Cognitive Computation (2024) 16:1268–1286
https://doi.org/10.1007/s12559-024-10261-9

A Mutual Information-Based Many-Objective Optimization Method for EEG Channel Selection in the Epileptic Seizure Prediction Task

Najwa Kouka[1] · Rahma Fourati[1,2] · Asma Baghdadi[1,3,4] · Patrick Siarry[4] · M. Adel[1,5]

Received: 20 February 2023 / Accepted: 25 February 2024 / Published online: 23 March 2024

Abstract

Epileptic seizure prediction using multi-channel electroencephalogram (EEG) signals is very important in clinical therapy. A large number of channels lead to high computational complexity with low model performance. To improve the performance and reduce the overfitting that arises due to the use of unrelevant channels, the present paper proposed a channel selection method to study the brain region activation related to epileptic seizure. Our method is bio-inspired and cognitive since it integrates the novel binary many-objective particle swarm optimization with a ConvLSTM model. The proposed method has two advantages. First, it performed a new initialization strategy based on channel weighting with mutual information, thereby promoting the fast convergence of the optimization algorithm. Second, it captures spatio-temporal information from raw EEG segments thanks to the ConvLSTM model. The selected sub-channels are optimized as many-objective optimization problem that includes maximizing F1-score, sensitivity, specificity, and minimizing the ratio rate of selected channels. Our results have shown a performance of up to 97.94% with only one EEG channel. Interestingly, when using all the EEG channels available, lower performance was achieved compared to the case when EEG channels were selected by our approach. This study revealed that it is possible to predict epileptic seizures using a few channels, which provides evidence for the future development of portable EEG seizure prediction devices.

Keywords Seizure prediction · Binary many-objective optimization · EEG channel selection · ConvLSTM

Introduction

An electroencephalogram, commonly referred to as an EEG, is a diagnostic test that records the electrical activity of the brain. This is done by attaching small metal electrodes to the scalp. The electrical impulses of the brain cells, which are constantly active even during sleep, appear as wave-like patterns on the EEG recording. The EEG test is widely used for the diagnosis of medical applications [1, 2] and epilepsy, in particular, [3, 4]. Epilepsy is a neurological disorder characterized by abnormal cerebral activity that can lead to seizures, strange behaviors, unusual sensations, and sometimes loss of consciousness [5].

The scientific community has approached three kinds of problems concerning epilepsy diagnosis, notably seizure detection [4, 6], seizure type classification [4, 7], and seizure prediction [3, 8, 9]. It is important to differentiate between the three problems in order to understand the context and applicability of each one. Seizure detection consists of discrimination between normal people and epileptic patients. For more precise diagnosis, seizure type classification aims at identifying the type of seizure in an epileptic patient. Therefore, this task includes only patients with seizures and not normal people. Focal non-specific, complex partial, and tonic clonic are some examples of seizures. On one side, choosing the appropriate treatment for epilepsy depends on accurately

✉ Najwa Kouka
najwa.kouka@enis.usf.tn

1 Research Groups in Intelligent Machines, National Engineering School of Sfax (ENIS), University of Sfax, BP 1173, 3038 Sfax, Tunisia

2 Université de Jendouba, Faculté des Sciences Juridiques, Economiques et de Gestion de Jendouba, 8189 Jendouba, Tunisie

3 Esprit School of Business, Computer Science and Applied Mathematics Department, Tunis, Tunisia

4 Laboratoire Images, Signaux et Systèmes Intelligents, Université Paris Est Créteil, 120 rue Paul Armangot, 94400 Vitry sur Seine, France

5 Department of Electrical and Electronic Engineering Science, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

identifying the type of seizure. Consequently, the identification of the seizure type has a more significant and immediate impact on treatment than detecting the seizure itself. On the other side, seizures should be delimited in time, but the borders of ictal (during a seizure), interictal (between seizures), and postictal (after a seizure) often are indistinct. Note that, the preictal state, characterized by an unusual pattern of neural activity, begins a few minutes prior to the start of a seizure. Predicting the onset of a seizure before it happens has the potential to be valuable. This could allow patients to protect themselves, or even prevent seizures through therapeutic interventions just before they occur.

In the field of seizure prediction, various techniques [3, 8, 10–14] have been developed to identify the onset of the preictal state. The preictal state is considered to be a positive classification, while the interictal state is deemed a negative classification. It is of utmost importance that the proposed method accurately predicts the preictal state in order to prevent seizures, while also avoiding false predictions. Thus, when evaluating a seizure prediction method, it is important to consider both its sensitivity and specificity. The ultimate goal is to distinguish the preictal brain state from the interictal state as accurately and early as possible so that it can be applied in real time. However, the online prediction of epileptic seizures remains a challenge, and further refinement is needed to achieve high sensitivity and low false prediction rates. The false-positive rate per hour (FPR/h), defined as the number of false alarms per hour, is a crucial metric for determining the accuracy of seizure prediction. This research focuses on identifying the channels that display clinical and physiological changes prior to a seizure, in order to understand the underlying mechanisms and possibly anticipate seizures.

Specifically, there is a high level of interest from both industry and the scientific community in creating portable medical support systems that are equipped with algorithms to detect the early stages of epileptic seizures or even predict them hours before they occur. This will assist ambulatory patients or their caregivers in avoiding potential harm caused by seizures. In designing these portable systems, it is crucial to implement computationally efficient prediction algorithms that use the least amount of channels possible in order to minimize power consumption and ensure extended battery life.

Indeed, the goal of EEG channel selection is to reduce the number of selected channels while maximizing the prediction accuracy, which can be regarded as a many-objective optimization problem. Thus, several kinds of meta-heuristics algorithms, inspired by various behaviors in nature, have been proposed for such type of problem [15–17]. These algorithms are referred to as wrapper approaches because they combine (1) the maximization of classification or prediction performance and (2) minimizing the channel subset size in the fitness function. In particular, there is an increasing number of studies that have applied particle swarm optimization (PSO) for the channel selection problem [18, 19].

The PSO algorithm is a metaheuristic stochastic algorithm that simulates the behaviors of birds flocking or fish schooling, proposed by Kennedy and Eberhart in 1995 [20]. The original version was proposed for solving continuous problems, which hinders its use for binary solution space. Thus, Kennedy and Eberhart [21] proposed the binary PSO (BPSO) in 1997 for solving binary problems such as feature selection problems. One limitation of the BPSO algorithm is its premature convergence. Moreover, the high dimensional data of EEG channels leads to the high dimensionality of search space with numerous feasible solutions (channel subsets). This paper aims to address this challenge by enhancing the binary particle swarm optimization (BPSO) algorithm to better suit the task of EEG channel selection.

The key findings of this paper can be summarized as follows:

- The EEG channel selection problem for seizure prediction includes (1) the minimizing of channel subset size and maximizing (2) F1-score, (3) the sensitivity, and (4) the specificity, respectively.
- A convolutional long short-term memory (ConvLSTM) model that captures spatio-temporal information within EEG signals is proposed for the prediction of epileptic seizures.
- Population initialization is based on mutual information (MI) to include the search space by critical channels related to an epileptic seizure. Inter-channel connection matrix is calculated based on MI, and channel reduction is conducted by using thresholding and connection matrix analysis.

For readability purpose, Table 1 presents the different abbreviations used in this paper and the corresponding full words.

Related Work

The prediction of epileptic seizures in advance will reduce the consequences for patients. Unfortunately, despite several attempts devoted to predicting seizures, reducing the channel number in this task remains a challenging problem.

On one side, most existing methods are patient-dependent where the train and test sets are extracted from the same patient data and have the same distribution over time. In other words, each patient has his own trained model. Such models fail against new patients having no similar characteristics and different seizure manifestations. On the other side, a patient-independent classification scheme trains a single model taking data from different patients. The advantage is

Table 1 List of notations and their descriptions

Notation	Description
A	Archive
BPSO	Binary particle swarm optimization
BMaOPSO-CA	Binary many-objective particle swarm optimization with cooperative agents
CNNs	Convolutional neural networks
ConvLSTM	Convolutional long short-term memory
EA	External archive
EEG	Electroencephalogram
F1	F1-score
FPR/h	False-positive rate per hour
FC	Fully connected
LA	Local archive
LDA	Linear discriminant analysis
LOPO	Leave-one-patient-out
MI	Mutual information
MI-BMaOPSO	Mutual information-based binary many-objective particle swarm optimization
MI-BMaOPSO-CA	Mutual information-based binary many-objective particle swarm optimization with cooperative agents
MOPs	Multi-objective optimization problems
MOPSO	Multi-objective particle swarm optimization
MAS	Multi-agent system
NMI	Normalized mutual information
NS	Number of sub-swarm
NG	Number of generations
pBest	Personal best position
gBest	Global best position in sub-swarm
GBest	Global best position in whole swarm
PP	Preictal period
PSO	Particle swarm optimization
SEN	Sensitivity
SF	Sequential forward selection
SPEC	Specificity
SVM	Support vector machine
TL	Transfer learning

that the model can generalize better due to the training using heterogeneous EEG data. The model learns different characteristics; thus, it performs well in detecting the preictal segments for a new patient.

This work deals with EEG channel selection for seizure prediction. This section overviews mainly the related works for channel selection. Since these methods are patient-specific and our method is patient-independent, we also overview related works on seizure prediction using all channels and trained in a patient-independent classification scheme.

Patient-Independent Seizure Prediction with All Channels

Previous research has shown promising results using convolutional neural networks (CNNs) for epileptic seizure prediction. For example, Hu et al. [22] used a CNN model and achieved an accuracy of 86.25% on the CHB-MIT dataset. The model demonstrated an improvement of nearly 12% in accuracy for identifying preictal samples closer to a seizure, compared to those further in time. Another study by Khan et al. [23] used a subset of 16 subjects from the CHB-MIT

dataset and achieved a ROC-AUC of 0.8660 using a 10-min preictal window.

The effectiveness of CNNs is further evidenced in the works by [24, 25], achieving accuracies of 80.00% and 92.02%, respectively. The convolutional operation can vary based on the kernel shape used, distinguishing between a 2D kernel for spatial convolution and a 1D kernel for temporal convolution. While [24, 25] proposed CNN models with spatial convolution, Mao et al. [26] introduced a CNN model with temporal convolution, achieving a sensitivity of 58.80% and a specificity of 63.70%.

