Wavelet ELM-AE Based Data Augmentation and Deep Learning for Efficient Emotion Recognition Using EEG Recordings

Berna Arı¹, Kamran Siddique², Senior Member, IEEE, Ömer F. Alcin³, Muzaffer Aslan⁴, Abdulkadir Şengür¹ and Raja Majid Mehmood², Senior Member, IEEE

¹Department of Electrical and Electronics Engineering, Firat, CO 23000 Turkey
²Information and Communication Technology Department, School of Computing and Data Science, Xiamen University Malaysia, Sepang 43900 Malaysia
³Department of Electrical and Electronics Engineering, Malatya Turgut Ozal, CO 44000 Turkey
⁴Muzaffer Aslan is with the Department of Electrical and Electronics Engineering, Bingol, CO 12000 Turkey

Corresponding author Raja Majid Mehmood (e-mail: rmex07@ieee.org).

This work has partially supported by the Xiamen University Malaysia Research Fund (XMUMRF) (Grant No: XMUMRF/2022-C9/IECE/0035).

ABSTRACT Emotion perception is critical for behavior prediction. There are many ways to capture emotional states by observing the body and copying actions. Physiological markers such as electrophysioencephalography (EEG) have gained popularity, as facial emotions may not always adequately convey true emotion. This study has two main aims. The first is to measure four emotion categories using deep learning architectures and EEG data. The second purpose is to increase the number of samples in the dataset. To this end, a novel data augmentation approach namely the Extreme Learning Machine Wavelet Auto Encoder (ELM-W-AE) is proposed for data augmentation. The proposed data augmentation approach is both simple and faster than the other synthetic data augmentation approaches. For deep architectures, large datasets are important for performance. For this reason, data multiplexing approaches with classical and synthetic methods have become popular recently. The proposed synthetic data augmentation is the ELM-W-AE because of its efficiency and detail reproduction. The ELM-AE structure uses wavelet activation functions such as Gaussian, GgW, Mexican, Meyer, Morlet, and Shannon. Deep convolutional architectures classify EEG signals as images. EEG waves are scalogram using Continuous Wavelet Transform (CWT). The ResNet18 architecture recognizes emotions. The proposed technique uses GAMEMO data collected during gameplay. Each of these states is represented in the GAMEMO data collection. The visual data set created from the signal was divided into two groups as 70% training and 30% testing. ResNet18 has been fine-tuned with augmented photos, training images only. It achieved 99.6% classification accuracy in tests. The proposed method is compared with the other approaches on the same dataset, and an approximately 22% performance improvement is achieved.

INDEX TERMS auto-encoder, data augmentation, deep learning, emotion recognition, emotion measurement.

I. INTRODUCTION Emotion is an abstract term that is described as a reaction to external and internal stimuli and expresses human consciousness. In recent years, emotion detection has been a popular research area. However, the concept of abstraction contained within it has made it difficult to comprehend people's emotional states [1]. When the studies are examined, there is a tendency on two headings in emotion recognition. The first is a discrete model that distinguishes between positive and negative emotions. Joy, anger, anticipation, surprise, disgust, sadness, confidence, and fear are the eight primary emotions in the discrete models [2]. The second one is the dimensional model. The dimensional model has been looked at under Russell and Plutchick model.
The first is Russell’s arousal-valence coordinate system, which divides emotions into four sections on the coordinate plane, with the left side representing negative emotions and the right side representing positive emotions. On the other side, the arousal axis shows the progression of emotions from sedentary to active [3]. The wheel of emotion described by Plutchick is another dimensional paradigm [4]. Emotions are ranked according to their intensity on the emotion wheel. In addition, in this concept, emotions combine to form complex emotions. However, the difficulty of some languages to convey every feeling generated by this wheel calls into question the model’s universality.

While several studies have been conducted on emotion detection based on facial expressions, its reliability has been questioned. The fact that facial expressions can be imitated, meaning that the real emotion felt on the inside is conveyed or silenced differently with gestures and facial expressions, has harmed the studies’ accuracy [5]. This has conducted emotion detection toward physiological signals like EEG and electrocardiogram (ECG) [6-8]. EEG records the electrical activity via electrodes on the scalp in the brain. The monitoring of conscious brain activities obtained from EEG signals and the identification of human emotions has been facilitated using Brain-Computer Interface (BCI) technology [9].

While the researchers have focused on deep learning in emotion recognition, the data sets’ scarcity has become a significant limitation. Today, studies in many disciplines are carried out on data sets obtained from research areas. In some cases, data sets are sufficient for modeling the situation, while sometimes they are insufficient in terms of the number of samples. The number of samples in the data set is significant for many classification methods [10]. This number has a significant impact on classifier generalization capacity and accuracy. The term “data augmentation” is come up to describe methods for creating iterative optimization or sampling algorithms using unobserved data or hidden variables as a result of this need [11]. Various methods for increasing data are available in the literature.

The amount of data available, particularly in medical fields, is limited, and obtaining it is expensive. Furthermore, the amount of data obtained is inadequate and inconsistent for most scientific fields due to factors such as the inability to access previously generated data and the time-consuming data collection process. In addition to these limitations, the amount of data is of great importance in achieving the desired goal in deep architectures, which have increased their popularity day by day. There are several different ways to acquire new data pieces using conventional data augmentation methodologies, such as taking the symmetry of the image based on various axes, cutting out a random sample piece from the image, changing the axes’ location, changing the color ratios in the image, noise addition and so on [12,13]. The model’s propensity to memorize has been avoided thanks to this multiplication/increment. Although these methods can be used to perform operations like object detection, they don’t always produce beneficial results in photos. Recently, some methods such as Variational Auto-Encoder (VAE), Generative Adversarial Nets (GAN), ELM-AE have attracted attention in terms of generating real data [11,14,15].

