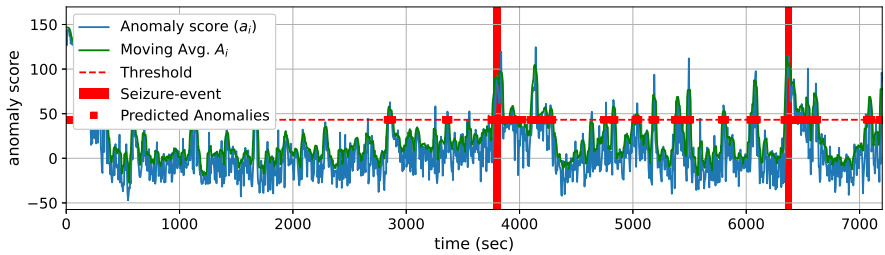
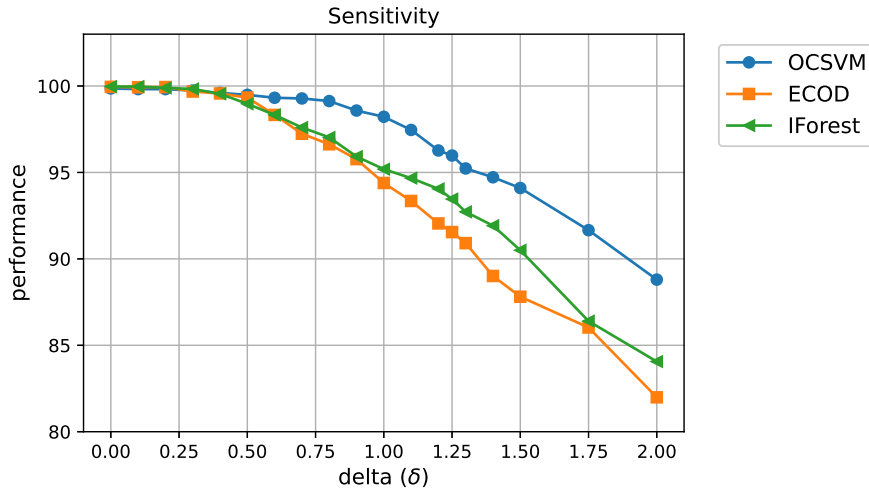


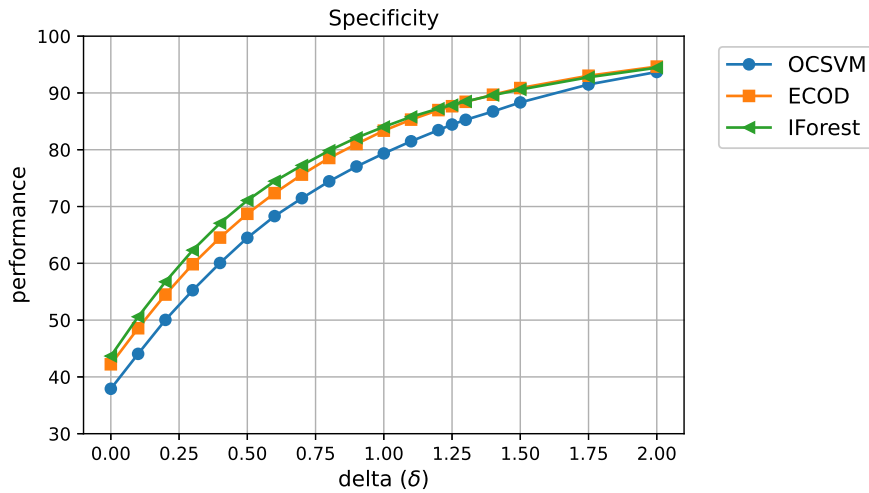
Figure 11: Anomaly score calculation of a single EEG recording by OCSVM method ($\delta = 1.0$)

line indicates the threshold with $\delta = 1.0$, and the red square is the anomaly segments labeled by the anomaly detection method. The purpose of the moving averaging filtering is to smooth the transition process from segment to segment. The main goal of anomaly detection is to filter out the normal EEG segments from the potential abnormal EEG segments for the seizure detection task. Here, the parameter δ plays a vital role in the overall model's performance. A low δ value will increase the sensitivity and the false positive rates in the detection task, making the final classification task more difficult by leveraging the class imbalance problem. Conversely, a higher δ value will decrease the false positive rates but end with lower sensitivity or higher false negative rates. Therefore, it is crucial to select the appropriate δ value for the application.

For the anomaly detection task, we also experimented with IForest and ECOD classifiers. Figure 12 shows the average sensitivity (Figure 12a) and specificity (Figure 12b) of the three anomaly detection methods: OCSVM, IForest, and ECOD. The average is taken over the performance of all the subjects from the Juntendo dataset. As discussed, the sensitivity is inversely proportional to the δ value, while the specificity shows the opposite trend. By nature, seizures are a rare event in most patients, and only a few times occur in long recordings. Therefore, from a seizure detection and machine learning viewpoint, the anomaly detection method should extract all seizure events (high sensitivity) and give false positives as low as possible. As shown in Figure 12a, OCSVM outperformed the IForest and ECOD methods, in terms of sensitivity, in the case of specificity (see Figure 12b), it's just the opposite case. Besides, it is seen in Figure 12 that the drop rate of sensitivity is significantly lower for the OCSVM compared to the other two methods (i.e., it can keep more seizure segments as anomalies compared to the others). However, the increase rate in specificity is almost the same across the methods with increasing δ values.



(a) Sensitivity



(b) Specificity

Figure 12: The average (a) Sensitivity, and (b) Specificity of the OCSVM, ECOD, and IForest methods in anomaly detection process

5.3 Ablation Study

The contribution of each stage to the final performance (Juntendo dataset) of the proposed method ($\delta = 1.0$) is shown in Figure 13. The specificity, precision, and F1–score increase in each step while the sensitivity decreases. Here, the

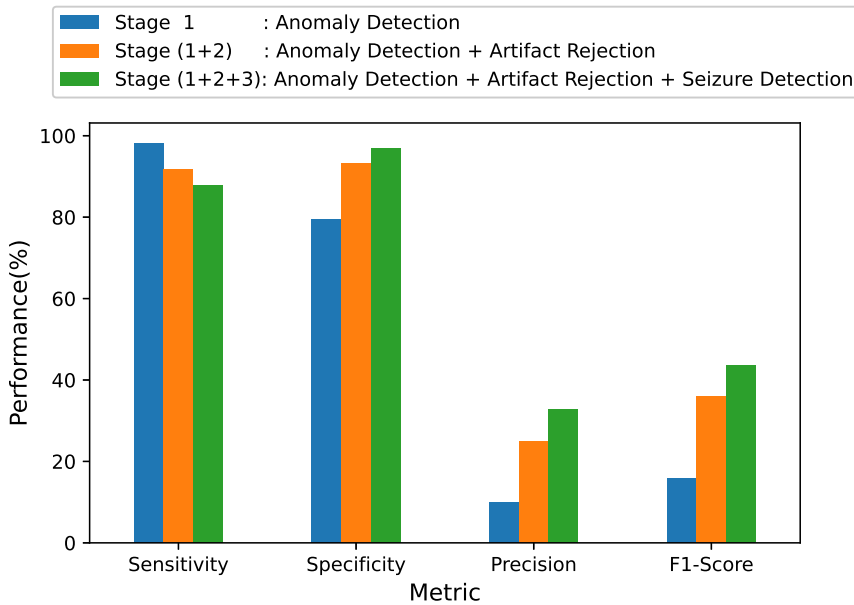


Figure 13: Contribution of each stage to the final performance (Classifier: OCSVM ($\delta = 1.0$) + LightGBM)

aim of the initial anomaly detection stage is to detect the seizure events as much as possible for further classification process. As the name implies, the next artifact rejection stage focuses on special and general types of artifacts while keeping the seizure event as much as possible. The results from this stage demonstrate high sensitivity but moderate specificity which is expected. By using the reference Riemannian potatoes from the artifact dataset, the final stage reduces false positives by improving specificity and precision, reflected in the F1–score improvement. In summary, the features of the anomaly detection stage has highest contribution in the sensitivity, whereas the features of the later two stages have more contribution to the improvement of the specificity. In Figure 14, we show the sensitivity of each stage with respect to the different δ values in the Juntendo dataset.