Dissanayake et al. [27] carried out a study on a multi-task siamese network for patient-independent epileptic seizure prediction, which showed a 91.51% accuracy for identifying seizures with a 1-h early prediction window, using 24 subjects from the CHB-MIT dataset [27]. Additionally, they proposed two methods of geometric deep learning for epileptic seizure prediction in their recent work [28]. The first method used a straightforward approach to graph computation based on the physical structure of the EEG grid, while the second method utilized deep neural networks to create graphs. Both techniques yielded state-of-the-art results on the CHB-MIT dataset, with accuracy rates reaching up to 95%.

While 10-fold cross-validation is considered in the works [27, 28], a recent work [13] achieved 91.96% using leave-one-patient-out (LOPO). The latter is more challenging since the model is tested on unseen patient and not only unseen data. In other words, the distribution of test data is very different from the training data distribution which makes the discrimination more difficult.

Channel Selection of Patient-Specific Seizure Prediction

For a selection method, it is important to know the horizon time or the preictal period, the number of selected channels, and the trial duration. Table 2 illustrates the recent channel selection methods with all details. The authors of [8] present a new, patient-specific method for selecting the most crucial EEG channels without the need for prior expert knowledge. The method is based on a cascading deep learning model that includes convolution blocks and attention layers. The data collected from the chosen EEG channels is processed by a deep learning model composed of a convolutional neural network and a gated recurrent neural network to predict epileptic seizures. This work supports the use of EEG sensors with a smaller number of electrodes that can be comfortably worn by patients for extended periods of time. The proposed method is tested on the CHB-MIT benchmark and shows superior performance compared to existing methods (e.g., random forest) by requiring fewer EEG channels, making it

Table 2 Existing works on EEG channel selection for seizure prediction

Year	Ref.	Selection method	#Ch	Preictal period (min)	Trial duration (s)	Classifier	Performance
2021	[11]	Channel elimination	6	10	8.0	CNN	Accuracy = 0.99 Sensitivity = 0.97 Specificity = 0.92 FPR/h = 0.0764
2021	[17]	SFS	2–7	23	3.9	MAML	Sensitivity = 0.92 FPR/h = 0.26
2021	[15]	GA	3–5	40	5.0	LR	Sensitivity = 0.38±0.19 FPR/h = 1.03±0.48
2021	[16]	KNN-GA	2–9	10	2.8	SVM	Accuracy = 0.74 Sensitivity = 0.69 Specificity = 0.73
2022	[12]	NSGA-II	4–7	45	5.0	LR	Sensitivity = 0.16±0.11 FPR/h = 0.21±0.08
2022	[8]	Attention score	2–6	30	30.0	CGRNN	F1-score = 0.76 Precision = 0.75 Recall = 0.78 FPR/h = 0.35
2023	[9]	NSGA-II	3	10	8.0	CNN	Accuracy = 0.96 Sensitivity = 0.96 Specificity = 0.96

possible to use only two or three EEG channels for accurate patient-specific seizure prediction. The results also suggest that the channel selection approach reduces the prediction time.

Jana and Mukherjee [11] proposed a new EEG channel selection algorithm. The algorithm assesses the importance of each channel individually and eliminates the least significant ones incrementally. The average classification accuracy was first calculated using all 22 channels. Then, one channel was removed at a time, and the average classification accuracy was re-calculated with the remaining 21 channels. If the removal of a channel did not negatively impact accuracy, it was deemed as less significant. This process was repeated until only the 6 most crucial channels remained.

In the study in [15], the authors designed a personalized search algorithm to predict seizures and identify the optimal preictal time. The algorithm uses evolutionary computation, where each individual in the population represents a set of five features. Essentially, the sets of features that perform the best in seizure prediction using a logistic regression classifier (fitness function) will survive and be selected for further propagation, similar to natural selection. This approach not only leverages the predictive power of the features and their combination but also aims to provide a deeper understanding of the seizure generation process by considering a sequence of time points instead of a single one, and by producing interpretable results.

In follow-up work, Pinto et al. [12] addressed the challenges in seizure prediction and developed an interpretable evolutionary model using NSGA-II that considers patient comfort while identifying the optimal set of features and searching for the preictal period [12]. This patient-specific approach offers interpretable insights, contributing to a better understanding of the seizure generation process and explaining the algorithm's decisions. The methodology was tested on 238 seizures and 3687 h of continuous data collected from 93 patients with various types of focal and generalized epilepsies. The data was recorded using scalp recordings.

Similarly, Jana and Mukherjee [9] proved the effectiveness of NSGA-II for the selection of only 3 channels from CHB-MIT dataset. The method is patient-dependent where a CNN model with 1D temporal convolution is trained for each patient.

In the research [16], the authors proposed a patient-specific optimization approach for EEG channel selection based on permutation entropy values and K nearest neighbors combined with a genetic algorithm. The method is applied to epileptic seizure prediction using the well-known support vector machine as the classifier and the CHB-MIT Scalp EEG dataset as the data source. The results from the 22 patients showed a significantly higher prediction rate (average 92.42%) compared to using all channels with the SVM classifier (71.13%). On average, the accuracy, sensitivity, and specificity improved by 10.58%, 23.57%, and 5.56%, respectively, with the selected channels. Four patient cases even achieved over 90% accuracy, sensitivity, and specificity rates with only a few selected channels. The standard deviations were also smaller compared to those obtained with all channels, demonstrating the robustness of the patient-specific channel selection approach in optimizing seizure prediction.

In the work by Romney and Manian [17], a feature selection method based on the sequential forward selection (SFS) algorithm was implemented. The SFS is a wrapper-based feature selection method, meaning it utilizes a machine-learning algorithm to identify the most relevant attributes. In this work, the SFS was combined with a linear discriminant analysis (LDA) for feature selection.

In summary, the current literature on epileptic seizure prediction has several limitations. There is a lack of patient-independent studies that take into account the high inter-subject variability in EEG data. Despite the difficulty in designing patient-independent models, it is important to pursue such research as it offers a more practical solution to the problem. The aim is to find the optimal EEG channel set by evaluating the prediction accuracy and prediction horizon of various methods. This research is a continuous process that seeks to identify the most important channels through the analysis of raw EEG data using spatio-temporal representation learning techniques.

Theoretical Background of Particle Swarm Optimization

The PSO is a type of evolutionary search technique that is inspired by the social behavior of animals like birds in flocking. In PSO, each particle represents a potential solution to a problem, and all particles navigate through the N-dimensional search space to find the optimal solution by updating their position vector $\vec{X}_i = (\vec{X}_{i1}; \vec{X}_{i2}, ..., \vec{X}_{iD})$ and velocity $\vec{V}_i = (\vec{V}_{i1}; \vec{V}_{i2}, ..., \vec{V}_{id})$ as per the Eqs. (1) and (2), respectively.

$$\begin{aligned}\vec{V}_{id}(t+1) = w\vec{v}_{id}(t) + c1r1(pBest_{id}(t) - \vec{x}_i(t)) \\ + c2r2(gBest_{id}(t) - \vec{x}_i(t))\end{aligned} \tag{1}$$

$$\vec{X}_{id}(t+1) = \vec{X}_{id}(t) + \vec{V}_{id}(t+1) \tag{2}$$

where t is the iteration number. The inertial weight w serves as a tool to regulate the speed and equilibrium of the algorithm's exploration and exploitation capabilities. If w is high, it maintains the particles at high velocity and discourages them from getting stuck in local optima. On the other hand, if w is low, the particles move at a slower pace and focus on exploiting the same area of search. The acceleration coefficients $c1$ and $c2$ determine the particle's tendency to move

towards either the best position found so far $pBest$ or the best global position found across the swarm $gBest$.

Multi-objective optimization problems (MOPs) are those which have more than one objective function that needs to be optimized. These objectives may sometimes be conflicting in nature. To handle such problems, multi-objective optimization algorithms are employed. These algorithms generate a set of optimal solutions that balance the conflicting objectives and these solutions are stored in a separate repository known as an external archive. This external archive is used in PSO in order to be suitable for optimizing $MOPs$ (multi-objective PSO, MOPSO). In the MOPSO algorithm, the external archive is updated in each iteration. When running the algorithm is terminated, the external archive provides the optimal solutions. The MOPSO flowchart is illustrated in Fig. 1.

MI Binary MOPSO for EEG Channel Selection

This section proposes the MI-BMaOPSO-CA algorithm for channel selection. First, basic elements for channel selection including solution encoding and fitness function are introduced. Then, the population-initialization phase which initializes the swarm using channel weighting results of normalized mutual information (NMI) is presented. Third, the cooperative multi-swarm is presented. Finally, the overall MI-BMaOPSO-CA approach is given.

Figure 2 illustrates the flowchart detailing our methodology. The process begins with dataset preparation, encompassing interictal and preictal trials, followed by partitioning into train and test sets. Subsequently, our algorithm is executed. The archive is then explored to identify the optimal solution, which comprises the selection of EEG channels deemed most effective for seizure prediction. Finally, the methodology concludes with an analysis of the contribution of brain regions.

