Adaptive Stepwise Feature Selection Approach for EEG-Based Epileptic Seizure Classification

Sunday Timothy Aboyeji[#]
CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
Shenzhen, China
timothy@siat.ac.cn

Wenfang Zhou[#]
Department of Otolaryngology, Luohu District People's Hospital
Shenzhen, China
wenfang_1985@126.com

Yuan Tao
Department of Otorhinolaryngology Peking University Shenzhen Hospital
Shenzhen, China
taop19772000@163.com

Mingxing Zhu
School of Electronics and Information Engineering Harbin Institute of Technology, Shenzhen
Shenzhen, China
bestarpc@126.com

Oluwarotimi Williams Samuel
School of Computing and Engineering, University of Derby
Derby, UK
o.samuel@derby.ac.uk

Ijaz Ahmad
CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
Shenzhen, China
ijaz@siat.ac.cn

Guoru Zhao
CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
Shenzhen, China
gr.zhao@siat.ac.cn

Ji Yi
Department of Ophthalmology, School of Medicine, Johns Hopkins University
Baltimore, USA
jiyi@jhu.edu

Michael Chi Fai Tong
Department of Otorhinolaryngology, Head and Neck Surgery The Chinese University of Hong Kong,
Hong Kong, China
mtong@ent.cuhk.edu.hk

Xin Wang*
CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
Shenzhen, China
wangxin@siat.ac.cn

Yi Guo*
Department of Neurology Shenzhen People's Hospital
Shenzhen, China
xuanyi_guo@163.com

Shixiong Chen*
School of Medicine, The Chinese University of Hong Kong, Shenzhen
Shenzhen, China
chenshixiong@cuhk.edu.cn

***Abstract*—Recent advancements in feature selection (FS) optimization algorithms have influenced the field of epileptic seizure classification. However, integrating these optimization algorithms into machine learning (ML) models often creates time complexity, limiting their clinical deployment. To address this issue, we propose an innovative adaptive stepwise FS method tailored for epileptic seizure detection (ESD). First, a discrete wavelet transform (DWT) was applied to the preprocessed signal to get three levels of the db4 wavelet family within the frequency range pertinent to epileptic seizure classification. Linear and nonlinear features are then extracted from each level of the DWT. The selected features are initially ranked using the minimum relevance, maximum redundancy (mRMR) FS technique. After that, a stepwise FS approach was applied to the ranked features to optimize the performance of Random Forest (RF), K-Nearest Neighbour (KNN), and Support Vector Machine (SVM) classifiers. The experiment was performed on a publicly accessible CHB-MIT datasets in a patient-independent approach. The model's performance was assessed using accuracy, sensitivity, and specificity. The results show an improved performance of the ML models with the integration of stepwise algorithm into the mRMR technique. Among the classifiers, RF exhibited superior performance with accuracy, sensitivity, and specificity of 87.69%, 91.53%, and 83.86%, respectively, when 12 features were selected. Our proposed stepwise feature selection method (PSFS) performs similarly to generalize forward feature selection (GFFS), with an average accuracy of 88.37% and 88.57%, respectively across selected features with less computation. This makes PSFS a very efficient and effective FS in epileptic seizure classification.**

Keywords*—*Electroencephalogram, Discrete wavelet transform, Machine learning, Minimum redundancy maximum relevance, Stepwise feature selection

I. INTRODUCTION

Epileptic seizure is a severe neurological disease that endangers the patients affected by it. Such individuals cannot take up certain tasks because of the fear of being attacked by the seizure when performing daily life activities. Approximately 50 million individuals globally are currently affected by epilepsy [1], and about 100 million of them have the probability of a seizure attack at least once a year [2]. Neurologists usually manage patients through continuous monitoring of electroencephalogram (EEG) signals from their brains. This approach proved labour-intensive and time-consuming as some patients' EEG recordings span two to three weeks. Therefore, computer-aided diagnosis systems (CADs) have been introduced to aid the process of identifying

This work was supported in part by the National Key R&D Program of China (2022YFE0197500), National Natural Science Foundation of China (#81927804, #62101538), Science and Technology Planning Project of Shenzhen (#JSGG20210713091808027, #JCYJ20230807093819039), and the Science and Technology Program of Guangdong Province (2022A0505090007). (Corresponding author: Xin Wang, Yi Guo, and Shixiong Chen.)