The key objective of this study is to present a reliable method for EEG-based emotion recognition. The proposed method is based on the conversion of signals to images, data augmentation, and deep learning. Additionally, the approach is predicated on two major scenarios. In the first scenario, signal-to-image conversion and deep learning steps are used to transform the original data. In the alternative scenario, images obtained following the signal-to-image conversion step are synthetically enhanced and then subjected to deep learning steps. Separate scenarios of retesting the data following synthetic enhancement were examined. The CWT method is used to convert the signal to the image. The synthetic data were obtained using the wavelet-based ELM-AE structure, which will be referred to as ELM-W-AE throughout the remainder of the article. In the experiments, the proposed approach is validated against a variety of wavelet functions, including Gaussian, GgW, Mexican, Meyer, Morlet, and Shannon. Six different wavelet functions were used to generate synthetic data, and the effect of wavelet structure on performance was investigated. For classification, deep learning is preferred. The advantage of synthetic data acquisition is that it enables learning and training for the desired deep architectural structure. The proposed method has been validated against the GAMEMO dataset. In the GAMEMO dataset, EEG signals were collected while subjects participated in emotional evaluation games. The GAMEMO dataset is used to detect emotions using the ResNet18 architecture. The ResNet18 model is used in a fine-tuned transfer learning format. The final three layers of the pre-trained ResNet18 model are altered to address our issue.

So, in this study, a novel data augmentation approach namely ELM-W-AE is proposed for data augmentation. The proposed data augmentation approach is both simple and faster than the other synthetic data augmentation approaches. Besides, with proposed data augmentation approach, the proposed method produces an improved accuracy score than the other methods.

The rest of the paper is organized as follows; Section II includes the literature review on emotion recognition and data augmentation. Section III consists of the most commonly used datasets in the literature and the dataset used in the study. Section IV includes a detailed explanation of the proposed method, Section V includes experimental studies, their results, and comparison of its with other studies, Section VI provides for the conclusion of the study.

VOLUME XX, 2017

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
II. RELATED WORKS

The related works are arranged in three sub-sections. The related works about emotion recognition are detailed in the first sub-section. The second sub-section lists AE works used data augmentation. Emotion recognition data sets are briefly described in the final sub-section.

A. Literature Review on Emotion Recognition

The authors of [16] wanted to see if brain signals might be exploited to discern emotions. 28 people were asked to play five-minute games with an EMOTIV EPOC+ wearable 14-channel EEG interface. These games monitored participants’ EEG data for 20 minutes and contained four emotions: boredom, calm, horror, and fear. The discrete wavelet transform was used to analyze the signals’ time-frequency (DWT). Second-order Daubechies filter was used to recover D1-D4 data and A4 approximation coefficients. Feature extraction included detrended fluctuation analysis, Shannon entropy, standard deviation, variance, and zero transitions. The EEG channels were categorized using Support Vector Machine (SVM), K-Nearest Neighbors (k-NN), and multilayer perceptron neural network (MLPNN) classifiers based on positive-negative emotion prediction and arousal-valence dimension condition. The best classification accuracy rates were 75.0 for k-NN, 72.2 for SVM, and 82.2 for MLPNN. In [5], the researchers used a 32-channel EEG equipment to capture signals from 44 subjects to establish a new dataset. The EEG signals were recorded throughout 12 films, three of which were happy, scary, sad, or neutral. The signals are then adjusted to zero mean and unit variance using a DWT. The retrieved parameters such as average amplitude change, absolute square root sum, and root mean square were used to classify the four emotions. The Fp1-F7 channel’s gamma sub-band had the best ELM performance, with 94.7% accuracy. [17] used an online semi-supervised learning method to recognize emotions from EEG signals. 14-channel EEG data from 28 subjects playing four distinct video games were analyzed using the Fourier spectrum. The collected features showed the Evolving Gaussian Fuzzy Classification (eGFC)’s efficiency in real-time learning of EEG data, with 72.2% performance for four category classifications using the arousal-valence method. According to [9], BCI technologies are employed as an interface between sensors and the brain. A Spiking Neural Network (SNN) was used to analyze DEAP data and 60 EEG samples. The method is recommended because the SNN neuron structure is more realistic than the Artificial Neural Network (ANN). On average, 84.6% of the valence mood level was correctly identified using the SNN architecture NeuCube. [18] created an EEG-based emotion identification system using fractal pattern feature extraction. The 14-channel GAMEEMO data collection was decomposed using a fractal design and Tunable Q-factor Wavelet Transform (TQWT). An Iterative Chi-square Selector (IChi2) was utilized for feature extraction. The model was tested using 10-fold cross-validation using linear discriminant analysis (LDA), k-NN, and SVM. The SVM classifier had the greatest accuracy of 99.8%. [19] proposes an arousal-valence-based real-time emotion classification system for four emotional classes. DEAP’s 10-channel EEG recordings were acquired by first dividing them into overlapping intervals of 2-4 seconds duration. With k-NN (k = 3), arousal classification accuracy is 86.8% and valence is 84.1. The study also claims it is more accurate than higher frequency bands, especially the gamma band. The authors used EEG signals from their own GAMEEMO data set in [20]. The data were utilized to distinguish good and negative emotions. The investigation began with determining the signals’ spectral entropy. Then the classifier gets the values. The deep bidirectional long-short term memory (BiLSTM) architecture was employed as a classifier. The approach yielded 76.91 percent accuracy and a 90% ROC score.

