A Comparative Analysis of Artificial Intelligence Optimization Algorithms for the Selection of Entropy-based Features in the Early Detection of Epileptic Seizures

Yusuf Bahri Özçelik and Aytaç Altan

Zonguldak Bülent Ecevit University, Department of Electrical and Electronics Engineering, Zonguldak, Turkey
ybahri.ozcelik@beun.edu.tr, aytacaltan@beun.edu.tr

Abstract

Epilepsy, a neurological condition colloquially known as a seizure disorder, causes involuntary muscle contractions and cognitive changes through sudden, uncontrolled neuronal discharges in the brain. The recurrent, unpredictable nature of these seizures poses the threat of potentially fatal or irreversible brain damage, underscoring the critical importance of early detection of epilepsy seizures. This study extracts informative features from medical records to improve early epilepsy seizure diagnosis. Employing bio-inspired optimization algorithms, it performs feature selection and constructs two different machine learning models, both equipped with optimization algorithms for epilepsy seizure diagnosis. Evaluation encompasses comprehensive metrics including accuracy, precision, F1 score and computational cost, with convergence graphs highlighting the impact of the optimization algorithms. Encouragingly, the results show that the model, using just five selected features, achieves an impressive 95.28% accuracy in diagnosing epileptic seizures. This highlights the suitability of the proposed model for real-time applications, characterized by its streamlined parameter complexity.

1. Introduction

Epilepsy is a neurological brain disorder associated with prolonged seizures resulting from sudden impairment of the central nervous system [1]. Epilepsy, one of the most prevalent brain disorders today, affects approximately 70 million individuals globally [2]. With epilepsy estimated to affect approximately 1% of the world's population, recurrent epileptic seizures pose a risk to people worldwide [3-4]. In some cases, seizures can cause permanent brain damage and in others they can be fatal. Early diagnosis of recurrent seizures is therefore vital to identify and treat people with this life-threatening condition.

Epileptic seizures, which are abrupt and transient, can be quantified from synchronized electrical activity in the brain. Numerous methods have been proposed in the literature to measure epileptic signals, including electroencephalography (EEG), positron emission tomography (PET), single photon emission computed tomography (SPECT), magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI) [5]. However, data from PET, SPECT, MRI and fMRI are limited due to their lengthy and costly nature. In contrast, EEG, a technique in which electrodes are placed on the skull to measure brain signals, provides a cost-effective and easily accessible means of data collection and is widely used in the diagnosis of epilepsy [6]. Patient examinations are generally conducted by neurology specialists to establish an epilepsy diagnosis. However, diagnosing epilepsy based on EEG signal analysis is a time-consuming and labor-intensive process for neurologists [7]. Therefore, the use of automated systems to assist neurologists in the early diagnosis of epilepsy is of paramount importance. Automated analysis of EEG signals for epileptic seizure detection involves feature extraction from EEG signals using classical signal processing techniques, followed by classification of these features using machine learning algorithms [8-10]. With the rapid development of deep learning algorithms, convolution-based end-to-end learning approaches have also proven successful in automatic analysis of EEG signals for early diagnosis of epilepsy [11, 12].

This study presents models for the early diagnosis of epileptic seizures from EEG signals using machine learning and bio-inspired optimization algorithms. The performance of these models has been evaluated through metrics such as accuracy (Acc), precision (Prec), F1-score, and computational cost (C-Cost), and their reliability has been tested using a 10-fold cross-validation method. The rest of the study is organized as follows: Section 2 introduces the techniques used in the proposed model structure for the diagnosis of epileptic seizures. Section 3 presents and discusses the experimental results obtained for the proposed model. The conclusions of the study are highlighted in Section 4.