#The first two authors contribute equally to this work.

epileptiforms from long-recorded EEGs. This procedure involves the classification of EEG signals into ictal (seizure) and interictal (normal) states by machine learning (ML). ML classifiers cannot directly learn EEG signal patterns; therefore, preprocessing is needed. After the noise and other artifacts have been removed from the EEG signals, feature extraction processes are the next stages. Features can be extracted directly from time domain or frequency domain using fast Fourier transform (FFT) or time-frequency domain utilizing discrete wavelet transform (DWT). Moreover, multiple linear and nonlinear features can be extracted from these domains, especially when considering multichannel EEG signals. For example, Abou-Abbas et al. [3] extracted 856 features from 19 EEG signal channels by considering different wavelet transform decomposition levels. In addition, Yang et al. [4] likewise extracted 638 features from 22 channels of EEG signal. The significant number of features often extracted from multichannel EEG signals makes the machine learning models computationally intensive and, in most cases, introduces redundancy into the system. Therefore, various feature selection techniques have been proposed to select the clinically representative features for epileptic seizure classifications. Among the commonly used FS is the correlation FS introduced by [5]. The authors extracted relevant features from time-domain, entropy-based, and DWT features. The stability of various feature selection techniques, such as the Gini Index, minimum redundancy and maximum relevance (mRMR), unsupervised graph-based feature selection, etc., were assessed by [3]. The authors in [6] used correlation coefficient and distance correlation to perform linear and nonlinear feature selection. Other commonly used feature selection methods for EEG-based seizure detection are principal component analysis (PCA) [7], mutual information base FS [8], etc. However, FS techniques are biased to inherent properties (linear or nonlinear) of certain features and pay attention to those features during computation, affecting their optimum performance. In this case, features with relevant information that could contribute to the model's overall accuracy are lost. A recursive feature elimination (RFE) method is often used to continuously exclude features with low feature important scores from the subset to overcome the information loss due to early feature vector reduction. However, RFE is computationally intensive and unsuitable for large feature-size datasets. Therefore, this study proposed a stepwise feature selection technique for epileptic seizure detection. A base feature selection technique such as mRMR is first used to rank

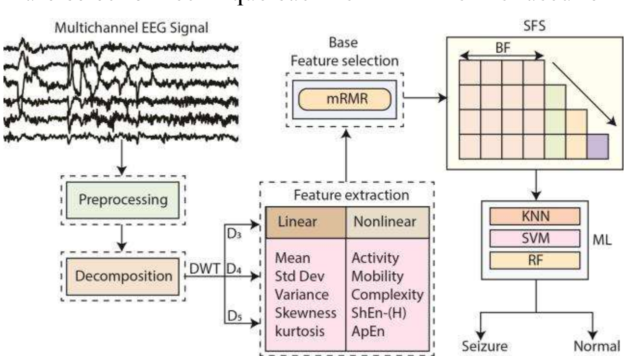

Fig. 1. Proposed framework of minimum redundancy maximum relevance stepwise FS (mRMRSFS). BF represents base features selected by mRMR.

the features according to their relevance in classifying seizure. After that, a number of initial features are selected, and the stepwise algorithm is implemented over a specified range to provide a cost-effective feature selection algorithm. In this way, the time complexity of the model is reduced for improved performance for epileptic seizure classification.

II. METHODS

A. EEG datasets description and preprocessing

In this study, the publicly accessible CHB-MIT datasets from Boston Children's Hospital were adopted. CHB-MIT database is a multichannel scalp EEG recording of 22 pediatric subjects with intractable seizures [9]. Recordings were acquired at 256 samples per second with 16-bit resolution. The bipolar montage international 10-20 channel configuration system was utilized to acquire the EEG datasets. The first ten patients of the CHB-MIT datasets with similar 23 channels are combined to achieve patient-independent feature selection analysis in this study. The multichannel EEG signals of the ictal interval of each patient with the corresponding preictal interval are acquired to achieve binary classification. Hamming window with low and high passband edges of 0.5 Hz and 40 Hz, respectively, was used to remove the noise and other artifacts.

B. Decomposition and feature extraction

As shown in Fig. 1, DWT was performed to decompose the signal into relevant frequencies of interest. DWT serves as a technique in signal processing that breaks down a signal into various frequency bands, offering insights into both the time and frequency aspects of the signal [10]. The full description of DWT decomposition technique can be found in [5]. After the decomposition, three significant levels corresponding to the frequency bands of interest of the Daubechies (db4) wavelet function family are extracted. Since the resulting EEG signal is already preprocessed in the frequency range of 0.5 Hz to 40 Hz, levels D5, D4, and D3 capture frequency ranges of 0 – 8 Hz, 8 –16 Hz, 16 – 32 Hz, respectively, covering the most seizure activities (3 – 30 Hz) according to [11]. As shown in Fig. 1, five linear and five nonlinear features are extracted from each decomposition level. Thereafter, ten relevant features are extracted from each channel. Mean, variance, standard deviation, skewness, and kurtosis describe the linear property of EEG signal while Hjorth parameters (activity, mobility, and complexity), Shannon entropy (*H(x)*) and approximate entropy (*ApEn*) describe the complexity and nonlinear nature of EEG signals. The process of obtaining linear features and Hjorth parameters are described in [3], [4].