B. Literature Review on Data Augmentation

In [21], three alternative methods were used to attempt to tackle the problem of insufficient data in EEG emotion recognition: VAE, GAN, and classical. According to the authors, they achieved the best results with GAN in trials using the DEAP and SEED datasets. They improved performance by over 10% during the test phase of their networks trained with SVM and deep architectures. In [22], they aimed to generate more data to increase the performance of the classifier. They proposed a model that duplicates data from ELM-AE images in their paper. They stated that they chose the auto-encoder approach over other data augmentation strategies because it was simpler and more efficient. They tested their methods on the JAFFE database, which contains Japanese female facial expressions, and looked at the impact of data enhancement on results, using k-NN, SVM, and Convolutional Neural Networks (CNN). They stressed that their approach was a viable alternative for data enhancement tasks and that it produces better results in most cases than other common strategies, according to the findings. In [23], a data enhancement and feature extraction method using a Variational Automatic Encoder (VAE) for acoustic modeling is described. The authors declared that the VAE was a helpful model based on variational Bayesian learning and a deep learning framework. A VAE can generate new information by extracting hidden values of input variables. VAE was a popular method for building images and sentences. A VAE was used to improve speech structure data for acoustic modeling and feature vector extraction from their research. The size of a speech ensemble was doubled by using a VAE system to encode the hidden variables extracted from the original expressions. Latent variables inferred from speech waveforms are said to have concealed "meanings" of the waveforms, allowing them to be used as acoustic properties for automatic speech recognition (ASR). They used a VAE system to show the efficacy of data augmentation and that latent variable-based features can be used in ASR in their research. In [24], a two-stage model is proposed to improve
recognition rate by examining a data set of documents containing Japanese letters. The model's first step was to figure out how to distribute data and fit the shape vectors of the characters on the page, while the second is to generate new examples. The study's VAE structure was created to divide data diversity into regions, create simple examples of in-class multi-modality, and avoid mode reduction. They accomplished this by organizing the VAE model and proposing a gradual and unregulated feature extraction structure for the VAE model. The CNN-based classification network achieved 94.02% efficiency for the non-multiplexed data set, while this rate could be improved to 95.56% for the multiplexed data set. It is viewed as the study's focal point, where the recognition rate was increased by using an enriched data collection.

C. Datasets about Emotion Recognition

With three primary stimuli aspects, it is possible to manage the emotion recognition system, whose research and the application field are expanding by the day. Audio, visual, and audio-visual are the three types. These categories were used to build datasets, and the studies’ accuracy was comparable.

- The Belfast facial expression dataset was developed to explore gender, cultural, and individual variations in emotion interpretation from TV shows and interview recordings [25].
- Their voices are sad, funny, etc. IADS (International Affective Digitalized Sounds) public dataset, which aims to estimate emotion according to hearing sense by labeling it according to its condition, was created [26]. In auditory stimuli, sounds are generally applied to stimulate the person’s emotions by affecting the sensation.
- The HUMAINE dataset has been developed, including various scenarios in terms of emotion recognition, audio-visual recordings, and an expanded version of Belfast [27].
- The MANHOB physiological dataset was developed, including sound signals, mimics, and EEG signals [28].
- A 32-channel ECG and a MAHNOB HCI dataset containing EEG signals, which evaluate the participants' feelings after the movie according to the valence-arousal scale, were created [29].
- The IEMOCAP audio-visual dataset was developed, which communicates emotional states such as happiness, frustration, sadness, disappointment, and neutrality and was collected from participants due to a double session [30].
- The VAM (Vera Am Mittag) dataset was generated using audio-visual recordings, including the participants' natural responses during a TV show [31].
- The DEAP audio-visual dataset was created based on 32-channel valence-arousal scale using music clips in the environment [32].
- The eINTERFACE dataset has been developed, which contains audio-visual recordings from various countries and includes tags for enjoyment, rage, sadness, surprise, disgust, and panic [33].

III. PROPOSED METHOD

In this paper, a novel approach is proposed for EEG-based emotion recognition. Fig. 1 shows the arousal-valence emotion model that is considered. As seen in Fig.1, excited, happy, and pleased emotions are in the region that is indicated by ‘HAPV’. Annoying, angry, and nervous emotions are in the region ‘HANV’ and similarly, sad, bored and sleepy emotions are located in the region ‘LANV’. Lastly, calm, peaceful, and relaxed emotions are in the area ‘LAPV’. Through the arousal axis, the feeling of emotions changes from high to low. Likewise, in the valence axis, the emotions are changed through negative to positive. In addition, the four games labeled boring (B), calm (C), horror (H) and funny (F) in the dataset used correspond to LANV, LAPV, HANV and HAPV, respectively.

FIGURE 1. Arousal-Valence emotion model

In Fig.2, the illustration of the ELM-W-AE method is given. As seen in Fig. 2, the input EEG signals are initially converted to CWT scalogram images. The following formula is used to transform a function x(t) given a mother wavelet ψ(t) through CWT:

\[ x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi \left( \frac{t-b}{a} \right) x(t) dt \]  

where a denotes the scale or dilation parameter, and b is the shifting parameter, which denotes the time information in the transform [34].

The wavelet used for CWT is the analytic Morse wavelet as it has better time-frequency localization. For Morse wavelet symmetry parameter (gamma) and time bandwidth product were kept 3 and 60 respectively. Voices per octave were kept 10 [35].