2. Materials and Methods

This section presents the materials and methods used in the proposed model structure for the diagnosis of epileptic seizures. The framework of the proposed model is shown in Fig. 1. It is clear from Fig. 1 that the model structure includes the following stages: data set preparation, entropy-based feature extraction, optimization-based feature selection, model training and model evaluation. In this study, a model is proposed that uses bio-inspired human learning optimization (HLO) algorithm to select the features that can provide the highest model performance among the extracted features, with the aim of achieving the best model performance at the lowest C-Cost. The aim is to demonstrate that a model with the highest accuracy using the fewest number of features can be effectively used in real-time applications.

2.1. Epileptic Seizure Recognition Dataset

In this study, an open-access dataset from the *Epileptic Seizure Recognition* database [13] was used. This dataset, consisting of EEG signals, was obtained from 500 subjects. EEG recordings of 23,6 seconds duration were collected from each subject. The time series obtained from the recordings were sampled into 4097 data points. To enable segmentation of data points into 1-second segments, the data points were divided into 23 parts.

Consequently, the dataset consists of a total of 11.500 data points, each containing 178 data points representing 1 second (23 segments × 500 subjects). The data were collected using 5 different labels (1, 2, 3, 4, 5).

Looking at the labels, label 1 represents seizure activity. Label 2 corresponds to EEG recordings from the tumor region. Label 3 represents data from a healthy brain region. Label 4 indicates measurements taken with the subject's eyes closed, while label 5 represents recordings made with the subject's eyes open. As can be seen from the dataset, except for the data represented by label 1, the others refer to healthy individuals in the context of epilepsy. Therefore, to enable a binary classification of epilepsy in this study, the data represented by labels 2, 3, 4, and 5 were combined into a single label.

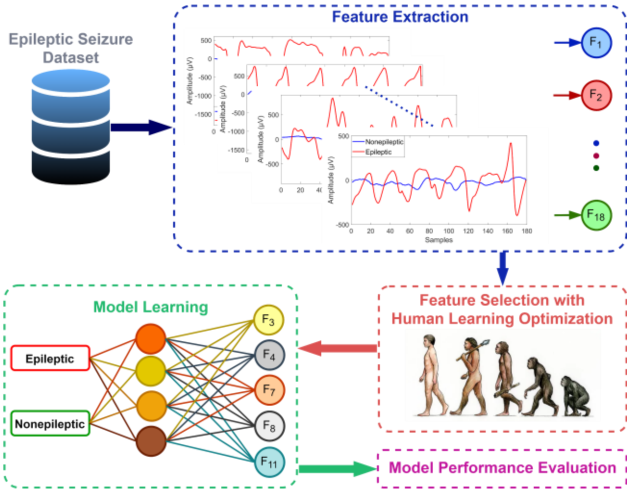

Fig. 1. Framework of the proposed model.

Upon further examination of the dataset, it's noted that each label contains 2300 data points. Therefore, while the dataset includes 2300 epileptic seizure signals, it also contains 9200 signals from healthy individuals. This data distribution can potentially lead to overfitting and biased results in the machine learning model to be trained. Hence, in the experimental study, 2300 normal and 2300 epileptic seizure data points were used, with each label (2, 3, 4, and 5) contributing 575 data points. The study was conducted with a total of 4600 data points. The EEG signals containing epileptic seizures and healthy EEG signals from the dataset are illustrated in Fig. 2.

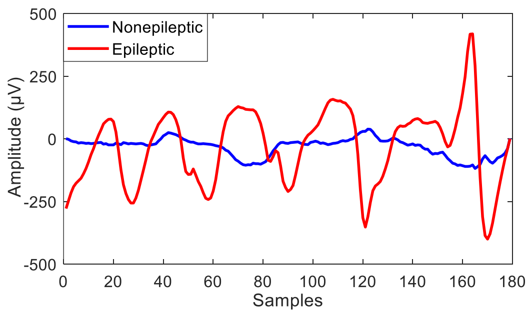

Fig. 2. EEG signals with healthy and epileptic seizure.