5.4 Seizure Classification

The main goal of this stage is to maintain high sensitivity while reducing false positives and improving specificity. As discussed in Section 4.3, the performance of the classifier can be tuned with the predicted probability threshold $\mathcal{P}_{\text{th}}$.

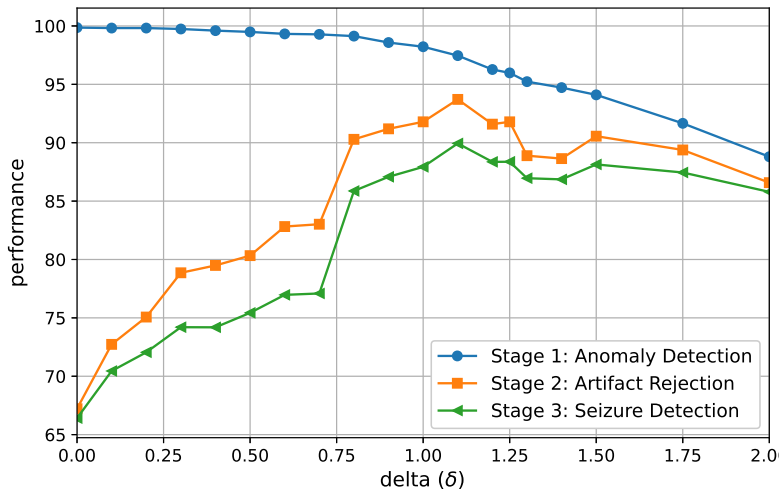
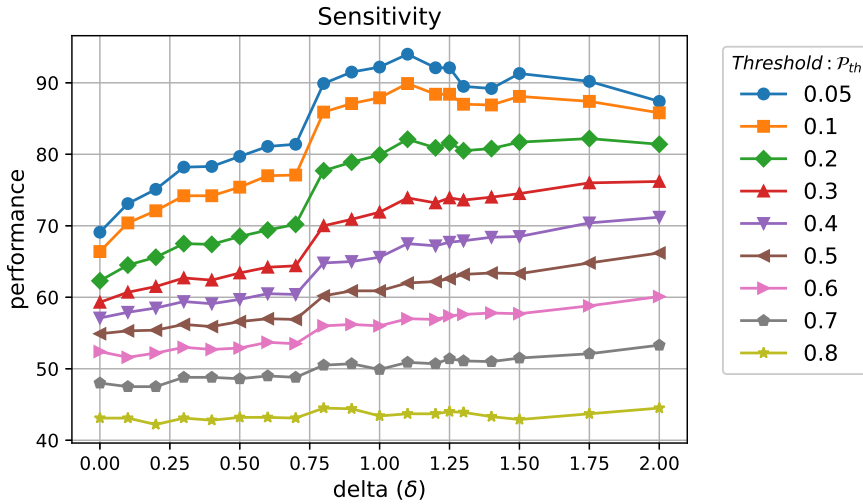


Figure 14: Contribution of the proposed method's each stage in terms of sensitivity

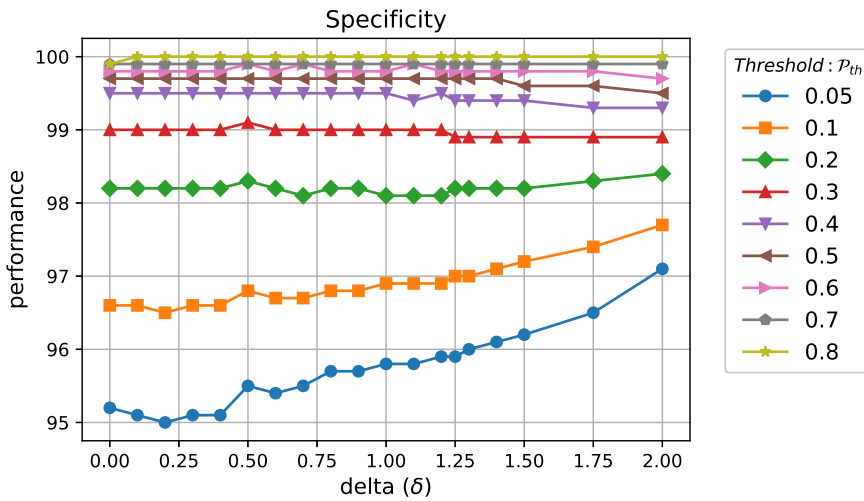
Figure 15 shows the effects of $\mathcal{P}_{\text{th}}$ with respect to the performance of sensitivity and specificity in the Juntendo dataset. As seen in Figure 15a, the sensitivity of the method is negatively correlated with $\mathcal{P}_{\text{th}}$, where the specificity (see Figure 15b) is positively correlated. Additionally, from $\delta = 0.8$ to 1.75, we see that the sensitivities for $\mathcal{P}_{\text{th}} = 0.05$ and 0.1 give us a normal distribution-like shape and have the peak performances in terms of sensitivity and optimal performance in specificity metric. Here, the selection of $\mathcal{P}_{\text{th}}$ might depend on the specific application and its focus or priorities. In this study, we considered a sensitivity of around 90% or more as an optimal solution for most of the applications with moderate specificity, which can be achieved in $\mathcal{P}_{\text{th}} = 0.1$ with less number of false positives compared to $\mathcal{P}_{\text{th}} = 0.05$.

Figure 16 shows the overall performance (using the Juntendo dataset) of the proposed method for different δ values with the LightGBM classifier, where OCSVM calculates the anomaly score. Here, the x and y-axes represent the δ and average performance of the proposed method, respectively. Due to the high imbalance class in the dataset, the specificity and accuracy are almost identical. Besides, the growth rate of the sensitivity is much higher, up to $\delta = 1.1$, before gradually degrading.