Problem Formulation

As previously stated, the channel selection problem involves selecting a limited number of relevant channels that can achieve comparable or even improved performance compared to using all channels. In this paper, four main conflicting objectives have been considered, which are represented by the F1-score ($f1$), sensitivity (SEN), and specificity ($SPEC$), and are defined as follows:

$$SEN = \frac{TP}{TP + FN} \tag{3}$$

$$SPEC = \frac{TN}{TN + FP} \tag{4}$$

$$F1 - score = \frac{TP}{TP + (\frac{FN+FP}{2})} \tag{5}$$

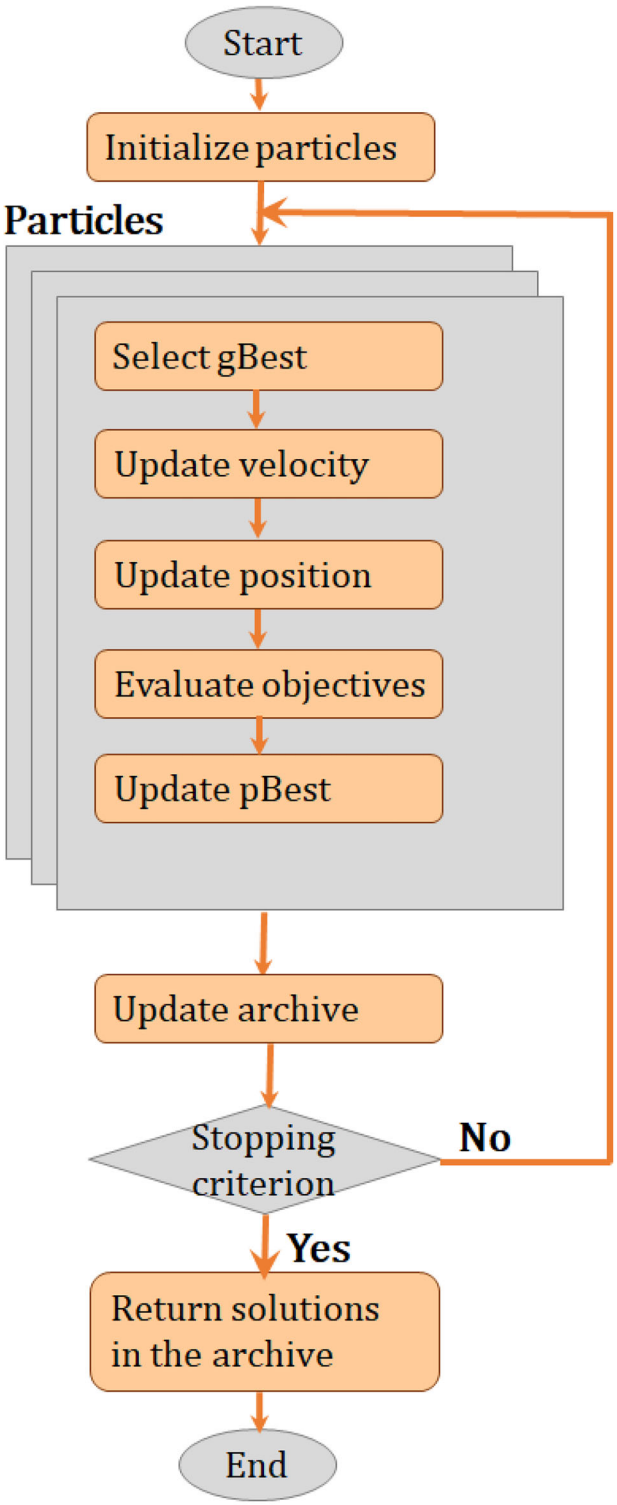


Fig. 1 The flowchart of MOPSO algorithm

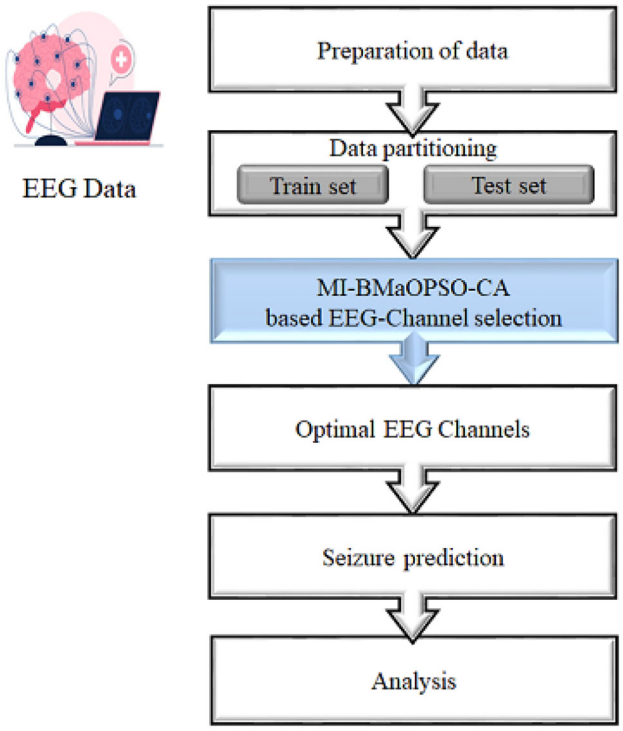


Fig. 2 Flowchart of the proposed methodology for EEG seizure prediction

To balance these performance metrics and the selected channels, a fitness function based on the linear weighting method is used. The considered fitness function is described as follows:

$$fitness(p_i) = \sum_{j=0}^{M} w_j f_j(X_i) \tag{6}$$

where particle's position p_i is defined by a dimensional vector X_i, j denotes the objective function, M denotes the number of objectives, f_j denotes the objective function (SEN, $SPEC$, and $F1$), and w_j represents the weight of objective function j. We equally set the weight ($w = 0.3$) for SEN, SPEC, and F1, respectively, and $w = 0.1$ for the ratio of selected channels. The weight of performance metrics was higher because they are more critical than the selected channels.

Particle Encoding

In the channel selection, a solution is encoded as a binary vector $X = [x_1 x_2 ... x_D]$ with dimension D dimensions. Each bit $X_i \in 0, 1$, where 1 denotes the ith channel is selected, and 0 denotes it is not selected. To map the solution into discrete space, the widely used Sigmoid function is adopted. The details of this function are presented as follows:

$$S(\vec{V}_{id}(t+1)) = \frac{1}{1+\exp(\vec{V}_{id}(t+1))} \tag{7}$$

$$\vec{X}_{id}(t) = \begin{cases} 1, & if \quad rand \le \vec{V}_{id}(t) \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where $\vec{X}$ is the binary solution of the channel selection problem, $\vec{V}_{id}$ denoted the velocity of solution i for dimension d, and $rand$ is a random number uniformly distributed between 0 and 1 and it is used as the threshold to decide whether the corresponding channel is selected. Correspondingly, Fig. 3 shows a binary solution for channel selection, in which if the position X_i equals 1 then this channel will be used in the prediction process.

Feature Learning Based on ConvLSTM Model

The ConvLSTM model is a captivating deep learning approach utilized to forecast subsequent video or image frames. It was first introduced in a research project focused on precipitation nowcasting [29]. Instead of the standard matrix multiplication, ConvLSTM incorporates convolution operations in each gate within the LSTM cell, thus enabling it to detect local spatio-temporal correlations.

The information processing by the ConvLSTM2D cell is shown in Eqs. (9) to (13), where $*$ denotes the convolution operator and "$\circ$" denotes the Hadamard product.

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \tag{9}$$

$$\begin{aligned} f_t = \sigma(&W_{xf} * X_t + W_{hf} * H_{t-1} \\ &+ W_{cf} \circ C_{t-1} + b_f) \end{aligned} \tag{10}$$

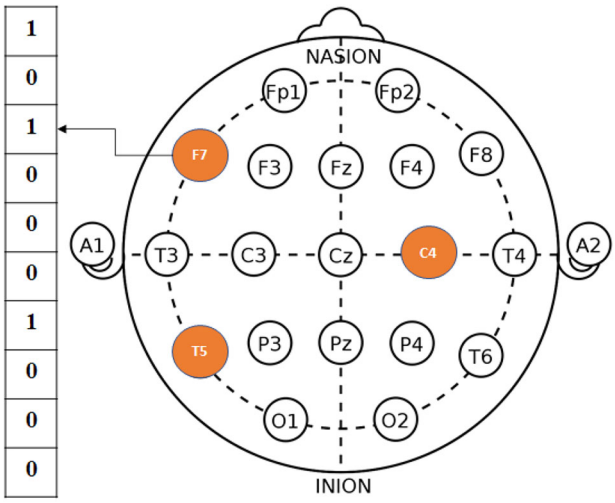


Fig. 3 Example of binary solution for the channel selection problem

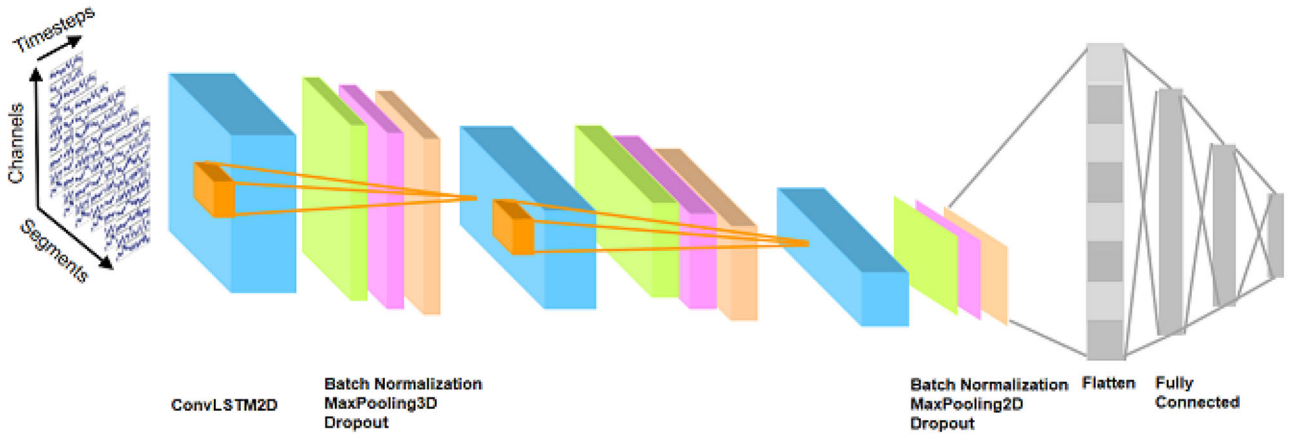


Fig. 4 The architecture of ConvLSTM model for epileptic seizure prediction

$$C_t = f_t \circ C_{t-1} + i_t \circ tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \tag{11}$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \tag{12}$$

$$H_t = o_t \circ tanh(C_t) \tag{13}$$

where $X_1, ..., X_t$ are the inputs, $C_1, ..., C_t$ are cell outputs, $H_1, ..., H_t$ are hidden states, and i_t, f_t, and o_t are gates of the ConvLSTM2D cell. Note that, a ConvLSTM2D layer accepts a tensor with 5D shape (batch_size, nb_segments, n_timesteps, n_channels, 1).