2.2. Entropy-based Feature Extraction

Feature extraction refers to the process of transforming raw data into a set of observable features that can be analyzed independently of the original data [8]. In the literature, there are many options for feature extraction, such as statistical, frequency-domain, and entropy-based feature extraction methods [14]. In this study, considering the chaos within the brain signals obtained through EEG, feature extraction based on entropy has been preferred. To effectively deal with the nonlinear dynamics in EEG signals for the diagnosis of epileptic seizures, 18 entropy-based features, as shown in Fig. 3, have been extracted and used in the proposed model structure.

Approximate Entropy	Sample Entropy	Fuzzy Entropy
Kolmogorov Entropy	Permutation Entropy	Conditional Entropy
Distribution Entropy	Spectral Entropy	Dispersion Entropy
Symbolic Dynamic Entropy	Increment Entropy	Cosine Similarity Entropy
Phase Entropy	Slope Entropy	Bubble Entropy
Gridded Distribution Entropy	Entropy of Entropy	Attention Entropy

Fig. 3. Entropy-based features extracted from EEG signals.

2.3. Feature Selection with Optimization Algorithms

Feature selection using optimization algorithms is a technique used to effectively improve the performance of a machine learning model while minimizing its complexity. This approach aims to simplify the model by eliminating features that do not contribute to improving its performance. This process addresses critical issues for real-time model deployment, such as parameter uncertainty, model complexity, and C-Cost. The choice of optimization algorithm for feature selection can significantly affect the efficiency of this process. Therefore, in this study, the HLO algorithm was used for feature selection, as it can adapt to the changing characteristics of the optimization problem during the feature selection process.

A notable advantage of the HLO algorithm is its low computational complexity, which makes it particularly useful for overcoming problems associated with real-time model use. In addition, HLO is less sensitive to initial conditions and parameter settings than many other optimization algorithms. It achieves this by actively avoiding local optima and thus striving to capture global optima in fewer iterations. As a result, this approach facilitates faster selection of the minimum number of features required for optimal model performance [15].

The HLO algorithm is a bio-inspired optimization technique that mimics the learning and problem-solving behaviour of humans. Its core principle is to adaptively search for optimal solutions within a complex problem space. HLO uses a simple mimicry of mechanisms found in human learning processes to facilitate easy implementation. It attempts to find optima using individual learning, social learning, random exploration learning, and re-learning operators. Individual learning refers to self-directed learning, while social learning is defined as learning with the help of others. Random exploration learning explains learning by randomly developing new methods due to a lack of knowledge during the learning process. In cases where relearning with new methods becomes necessary, this is defined as relearning, which prevents the algorithm from getting stuck [16]. Here's a step-by-step explanation of how the HLO algorithm works:

i. Initialization: The algorithm begins by initializing a population of potential solutions. These solutions represent different points in the problem space. The HLO employs a binary coding technique for problem solving. Therefore, each solution in the population is represented in a binary array format, initialized as '0' or '1', depending on whether the information related to the problem is present or not, as represented by

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1M} \\ \vdots & x_{ij} & \vdots \\ x_{1N} & \cdots & x_{NM} \end{bmatrix} \tag{1}$$

with the condition that $x_{ij} \in \{0,1\}$, $1 \leq i \leq N$, $1 \leq j \leq M$, where x_{ij} represents the ith individual, N denotes the number of individuals in the population, and M indicates the number of components contained in each information, the initialization population of the algorithm is obtained randomly using the matrix generated in (1).

ii. Evaluation: Each solution in the population is evaluated against a predefined objective or fitness function. This function quantifies how good each solution is at solving the given problem. The goal is usually to maximize or minimize this objective function (OF).

iii. Learning and Exploration: Inspired by human learning, the HLO algorithm uses a combination of learning and exploration strategies. It maintains a balance between using the best solutions found so far (learning) and exploring new regions of the problem space (exploration).