The analytic Morse wavelet is utilized for CWT because it has greater time-frequency localization [34]. The symmetry parameter (gamma) and time bandwidth product for Morse wavelets were preserved at 3 and 60, respectively. The number of voices per octave was chosen as 18 which is best in case of our EEG emotion recognition experiments. Thus, a scalogram image 222x38252 sized was constructed from an EEG signal.
Then, the scalogram images are normalized and resized to 224×224 for being appropriate for the input of the next building block. Initially, the scalogram images are in grayscale and are then converted to the color images by assigning the grayscale image in the color channels of a new image. After the previous procedure, the dataset is constructed. A data augmentation procedure comes after the data construction. To this end, the ELM-W-AE is considered. Various wavelet kernels are used in ELM-AE architecture. The wavelet functions are briefly explained in section III.B.

After data augmentation, deep transfer learning is used in the classification stage of the proposed method. The pre-trained ResNet18 model, which has 18 layers, is further trained (fine-tuned) in the classification procedure. The flow of the process is given in Fig. 3.

### A. Wavelet Based Extreme Learning Machine Auto-Encoder

The AE is an unsupervised learning system in which the input data are often used as output data [39]. It is made up of two parts: encoder and decoder. The input data are projected to the hidden layer in the encoder section, and an estimate of the input data is obtained in the decoder part. In the AE, the input $X_{NW}$ is initially encoded to a higher-level space and then an approximation of the input $X'_{NW}$ is obtained by using the encoded input $X$. Figure 4 shows the architecture of the ELM-AE structure.

![Wavelet Based Extreme Learning Machine Auto-Encoder Diagram](image)

**FIGURE 4.** An example of EEG to CWT scalogram images and ELM-W-AE data augmentation

### B. ResNet18 Architecture

In image classification, ResNet18, which was introduced for the problem of performance degradation with increasing depth, is frequently preferred [42]. ResNet18 could develop deep network architectures with this property by using residual units depicted in Figure 5.

Assuming that the neural network’s input parameter is $x$ and the target output is $h(x)$, the target output is likely to be extremely complex. In this case, residual $F(x)=H(x)-x$ changes

### Table I

| Wavelet Kernel Type | Function |
|---------------------|----------|
| Morlet $\psi(t) = \cos(1.75t)e^{\frac{it^2}{2}}$ |
| Gaussian $\psi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$ |
| Mexican $c = \frac{3}{\sqrt{7}} \pi^{-\frac{1}{2}} \psi(t) = c(1-t^2)\exp\left(\frac{t^2}{2}\right)$ |
| Shannon $\psi(t) = \frac{\sin(\pi t)-\sin(\pi(t-1/2))}{\pi(t-1/2)}$ |
| Meyer $\psi(t) = 35t^4 - 84t^5 + 70t^6 - 20t^7$ |
| GGW $\psi(t) = \sin(3t) + \sin(0.3t) + \sin(0.03t)$ |

Similar to the ELM algorithm, the output weights $\beta$ of the ELM-W-AE is calculated by using the Moore-Penrose inverse [36,41]:

$$\beta = H'X$$

**FIGURE 3.** ELM-AE architecture representation

By using the wavelet kernel $\psi$, the hidden layer output matrix $H_\psi$ of size $N \times K$ is re-defined as:

$$H_\psi = \begin{bmatrix} \psi(a_1^T x_1 + b_1) & \cdots & \psi(a_K^T x_1 + b_K) \\ \vdots & \ddots & \vdots \\ \psi(a_1^T x_N + b_1) & \cdots & \psi(a_K^T x_N + b_K) \end{bmatrix}_{N \times K}$$ (2)

The Morlet, Gaussian, Mexican, Shannon, Meyer and GGW wavelet activation functions are defined in Table 1 [40,41].
the goal output to $F(x)+x$ to avoid the performance deterioration problem caused by too many convolution layers. This is referred to as a short link. These mentioned linkages, which can perform identity matching by avoiding two or more levels, are defined as [43];

\[ x_{l+1} = f[x_l + F(x_l, k_t)] \]  

(4)

**FIGURE 4. The residual unit**

In this equation, $x_l$ and $x_{l+1}$ represent the input and output of the first residual unit, respectively, the active function $f$, the residual function $F$, and the convolution kernel $k$.

**IV. EXPERIMENTAL WORKS AND RESULTS**

**A. GAMEEMO Dataset**

The GAMEEMO dataset was developed by Alakus et al. using EEG signals collected from participants while playing video games using a wearable-portable system with 14 EEG channels [16]. The GAMEEMO dataset will be used to test the methods used in the analysis to achieve the emotion recognition objective. EEG signals were collected with the EMOTIV EPOC+ Mobile EEG device from 28 students between the ages of 20-27 in the Software Engineering Department of Firat University Faculty of Technology. On the scalp, EEG electrodes were placed in 16 different regions as AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, P3, and P4. Since P3 and P4 are reference electrodes, a 14-channel EEG device is used. The device’s sampling rate was set to 128 Hz, and the signal bandwidth was set to 0.16 - 43 Hz. The participants played four different computer games categorized as boring, calm, horror, and funny for five minutes each, yielding a total of 20 minutes of EEG data from each participant.

The EEG signal is containing 38,252 samples within the period. In addition, the four games labeled boring (B), calm (C), horror (H) and funny (F) in the dataset used correspond to LANV, LAPV, HANV and HAPV, respectively.

**B. Results**

The experimental works were carried out on MATLAB. The 14-channel EMOTIV EPOC+, a wearable and compact EEG unit, was used to capture EEG signals from 28 different subjects. Subjects played four separate video games for five minutes each to catch emotions (boring, quiet, horror, and funny), with a total of 20 minutes of EEG data available for each subject. The participants scored every video game on a scale of arousal and valence. The participants used the SAM type to score each video game on a scale of arousal and valence. During the scalogram image construction, the scale parameter of the CWT is chosen as twelve. The obtained dataset was randomly divided into two parts, where 70% of it was used for training, and the rest 30%, was used for testing the proposed approach and stayed constant in the process of the experiments. The training images were only augmented with ELM-W-AE. Gauss, GgW, Mexican, Meyer, Morlet, and Shannon activation functions were considered in Wavelet ELM-W-AE.