The model is composed of 3 ConvLSTM2D layers as depicted in Fig. 4. The feature maps are flattened and embedded into 128 elements. This block is responsible for the feature learning step. The last layer with softmax activation is responsible for the classification step.

Population Initialization Based on Normalized Mutual Information

The swarm initialization is a critical step in the search process to find rapidly the optimal solutions. Basically, in the PSO/BPSO variants, the swarm is randomly initialized with a random position in the search space. To be specific for BPSO, the initialization of the particle's position to 0 or 1 depends on predefined probability. Random initialization does not take into account the inherent features of data, which can offer valuable information to initialize BPSO with a higher-quality swarm.

In the present work, the initial swarm is initialized with critical channels that have higher information quality. For this purpose, mutual information is adopted here to measure the quality level of each channel. Without loss of generality, the mutual information measures how dependent two variables are. If two variables are independent, the MI is zero. Conversely, higher values of MI indicate a stronger dependence between the variables. The MI is a unique dependency measure due to its two key properties: (1) the ability to measure any type of relationship between variables and (2) its invariance under changes in the space of the variables. Mathematically, the MI is defined as:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \tag{14}$$

where $H(X)$ and $H(Y)$ are the marginal entropies, and $H(X, Y)$ is the joint entropy of the two variables X and Y.

For each sample s in the dataset, a MI_s connection matrix is created as follows:

$$MI_s = \begin{pmatrix} MI(X_0, Y_0) & \dots\dots & MI(X_{NB}, Y_0) \\ MI(X_0, Y_1) & \dots\dots & MI(X_{NB}, Y_0) \\ & \dots\dots & \dots \\ MI(X_0, Y_{NB}) & \dots\dots & MI(X_{NB}, Y_{NB}) \end{pmatrix} \tag{15}$$

where NB denotes the channel number. Once the MI matrix is generated, the NMI is calculated as follows:

$$NMI = \left\{ \sum_{i,j}^{NB} (MI_s(i, j)), s \in samples \right\} \tag{16}$$

To this end, the channel with a larger weight than a threshold has a higher probability to be set as 1.

MI-BMaOPSO Algorithm

The pseudocode of MI-BMaOPSO is demonstrated in Algorithm 1. The algorithm is further explained by the following steps.

- **Step 1:** Initially, the data is divided into train and test data.

- **Step 2:** The population of particles is initialized based on MI. Based on this initialization, each particle is evaluated.
- **Step 3:** Based on the population initialization, the key parameters are initialized such as velocity, p_{best}, and g_{best}.
- **Step 4:** Once the initialization is finished, the external archive is updated with the non-dominated solutions.
- **Step 5:** In each iteration, for each particle the velocity and position are updated using Eqs. (1) and (2), respectively.
- **Step 6:** Mutate each particle.
- **Step 7:** In the next step, the objectives functions are evaluated and the fitness function is updated.
- **Step 8:** At the end of each iteration, the best personnel experience for each particle is updated, and the archive is also updated with the new non-dominated solutions found so far.
- **Step 9:** If the stopping criteria are reached, then the archive is returned as the best optimal channel. Otherwise, the optimization process returned to Step 5 for a loop.

Algorithm 1 MI-BMaOPSO algorithm

Input: Population P with size N
Output: Archive A

for $i = 1$ to N **do**
  Divide dataset into train and test data
  Initialize $\vec{V}_i$
  Initialize $\vec{X}_i$ based on NMI
  Evaluate fitness for P_i
  Initialize $pBest_i = P_i$
end for
Save non-dominated solutions of P in the archive
$g = 0$
while $g \leq g_{max}$ **do**
  for $i = 1$ to N **do**
    Select leader
    Update velocity
    Update position
    Mutation
    Evaluate objectives with ConvLSTM Model
    Evaluate fitness function
    Update $pBest$
    Update archive A
  end for
  $g = g + 1$
end while

Cooperative Multi-swarm

The multi-swarm topology can increase swarm exploration and avoid premature convergence. During the search process, each sub-swarm is evaluated independently of the others. Even if one sub-swarm is trapped in the local optimum, other sub-swarms can still search for the global optimal solutions. Although the multi-swarm aims to increase diversity, it is slow in convergence speed exploration ability. In order to balance exploitation and exploration, the sharing of information and the interaction between sub-swarms must be provided. With the interaction, the sub-swarm that converges to local optima can rapidly jump out by learning from the good information learned by other sub-swarms.

Various cooperative interaction methods have been suggested to improve search efficiency through the exchange of useful information, such as elite learning. This approach involves the top particle of each swarm learning from other sub-swarms, and the valuable information found by one sub-swarm being shared with others. How to select elite particles to share good information and how to increase the exploitation by learning from the elite particles' abilities are two issues that need to be solved.

- **Elite particles selection**: the selection of elites is based on its improvement rate. The particle that has the best fitness value with minimum distance (minimum selected channels number). The improvement rate is mathematically calculated as follows:

$$I(X_i) = \frac{f(x_i^t) - f(x_i^{t-1})}{\left\| x_i^t - x_i^{t-1} \right\|} \tag{17}$$

where $\left\| x_i^t - x_i^{t-1} \right\|$ denotes the distance between particle at t and $t-1$. Since the position of the particle is defined in binary terms, the distance is measured by Hamming distance which represents the number of different bits between two binary positions as defined in 18. However, the greater I means that a current particle achieves a higher improvement with a relatively smaller fly distance than the previous one.

$$d(X_t, X_{t-1}) = \sum_{k=1}^{D} \begin{cases} 1, & if \ X_t(k) \neq X_{t-1}(k) \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

- **Learning strategy**: By using, the good information of elite particles of each sub-swarm, the particles can search for better positions. To find out all the knowledge discovered by elite particles, the crossover and mutation operators are applied. Both of these genetic operators have the ability to strengthen local exploitation and global exploration, respectively. To be specific, the crossover operator enables exchanging information among solutions and the mutation operator can increase the population diversity.
  In the learning strategy, two parents are first selected from the elite particles. Then, the crossover and mutation are applied to generate new solutions. These solutions inherit much more useful information from different elite particles. In every interaction between sub-swarms, the best

solution of generated solutions $GBest$ replaced the worst solution in each sub-swarm. Furthermore, this solution $GBest$ presents a new search direction in the search space that is applied in the velocity equation to increase the diversity. In this way, each particle will perform its flight based on its best personal experience $pBest$, the experience of its neighbors $gBest$, and the experience of the best global solution $GBest$. The velocity equation is defined as follows:

$$\vec{V}_{id}(t+1) = w\vec{v}_{id}(t) + c1r1(\vec{x}_{pBest_{id}}(t) - \vec{x}_{id}(t)) + , \\ c2r2(\vec{x}_{gBest_{id}}(t) - \vec{x}_{id}(t) + c3r3(\vec{x}_{GBest_{id}}(t) - \vec{x}_{id}(t)) \quad (19)$$

where t is the iteration number; W is the inertia weight; $c1$, $c2$, and $c3$ are learning factors; and $r1$, $r2$, and $r3$ are the random values uniformly distributed in [0, 1].

Algorithm Overview

The proposed approach as shown in Fig. 5 first uses the NMI method to initialize the swarm of N particles, and the velocity of each particle is initialized. Once the initialization is performed, the external archive is initialized by the founded optimal solution set. Next, the swarm is evaluated by the MI-BMaOPSO algorithm which is detailed in the "MI-BMaOPSO Algorithm" subsection, until the number of evaluations reaches half of the maximum number of evaluations. Afterward, the swarm is sorted based on its fitness values, then it is equally divided into NS sub-populations, where NS is the number of sub-swarm. For example, the entire swarm has 20 particles. It is divided into 4 sub-swarms with 5 particles in each one. The first sub-swarm contains the No. 1 to No. 5 particles. The second sub-swarm consists of the No. 6 to No. 10 particles and so forth. These sub-swarms are modeled as multi-agent system (MAS) and they search for better solutions corporately. Each agent performs the optimization of its own sub-swarm by the MI-BMaOPSO algorithm until attending the maximum number of evaluations. In order to increase the diversity of solutions, sharing knowledge is performed after a certain number of generations (NG). This step is carried out by the learning strategy as demonstrated in the "Cooperative Multi-Swarm" subsection. Once the end condition is reached, the external archive is updated by the different outputs of each sub-swarm. Finally, the external archive is reported as the best optimal channel subset.

The pseudocode of MI-BMaOPSO-CA is demonstrated in Algorithm 2. The different steps explained above are presented in the algorithm to show the global behavior. The stopping criterion is the number of itertaions.

Algorithm 2 MI-BMaOPSO-CA Algorithm

Require: Population P with size N
Ensure: External Archive EA
for $i = 1$ to N **do**
  Initialize particles positions x_i with NMI method
  Evaluate f_j, $j = 1, ..., k$
  Save non-dominated solutions of P in the archive EA
end for
K sub-swarm = Ranking (P)
for $i = 1$ to K **do**
  while $g \leq G_{max}$ **do**
    Select leader
    Update velocity
    Update position
    Mutation
    Evaluation with ConvLSTM Model
    Update Fitness function
    if $I(X_i^t) \geq I(pBesti^{t-1})$ **then**
      $pBesti = p_i$
    end if
    $g = g + 1$
    Save solutions in LA_i
    Learning strategy
  end while
  EA = EA $\cup$ LA_i
end for

Experiment Design

Data Preparation

The CHB-MIT scalp long-term EEG dataset [30] consists of recordings from 23 pediatric patients with intractable seizures at Boston Children's Hospital. The data was annotated by experts to identify the start and end of seizures and consists of 983 h of EEG recordings. However, the seizures are relatively short compared to the total EEG duration, resulting in an imbalanced data distribution that makes classification difficult. The recordings have a sampling frequency of 256 Hz with a 16-bit resolution, and were taken using the international 10-20 system for electrode positioning. To ensure consistency, 18 identical channels were selected from the recordings, including FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, and CZ-PZ.