$$H(x) = -\sum_i P(x_i)\log_2 P(x_i) \tag{1}$$

$$ApEn(m,r,N) = \phi^m(r) - \phi^{m+1}(r) \tag{2}$$

C. Feature selection and classification

The feature selection approach is a standard procedure in machine learning, especially in seizure detection algorithms, to reduce high-dimensional feature vectors extracted from multichannel EEG signals to low-dimensional subspaces. First, the minimum redundancy and maximum relevance feature selection were used to rank 690 features extracted from 23 channels from each of the three decomposition levels. In this study, the mRMR FS largely depends on mutual information between the features and the target variables (seizure and normal). It also depends on the correlation among the features to assess their redundancy. The calculation of mutual information follows the process described in [12]. Given n-dimensional data $X = \{X_1, X_2, ..., X_n\}$, the entropy of random variable X is given by Eq. 1. The conditional entropy of two random variables X and Y is

$$H(Y \mid X) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(y \mid x)\,. \qquad (3)$$

The mutual information (MI) between the two random variable can be obtained as:

$$MI(X;Y) = H(Y) - H(Y \mid X) \qquad (4)$$

The Pearson correlation coefficient between two features X and Y is

$$r = \frac{\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}}\,. \qquad (5)$$

For a given feature X_j and other feature X_i, if n is the number of other features, then, the redundancy is given as:

$$\text{Redundancy}(X_j) = \sum_{i=1}^{n}\frac{|r(X_j, X_i)|}{n} \qquad (6)$$

Finally, the mRMR that determines relevant subset of features to target variable is

$$mRMR(X_j) = MI(X_j, Y) - \text{Redundancy}(X_j)\,. \qquad (7)$$

The mRMR stepwise FS is described in algorithm 1.

Three state-of-the-art machine learning classifiers, such as Random Forest (RF), K-Nearest Neighbour (KNN), and Support Vector Machine (SVM), are selected in this study based on their remarkable success in classifying epileptic seizures [1], [3], [4], [5]. We assess the performance of the proposed mRMRSFS algorithm with the classifiers using notable metrics such as accuracy, sensitivity, and specificity through 10-fold cross-validation.

Algorithm 1: mRMR stepwise feature selection

Input: mRMR R ranked features, number of first selected k features, number of features to select up to n
Output: Best features X
Initialization: initial feature set $\Omega \leftarrow R[1{:}k]$, next feature set $\Phi \leftarrow R[k{:}n]$, best accuracy $\Lambda \leftarrow$ model accuracy on Ω
Best features $X \leftarrow$ set initial features Ω to best features
Current features $\Psi \leftarrow$ set best features X to current features
  for $i = 1$ to length(Φ) **do**
    $\Psi \leftarrow \Psi + \Phi[i]$
    $\Xi \leftarrow$ get model accuracy for the set of feature Ψ
    if $\Xi > \Lambda$ then
      $\Lambda \leftarrow \Xi$
      $X \leftarrow \Psi$
    end if
    if length(Ψ) $\geq n$ **then**
      Break
    end if
  end for
Return best features X

III. RESULTS AND DISCUSSION

A. Performance of the proposed mRMRSFS method on CHB-MIT datasets using three ML models

Table I shows the results using only the mRMR feature selection technique. Here, the first-specified relevant features are selected according to their mRMR ranking. The selected features are 15, 30, 40, 50 and 60. It can be observed that the performance of the classifiers increase when 50 features are selected and then decreases when 60 features are selected for RF and KNN. However, SVM performance decreases after 50 features are selected and then increases again for 60 features. RF shows a significantly better performance than SVM and KNN. This can be attributed to RF sensitivity values being higher than the sensitivity of the other two ML classifiers.

Table I: Performance of machine learning model base on mRMR. The scores are in percentage and "Feat" represents the number of selected features for typical experiment.