- *Individual Learning Operator*: The individual learning operator is an operator that enables each individual in the population to solve problems based on their own experience. In this context, experience refers to the information stored in each individual's individual knowledge database (IKD). The IKD is defined as

$$IKD_i = \begin{bmatrix} ik_{i11} & \cdots & ik_{i1M} \\ \vdots & ik_{ipj} & \vdots \\ ik_{iG1} & \cdots & ik_{iGM} \end{bmatrix} \tag{2}$$

with the condition that $1 \leq i \leq N$, $1 \leq p \leq G$, $1 \leq j \leq M$. Here, IKD_i represents the individual knowledge dataset of individual i, G specifies the dimension of individual knowledge, and ik_{ipj} denotes the pth best solution of individual i. Also, p is a random integer that selects which individual from the IKD is used in individual learning.

- *Social Learning Operator*: In order for the HLO algorithm to have effective search capabilities, the social learning operator, when generating a new solution, involves each individual in a probabilistic examination of the information stored in the social knowledge database (SKD), which is structured similarly to the IKD, and copies the bits corresponding to the best solution. The SKD is described as

$$SKD = \begin{bmatrix} sk_{11} & \cdots & sk_{1M} \\ \vdots & sk_{qj} & \vdots \\ sk_{h1} & \cdots & sk_{hM} \end{bmatrix} \tag{3}$$

with the condition that $1 \leq q \leq h$, $1 \leq j \leq M$. Here, h represents the size of the SKD. The newly generated candidate x_{ij} randomly selects one of the best solutions stored in the SKD and copies the corresponding bit.

- *Random Exploration Learning Operator*: In the process of human learning, individuals may not always be able to reproduce their individual knowledge and social knowledge due to various factors such as forgetting or they may attempt to try new strategies to improve their performance. As a result, a random learning situation arises. In the HLO algorithm, this situation is denoted as

$$x_{ij} = RE(0,1) = \begin{cases} 0, rand \leq 0.5 \\ 1, else \end{cases} \tag{4}$$

where $rand$ represents the generation of a random number between 0 and 1.

- *Re-learning Operator*: In HLO, an individual is considered to have encountered a bottleneck if its fitness does not improve over a certain number of generations. In such a scenario, the re-learning operator is triggered. This operator clears the individual's IKD, allowing it to gain new experience and relearn in subsequent generations. This process can help the HLO to escape local optima and achieve improved performance, similar to individuals adopting a new approach when faced with a bottleneck in their own learning experience.

iv. *Implementation of HLO:* The HLO algorithm achieves the discovery of a new solution by dynamically balancing individual learning, social learning, and random exploration learning to a certain degree, mathematically denoted by

$$x_{ij} = \begin{cases} RE(0,1), & 0 \leq rand < pr \\ ik_{ipj}, & pr \leq rand < pi \\ sk_{qj}, & else \end{cases} \tag{5}$$

where pr represents the probability of random exploration learning, the individual learning rate is expressed as ($pi - pr$), and the social learning rate is expressed as ($1 - pi$). When an individual's learning process reaches a bottleneck, the re-learning operator comes into play. It facilitates the individual's re-learning by updating his IKD value independently of his past state. Furthermore, the IKD values are updated if they provide a better solution than the worst solution in the IKD, based on fitness values. The same processes are applied to the SKD values. All these operations continue iteratively until the stopping criteria of the optimization algorithm are satisfied.

In this study, the effectiveness of the HLO algorithm was evaluated by comparing the results obtained with the genetic algorithm (GA) [17]. Both the HLO and the GA algorithms aim to select the minimum number of features from the 18 features extracted using a recursive approach with the k-nearest neighbor technique that provides the highest performance and fastest response time, as measured by the OF below:

$$OF = \alpha \times e_r + \beta \times \left(\frac{Selected\ features\ subsets}{Total\ features\ number} \right) \tag{6}$$

Here, e_r represents the classification error, where $\alpha \epsilon$ [0,1] and $\beta = 1 - \alpha$ denote the importance of classification quality and subset size, respectively. In this study, to achieve the highest performance with minimum feature selection, the values of α and β were set to 0.99 and 0.01, respectively. These defined values for alpha and beta effectively suppress a high classification error rate, allowing a minimum number of features to be selected.