All files have been annotated for ictal states and provided information about the start and end of the seizure state. The preictal state can be assumed as the state before the start of the ictal state. For the purpose of seizure prediction, we extracted preictal and interictal segments from the 24 cases included in the dataset. Since every case can have one or multiple seizures, we extracted all existent seizures for each case. The number of trials in the preictal class depends on the selected preictal period. Regarding our purpose of channel optimization and the search for the optimal preictal period, the number of trials in each class differs within the solution search. Basically, we prepared the data into sub-datasets. For

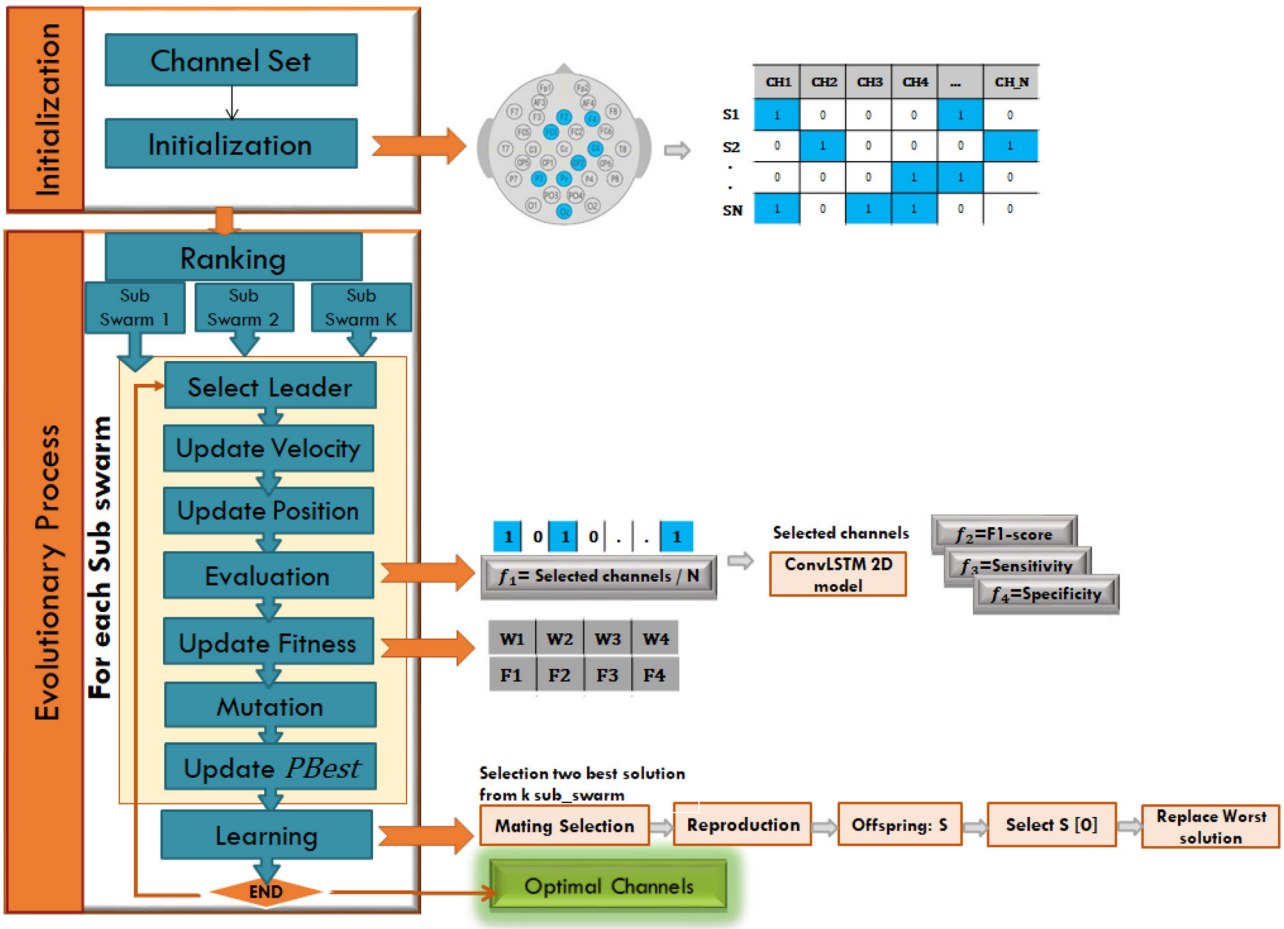


Fig. 5 Mutual information-based binary many-objective PSO with cooperative agents

a PP = 5 min, we have extracted 5 min before the onset as preictal trials and all the data after the onset's end as interictal trials. same for the 10 and 15 min PP. Then, we segmented the trials into 5-s segments, the final dimension of one segment is $(18x1280)$.

ConvLSTM Hyper-parameters

The ConvLSTM model depicted in Table 3 takes as input a 5D tensor with the shape $(batch\_size \times nb\_segments \times timesteps \times eeg\_channels \times color)$. The EEG data is considered as a video. Therefore, in our case, the last dimension color is equal to 1, since it is not significant. In our case, the initial trial of shape (18×1280) is changed to $(10 \times 128 \times 18 \times 1)$. The model is composed of 3 ConvLSTM2D layers. The first two layers are created with the attribute $return\_sequences = True$ to keep the 5D tensor shape. Actually, it allows to collect of activations of each ConvLSTM cell in the layer. The third ConvLSTM2D layer with the attribute $return\_sequences = False$ to reduce the 5^{th} dimension and obtain a tensor with 4D shape. The three ConvLSTM2D layers have 16, 32, and 64 filters, respectively, that are 3×3 pixels in size. The max-pooling layer is performed after each convolution for data downsampling with $pool\_size = 1 \times 2 \times 2$, $pool\_size = 1 \times 2 \times 2$, $pool\_size = 2 \times 2$, respectively. Then, a dropout layer with a probability of 0.2 is added to prevent overfitting. The latter is flattened to obtain a 1D tensor and then inputted to a fully connected (FC) layer with 1024 neurons, and then to another FC layer with 128 neurons. The last FC layer is activated using *softmax* since CHB-MIT has 2 labels (interictal and preictal). To be specific, the network is trained from scratch using the Adam optimizer, with a learning rate of 0.0001. The categorical cross entropy has been chosen as the loss function.

Selection of Preictal Period

Table 4 depicts the results obtained by varying the preictal duration used to create the dataset. The data are partitioned into 60% for training, 10% for validation, and 30% for the test set. The proposed ConvLSTM network demonstrates

Table 3 Architecture of the ConvLSTM full model

Layer	Filter	Size	Stride	Parameter
Input			10*x*128*x*18*x*1	
ConvLSTM2D	16	3*x*3	1	return_sequences = True
BatchNormalization	—	—	—	—
MaxPooling3D	—	1*x*2*x*2	—	—
Dropout	—	—	—	rate=0.2
ConvLSTM2D	32	3*x*3	1	return_sequences = True
BatchNormalization	—	—	—	—
MaxPooling3D	—	1*x*2*x*2	—	—
Dropout	—	—	—	rate=0.2
ConvLSTM2D	64	3*x*3	1	return_sequences = False
BatchNormalization	—	—	—	—
MaxPooling2D	—	2*x*2	—	—
Dropout	—	—	—	rate=0.2
Flatten	—	—	—	—
FC	1024	—	—	—
FC	128	—	—	—
FC	2	—	—	activation = "softmax"

performance when evaluated on shorter preictal durations, closer to the seizure onset. Consequently, in all subsequent experiments, we considered a preictal period of 5 min. These findings align with results achieved in the works [31, 32].

Results and Discussion

NMI Contribution in MI-BMaOPSO-CA

In this study, we computed the NMI connection matrix across different channels for predicting epileptic seizures, as depicted in Fig. 6. The NMI value ranges from 0 to 1, where 0 indicates a complete lack of correlation between two channels, while 1 signifies complete correlation.

Moreover, we have set the threshold value at 0.5, designating channels with an NMI value greater than 0.5 as critical channels ($CH_{critical}$). These critical channels are identified during the initialization phase, and the initial swarm starts with random initialization around these $CH_{critical}$ points.

Table 4 Performance variation (%) with different preictal period (min)

Preictal period	F1-score	Sensitivity	Specificity
5	**91.83**	**91.96**	**91.06**
10	80.51	78.25	77.94
15	74.68	73.77	73.91

Values in bold indicate best results

To get a better understanding of how the incorporation of NMI-based initialization improves the performance of our algorithm, we experiment with two variants: MI-BMaOPSO-CA and BMaOPSO-CA. The comparison results of these two variants are illustrated in Fig. 7 in terms of $F1$, Sen, and Spe performance metrics, respectively. It is observed that the incorporation of NMI in the initialization phase leads to better results in terms of $F1$, Sen, and Spe, respectively.

In addition, the first initial swarm for MI-BMaOPSO-CA and BMaOPSO-CA is presented in Fig. 8 to show the effectiveness of NMI to start the optimization by a high solu-

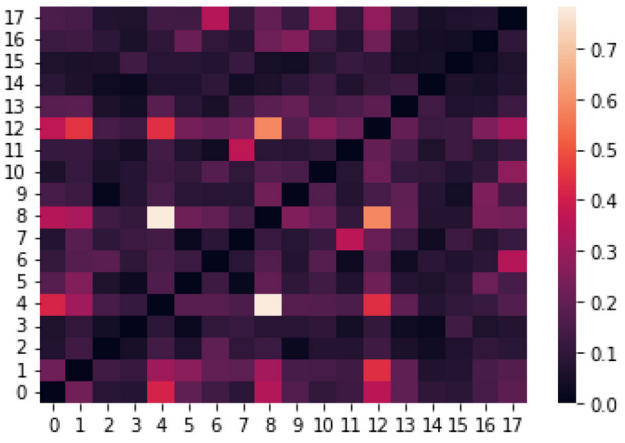


Fig. 6 NMI-based channel matrix correlation

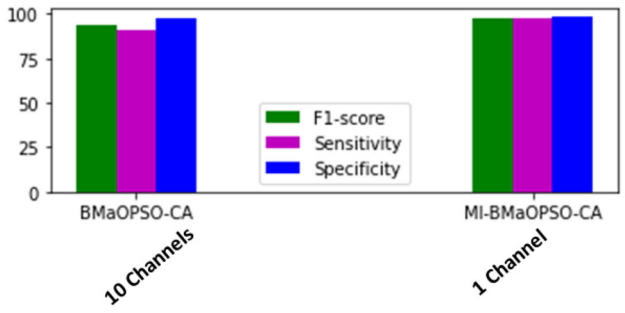


Fig. 7 Validation results for the contribution of MI in terms of F1-score (%), sensitivity (%), and specificity (%)

tion's quality. Based on these results, it can be concluded that the NMI-based initialization has an advantage in improving exploitation ability.