	RF			KNN			SVM		
Feat	**Acc**	**Sen**	**Spec**	**Acc**	**Sen**	**Spec**	**Acc**	**Sen**	**Spec**
15	86.51	91.71	81.33	86.08	88.38	83.78	80.33	94.94	66.43
30	87.75	92.33	83.19	86.15	88.53	83.78	86.05	81.38	90.74
40	88.42	92.90	83.96	86.91	89.17	84.65	86.45	82.54	90.37
50	88.61	93.15	84.08	86.08	87.21	84.95	86.41	82.99	89.84
60	88.59	93.27	83.91	85.14	84.06	86.22	87.57	84.93	90.22

To further improve the performance of mRMR more efficiently, we added features to already ranked features in a stepwise manner. For each feature range (FR), the selected features are sequentially arranged in Table II. The results show that integrating SFS with mRMR allows the selection of feature combinations that improve the ML classifiers within the specified range. When 15 features are used, RF achieved 86.51%, but when optimized within the range (10-15), 12 features achieved an accuracy of 87.69%. Similar improvements are observed for KNN and SVM by using the proposed mRMRSFS. From Fig. 2, oscillation in the performance of the ML classifiers can be observed as the number of features increases. This shows that the number of features that improve the performance is within a specific range, and developing an optimization algorithm to select the combination of these features is essential.

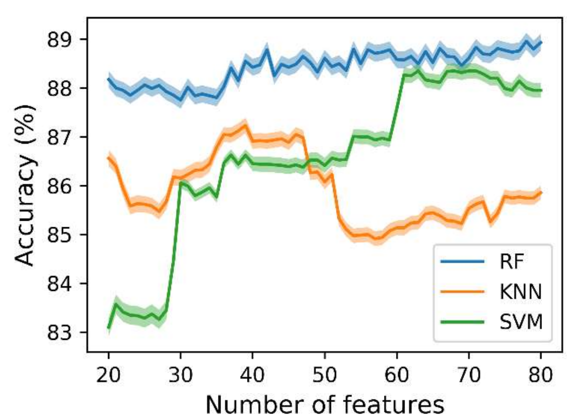

Fig. 2. Performance of the three ML classifiers as the number of selected features increase. The shaded region indicates the standard deviation

Table II: Performance of machine learning models base on mRMR + SFS. The numbers in the parentheses represents the features selected in each of the feature ranges (FR) for individual ML.

	RF (12, 25, 39, 42, 54)			**KNN** (12, 21, 39, 46, 51)			**SVM** (15, 30, 37, 48, 60)		
FR	**Acc**	**Sen**	**Spec**	**Acc**	**Sen**	**Spec**	**Acc**	**Sen**	**Spec**
10-15	87.69	91.53	83.86	86.67	88.68	84.68	80.33	94.94	66.43
20-30	88.06	92.40	84.17	86.40	88.85	83.96	86.05	81.38	90.74
30-40	88.55	93.12	83.98	87.23	90.02	84.45	86.62	82.34	90.91
40-50	88.78	93.39	84.18	87.05	89.42	84.68	86.52	82.84	90.22
50-60	88.79	93.52	84.08	86.23	87.19	85.27	87.57	84.93	90.22

KNN achieves optimum performance between 39 to 46 features and then decreases drastically. However, KNN started improving again after 55 features. SVM had a sharp increase to 86.05% when 30 features were selected. RF achieved little or no improvement after 60 features, while the performance of SVM began to drop after 70 features. However, with the integration of SFS into the mRMR technique, we were able to identify optimum features within different ranges. In addition, the accuracy, sensitivity, and specificity achieved with 25 features by RF is higher than the accuracy, specificity, and sensitivity of 30 features in [4] for patient-independent task.

B. Time complexity comparison of the proposed stepwise FS approach and generalized forward FS

We further compare the performance of our proposed stepwise feature selection (PSFS) with the generalized forward feature selection (GFFS). In GFFS, the ranked features based on mRMR are iteratively added to improve the model's performance, starting with no features. This differs from the PSFS, in which the features are added within the specified range. As shown in Fig. 3, PSFS has comparable performance with GFFS with slight differences. However, the proposed SFS algorithm is less computationally expensive as the maximum time spent is much lower than that for GFFS. For example, when 54 features were selected, the computational time of PSFS was one-seventh of that of GFFS. It can be noticed that the time complexity of the GFFS increases drastically as the number of features added increases. However, the computational time for PSFS has steadily increased with improved accuracy. The wide gap between the computational time of our PSFS and GFFS suggested its suitability as FS for ESD, where a fast and accurate system is required. In addition, we identified Shannon entropy, activity, mean, and variance as the most relevant features for seizure classification based on the first 54 ranked features from the proposed method. These features are selected across all the channels of CHB-MIT datasets, but channels 'FT9-FT10' and 'FT10-T8' have more occurrences.