2.4. Building the Classification Model

Classification models are machine learning based approaches that are used to categorize labelled data in a dataset based on the features extracted from it. In the literature, several successful classification models have been used for this purpose, including k-nearest neighbors, random forests and support vector machines [18]. However, fully connected layers at the end of convolutional layers have also demonstrated superior performance in classification problems. Therefore, in this study, a model consisting of 100 neurons in a fully connected layer and a classification layer was used to diagnose epileptic seizures with the features selected by HLO. The Softmax activation function was preferred for the classification process in the classification layer.

3. Empirical Results and Discussions

In this section, we present experimental studies that demonstrate the effectiveness of the proposed model for diagnosing epileptic seizures. The proposed model includes the steps outlined in Fig. 1. In the pre-processing step, min-max normalization was applied to the data, as the signals obtained from the dataset had been cleaned of noise. The entropy-based features shown in Fig. 3 were extracted from the scaled signals. To keep the complexity of the proposed model low and improve the classification performance, the features were selected using the HLO algorithm. The selected features were trained with a classifier composed of fully connected layers. During the model training, the 'learning rate', 'batch size' and 'epoch' values were set to 10^{-2}, 256 and 100, respectively. The dataset was randomly and independently divided into two groups, training (70%) and testing (30%), using the hold-out method. Both models created in the study were validated using a 10-fold cross-validation approach.

All epileptic seizure diagnosis experiments were performed on a personal computer equipped with an Intel Core i7-12700H CPU, a 6 GB NVIDIA GeForce RTX 3060 graphics card and 16 GB RAM. All code was compiled using MATLAB 2022b. When Fig. 4 is analyzed, it can be seen that the value of the OF in feature selection with GA is measured as 0.0540 at the 100th iteration, while the value of the OF in feature selection with HLO algorithm is calculated as 0.0394. This observation indicates a remarkable improvement of 27.03% in the feature selection and in the classification performance of the model achieved by the HLO algorithm compared to the GA. It's also worth noting that the HLO algorithm reached the best solution in the 41st iteration and converged to the optimal solution faster than the GA. In contrast, GA did not converge to the optimal solution even after 99 iterations. It's also clear that the C-Cost of the HLO algorithm is lower than that of the GA. This comparison highlights the superior performance of the HLO algorithm in feature selection and its ability to quickly approach the optimal solution, making it a promising choice for applications in which computational efficiency and effective feature selection are essential.

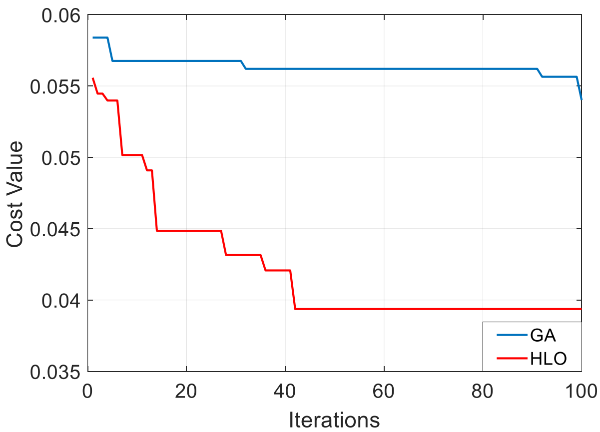

Fig. 4. Cost values of GA and HLO algorithms.

The classification performances for each layer of both the proposed model and the model generated by GA are presented in Tables 1 and 2, respectively. For the proposed model, the average values of Acc, Prec, F1-score, and C-Cost for all layers were measured as %95.28, %95.30, %95.29, and 11.8 seconds, respectively. On the other hand, the model built with GA was tested and the average values of Acc, Prec, F1-score and C-Cost for all layers were measured as %94.65, %94.70, %94.68, and 32.7 seconds respectively.