**FIGURE 5. The illustration of the proposed method**

This article has been accepted for publication in IEEE Access. This is the author’s version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3181887
The training images increased by six times, which is including the originals. The ResNet18 was fine-tuned by using the stochastic gradient descent optimizer method (SGDM), where the input batch size, number of epochs, and the initial learning rate parameters were set to 32, 30, and 0.05, respectively. The achievement of the proposed method was evaluated based on the various evaluation metrics, such as accuracy, sensitivity, specificity, precision, F1-Score, Mathew correlation (MCC), and kappa, respectively [44,45].

Also, the receiver operating characteristic (ROC) curve, which is defined as a plot of test sensitivity or true positive rate (TPR) as the y-axis versus its 1-specificity or false positive rate (FPR) as the x-axis, is a useful tool for assessing the quality or performance of medical diagnostic tests. It’s widely used in radiology to assess the performance of many classifiers tests [46].

AUC is a measure of a diagnostic test’s overall performance, defined as the average value of sensitivity for all conceivable values of specificity. AUC can have any value between 0 and 1, with a higher value indicating greater overall diagnostic test performance [47].

The obtained results were represented in Table 2. The columns of Table 2 show the performance metrics, and the rows show the methods that were used. In the first row of Table 2, the results, which were obtained without data augmentation, were given. As seen in Table 2, without data augmentation, 77.66% accuracy, 77.66% sensitivity, 92.55% specificity, 77.96% precision, 0.78 F1-score, 0.70 MCC and 0.40 kappa values were obtained. In the other rows of Table 2, the results with data augmentation were given. As seen in the other rows, various activation functions were examined in the data augmentation case.

### TABLE II

| Method  | Acc. (%) | Sens. (%) | Spec. (%) | Prec. (%) | F1. | MCC | Kappa |
|---------|----------|-----------|-----------|-----------|-----|-----|------|
| Original | 77.66    | 77.66     | 92.55     | 77.96     | 0.78| 0.70| 0.40 |
| Gauss   | 98.72    | 98.72     | 99.57     | 98.71     | 0.99| 0.98| 0.97 |
| GgW     | 99.57    | 99.62     | 99.86     | 99.53     | 0.99| 0.99| 0.99 |
| Mexican | 98.51    | 98.52     | 99.51     | 98.45     | 0.98| 0.98| 0.96 |
| Meyer   | 99.36    | 99.44     | 99.78     | 99.42     | 0.99| 0.99| 0.98 |
| Morlet  | 98.51    | 98.69     | 99.52     | 98.37     | 0.98| 0.98| 0.96 |
| Shannon | 98.30    | 98.37     | 99.4      | 98.28     | 0.98| 0.98| 0.96 |

Acc.= Accuracy, Sens.= Sensitivity, Spec.= Specificity, Prec.= Precision, F1.= F1-Score

When one examined the ELM-W-AE based data augmentation of results, obvious performance improvements were noticeable. The improvement ratio was almost over 20% for all activation functions. The best evaluation scores, 99.574% accuracy, 99.621% sensitivity, 99.862% specificity, 99.53% precision, and 0.994 F1-score, and 0.994 MCC and 0.989 Kappa values obtained for the GgW activation function. Besides, the Meyer activation function produced the second-best evolution scores, where 99.362% accuracy, 99.444% sensitivity, 99.780% specificity, 99.421% precision, and a 0.994 F1-score 0.992 MCC and 0.983 Kappa values were obtained. It is worth mentioning that the Shannon kernel produced the worst evaluation metrics.

In Fig. 6 and 7, the ROC curve presentation and the confusion matrix were given for the original dataset (without data augmentation). In Fig. 6, while the x-axis shows the false-positive rates, the y-axis shows the true-positive rates, respectively. Each ROC curve shows an emotion where different colors were used.

![Figure 6. The ROC curves for original data](image)

As seen in Fig. 6, all ROC curves raised through to the 0.8 true positive rate value in the 0-0.1 range of the false positive rate. Then, they went to the one true positive rate value when the false positive rate was one.

As mentioned earlier, Fig. 7 shows the confusion matrix for the original dataset. For the confusion matrix, the rows show the true class, and the columns show the predicted class. As seen in Figure 7, 92, 92, 93, and 88, test samples from the HANV, HAPV, LANV, and LAPV classes were correctly predicted. In addition, 27, 28, 34, and 16 test samples were wrongly predicted for the HANV, HAPV, LANV, and LAPV classes, respectively.

ROC curves for the HANV, HAPV, LANV, and LAPV classes are given in Fig. 8. The GgW activation function was used in ELM-W-AE. Besides, the confusion matrix for GgW activation function was given in Fig. 9. The illustrations for GgW activation function were given as that activation function yielded the best performance among the other activation functions. When Fig. 8 was examined, it was seen the false positive rate was around zero, while the curves increased by one true positive rate for all four classes. This shows an huge area under ROC, which means high performance. It was observed that the ROC curves immediately rose along the true positive velocity axis for all classes, and almost all classes were classified with 100%...
accuracy, creating a nearly 90-degree gap with the false-positive velocity axis.