Optimization Results of Channel Selection

This section is elaborated to compare our approach with two optimization algorithms: NSGA-III and BMOPSO. The performance of peer competitor algorithms is calculated based on $F1$, SEN, and $SPEC$. Table 5 shows the comparison results between algorithms, where the best value is highlighted in boldface.

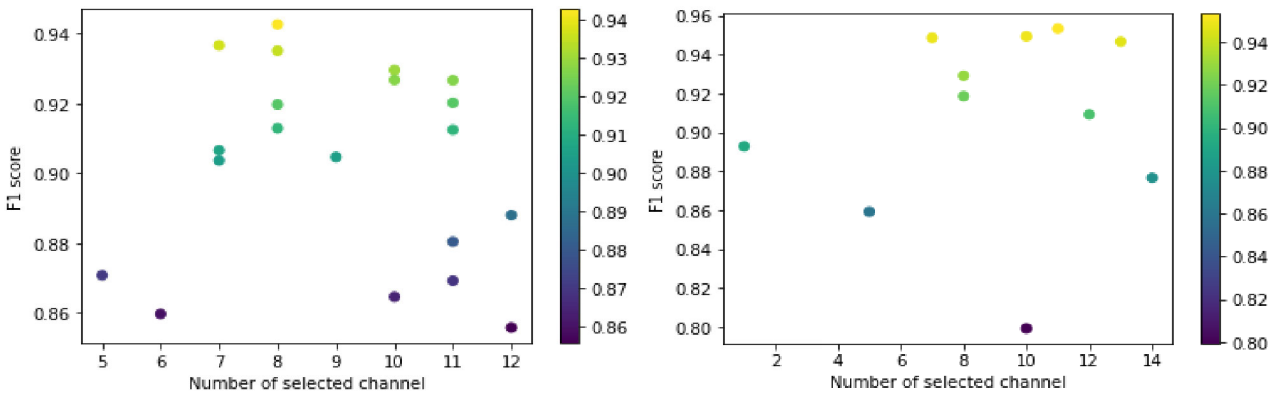
Figure 9 illustrates the confusion matrix of the best solution for each considered algorithm as well as the baseline model. Accordingly, when using all channels (18 channels), the proposed ConvLSTM model obtains $F1 = 91.38\%$, $SEN = 91.96\%$, and $SPEC = 91.06\%$. On the other hand, with the optimization algorithms, the BMOPSO obtains $F1 = 94.63\%$, $SEN = 97.21\%$, and $SPEC = 92.25\%$ with 14 channels, while the NSGA-III obtains $F1 = 95.72\%$, $SEN = 97.54\%$, and $SPEC = 94.21\%$ with 14 with 8 channels. The MI-BMaOPSO-CA achives $F1 = 97.94\%$, $SEN = 97.44\%$, and $SPEC = 98.30\%$ with only one channel. It is observed that the number of EEG channels can be reduced from 18 to 14, 8, and 1 channels by using BMOPSO, NSGA-III, and BMaOPSO-CA, respectively.

It is not surprising that NSGA-III is better than BMOPSO since it is well known for its fast convergence. The miss-classification rate of the ConvLSTM model with 8 channels selected by NSGA-III is 1.37% for the interictal state and 3.54% for the preictal state. For BMOPSO, the miss-classification rate of 14 channels ConvLSTM model reached 2.77% for the interictal state and 6.12% for the preictal state. Our proposed optimization method converged to C3-P3 channel which reduced the miss-classification rate to 0.79% and 1.07% for interictal and preictal states, respectively.

The experimental results on the CHB-MIT dataset showed that it is possible to reduce the number of selected channels from 18 to 1 without a significant decrease in the classification performance while maintaining a good balance between sensitivity and specificity.

The location of EEG channels for epileptic seizure after channel selection using optimization algorithms is illustrated in Fig. 10. It reveals that epileptic seizure prediction involves a different combination of EEG channels.

For BMOPSO, the selected 14 channels are distributed among frontal: FP1-F3, FP2-F8, FP2-F4; frontal-temporal: F7-T7, F8-T8; temporal-parietal: T7-P7, T8-P8; central-parietal: C3-P3, C4-P4, CZ-PZ; parietal-occipital: P3-O1, P8-O2; and frontal-central: F4-C4, FZ-CZ. The selected 8 channels for NSGA-III are frontal: FP2-F8; frontal-temporal: F7-T7, F8-T8; central-parietal: C3-P3, C4-P4; temporal-parietal: T8-P8; and parietal-occipital: P3-O1, P8-O2. As noted above, the proposed approach obtains the best $F1$ 97.94% with only one channel, which is central-parietal: C3-P3. Further, it achieves also a comparable performance ($F1$ = 96.72%) with 4 channels that are C3-P3, FP2-F4, F4-C4, and FZ-CZ.



(a) BMaOPSO-CA (b) MI-BMaOPSO-CA

Fig. 8 Variation of F1-score and channel number objectives in the first generation

Table 5 Performance metrics (%) of the Pareto-front

Algorithm	#Ch	F1-score	Sensitivity	Specificity
Baseline	**18**	**91.83**	**91.96**	**91.06**
BMOPSO	**14**	**94.63**	**97.21**	**92.25**
	10	93.75	93.35	94.54
	17	91.88	87.23	97.57
	11	91.64	97.87	85.29
	13	91.46	87.81	96.12
	11	91.26	95.66	86.82
	6	90.65	88.79	93.42
	13	90.42	84.49	97.75
	5	89.87	92.21	87.73
	9	89.79	96.59	82.53
NSGA-III	**8**	**95.72**	**97.54**	**94.21**
	7	94.96	96.38	93.88
	9	94.07	92.01	96.80
	5	93.40	94.55	92.68
	7	92.63	90.24	95.57
	7	92.24	90.32	95.01
	9	91.91	87.89	96.98
	6	91.74	96.60	87.07
	10	91.50	86.05	98.09
	6	90.81	88.26	94.33
MI-BMaOPSO-CA	**1**	**97.94**	**97.44**	**98.30**
	4	96.77	95.72	97.62
	6	95.90	96.57	94.93
	2	95.64	98.02	93.20
	2	94.25	97.15	91.39
	6	93.23	88.89	97.59
	3	92.02	86.22	98.27
	9	91.15	94.59	87.92
	6	90.57	85.53	95.33
	5	89.76	81.99	98.85

Values in bold indicate best results

According to these results, we can observe that the commonly selected channel is C3-P3 located in the central-parietal lobe. This region is consistent with the physiological principles of epileptic seizure [33].

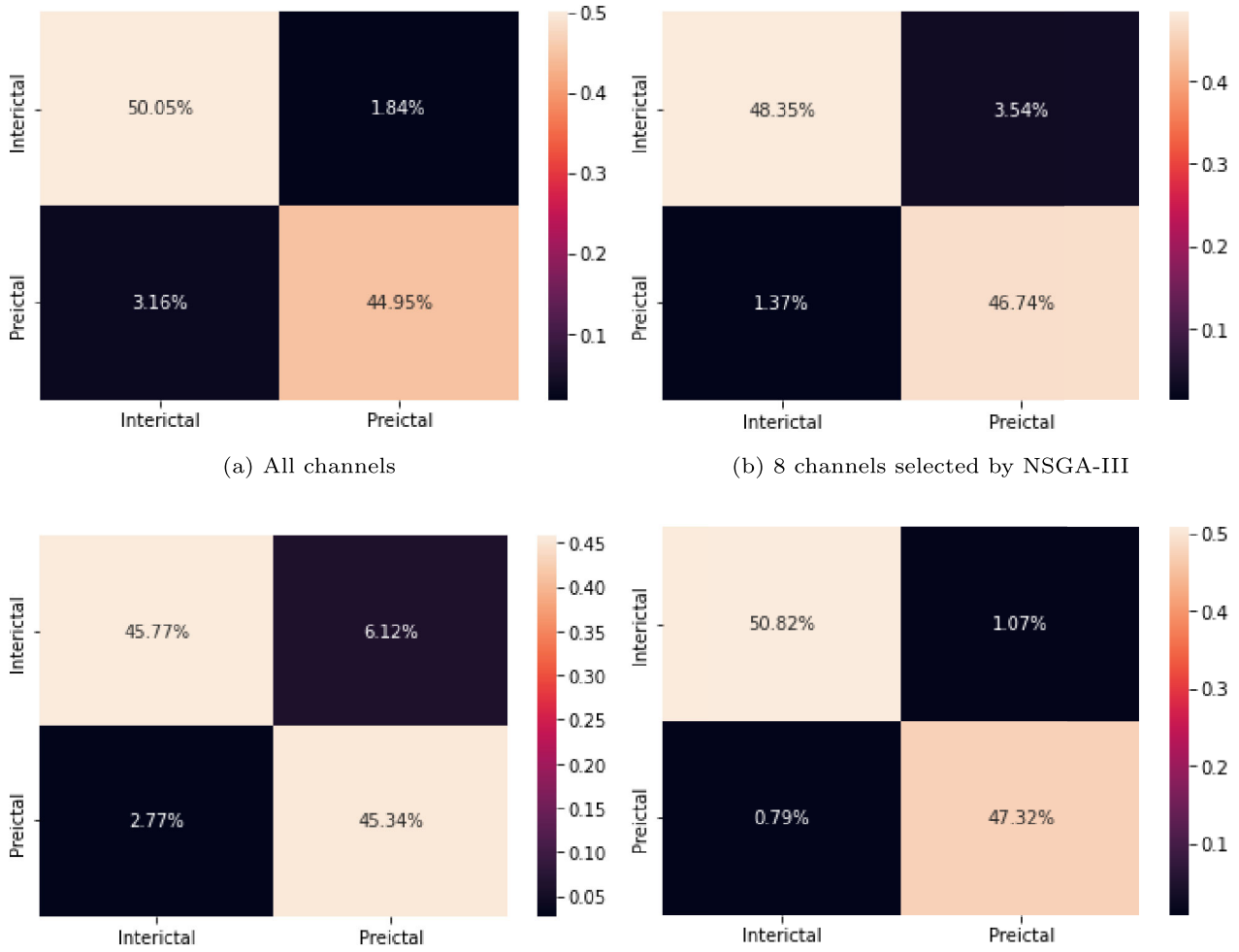
Comparison with Other Approaches

Table 6 shows a comparison between our channel selection method for epileptic seizure prediction and some previous studies that also deal with channel selection and consider the same dataset.