Table 5 summarizes the performance of the proposed method (OCSVM + LightGBM) for each patient in the Juntendo dataset with $\delta = 1.1$. As can be seen, for most patients, owing to the high label imbalance, the accuracy is very close to and dominated by specificity. Except for patient Pt–08, the sensitivities



(a) Sensitivity



(b) Specificity

Figure 15: The effects of the predicted probability thresholds vs (a) Sensitivity, and (b) Specificity

are over 70% for all other patients and over 90% for 15 patients. The method also faced low performance for some patients like Pt–04, Pt–14, Pt–17, and Pt–22 in terms of precision and F1–score, which is a common scenario in seizure

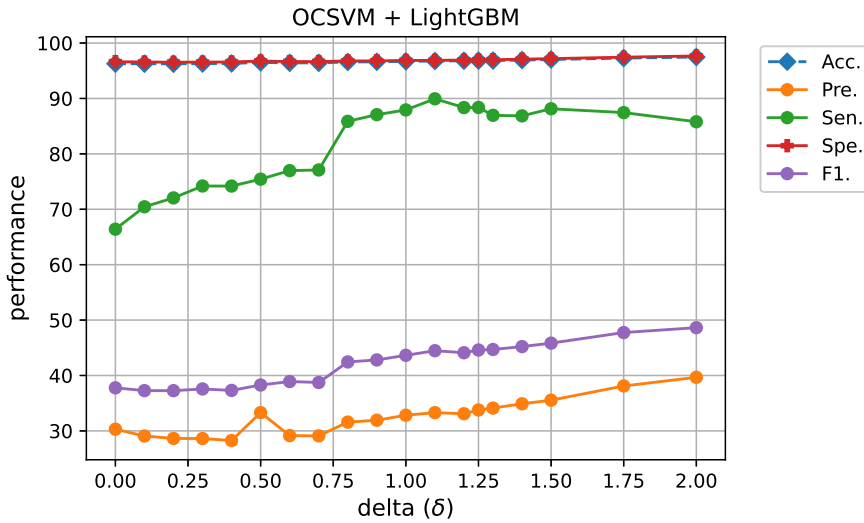


Figure 16: The average classification performance of the proposed method obtained from the Juntendo dataset (OCSVM + LightGBM) with different δ values

detection tasks, particularly when using leave-one-patient-out cross-validation approaches.

5.5 Comparative Study

In Table 6, we compare the performance of our proposed method with two state-of-the-art models, EEGNet-8,2 and EEGWaveNet. It shows that our proposed model's performance is superior (in terms of sensitivity) in detecting epileptic seizures compared to the other state-of-the-art end-to-end approaches. Sensitivity is particularly important in the medical domain because of its rare occurrence in long EEG recordings. We have provided patient-wise performance results for the end-to-end models in Appendix C, with the individual results for EEGNet-8,2 and EEGWaveNet (detailed in Tables 12, and 13), respectively. Additionally, the number of parameters in our proposed method, compared to those of the end-to-end approaches, is listed in Table 7. Notably, our model uses significantly fewer parameters than the other methods. The total parameter count for our method includes those for OCSVM and lightGBM, as well as the parameters δ, ρ, η, γ and $\mathcal{P}_{\text{th}}$. The total number of parameters is an important consideration when training a model because models with a higher parameter count require larger datasets. Therefore, reducing the number of parameters is advantageous whenever possible.

Table 5: Patient-wise performance of the proposed method in the Juntendo dataset

Pt. ID	Acc.	Pre.	Sen.	Spe.	F1
Pt-01	97.44	37.03	91.25	97.54	52.68
Pt-02	98.15	48.47	99.08	98.14	65.09
Pt-03	95.56	22.46	99.89	95.51	36.67
Pt-04	97.77	11.12	100.0	97.76	20.02
Pt-05	98.25	54.63	95.40	98.31	69.48
Pt-06	97.37	42.12	94.36	97.43	58.25
Pt-07	96.30	23.25	96.85	96.29	37.50
Pt-08	96.26	19.28	39.05	97.23	25.81
Pt-09	99.44	98.15	96.47	99.68	97.45
Pt-10	97.38	17.88	88.98	97.43	29.78
Pt-11	98.50	61.50	99.51	98.47	76.02
Pt-12	99.51	94.20	99.34	99.52	96.70
Pt-13	96.26	29.52	71.73	96.72	41.82
Pt-14	91.04	8.86	96.00	91.00	16.23
Pt-15	93.44	11.37	92.31	93.45	20.25
Pt-16	96.60	27.99	100.0	96.56	43.73
Pt-17	93.36	4.80	79.33	93.41	9.05
Pt-18	98.45	17.86	75.50	98.55	28.88
Pt-19	97.49	40.66	96.17	97.52	57.16
Pt-20	98.17	29.66	98.00	98.17	45.54
Pt-21	95.45	20.34	79.10	95.68	32.36
Pt-22	96.36	9.51	90.37	96.39	17.20
Mean	**96.75**	**33.30**	**89.94**	**96.87**	**44.47**

Table 6: Performance comparison with state-of-the-art end-to-end methods on Juntendo dataset

Methods	Acc.	Pre.	Sen.	Spe.	F1.
EEGNet-8,2	99.19	65.76	45.96	99.88	50.28
EEGWaveNet	94.86	33.36	76.75	95.07	42.14
Proposed Method	**96.75**	**33.30**	**89.94**	**96.87**	**44.47**

Table 7: Number of parameters in different methods

Methods	Number of parameters
EEGNet-8,2	7.49K
EEGWaveNet	46.39K
Proposed Method	≈ 750

We also verified the scalability of our proposed method using another publicly available epileptic dataset, the "Siena" EEG dataset (described in 3.1.2). The performance of our method is reported and compared with existing works in Table 8. Using the leave-one-patient-out cross-validation technique

on the Siena dataset, we achieved a mean sensitivity of 63.8%, specificity of 98.7%, precision of 32.4%, and an F1–score of 40.5%. The patient-wise performance is detailed in Appendix D. These results demonstrate that our method is scalable to other datasets and can achieve satisfactory performance with different recording setups.

Table 8: Performance comparison of the proposed method with Siena dataset.

Methods	Acc.	Sen.	Spe.
Time-domain + Frequency-domain + Entropy (Sigsgaard *et al.*, 2023) [70]	–	81.43	99.38
Proposed method	**98.42**	**63.75**	**98.67**

Table 9 shows a comparative performance study with recent state-of-the-art approaches. Owing to the different experimental setup and recording conditions, it is challenging to compare the results with those of other studies with the same parameters. We used the Juntendo dataset to conduct this comparative study and demonstrate how our method performs in the seizure detection task compared to other methods. All the studies listed in Table 9 used scalp EEG recordings and the leave-one-patient-out cross-validation approach to validate performance. In the Juntendo dataset, our proposed method outperformed the existing methods in terms of sensitivity (also known as recall), while maintaining moderate specificity.

Table 9: Performance comparison of the proposed method with other recent approaches.

Methods	Datasets	Acc.	Sen.	Spe.
Single-channel EEG + wavelet features + SVM (Janjarasjitt, 2017) [33]	CHB-MIT	96.87	72.99	98.13
DWT + CNN + BLSTM (Liu *et al.*, 2022) [53]	CHB-MIT	97.51	83.11	97.58
Multiscale CNN (Thuwajit *et al.*, 2022) [75]	CHB-MIT	96.74 ±4.04	75.32 ±20.46	95.96 ±4.38
Time-domain + Frequency-domain + Entropy (Sigsgaard *et al.*, 2023) [70]	Siena	–	81.43	99.38
Riemannian Manifold + SVM (Orihara *et al.* 2023) [61]	CHB-MIT	86.1	86.8	85.4
Proposed method	**Juntendo**	**96.75**	**89.94**	**96.87**

However, it is evident from the performance of different methods that the results are not satisfactory for some patients. In Appendix C, we present a patient-wise comparison of precision and sensitivity across various end-to-end methods, along with our proposed method, in Figures 17 and 18, respectively.