Table 1. Performance of the model built with HLO.

Model built with HLO	**Selected Features:** *Fuzzy, Kolmogorov, Distribution, Spectral, Increment Entropies*				
		Acc (%)	Prec (%)	F1-score (%)	C-Cost (sec)
	Fold 1	94.78	94.92	94.85	18.2
	Fold 2	96.09	96.10	96.09	12.2
	Fold 3	94.35	94.36	94.35	11.4
	Fold 4	97.61	97.62	97.61	11.2
	Fold 5	96.30	96.31	96.31	11.1
	Fold 6	95.00	95.02	94.99	11.0
	Fold 7	95.00	94.94	94.97	10.9
	Fold 8	94.13	94.15	94.14	10.8
	Fold 9	94.35	94.36	94.35	10.6
	Fold 10	95.22	95.22	95.22	10.6
	Average	**95.28**	**95.30**	**95.29**	**11.8**

Table 2. Performance of the model built with GA.

Model built with GA	**Selected Features:** *Fuzzy, Kolmogorov, Cosine Similarity, Dispersion, Conditional, Distribution, Phase, Slope, Bubble, Gridded Distribution, Attention Entropies and Entropy of Entropy*				
		Acc (%)	Prec (%)	F1-score (%)	C-Cost (sec)
	Fold 1	95.00	95.01	95.00	45.0
	Fold 2	94.13	94.27	94.20	31.2
	Fold 3	94.78	94.79	94.79	30.8
	Fold 4	94.78	94.84	94.81	26.4
	Fold 5	95.00	95.07	95.03	38.5
	Fold 6	93.26	93.26	93.26	35.2
	Fold 7	94.35	94.41	94.38	31.9
	Fold 8	95.87	95.90	95.89	27.5
	Fold 9	93.91	93.95	93.93	29.7
	Fold 10	95.43	95.44	95.44	30.8
	Average	94.65	94.69	94.67	32.7

When Tables 1 and 2 are considered together, it can be seen that both models have classification performances above 90%, with a difference of about 1% between them. However, if we consider the C-Cost criterion, there is a difference of about three times between them. The model built with GA achieved this performance with 12 features and an average computation time of 32.7 seconds. In contrast, the proposed model achieved the same performance using only 5 features with an average cost of 11.8 seconds. It is evident that a model with real-time applicability, low C-Cost and a classification performance of around 95% could be effectively used by neurologists in the diagnosis of epileptic seizures.

4. Conclusions

This study focused on the early diagnosis of epileptic seizures using EEG signals. We used an approach that involved extracting entropy-based features from these EEG signals and then selecting the most relevant features using the HLO algorithm. The resulting model achieved an impressive accuracy rate of 95.28% at a relatively low C-Cost of 11.8 seconds. Several key criteria were used to select the optimal model, including Acc, Prec, F1-score, and C-Cost. Rigorous evaluation showed that the model utilizing the HLO algorithm for feature selection outperformed other models in terms of diagnosing epileptic seizures. Not only did it demonstrate superior performance, but it also effectively addressed the issue of computational efficiency.

The proposed model has significant implications for real-time epilepsy diagnosis. By enabling the timely detection of epileptic seizures, it has the potential to greatly assist neurologists in making early diagnoses and thus initiating timely medical interventions. This research contributes to both the field of epilepsy diagnosis and the wider field of healthcare, highlighting the important role that machine learning and optimization algorithms can play in improving medical diagnosis and patient care.

5. References

[1] E. Beghi, "The Epidemiology of Epilepsy", *Neuroepidemiology*, vol. 54, no. 2, pp.185-191, 2020.

[2] R. D. Thijs, R. Surges, T. J. O'Brien, and J. W. Sander, "Epilepsy in Adults", *The Lancet*, vol. 393, no. 10172, pp. 689-701, 2019.