![Confusion Matrix for Original Images](image1.png)

**FIGURE 7.** The confusion matrix for original images

![ROC Curve for GgW Activation Function](image2.png)

**FIGURE 8.** The ROC curves for data augmentation with GgW activation function

![Confusion Matrix for Data Augmentation with GgW Activation Function](image3.png)

**FIGURE 9.** The confusion matrix for data augmentation with GgW activation function

When the confusion matrix that was given in Fig. 9 was analyzed, it was seen that HANV and HAPV classes were classified with 100% accuracy scores. Besides, for LANV and LAPV classes only one samples were wrongly classified.

In Tables 3-9, the evaluation metrics for each class for original and the examined activation functions were given. While Table 3 shows the results for the original dataset, Tables 4-9 show the performance evaluation metrics for an individual class for Gauss, GgW, Mexican, Meyer, Morlet, and Shannon activation functions, respectively. In Table 3, the results for the original dataset were given. As seen in Table 3, 78.63%, 77.97%, 78.81%, and 75.21% correct classification rates were obtained for HANV, HAPV, LANV, and LAPV classes, respectively. Table 3 also gives the other evaluation metrics for the mentioned individual classes.

| CLASS | Acc. (%) | Sens. (%) | Spec. (%) | Prec. (%) | F1 | MCC | Kappa |
|-------|----------|-----------|-----------|-----------|----|-----|-------|
| HANV  | 78.63    | 78.63     | 92.35     | 77.31     | 0.78 | 0.71 | 0.53  |
| HAPV  | 77.97    | 77.97     | 92.05     | 76.67     | 0.77 | 0.70 | 0.53  |
| LANV  | 78.81    | 78.81     | 90.34     | 73.23     | 0.76 | 0.68 | 0.52  |
| LAPV  | 75.21    | 75.21     | 95.47     | 84.62     | 0.79 | 0.74 | 0.56  |

Table 4 indicated the evaluation scores for each class for the Gauss activation function. As seen in Table 4, data augmentation highly improved the obtained results, where 100%, 97.83%, 97.87% and 99.19% correct classification rates were obtained for HANV, HAPV, LANV, and LAPV classes. The other evaluation metrics were also increased when the ELM-W-AE based data augmentation was considered.

| CLASS | Acc. (%) | Sens. (%) | Spec. (%) | Prec. (%) | F1 | MCC | Kappa |
|-------|----------|-----------|-----------|-----------|----|-----|-------|
| HANV  | 100.0    | 100.0     | 99.44     | 98.26     | 0.99 | 0.99 | 0.52  |
| HAPV  | 97.83    | 97.83     | 99.74     | 98.90     | 0.98 | 0.98 | 0.61  |
| LANV  | 97.87    | 97.87     | 99.69     | 99.28     | 0.99 | 0.98 | 0.41  |
| LAPV  | 99.19    | 99.19     | 99.42     | 98.40     | 0.99 | 0.98 | 0.47  |

Table 5 indicated the evaluation scores for each class for the GgW activation function. As seen in Table 5, using the GgW activation function, the results of the original dataset were highly improved. Besides, the GgW activation function produced better results than the Gauss activation function. 100%, 100%, 99.29%, and 99.19% correct classification rates were obtained for HANV, HAPV, LANV, and LAPV classes.
Table 6 indicated the evaluation scores for each class for the Mexican activation function. Using Mexican activation function, 100%, 97.82%, 97.87%, and 98.39% correct classification rates were obtained for HANV, HAPV, LANV, and LAPV classes. The results obtained by the Mexican activation function were also worse than the achievements of the GgW activation function.

Table 7 shows the evaluation scores for each class for the Meyer activation function. The Meyer activation function produced similar scores to the GgW activation function where 100%, 100%, 98.58%, and 99.19% correct classification rates were obtained for HANV, HAPV, LANV, and LAPV classes.

Table 8 shows the evaluation scores for each class for the Morlet activation function. With the Morlet activation function 99.12%, 100%, 96.45%, and 99.19%, correct classification rates were obtained for HANV, HAPV, LANV, and LAPV classes.

Lastly, Table 9 shows the evaluation scores for each class for the Shannon activation function. By using the Shannon activation function, 99.12%, 98.91%, 97.8% and 97.58% correct classification rates were obtained for HANV, HAPV, LANV, and LAPV classes.

C. Comparison with Other Methods
Existing studies using the same dataset have been compared to our suggested technique in this section. Alakus et al. [16] proposed the GAMEEMO dataset used in this study. We compared three research publications utilizing the data set with our technique [16, 18, 20]. Table 12 shows performance comparison results.

Alakus et al. [16] was investigated classification accuracy at the channel level and obtained that the average accuracy rates for k-NN, SVM, and MLPNN were 75.0%, 72.2%, and 82.2%, respectively. Tuncer et al. reported an average accuracy of 98.3% for k-NN, 87.2% for LDA, and 98.9% for SVM in their channel-based research. Alakus and Turkoglu said 76.9% success with the

\[ \text{TABLE V} \]

**INDIVIDUAL CLASS RESULTS FOR ELM-W (GGW) –AE IMAGES**

| Class | Acc. (%) | Sens. (%) | Spec. (%) | Prec. (%) | F1. (%) | MC | Kapp |
|-------|----------|-----------|-----------|-----------|--------|----|-----|
| HANV  | 99.12    | 99.12     | 99.44     | 98.25     | 0.98   | 0.98| 0.52 |
| HAPV  | 100.0    | 100.0     | 99.27     | 96.84     | 0.98   | 0.98| 0.60 |
| LANV  | 96.45    | 96.45     | 100.0     | 100.0     | 0.98   | 0.98| 0.42 |
| LAPV  | 99.19    | 99.19     | 99.42     | 98.40     | 0.98   | 0.98| 0.47 |

\[ \text{TABLE VI} \]