As shown in Figure 17, precision is low across all methods for certain patients. Additionally, while the EEGNet-8,2 shows higher precision compared to the other two methods, it suffers from low sensitivity (see Figure 18). This is because low sensitivity indicates that fewer true class samples are detected, which also reduces the probability of false positives, thereby increasing precision according to its definition.

5.6 Limitations and Future Directions

The main limitation of the proposed method is its low performance (specifically low precision and F1–score) for certain patients, which requires further study to develop a better-generalized solution. In practice, EEG signals can vary significantly between subjects due to factors such as age, gender, and individual brain anatomy. According to our analysis, the most probable reason for the low performance for certain patients is the lack of sufficient similar samples in the training dataset, i.e., they are statistically far away from most of the patients' EEG data. Usually, the models are more biased to the common seizure types, and show low performance to the patients with uncommon or unique seizure types due to the training process. Moreover, a few patients' EEG show higher class-overlapping characteristics between artifacts and seizures than others. Additionally, it is quite common to observe that some patients' EEGs also contain lots of noise because of the noise-prone recording process. Any of the aforementioned issues or their combinations could be a possible reason for the low performance, it requires further study to improve the frameworks' robustness in the future.

The class overlapping issue of artifact and seizure events is a challenging task. According to our observation, the context-based EEG analysis could help solve the class-overlapping problems in near future. Additionally, zero-shot, one-shot, and few-shot-based leanings have gained popularity in recent times. Those approaches could open some paths to solving the unseen seizure-type problem in the model training process.

As mentioned before, the parameter values that we have used are selected using grid-search method and we have found them optimal for our experimental dataset. However, more study needed on other dataset to confirm their effectiveness on seizure detection task. The parameters might need to tune in a certain extent to get the best performance with different dataset settings. Those parameter values should also tuned based on the target application.

6 Conclusion

Detection of epileptic seizures by using the scalp EEG alone is always challenging due to its non-linearity and artifact-prone recording process. This study

proposed a framework to automate the seizure detection process from scalp EEG recordings with anomaly detection, artifact rejection, and seizure detection techniques using the Riemannian manifold and transfer-learning-based approaches. The proposed method comprises three major stages: the initial pre-screening of EEG for anomaly identification, eliminating extra-physiological and general artifacts, and the detection of seizures from the remaining abnormal EEG with classification. The Riemannian manifold-based features were the backbone of the feature calculation combined with the power-based features. The proposed method was evaluated using a leave-one-patient-out cross-validation approach on private and public epileptic EEG datasets. The data-adaptive method used only a few parameters that can be tuned based on application requirements or priorities for performance metrics, such as sensitivity or specificity. Despite the low performance in some cases, the overall performance of the proposed method is still quite promising, demonstrating the potential for implementation in real-world applications with further development.

A EEGNet-8,2 Architecture

In Table 10, $M(= 19)$ denotes the number of channels, $N(= 2500)$ is the number of time points, and ELU denotes "Exponential Linear Unit". Here, the dropout rate is 0.25.

Table 10: EEGNet-8,2 Architecture

Layer	Type	Kernel	Output	Activation
Input Layer	–	–	(1, M, N)	–
Temporal Convolution	Conv2D	(1, 64)	(8, M, N-63)	Linear
Batch Normalization	BatchNorm2D	–	(8, M, N-63)	–
Depthwise Convolution	DepthwiseConv2D	(M, 1)	(16, 1, N-63)	Linear
Batch Normalization	BatchNorm2D	–	(16, 1, N-63)	–
Activation	ELU	–	(16, 1, N-63)	ELU
Average Pooling	AvgPool2D	(1, 4)	(16, 1, (N-63)//4)	–
Dropout	Dropout	–	(16, 1, (N-63)//4)	–
Separable Convolution	SeparableConv2D	(1, 16)	(16, 1, (N-63)//4-15)	Linear
Batch Normalization	BatchNorm2D	–	(16, 1, (N-63)//4-15)	–
Activation	ELU	–	(16, 1, (N-63)//4-15)	ELU
Average Pooling	AvgPool2D	(1, 8)	(16, 1, ((N-63)//4 - 15)//8)	–
Dropout	Dropout	–	(16, 1, ((N-63)//4 - 15)//8)	–
Flatten	–	–	(Flattened Output)	–
Dense Layer	Dense	–	2	Softmax

B EEGWaveNet Architecture

In Table 11, $M(= 19)$ denotes the number of channels, $N(= 2500)$ is the number of time points, and k is the order of layers in Module B.

Table 11: EEGWaveNet Architecture

Module	Layer	Kernel	Output	Activation
A	Input		(M, N)	Linear
	Conv1D	kernel 2, stride 2, group M	(M, N/2)	Linear
	Conv1D	kernel 2, stride 2, group M	(M, N/4)	Linear
	Conv1D	kernel 2, stride 2, group M	(M, N/8)	Linear
	Conv1D	kernel 2, stride 2, group M	(M, N/16)	Linear
	Conv1D	kernel 2, stride 2, group M	(M, N/32)	Linear
	Conv1D	kernel 2, stride 2, group M	(M, N/64)	Linear
B	Input		(M, $N/2^k$)	
	Conv1D	(32×4)	(32, $N/2^k - 3$)	
	BatchNorm1D			
	Activation			Leaky ReLU
	Conv1D	(32×4)	(32, $N/2^k - 6$)	
	BatchNorm1D			
	Activation			Leaky ReLU
	Global average pooling		32	
C	Input		160	
	Fully Connected		64	Leaky ReLU
	Fully Connected		32	Sigmoid
	Fully Connected		2	Linear
	Classifier		1	Log Softmax

C Juntendo Dataset Patient-wise Results: End-to-End Models

The patient-wise performances of the EEGNet-8,2 and EEGWaveNet are shown in Tables 12 and 13, respectively. Additionally, the comparative performance in terms of "precision" and "sensitivity" for all the patients across different methods are shown in Figures 17 and 18, respectively.

Table 12: EEGNet-8,2 Patient-wise Performance

Pt.ID	Acc.	Pre.	Sen.	Spe.	F1
Pt-01	99.97	97.83	100.0	99.96	98.9
Pt-02	98.30	100.0	2.0	100.0	3.92
Pt-03	99.03	78.12	33.78	99.88	47.17
Pt-04	99.72	0.0	0.0	100.0	0.0
Pt-05	98.38	73.81	34.44	99.74	46.97
Pt-06	98.89	95.0	45.24	99.95	61.29
Pt-07	99.51	95.24	60.61	99.96	74.07
Pt-08	98.58	70.59	25.0	99.82	36.92
Pt-09	98.47	100.0	90.39	100.0	94.95
Pt-10	99.32	30.0	6.67	99.9	10.91
Pt-11	99.03	91.84	65.22	99.86	76.27
Pt-12	99.72	100.0	96.2	100.0	98.06
Pt-13	98.61	88.89	29.63	99.93	44.44
Pt-14	99.03	31.25	6.41	99.87	10.64
Pt-15	99.1	0.0	0.0	100.0	0.0
Pt-16	99.79	100.0	84.21	100.0	91.43
Pt-17	99.93	85.71	100.0	99.93	92.31
Pt-18	99.48	0.0	0.0	99.9	0.0
Pt-19	99.44	75.76	100.0	99.43	86.21
Pt-20	99.97	95.74	100.0	99.97	97.83
Pt-21	98.32	36.9	31.31	99.25	33.88
Pt-22	99.58	0.0	0.0	100.0	0.0
Mean	**99.19**	**65.76**	**45.96**	**99.88**	**50.28**