There are two types of selection: stable and variable. Most existing methods [8, 12, 15–17] opt for the variable type, selecting optimal channels tailored to specific patients. However, this approach necessitates designing a personalized EEG headcap for each patient, which is not practical. In our view, stable selection proves more suitable as it identifies optimal channels applicable across all patients, fostering the development of a comfortable, portable EEG headcap

Our approach stands out as the sole method employing a patient-independent scheme to select optimal channels. We introduced a ConvLSTM model that learns common features across all patients rather than patient-specific ones. Remarkably, our model using only one channel achieved 0.97%, 0.97%, and 0.98% for the $F1$, SEN, and $SPEC$ metrics, respectively, demonstrating robust generalization across all patients.

To assess the effectiveness of our ConvLSTM model, we conducted a comparative analysis with recent subject-independent methods, as outlined in Table 7. Notably, comparing different methods requires a consistent experimental setup for a fair assessment. However, existing methods vary in terms of the chosen preictal period and evaluation scheme. For example, [22, 27, 28] used a 60-min duration. Specifying



(a) All channels

(b) 8 channels selected by NSGA-III

(c) 14 channel selected by BMOPSO

(d) One channel selected by MI-BMaOPSO-CA

Fig. 9 Confusion Matrix of ConvLSTM model with **a** all channels and **b** selected channel

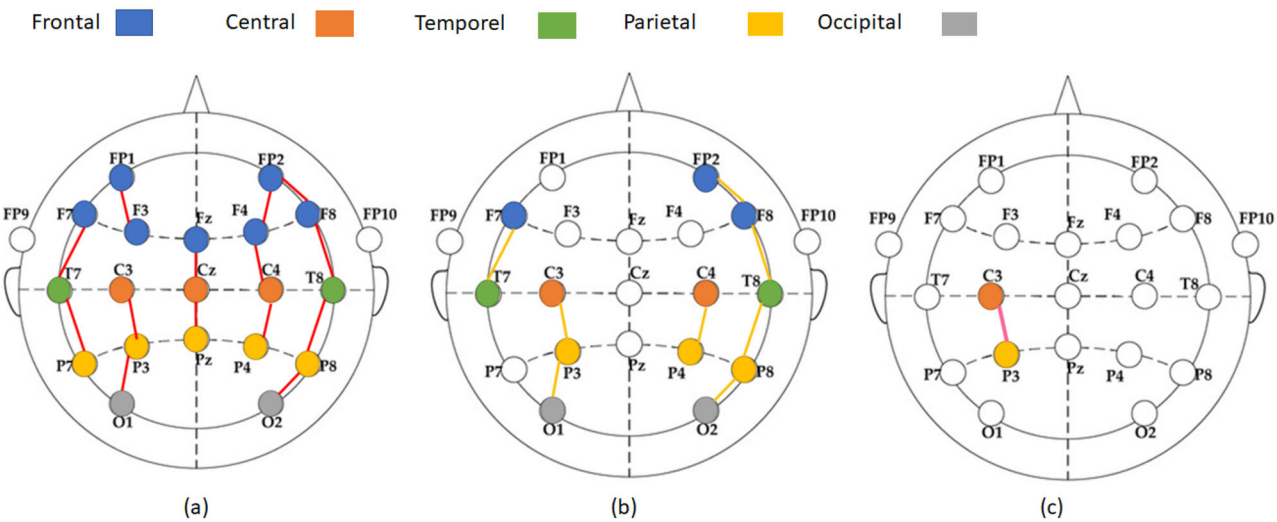


Fig. 10 EEG optimal channels location of compared algorithms: **a** BMOPSO (14 channels), **b** NSGA-III (8 channels), **c** MI-BMaOPSO-CA (1 channel)

Table 6 Channel selection findings on CHB-MIT dataset

Year	Ref.	Selection type[a]	Selection method	Selected channels	Performance
2021	[11]	Stable	Channel elimination	P3-O1, FP2-F8, P8-O2, P7-T7, T7-FT9, FT10-T8	Accuracy = 0.99 Sensitivity = 0.97 Specificity = 0.92 FPR/h = 0.0764
2021	[17]	Variable	SFS	—	Sensitivity = 0.92 FPR/h = 0.26
2021	[15]	Variable	GA	—	Sensitivity = 0.38±0.19 FPR/h = 1.03±0.48
2021	[16]	Variable	KNN-GA	—	Accuracy = 0.74 Sensitivity = 0.69 Specificity = 0.73
2022	[12]	Variable	NSGA-II	—	Sensitivity = 0.16±0.11 FPR/h = 0.21±0.08
2022	[8]	Variable	Attention score	—	F1-score = 0.76 Precision = 0.75 Recall = 0.78 FPR/h = 0.35
2023	[9]	Stable	NSGA-II	C4-P4, P4-O2, FP2-F8	Accuracy = 0.96 Sensitivity = 0.96 Specificity = 0.96
2023	Ours	Stable	BMOPSO	F7-T7, T7-P7, FP1-F3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ	F1-score = 0.94 Sensitivity = 0.97 Specificity = 0.92
2023	Ours	Stable	NSGA-III	F7-T7, C3-P3, P3-O1, C4-P4, P3-O1, FP2-F8, F8-T8, T8-P8, P8-O2	F1-score = 0.95 Sensitivity = 0.97 Specificity = 0.94
2024	**Ours**	**Stable**	**MI-BMOPSO-CA**	**C3-P3**	**F1-score = 0.97** **Sensitivity = 0.97** **Specificity = 0.98**

Values in bold indicate best results

[a]**Variable**: selected channels per patient. **Stable**: selected channels for all patients

Table 7 Comparison with previous studies on CHB-MIT dataset in patient-independent scheme

Year	Ref.	Preictal period (min)	Evaluation	Accuracy (%)	Sensitivity (%)	Specificity (%)
2017	[23]	10	10-fold CV	—	87.80	—
2019	[22]	60	Train/test	86.25	81.20	—
2021	[27]	60	10-fold CV	91.54	92.45	89.94
2022	[28]	60	10-fold CV	95.38	94.47	94.16
2022	[24]	8	Train/test	80.00	—	—
2023	[25]	10	Leave one trial out	—	92.02	—
2023	[26]	1	Leave one trial out	—	58.80	63.70
2024	Ours	5	Train/test	93.24	91.96	91.06

Table 8 Performance results (%) of LOPO vs TL: each row presents the classification with full channels in the first line as well as one channel in the second line

Patient	Leave-one-patient-out		Transfer learning	
	F1-score	FPR/h	F1-score	FPR/h
CH01	82.84	0.0963	99.26	0.0037
	81.96	0.1452	**99.62**	**0.0023**
CH02	65.15	0.4082	**98.88**	**0.0435**
	66.18	0.3706	76.72	0.1166
CH03	77.94	0.2642	**98.93**	**0.0087**
	83.57	0.1187	97.49	0.0102
CH04	87.23	0.0979	95.33	0.0224
	87.37	0.0559	**96.53**	**0.0218**
CH05	72.31	0.2179	88.89	0.0946
	83.44	0.1320	**98.20**	**0.0636**
CH06	89.15	0.0398	93.80	0.0769
	89.97	0.0139	**97.04**	**0.0272**
CH07	78.68	0.2722	90.39	0.0907
	80.29	0.1105	**99.58**	**0.0563**
CH08	72.32	0.1103	90.20	0.0181
	77.71	0.0965	**96.00**	**0.0114**
CH09	87.29	0.0808	95.77	0.0562
	79.56	0.0916	**99.15**	**0.0258**
CH10	98.47	0.0094	**99.83**	**0.0006**
	78.94	0.0221	97.63	0.0883
CH11	67.56	0.5689	85.13	0.4186
	70.41	0.2934	**99.10**	**0.1023**
CH12	70.02	0.1527	**96.53**	**0.0005**
	70.80	0.1169	77.59	0.1023
CH13	86.65	0.0905	**95.90**	**0.0140**
	77.51	0.1124	94.60	0.0615
CH14	98.85	0.0023	**99.76**	**0.0002**
	93.68	0.0467	98.00	0.0615
CH15	82.25	0.0159	91.11	0.0102
	82.69	0.0240	**91.95**	**0.0480**
CH16	67.99	0.1745	84.38	0.1033
	85.07	0.0693	**91.25**	**0.0415**
CH17	61.56	0.3716	84.07	0.0599
	87.54	0.0326	**97.89**	**0.0378**
CH18	93.37	0.0374	96.69	0.0615
	92.91	0.03015	**97.57**	**0.0164**
CH19	72.09	0.4138	**99.21**	**0.0012**
	91.99	0.0343	98.63	0.0061
CH20	97.69	0.0037	**99.06**	**0.0043**
	76.53	0.0376	95.73	0.0502
CH21	94.21	0.0076	**98.29**	**0.06151**
	94.67	0.0322	98.41	0.0099
CH22	79.48	0.2975	98.68	0.0615
	88.10	0.0332	**98.81**	**0.0151**
CH23	94.40	0.0290	95.26	0.0615
	82.01	0.0474	**97.03**	**0.0246**
CH24	62.22	0.0162	**96.15**	**0.0023**
	82.01	0.0474	89.52	0.01468

Values in bold indicate best results

the data partition is also an important step in the experiment and affects the global performance. In some works [25, 26], the evaluation is done using N-1 trials for training and one trial for testing. In our study, the ConvLSTM model achieved an accuracy of 93.24%, slightly below the method proposed in [28]. Remarkably, our model using only one channel achieved a state-of-the-art accuracy of 98.01%. This investigation represents the first exploration of EEG channel reduction in a patient-independent scheme, prompting further exploration to unravel the critical factors contributing to the model's exceptional performance.