**INDIVIDUAL CLASS RESULTS FOR ELM-W (MEXICAN) –AE IMAGES**

| Class | Acc. (%) | Sens. (%) | Spec. (%) | Prec. (%) | F1. (%) | MC | Kapp |
|-------|----------|-----------|-----------|-----------|--------|----|-----|
| HANV  | 99.70    | 99.70     | 99.70     | 99.70     | 0.98   | 0.98| 0.61 |
| HAPV  | 99.27    | 99.27     | 99.27     | 99.27     | 0.98   | 0.98| 0.61 |
| LANV  | 98.94    | 98.94     | 98.94     | 98.94     | 0.98   | 0.98| 0.61 |
| LAPV  | 99.41    | 99.41     | 99.41     | 99.41     | 0.98   | 0.98| 0.61 |

\[ \text{TABLE VII} \]

**INDIVIDUAL CLASS RESULTS FOR ELM-W (MEYER) –AE IMAGES**

| Class | Acc. (%) | Sens. (%) | Spec. (%) | Prec. (%) | F1. (%) | MC | Kapp |
|-------|----------|-----------|-----------|-----------|--------|----|-----|
| HANV  | 100.0    | 100.0     | 100.0     | 100.0     | 1.00   | 1.00| 0.52 |
| HAPV  | 100.0    | 100.0     | 100.0     | 100.0     | 1.00   | 1.00| 0.61 |
| LANV  | 98.58    | 98.58     | 99.69     | 99.29     | 0.98   | 0.98| 0.41 |
| LAPV  | 99.19    | 99.19     | 99.42     | 98.40     | 0.98   | 0.98| 0.47 |

\[ \text{TABLE VIII} \]

**INDIVIDUAL CLASS RESULTS FOR ELM-W (MORLET) –AE IMAGES**

| Class | Acc. (%) | Sens. (%) | Spec. (%) | Prec. (%) | F1. (%) | MC | Kapp |
|-------|----------|-----------|-----------|-----------|--------|----|-----|
| HANV  | 99.12    | 99.12     | 99.44     | 98.25     | 0.98   | 0.98| 0.52 |
| HAPV  | 100.0    | 100.0     | 99.27     | 96.84     | 0.98   | 0.98| 0.60 |
| LANV  | 96.45    | 96.45     | 100.0     | 100.0     | 0.98   | 0.98| 0.42 |
| LAPV  | 99.19    | 99.19     | 99.42     | 98.40     | 0.98   | 0.98| 0.47 |

Table 9 shows the evaluation scores for each class for the Shannon activation function. By using the Shannon activation function, 99.12%, 98.91%, 97.8% and 97.58% correct classification rates were obtained for HANV, HAPV, LANV, and LAPV classes.
individually, yielded a classification success rate of 99.6% using the ResNet18 design.

V. CONCLUSIONS

In this paper, a novel approach was proposed for data augmentation. The proposed method was based on the ELM-W-AE. Various activation functions were examined for the EEG-based emotion classification. The following conclusions are inferred from the experimental works.

1-) Initially, it was observed that the data augmentation was quite effective in EEG-based emotion classification. With data augmentation, almost 20% improvement in accuracy score was observed.

2-) When the achievements of the various activation functions were examined in the ELM-W-AE structure, it was seen that GgW activation function produced the best evaluation scores.

3-) Except Morlet and Shannon activation functions, for all other activation functions, the HANV class was classified with a 100% accuracy score. As is seen that the best emotion prediction is for the HANV class, which includes annoying, angry, and nervous emotions. It is seen that the next best class prediction is in the HAPV class, which includes excited, happy, and pleased emotions.

REFERENCES

[1] K. Michalopoulos, N. Bourbakis, Application of multiscale entropy on EEG signals for emotion detection, Proceedings – 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (2017) 341–344.

[2] P. Ekman, An argument for basic emotions, Cogn. Emot. 6 (34) (1992) 169–200.

[3] J.A. Russell, Culture and the categorization of emotions, Psychol. Bull. 110 (3) (1991) 426–450.

[4] R. Plutchik, The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice, Am. Sci. 89 (4) (2001) 344–350.

[5] Seal, A., Reddy, P. P. N., Chaitanya, P., Meghana, A., Jahnavi, K., Kreckar, O., & Hudak, R. (2020). An EEG Database and Its Initial Benchmark Emotion Classification Performance. Computational and mathematical methods in medicine, 2020.

[6] S. Katsigianniadis N. Ramzan, ‘Dreamer: a database for emotion recognition through EEG and ecg signals from wireless low-cost off-the-shelf devices,” IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 1, pp. 98–107, 2018.

[7] Mehnoood, R. M., Yang, H. J., & Kim, S. H. (2020). Children emotion regulation: Development of neural marker by investigating human brain signals. IEEE Transactions on Instrumentation and Measurement, 70, 1-11.

[8] Kim, S. H., Yang, H. J., Nguen, N. A. T., Mehnoood, R. M., & Lee, S. W. (2020). Parameter estimation using unscented Kalmam filter on the gray-box model for dynamic EEG system modeling. IEEE Transactions on Instrumentation and Measurement, 69(9), 6175-6185.

[9] Al-Nafjan, A., Alharthi, K., & Kurdi, H. (2020). Lightweight Building of an Electroencephalogram-Based Emotion Detection System. Brain Sciences, 10(11), 781.

[10] Najafabadi, M. M., Villanustre, F., Khooshgoftaar, T. M., Seliya, N., Wald, R., & Mahemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of big data, 2(1), 1-21.

[11] Kasun, I. L. C, H M Zhou, G B Huang, and C M Yong. “Representational Learning with ELMs for Big Data.” IEEE Intelligent Systems 28(6) (2013): 31–34.