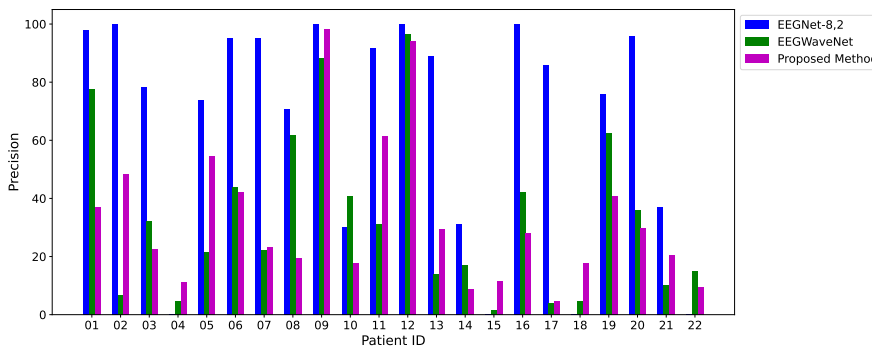


Figure 17: Patient-wise "precision" comparison across different methods.

Table 13: EEGWaveNet Patient-wise Performance

Pt.ID	Acc.	Pre.	Sen.	Spe.	F1
Pt-01	99.55	77.59	100.0	99.54	87.38
Pt-02	80.21	6.67	80.0	80.21	12.31
Pt-03	97.53	32.29	83.78	97.71	46.62
Pt-04	98.33	4.55	25.0	98.54	7.69
Pt-05	94.0	21.36	70.0	94.52	32.73
Pt-06	97.64	43.92	77.38	98.04	56.03
Pt-07	96.22	22.06	90.91	96.28	35.5
Pt-08	98.82	61.67	77.08	99.19	68.52
Pt-09	97.43	88.4	96.51	97.61	92.28
Pt-10	99.21	40.62	57.78	99.47	47.71
Pt-11	95.56	31.21	71.01	96.16	43.36
Pt-12	99.65	96.59	98.73	99.73	97.65
Pt-13	91.25	13.87	70.37	91.65	23.17
Pt-14	97.11	17.18	57.69	97.47	26.47
Pt-15	72.88	1.54	46.15	73.13	2.98
Pt-16	98.19	42.22	100.0	98.17	59.38
Pt-17	90.14	4.05	100.0	90.1	7.79
Pt-18	93.54	4.69	75.0	93.62	8.82
Pt-19	98.96	62.5	100.0	98.94	76.92
Pt-20	98.61	36.0	100.0	98.6	52.94
Pt-21	93.83	10.16	44.44	94.52	16.54
Pt-22	98.26	14.81	66.67	98.4	24.24
Mean	**94.86**	**33.36**	**76.75**	**95.07**	**42.14**

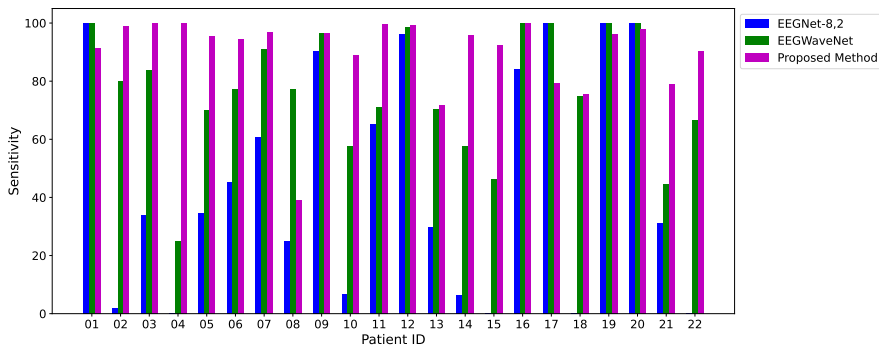


Figure 18: Patient-wise "sensitivity" comparison across different methods.

D Siena Dataset Patient-wise Performance: Proposed Method

The patient-wise performance of the proposed method on the "Siena" dataset is shown in Table 14.

Table 14: Patient-wise performance of the proposed method in the Siena dataset

Pt. ID	Acc.	Pre.	Sen.	Spe.	F1
PN00	98.90	89.71	68.60	99.77	77.75
PN01	99.34	25.58	88.25	99.36	39.66
PN03	98.96	12.97	47.06	99.11	20.34
PN05	99.54	43.85	72.29	99.65	54.58
PN06	98.08	19.28	71.38	98.24	30.36
PN07	95.80	00.20	04.17	95.97	00.38
PN09	98.94	43.26	87.02	99.04	57.79
PN10	98.86	26.56	62.35	99.06	37.25
PN11	96.23	06.16	34.73	96.62	10.46
PN12	98.91	46.63	60.60	99.30	52.71
PN13	99.06	46.95	82.19	99.20	59.76
PN14	98.64	18.92	78.78	98.72	30.51
PN16	97.79	27.41	58.19	98.24	37.27
PN17	99.04	45.82	76.77	99.23	57.39
Mean	**98.42**	**32.37**	**63.75**	**98.67**	**40.45**

References

[1] D. Abasolo, C. J. James, and R. Hornero, "Non-linear Analysis of Intracranial Electroencephalogram Recordings with Approximate Entropy and Lempel-Ziv Complexity for Epileptic Seizure Detection", in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, 1953–6, DOI: 10.1109/IEMBS.2007.4352700.

[2] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals", en, *Computers in Biology and Medicine*, 100 (September), September 2018, 270–8, ISSN: 00104825, DOI: 10.1016/j.compbiomed.2017.09.017.

[3] M. S. Akter, M. R. Islam, Y. Iimura, H. Sugano, K. Fukumori, D. Wang, T. Tanaka, and A. Cichocki, "Multiband entropy–based feature–extraction method for automatic identification of epileptic focus based on high-frequency components in interictal iEEG", *Scientific reports*, 10(1), 2020, 1–17, DOI: https://doi.org/10.1038/s41598-020-62967-z.

[4] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs", *Pattern Recognition*, 45(6), 2012, 2137–44, Brain Decoding, ISSN: 0031-3203, DOI: https://doi.org/10.1016/j.patcog.2011.04.018.

[5] S. K. Arjaria, G. Chaubey, and N. Shukla, "Hjorth Parameter based Seizure Diagnosis using Cluster Analysis", *Journal of Physics: Conference Series*, 1998(1), August 2021, 012020, DOI: 10.1088/1742-6596/1998/1/012020.

[6] A. Barachant, A. Andreev, and M. Congedo, "The Riemannian Potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry", *TOBI Workshop IV*, January, January 2013, 19–20, https://hal.science/hal-00781701/document.

[7] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry", *IEEE Transactions on Biomedical Engineering*, 59(4), 2011, 920–8, DOI: 10.1109/TBME.2011.2172210.

[8] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications", *Neurocomputing*, 112, 2013, 172–8, Advances in artificial neural networks, machine learning, and computational intelligence, ISSN: 0925-2312, DOI: https://doi.org/10.1016/j.neucom.2012.12.039.