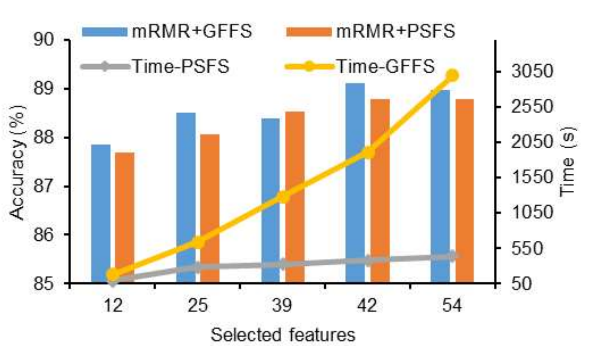

Fig. 3. Comparison of the proposed stepwise feature selection with commonly used generalized forward feature selection.

IV. CONCLUSION

This study presents a cost-efficient and computationally efficient optimization FS algorithm for ESD. Initially, the extracted features are ranked using mRMR. Identifying feature subsets of these ranked features that enhance ML classifiers' performance is crucial for improving seizure classification. However, finding possible feature subset combinations for optimum performance is computationally intensive. Therefore, we have stepwisely added new features to initially selected mRMR features. This strategy resulted in a computationally efficient FS method. The model achieved comparable performance with the GFFS approach and used relatively less time. In the future, we plan to consider the fusion of different feature subsets within the defined range in an efficient optimization algorithm. Finally, the proposed FS technique can be integrated into computer-aided diagnosis systems to enhance the efficiency of epileptic seizure classification.

REFERENCES

[1] M. K. Siddiqui, R. Morales-Menendez, X. Huang, and N. Hussain, "A review of epileptic seizure detection using machine learning classifiers," Brain Inform, vol. 7, no. 1, p. 5, Dec. 2020, doi: 10.1186/s40708-020-00105-1.

[2] G. Alarcón and A. Valentín, Introduction to epilepsy. Cambridge University Press, 2012.

[3] L. Abou-Abbas, K. Henni, I. Jemal, A. Mitiche, and N. Mezghani, "Patient-independent epileptic seizure detection by stable feature selection," Expert Syst Appl, vol. 232, p. 120585, Dec. 2023, doi: 10.1016/j.eswa.2023.120585.

[4] S. Yang et al., "Selection of features for patient-independent detection of seizure events using scalp EEG signals," Comput Biol Med, vol. 119, p. 103671, Apr. 2020, doi: 10.1016/j.compbiomed.2020.103671.

[5] M. Mursalin, Y. Zhang, Y. Chen, and N. V Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier," Neurocomputing, vol. 241, pp. 204–214, 2017.

[6] I. Ahmad et al., "An efficient feature selection and explainable classification method for EEG-based epileptic seizure detection," Journal of Information Security and Applications, vol. 80, p. 103654, Feb. 2024, doi: 10.1016/j.jisa.2023.103654.

[7] E. M. Imah and A. Widodo, "A comparative study of machine learning algorithms for epileptic seizure classification on EEG signals," in 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE, Oct. 2017, pp. 401–408. doi: 10.1109/ICACSIS.2017.8355065.

[8] K. M. Hassan, Md. R. Islam, T. T. Nguyen, and Md. K. I. Molla, "Epileptic seizure detection in EEG using mutual information-based best individual feature selection," Expert Syst Appl, vol. 193, p. 116414, May 2022, doi: 10.1016/j.eswa.2021.116414.

[9] A. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. ," Circulation [Online], vol. 101, no. 23, pp. 215–220, 2000.

[10] K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, and H. S. A. Shatnawi, "Effective Early Detection of Epileptic Seizures through EEG Signals Using Classification Algorithms Based on t-Distributed Stochastic Neighbor Embedding and K-Means," Diagnostics, vol. 13, no. 11, p. 1957, Jun. 2023, doi: 10.3390/diagnostics13111957.

[11] S. Grewal and J. Gotman, "An automatic warning system for epileptic seizures recorded on intracerebral EEGs," Clinical Neurophysiology, vol. 116, no. 10, pp. 2460–2472, Oct. 2005, doi: 10.1016/j.clinph.2005.05.020.

[12] Kai Keng Ang and Chai Quek, "Rough set-based neuro-fuzzy system," in The 2006 IEEE International Joint Conference on Neural Network Proceedings, IEEE, 2006, pp. 742–749. doi: 10.1109/IJCNN.2006.246758