Transfer Learning for Seizure Prediction

As mentioned above, this study's main objective is to create a seizure prediction model that is not dependent on a specific patient. Additionally, we assess the creation of personalized classifiers through transfer learning.

In this experiment, we remove instances (di) extracted from the selected subject i (i ∈ [1, 2,..., 24]) from the dataset (both training and evaluation). Then, we train the model (Mi) using the rest of the available data. After training the model, we tested two different classification schemes: leave-one-patient-out (LOPO) and transfer learning (TL) as depicted in Table 8. In the formal, the test set contains all instances di of the discarded subject i. In the latter, 30% of subject instances di are taken as train set for fine-tuning the learned model (Mi), and the remaining 70% is considered as a test set. We perform this evaluation for all 24 patients, and train models for 20 epochs.

Referring to Table 8, the transfer learning classification outperforms the leave-one-patient-out mode. This superiority stems from fine-tuning the model with 30% of the test data, enabling the model to grasp the data distribution. In the leave-one-patient-out mode, the model struggles to generalize to unseen data from new patients due to the high variability of EEG signals.

Specifically, in the leave-one-patient-out mode, the one-channel model achieved higher performance for 15 out of 24 patients. Additionally, in the transfer learning mode, the one-channel model showed better performance for 13 out of 24 patients. These findings underscore the significance of reducing EEG channels.

Conclusion

The primary objectives of this study were twofold: to conceive a deep learning model for patient-independent epileptic seizure prediction and to identify the most relevant EEG channels.

Our proposed optimization method (MI-BMaOPSO-CA) comprises two modules. First, a neural module based on ConvLSTM layers is developed to extract discriminative features from raw EEG data. Second, an evolutionary process with four objectives is employed to determine the most pertinent EEG channels. Rigorous validation on CHB-MIT using various evaluation methods reveals that even a single channel can yield promising performance results.

Our approach stands out from prior methods as it automatizes both channel selection and feature learning, eliminating the need for manual feature extraction. To our knowledge, this is the pioneering work focusing on EEG channel reduction for patient-independent epileptic seizure prediction. Proposing a universal model applicable to all patients holds significance, facilitating clinicians to directly test it on new patients. The model operates with raw EEG data as input, further enhancing its ease of use in real-time clinical applications.

In future investigations, our aim is to identify the optimal preictal period for seizure prediction. Our study suggests that the choice of preictal period significantly impacts the overall model performance. Therefore, our next focus will center on exploring the preictal period and fine-tuning it concerning FPR/h in our optimization method.

Data Availability The CHB-MIT dataset in our experiments is public. It is available online.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Not applicable.

Conflict of Interest The authors declare no competing interests.

References

1. Hazarika BB, Gupta D, Kumar B. EEG signal classification using a novel Universum-based twin parametric-margin support vector machine. Cogn Comput. 2023. https://doi.org/10.1007/s12559-023-10115-w.
2. Baghdadi A, Aribi Y, Fourati R, et al. Psychological stimulation for anxious states detection based on EEG-related features. J Ambient Intell Humaniz Comput. 2021;12(8):8519–33.
3. Baghdadi A, Fourati R, Aribi Y, et al. Robust feature learning method for epileptic seizures prediction based on long-term EEG signals. In: 2020 International Joint Conference on Neural Networks (IJCNN). 2020. p. 1–7. https://doi.org/10.1109/IJCNN48605.2020.9207070.
4. Baghdadi A, Fourati R, Aribi Y, et al. A channel-wise attention-based representation learning method for epileptic seizure detection and type classification. J Ambient Intell Humaniz Comput. 2023;14(7):9403–18.
5. World Health Organization. Epilepsy. 2022. https://www.who.int/news-room/fact-sheets/detail/epilepsy. Accessed 3 Feb 2023.
6. Visalini K, Alagarsamy S, Nagarajan D. Neonatal seizure detection using deep belief networks from multichannel EEG data. Neural Comput Appl. 2023;35:10637–47.
7. Wu D, Li J, Dong F, et al. Classification of seizure types based on multi-class specific bands common spatial pattern and penalized ensemble model. Biomed Signal Process Control. 2023;79:104118. https://doi.org/10.1016/j.bspc.2022.104118. https://www.sciencedirect.com/science/article/pii/S1746809422005742.
8. Affes A, Mdhaffar A, Triki C, et al. Personalized attention-based EEG channel selection for epileptic seizure prediction. Expert Syst Appl. 2022;206:117733. https://doi.org/10.1016/j.eswa.2022.117733. https://www.sciencedirect.com/science/article/pii/S0957417422010144.
9. Jana R, Mukherjee I. Efficient seizure prediction and EEG channel selection based on multi-objective optimization. IEEE Access. 2023;11:54112–21. https://doi.org/10.1109/ACCESS.2023.3281450.
10. Mormann F, Kreuz T, Rieke C, et al. On the predictability of epileptic seizures. Clin Neurophysiol. 2005;116(3):569–87.
11. Jana R, Mukherjee I. Deep learning based efficient epileptic seizure prediction with EEG channel optimization. Biomed Signal Process Control. 2021;68(102):767. https://doi.org/10.1016/j.bspc.2021.102767. https://www.sciencedirect.com/science/article/pii/S1746809421003645.
12. Pinto M, Coelho T, Leal A, et al. Interpretable EEG seizure prediction using a multiobjective evolutionary algorithm. Sci Rep. 2022;12(1):1–15.
13. Wang Y, et al. A spatiotemporal graph attention network based on synchronization for epileptic seizure prediction. IEEE J Biomed Health Inform. 2023;27(2):900–11. https://doi.org/10.1109/JBHI.2022.3221211.
14. Ra JS, Li T, YanLi. A novel epileptic seizure prediction method based on synchroextracting transform and 1-dimensional convolutional neural network. Comput Methods Programs Biomed. 2023;240:107678. https://doi.org/10.1016/j.cmpb.2023.107678. https://www.sciencedirect.com/science/article/pii/S0169260723003437.
15. Pinto M, Leal A, Lopes F, et al. A personalized and evolutionary algorithm for interpretable EEG epilepsy seizure prediction. Sci Rep. 2021;11(1):1–12.
16. Ra JS, Li T, Li Y. A novel permutation entropy-based EEG channel selection for improving epileptic seizure prediction. Sensors. 2021;21(23):7972.
17. Romney A, Manian V. Optimizing seizure prediction from reduced scalp EEG channels based on spectral features and MAML. IEEE Access. 2021;9:164348–57. https://doi.org/10.1109/ACCESS.2021.3134166.
18. Li R, Ren C, Zhang X, et al. A novel ensemble learning method using multiple objective particle swarm optimization for subject-independent EEG-based emotion recognition. Comput Biol Med. 2022;140:105080. https://doi.org/10.1016/j.compbiomed.2021.105080. https://www.sciencedirect.com/science/article/pii/S001048252100874X.
19. Sheoran P, Saini J. Optimizing channel selection using multi-objective FODPSO for BCI applications. Brain-Computer Interfaces. 2022;9(1):7–22. https://doi.org/10.1080/2326263X.2021.1966985.

20. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks, vol. 4. 1995. p. 1942–8. https://doi.org/10.1109/ICNN.1995.488968.
21. Kennedy J, Eberhart R. A discrete binary version of the particle swarm algorithm. In: 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, vol. 5. 1997. p. 4104–8. https://doi.org/10.1109/ICSMC.1997.637339.
22. Hu W, Cao J, Lai X, et al. Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks. J Ambient Intell Human Comput. 2023;14:15485–95. https://doi.org/10.1007/s12652-019-01220-6.
23. Khan H, Marcuse L, Fields M, et al. Focal onset seizure prediction using convolutional networks. IEEE Trans Biomed Eng. 2018;65(9):2109–18. https://doi.org/10.1109/TBME.2017.2785401.
24. Halawa RI, Youssef SM, Elagamy MN. An efficient hybrid model for patient-independent seizure prediction using deep learning. Appl Sci. 2022;12(11):5516.
25. Nazari J, Nasrabadi AM, Menhaj MB, et al. Epileptic seizure prediction using multi-channel raw EEGs with convolutional neural network. J Robot Syst. 2023;16(2):26–35.
26. Mao T, Li C, Zhao Y, Song R, Chen X. Online test-time adaptation for patient-independent seizure prediction. IEEE Sens J. 2023;23(19):23133–44. https://doi.org/10.1109/JSEN.2023.3307223.
27. Dissanayake T, Fernando T, Denman S, et al. Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals. IEEE Sens J. 2021;21(7):9377–88.
28. Dissanayake T, Fernando T, Denman S, et al. Geometric deep learning for subject independent epileptic seizure prediction using scalp EEG signals. IEEE J Biomed Health Inform. 2022;26(2):527–38.
29. Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Cortes C, Lawrence ND, Lee DD, et al., editors. Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7–12, 2015, Montreal, Quebec, Canada. 2015. p. 802–10. http://papers.nips.cc/paper/5955-convolutional-lstm-network-a-machine-learning-approach-for-precipitation-nowcasting.
30. Shoeb A. CHB-MIT Scalp EEG Database. 2010. https://physionet.org/content/chbmit/1.0.0/. Accessed 3 Jan 2022.
31. Bandarabadi M, Rasekhi J, Teixeira CA, et al. On the proper selection of preictal period for seizure prediction. Epilepsy Behav. 2015;46:158–66.
32. Jiang X, Liu X, Liu Y, et al. Epileptic seizures detection and the analysis of optimal seizure prediction horizon based on frequency and phase analysis. Front Neurosci. 2023;17:1191683.
33. Bergil E, OCBozkurt MR. An evaluation of the channel effect on detecting the preictal stage in patients with epilepsy. Clin EEG Neurosci. 2021;52(5):376–85.