[9] Q. Barthelemy, L. Mayaud, D. Ojeda, and M. Congedo, "The Riemannian Potato Field: A Tool for Online Signal Quality Index of EEG", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(2), February 2019, 244–55, ISSN: 1534-4320, 1558-0210, DOI: 10.1109/TNSRE.2019.2893113.

[10] C. Baumgartner and J. P. Koren, "Seizure detection using scalp-EEG", *Epilepsia*, 59(S1), 2018, 14–22, DOI: https://doi.org/10.1111/epi.14052.

[11] P. Berg and M. Scherg, "Dipole modelling of eye activity and its application to the removal of eye artefacts from the EEG and MEG", *Clinical Physics and Physiological Measurement*, 12(A), January 1991, 49, DOI: 10.1088/0143-0815/12/A/010.

[12] K. J. Blinowska, R. Kus, and M. Kaminski, "Granger causality and information flow in multivariate processes", *Phys. Rev. E*, 70(5), 5 November 2004, 050902, DOI: 10.1103/PhysRevE.70.050902.

[13] S. Casarotto, A. M. Bianchi, S. Cerutti, and G. A. Chiarenza, "Principal component analysis for reduction of ocular artefacts in event-related potentials of normal and dyslexic children", *Clinical Neurophysiology*, 115(3), March 2004, 609–19, ISSN: 1388-2457, DOI: 10.1016/J.CLINPH.2003.10.018.

[14] X. Chen, Z. J. Wang, and M. McKeown, "Joint Blind Source Separation for Neurophysiological Data Analysis: Multiset and multimodal methods",

IEEE Signal Processing Magazine, 33(3), 2016, 86–107, DOI: 10.1109/MSP.2016.2521870.

[15] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review", *Brain-Computer Interfaces*, 4(3), 2017, 155–74, DOI: 10.1080/2326263X.2017.1297192.

[16] J. Craley, E. Johnson, C. Jouny, and A. Venkataraman, "Automated inter-patient seizure detection using multichannel Convolutional and Recurrent Neural Networks", *Biomedical Signal Processing and Control*, 64, 2021, 102360, DOI: 10.1016/j.bspc.2020.102360.

[17] W. D. Clercq, A. Vergult, B. Vanrumste, W. V. Paesschen, and S. V. Huffel, "Canonical Correlation Analysis Applied to Remove Muscle Artifacts From the Electroencephalogram", *IEEE Transactions on Biomedical Engineering*, 53(12), 2006, 2583–7, DOI: 10.1109/TBME.2006.879459.

[18] M.-J. Darvishi-Bayazi, M. S. Ghaemi, T. Lesort, M. R. Arefin, J. Faubert, and I. Rish, "Amplifying pathological detection in EEG signaling pathways through cross-dataset transfer learning", en, *Computers in Biology and Medicine*, 169 (February), February 2024, 107893, ISSN: 00104825, DOI: 10.1016/j.compbiomed.2023.107893.

[19] J. Dauwels, F. Vialatte, and A. Cichocki, "Diagnosis of Alzheimer's Disease from EEG Signals: Where Are We Now?", *Current Alzheimer Research*, 7(6), 2010, 674–93, DOI: 10.2174/156720510792231720.

[20] A. Emami, N. Kunii, T. Matsuo, T. Shinozaki, K. Kawai, and H. Takahashi, "Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images", *NeuroImage: Clinical*, 22, 2019, 101684, ISSN: 2213-1582, DOI: https://doi.org/10.1016/j.nicl.2019.101684.

[21] A. Flexer, H. Bauer, J. Pripfl, and G. Dorffner, "Using ICA for removal of ocular artifacts in EEG recorded from blind subjects", *Neural Networks*, 18(7), 2005, 998–1005, ISSN: 0893-6080, DOI: https://doi.org/10.1016/j.neunet.2005.03.012.

[22] S. Fordington and M. Manford, "A review of seizures and epilepsy following traumatic brain injury", en, *Journal of Neurology*, 267(10), October 2020, 3105–11, ISSN: 0340-5354, 1432-1459, DOI: 10.1007/s00415-020-09926-w.

[23] J. Gotman, "Automatic seizure detection: improvements and evaluation", *Electroencephalography and Clinical Neurophysiology*, 76(4), 1990, 317–24, ISSN: 0013-4694, DOI: https://doi.org/10.1016/0013-4694(90)90032-F.

[24] L. Guo, D. Rivero, and A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks", en, *Journal of Neuroscience Methods*, 193(1), October 2010, 156–63, ISSN: 01650270, DOI: 10.1016/j.jneumeth.2010.08.030.

[25] K. M. Hassan, M. R. Islam, T. T. Nguyen, and M. K. I. Molla, "Epileptic seizure detection in EEG using mutual information-based best individual feature selection", en, *Expert Systems with Applications*, 193 (May), May 2022, 116414, ISSN: 09574174, DOI: 10.1016/j.eswa.2021.116414.

[26] K. M. Hassan, M. R. Islam, T. Tanaka, and M. K. I. Molla, "Epileptic Seizure Detection from EEG Signals Using Multiband Features with Feedforward Neural Network", in *2019 International Conference on Cyberworlds (CW)*, 2019, 231–8, DOI: 10.1109/CW.2019.00046.

[27] K. M. Hassan, X. Zhao, H. Sugano, and T. Tanaka, "Detection of Epileptic Seizures in Long EEG Recordings Using an Anomaly Detector with Artifact Rejection", in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, 2230–4, DOI: 10.1109/ICASSP48485.2024.10447376.

[28] P. He, M. Kahle, G. Wilson, and C. Russell, "Removal of ocular artifacts from EEG: A comparison of adaptive filtering method and regression method using simulated data", *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, vol. 7, 2005, ISSN: 05891019, DOI: 10.1109/iembs.2005.1616614.

[29] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory", *Physica D: Nonlinear Phenomena*, 31(2), 1988, 277–83, ISSN: 0167-2789, DOI: https://doi.org/10.1016/0167-2789(88)90081-4.

[30] S. A. Hillyard and R. Galambos, "Eye movement artifact in the CNV", *Electroencephalography and Clinical Neurophysiology*, 28(2), 1970, 173–82, ISSN: 0013-4694, DOI: https://doi.org/10.1016/0013-4694(70)90185-9.

[31] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications", *Neural Networks*, 13(4), 2000, 411–30, ISSN: 0893-6080, DOI: https://doi.org/10.1016/S0893-6080(00)00026-5.

[32] T. Itakura and T. Tanaka, "Epileptic focus localization based on bivariate empirical mode decomposition and entropy", in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Kuala Lumpur: IEEE, December 2017, 1426–9, ISBN: 9781538615423, DOI: 10.1109/APSIPA.2017.8282255.

[33] S. Janjarasjitt, "Epileptic seizure classifications of single-channel scalp EEG data using wavelet-based features and SVM", en, *Medical & Biological Engineering & Computing*, 55(10), October 2017, 1743–61, ISSN: 0140-0118, 1741-0444, DOI: 10.1007/s11517-017-1613-2.

[34] X. Jiang, G. Bian, and Z. Tian, "Removal of Artifacts from EEG Signals: A Review", en, *Sensors*, 19(5), February 2019, 987, ISSN: 1424-8220, DOI: 10.3390/s19050987.