[3] A. H. Osman and A. A. Alzahrani, "New Approach for Automated Epileptic Disease Diagnosis using an Integrated Self-Organization Map and Radial Basis Function Neural Network Algorithm", *IEEE Access*, vol. 7, pp. 4741-4747, 2018.

[4] I. Ahmad, X. Wang, D. Javeed, P. Kumar, O. W. Samuel, and S. Chen, "A Hybrid Deep Learning Approach for Epileptic Seizure Detection in EEG Signals", *IEEE Journal of Biomedical and Health Informatics*, doi:10.1109-JBHI.2023.3265983, 2023.

[5] S. S. Spencer, "The Relative Contributions of MRI, SPECT, and PET Imaging in Epilepsy", *Epilepsia*, vol. 35, pp. 72-89, 1994.

[6] G. Wiegand, N. Japaridze, K. Gröning, U. Stephani, and N. E. Kadish, "EEG-Findings during Long-Term Treatment with Everolimus in TSC-Associated and Therapy-Resistant Epilepsies in Children", *Seizure*, vol. 103, pp. 101-107, 2022.

[7] M. C. Fernando-Dongas, R. A. Radtke, K. E. Van Landingham, and A. M. Husain, "Characteristics of Valproic Acid Resistant Juvenile Myoclonic Epilepsy", *Seizure*, vol. 9, no. 6, pp. 385-388, 2000.

[8] S. Pattnaik, N. Rout, and S. Sabut, "Machine Learning Approach for Epileptic Seizure Detection using the Tunable-Q Wavelet Transform based Time–Frequency Features", *International Journal of Information Technology*, vol. 14, no. 7, pp. 3495-3505, 2022.

[9] R. Sarić, D. Jokić, N. Beganović, L. G. Pokvić, and A. Badnjević, "FPGA-based Real-Time Epileptic Seizure Classification using Artificial Neural Network", *Biomedical Signal Processing and Control*, vol. 62, 102106, 2020.

[10] R. Ramos-Aguilar, J. A. Olvera-López, I. Olmos-Pineda, and S. Sánchez-Urrieta, "Feature Extraction from EEG Spectrograms for Epileptic Seizure Detection", *Pattern Recognition Letters*, vol. 133, pp. 202-209, 2020.

[11] Y. Li, Y. Liu, W. G. Cui, Y. Z. Guo, H. Huang, and Z. Y. Hu, "Epileptic Seizure Detection in EEG Signals using a Unified Temporal-Spectral Squeeze and Excitation Network", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp. 782-794, 2020.

[12] 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", *Computers in Biology and Medicine*, vol. 100, pp. 270-278, 2018.

[13] Epileptic Seizure Recognition Data Set. URL: https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition.

[14] Y. B. Özçelik and A. Altan, "Overcoming Nonlinear Dynamics in Diabetic Retinopathy Classification: A Robust AI-Based Model with Chaotic Swarm Intelligence Optimization and Recurrent Long Short-Term Memory", *Fractal and Fractional*, vol. 7, no. 8, 598, 2023.

[15] L. Wang, H. Q. Ni, R. X. Yang, M. R. Fei, and W. Ye, "A Simple Human Learning Optimization Algorithm", *Communications in Computer Information Science*, vol. 462, pp. 56–65, 2014.

[16] L. Wang, R. Yang, H. Ni, W. Ye, M. Fei, and P. M. Pardalos, "A Human Learning Optimization Algorithm and Its Application to Multi-Dimensional Knapsack Problems", *Applied Soft Computing*, vol. 34, pp. 736-743, 2015.

[17] S. M. Elsayed, R. A. Sarker, and D. L. Essam, "A New Genetic Algorithm for Solving Optimization Problems", *Engineering Applications of Artificial Intelligence*, vol. 27, pp. 57-69, 2014.

[18] İ. Yağ and A. Altan, "Artificial Intelligence-based Robust Hybrid Algorithm Design and Implementation for Real-Time Detection of Plant Diseases in Agricultural Environments", *Biology*, vol. 11, no. 12, 1732, 2022.