[35] A. M. Judith, S. B. Priya, and R. K. Mahendran, "Artifact Removal from EEG signals using Regenerative Multi-Dimensional Singular Value Decomposition and Independent Component Analysis", *Biomedical Signal*

Processing and Control, 74, 2022, ISSN: 17468108, DOI: 10.1016/j.bspc.2021.103452.

[36] T. P. Jung, C. Humphries, T. W. Lee, S. Makeig, M. McKeown, V. Iragui, and T. Sejnowski, "Removing electroencephalographic artifacts: comparison between ICA and PCA", in *Neural Networks for Signal Processing VIII. Proceedings of the 1998 IEEE Signal Processing Society Workshop (Cat. No.98TH8378)*, Cambridge, UK: IEEE, 1998, 63–72, ISBN: 9780780350601, DOI: 10.1109/NNSP.1998.710633.

[37] S. Kammerman, "Seizure disorders: Part 1. Classification and diagnosis", *Western Journal of Medicine*, 175(2), August 2001, 99–103, ISSN: 00930415, DOI: 10.1136/ewjm.175.2.99.

[38] M. J. Katz, "Fractals and the analysis of waveforms", *Computers in Biology and Medicine*, 18(3), 1988, 145–56, ISSN: 0010-4825, DOI: https://doi.org/10.1016/0010-4825(88)90041-8.

[39] I. Kaya, "A Brief Summary of EEG Artifact Handling", in *Brain-Computer Interface*, ed. V. Asadpour, Rijeka: IntechOpen, 2021, chap. 2, DOI: 10.5772/intechopen.99127.

[40] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: a highly efficient gradient boosting decision tree", in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, *NIPS'17*, Long Beach, California, USA: Curran Associates Inc., 2017, 3149–57, ISBN: 9781510860964, https://dl.acm.org/doi/10.5555/3294996.3295074.

[41] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards Zero Training for Brain-Computer Interfacing", en, *PLoS ONE*, 3(8), August 2008, e2967, ed. O. Sporns, ISSN: 1932-6203, DOI: 10.1371/journal.pone.0002967.

[42] P. S. Kumar, R. Arumuganathan, K. Sivakumar, and C. Vimal, "Removal of ocular artifacts in the EEG through wavelet transform without using an EOG reference channel", eng, *International Journal of Open Problems in Computer Science and Mathematics. IJOPCM*, 1(3), 2008, 189–200, https://api.semanticscholar.org/CorpusID:16228554.

[43] Y. Kumar, M. Dewal, and R. Anand, "Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine", en, *Neurocomputing*, 133 (June), June 2014, 271–9, ISSN: 09252312, DOI: 10.1016/j.neucom.2013.11.009.

[44] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals", en, *Human Brain Mapping*, 8(4), 1999, 194–208, ISSN: 1065-9471, 1097-0193, DOI: 10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C.

[45] K. G. A. Lakshmi, S. N. N. Surling, and O. Sheeba, "A novel approach for the removal of artifacts in EEG signals", *2017 International Confer-*

ence on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017, 2595–9, DOI: 10.1109/WiSPNET.2017.8300232.

[46] V. J. Lawhern, N. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces", *Journal of neural engineering*, 15(5), 2018, 056013, https://dx.doi.org/10.1088/1741-2552/aace8c.

[47] S. E. Leurgans, R. A. Moyeed, and B. W. Silverman, "Canonical Correlation Analysis When the Data are Curves", *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(3), July 1993, 725–40, ISSN: 0035-9246, DOI: 10.1111/j.2517-6161.1993.tb01936.x.

[48] M. Li and N. Zhang, "A dynamic directed transfer function for brain functional network-based feature extraction", en, *Brain Informatics*, 9(1), December 2022, 7, ISSN: 2198-4018, 2198-4026, DOI: 10.1186/s40708-022-00154-8.

[49] Y. Li, K. Wong, and H. De Bruin, "Electroencephalogram signals classification for sleep-state decision – a Riemannian geometry approach", en, *IET Signal Processing*, 6(4), 2012, 288, ISSN: 17519675, DOI: 10.1049/iet-spr.2011.0234.

[50] Y. Li and K. Wong, "Signal classification by power spectral density: An approach via Riemannian geometry", in *2012 IEEE Statistical Signal Processing Workshop (SSP)*, 2012, 900–3, DOI: 10.1109/SSP.2012.6319854.

[51] Z. Li, Y. Zhao, X. Hu, N. Botta, C. Ionescu, and G. Chen, "ECOD: Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions", *IEEE Transactions on Knowledge and Data Engineering*, 2022, 1–1, DOI: 10.1109/tkde.2022.3159580.

[52] J. Lin and A. Zhang, "Fault feature separation using wavelet-ICA filter", *NDT & E International*, 38(6), September 2005, 421–7, ISSN: 0963-8695, DOI: 10.1016/J.NDTEINT.2004.11.005.

[53] G. Liu, L. Tian, and W. Zhou, "Patient-Independent Seizure Detection Based on Channel-Perturbation Convolutional Neural Network and Bidirectional Long Short-Term Memory", en, *International Journal of Neural Systems*, 32(06), June 2022, 2150051, ISSN: 0129-0657, 1793-6462, DOI: 10.1142/S0129065721500519.

[54] S. Makeig, A. Bell, T. P. Jung, and T. J. Sejnowski, "Independent Component Analysis of Electroencephalographic Data", *Advances in Neural Information Processing Systems*, 8, 1995, https://dl.acm.org/doi/10.5555/2998828.2998849.

[55] N. Mammone, F. La Foresta, and F. C. Morabito, "Automatic Artifact Rejection From Multichannel Scalp EEG by Wavelet ICA", *IEEE Sensors Journal*, 12(3), 2012, 533–42, DOI: 10.1109/JSEN.2011.2115236.

[56] L. M. Manevitz and M. Yousef, "One-class svms for document classification", *Journal of Machine Learning Research*, 2 (March), March 2002, 139–54, ISSN: 1532-4435, https://dl.acm.org/doi/10.5555/944790.944808.

[57] C. C. Mayer, M. Bachler, M. Hörtenhuber, C. Stocker, A. Holzinger, and S. Wassertheurer, "Selection of entropy-measure parameters for knowledge discovery in heart rate variability data", *BMC bioinformatics*, 15(S6), 2014, S2, DOI: 10.1186/1471-2105-15-S6-S2.

[58] M. K. I. Molla, K. M. Hassan, M. R. Islam, and T. Tanaka, "Graph Eigen Decomposition-Based Feature-Selection Method for Epileptic Seizure Detection Using Electroencephalography", *Sensors*, 20(16), 2020, ISSN: 1424-8220, DOI: 10.3390/s20164639.

[59] M. Mursalin, Y. Zhang, Y. Chen, and N. V. Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier", en, *Neurocomputing*, 241 (June), June 2017, 204–14, ISSN: 09252312, DOI: 10.1016/j.neucom.2017.02.053.

[60] N. Nicolaou and J. Georgiou, "Detection of epileptic electroencephalogram based on Permutation Entropy and Support Vector Machines", en, *Expert Systems with Applications*, 39(1), January 2012, 202–9, ISSN: 09574174, DOI: 10.1016/j.eswa.2011.07.008.

[61] T. Orihara, K. M. Hassan, and T. Tanaka, "Active Selection of Source Patients in Transfer Learning for Epileptic Seizure Detection Using Riemannian Manifold", in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, 1–5, DOI: 10.1109/ICASSP49357.2023.10095272.

[62] T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert, "Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features", *Journal of Neural Engineering*, 14(4), May 2017, 046004, DOI: 10.1088/1741-2552/aa69d1.

[63] S. Saba-Sadiya, E. Chantland, T. Alhanai, T. Liu, and M. M. Ghassemi, "Unsupervised EEG Artifact Detection and Correction", *Frontiers in Digital Health*, 2, 2021, ISSN: 2673-253X, DOI: 10.3389/fdgth.2020.608920.

[64] S. Sanei and J. Chambers, *EEG Signal Processing*, Wiley, 2007, ISBN: 9780470511923, DOI: DOI:10.1002/9780470511923.

[65] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Champaign, IL: University of Illinois Press, 1998.

[66] A. Shariat, A. Zarei, S. A. Karvigh, and B. M. Asl, "Automatic detection of epileptic seizures using Riemannian geometry from scalp EEG recordings", en, *Medical & Biological Engineering & Computing*, 59(7-8), August 2021, 1431–45, ISSN: 0140-0118, 1741-0444, DOI: 10.1007/s11517-021-02385-z.

[67] X. Shi, Q. She, F. Fang, M. Meng, T. Tan, and Y. Zhang, "Enhancing cross-subject EEG emotion recognition through multi-source manifold metric transfer learning", en, *Computers in Biology and Medicine*, 174 (May), May 2024, 108445, ISSN: 00104825, DOI: 10.1016/j.compbiomed.2024.108445.

[68] L. Shoker, S. Sanei, and J. Chambers, "Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm", *IEEE Signal Processing Letters*, 12(10), 2005, 721–4, ISSN: 10709908, DOI: 10.1109/LSP.2005.855539.

[69] M. K. Siddiqui, R. Morales-Menendez, X. Huang, and N. Hussain, "A review of epileptic seizure detection using machine learning classifiers", en, *Brain Informatics*, 7(1), December 2020, 5, ISSN: 2198-4018, 2198-4026, DOI: 10.1186/s40708-020-00105-1.

[70] G. M. Sigsgaard and Y. Gu, "Improving the generalization of patient non-specific model for epileptic seizure detection", *Biomedical Physics & Engineering Express*, 10(1), December 2023, 15010, DOI: 10.1088/2057-1976/ad097f.

[71] Y. Song and P. Liò, "A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine", *Journal of Biomedical Science and Engineering*, 03(06), 2010, 556–67, ISSN: 1937-6871, 1937-688X, DOI: 10.4236/jbise.2010.36078.

[72] J. M. Stern, *Atlas of EEG Patterns*, Philadelphia, PA: Lippincott Williams & Wilkins, 2005.

[73] K. T. Sweeney, T. E. Ward, and S. F. McLoone, "Artifact Removal in Physiological Signals—Practices and Possibilities", *IEEE Transactions on Information Technology in Biomedicine*, 16(3), 2012, 488–500, ISSN: 10897771, DOI: 10.1109/TITB.2012.2188536.

[74] H. Takahashi, A. Emami, T. Shinozaki, N. Kunii, T. Matsuo, and K. Kawai, "Convolutional neural network with autoencoder-assisted multiclass labelling for seizure detection based on scalp electroencephalography", *Computers in Biology and Medicine*, 125, 2020, 104016, ISSN: 0010-4825, DOI: https://doi.org/10.1016/j.compbiomed.2020.104016.

[75] P. Thuwajit, P. Rangpong, P. Sawangjai, P. Autthasan, R. Chaisaen, N. Banluesombatkul, P. Boonchit, N. Tatsaringkansakul, T. Sudhawiyangkul, and T. Wilaiprasitporn, "EEGWaveNet: Multiscale CNN-based Spatiotemporal Feature Extraction for EEG Seizure Detection", *IEEE Transactions on Industrial Informatics*, 18(8), August 2022, 5547–57, ISSN: 1551-3203, 1941-0050, DOI: 10.1109/TII.2021.3133307.

[76] M. C. Tjepkema-Cloostermans, F. B. Van Meulen, G. Meinsma, and M. J. Van Putten, "A Cerebral Recovery Index (CRI) for early prognosis in patients after cardiac arrest", en, *Critical Care*, 17(5), October 2013, R252, ISSN: 1364-8535, DOI: 10.1186/cc13078.

[77] R. Tomioka and K. Aihara, "Classifying matrices with a spectral regularization", in *Proceedings of the 24th International Conference on Machine Learning, ICML '07*, Corvalis, Oregon, USA: Association for Computing Machinery, 2007, 895–902, DOI: 10.1145/1273496.1273609.

[78] A. Turnip and E. Junaidi, "Removal artifacts from EEG signal using independent component analysis and principal component analysis", in *2014 2nd International Conference on Technology, Informatics, Management, Engineering & Environment*, 2014, 296–302, DOI: 10.1109/TIME-E.2014.7011635.

[79] I. Ullah, M. Hussain, E. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach", en, *Expert Systems with Applications*, 107 (October), October 2018, 61–71, ISSN: 09574174, DOI: 10.1016/j.eswa.2018.04.021.

[80] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal—state-of-the-art and guidelines", *Journal of Neural Engineering*, 12(3), April 2015, 031001, ISSN: 1741-2560, DOI: 10.1088/1741-2560/12/3/031001.

[81] R. N. Vigario, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings", *IEEE Transactions on Biomedical Engineering*, 47(5), 2000, 589–93, DOI: 10.1109/10.841330.

[82] R. N. Vigario, "Extraction of ocular artefacts from EEG using independent component analysis", *Electroencephalography and Clinical Neurophysiology*, 103(3), 1997, 395–404, ISSN: 0013-4694, DOI: https://doi.org/10.1016/S0013-4694(97)00042-8.

[83] D. M. Vos, S. Riès, K. Vanderperren, B. Vanrumste, F. X. Alario, V. S. Huffel, and B. Burle, "Removal of muscle artifacts from EEG recordings of spoken language production", *Neuroinformatics*, 8(2), June 2010, ISSN: 15392791, DOI: 10.1007/s12021-010-9071-0.

[84] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series", *Physica D: Nonlinear Phenomena*, 16(3), 1985, 285–317, ISSN: 0167-2789, DOI: https://doi.org/10.1016/0167-2789(85)90011-9.

[85] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short data sets", *Annals of biomedical engineering*, 41(2), 2013, 349–65, DOI: 10.1007/s10439-012-0668-3.

[86] F. Yger, M. Berar, and F. Lotte, "Riemannian Approaches in Brain-Computer Interfaces: A Review", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10), October 2017, 1753–62, ISSN: 1534-4320, DOI: 10.1109/TNSRE.2016.2627016.

[87] X. Zhao, N. Yoshida, T. Ueda, H. Sugano, and T. Tanaka, "Epileptic seizure detection by using interpretable machine learning models", *Jour-*

nal of Neural Engineering, 20(1), February 2023, 015002, DOI: 10.1088/1741-2552/acb089.

[88] Y. Zhao, Z. Nasrullah, and Z. Li, "PyOD: A Python Toolbox for Scalable Outlier Detection", *Journal of Machine Learning Research*, 20(96), 2019, 1–7, http://jmlr.org/papers/v20/19-011.html.