[12] Wang, F., Zhong, S. H., Peng, J., Jiang, J., & Liu, Y. (2018, February). Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. In International Conference on Multimedia Modeling (pp. 82-93).

[13] S. Hussein, R. Gillies, K. Cao, Q. Song, and U. Bagic, "Tumornet:Lung nodule characterization using multi-view convolutional neural network with gaussian process," in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), April 2017, pp. 1007–1010.

[14] Huang, W. N., Zhang, Y., & Glass, J. (2017, December). Unsupervised domain adaptation for robust speech recognition via variational autoencoder-based data augmentation. In 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (pp. 16-23).

[15] Frid-Aidar, M., Klang, E., Amital, M., Goldberger, J., & Greenspan, H. (2018, April). Synthetic data augmentation using GAN for improved liver lesion classification. In 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) (pp. 289-293).

[16] T. B. Alakus, M. Gonen, and I. Turkoglu, "Database for an emotion recognition system based on eeg signals and various computer games–GAMEEMO," Biomedical Signal Processing and Control, vol. 60, p. 101951, 2020.

[17] Leite, D., Frigeri Jr Jr, V., & Medeiros, R. (2021). Adaptive Gaussian Fuzzy Classifier for Real-Time Emotion Recognition in Computer Games. arXiv preprint arXiv:2103.05489.

[18] Tuncer, T., Dogan, S., & Subasi, A. (2021). A new fractal pattern feature generation function based emotion recognition method using EEG. Chaos, Solitons & Fractals, 144, 110671.

[19] Mohammadi, Z., Frounchi, J., & Amirli, M. (2017). Wavelet-based emotion recognition system using EEG signal. Neural Computing and Applications, 28(8), 1895-1900.

[20] Alakus, T. B., & Turkoglu, I. (2020). Emotion recognition with deep learning using GAMEEMO data set. Electronics Letters, 56(25), 1364-1367.

[21] Luo, Y., Zhu, L. Z., Wan, Z. Y., & Lu, B. L. (2020). Data augmentation for enhancing EEG-based emotion recognition with deep generative models. Journal of Neural Engineering, 17(5), 056021.

[22] Ferreira, J., Ferro, M., Fernandez, B., Valenca, M., Bastos-Filho, C., & Barros, P. (2017, November). Extreme learning machine autoencoder for emotion recognition. In 2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (pp. 1-6).

[23] Nishizaki, H. (2017, December). Data augmentation and feature extraction using variational autoencoder for acoustic modeling. In 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 1222-1227). IEEE.

[24] Cao, G., & Kamata, S. I. (2019, September). Data augmentation for historical documents via cascade variational auto-encoder. In 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) (pp. 340-345).

[25] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, “Emotional speech: towards a new generation of databases,” Speech Communication, vol. 40, no. 1-2, pp. 33–60, 2003.

[26] A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, "Emotion recognition and its applications,” in Human-Computer Systems Interaction: Backgrounds and Applications 3, 60, 2003.

[27] P. Petta, C. Pelachaud, and R. Cowie, Emotion-Oriented Sys- tems: The HUMAINE Handbook, Springer, 2011.

[28] S. Petridis, B. Martínez, M. Pantic, The MANHOB laughter database, Image Vis. Comput. 31 (2) (2013) 186–202.

[29] M. Soyleymani, J. Lichtenaue, T. Pan, M. Pantic, A multimodal database for affect recognition and implicit tagging, IEEE Trans. Affect. Comput. 3 (1) (2012) 1–14.
Abdulkadir Şengür received the B.Sc. degree in electronics and computers education, the M.Sc. degree in electronics education, and the Ph.D. degree in electrical and electronics engineering from Firat University, Turkey, in 1999, 2003, and 2006, respectively. He became a Research Assistant at the Technical Education Faculty, Firat University, in February 2001. He is currently an Associate Professor with the Electrical Engineering Department, Malatya Turgut Ozal University. His research interests include signal processing, machine learning, medical image processing, and deep learning.

KAMRAN SIDDIQUE (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Dongguk University, South Korea. He is currently an Associate Professor with Xiamen University Malaysia. His research interests include big data, cybersecurity, machine learning, the IoT, and cloud computing.

Ömer Faruk Alçin received a bachelor’s degree in Electronics and Computers Education, a master’s degree in Electronics Education, a Ph.D. degree in Electrical- Electronics Engineering from Firat University, Turkey, in 2007, 2011, and 2015, respectively. He became a research assistant in the Technical Education Faculty of Firat University in February 2009. He is currently an Associate Professor with the Electrical Engineering Department, Malatya Turgut Ozal University. His research interests include signal processing, machine learning, medical image processing, and deep learning.

Berna Ari received the B.Sc. degree in computer technology and M.Sc. degree in electrical and electronics engineering from Firat University, Turkey, in 2013 and in 2017, respectively. She is currently an Ph.D. student in electrical and electronics engineering from Firat University. Her main research interests include data augmentation, machine learning, artificial intelligence.

Abdulkadir Şengür received the B.Sc. degree in electronics and computers education, the M.Sc. degree in electronics education, and the Ph.D. degree in electrical and electronics engineering from Firat University, Turkey, in 1999, 2003, and 2006, respectively. He became a Research Assistant at the Technical Education Faculty, Firat University, in February 2001. He is currently a Professor with the Technology Faculty, Firat University. His research interests include signal processing, image segmentation, pattern recognition, medical image processing, and computer vision.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1), 1-13.

Allouche, O., Tsour, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, 43(6), 1223-1232.

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine*, 4(2), 627.

Kumar, R., & Indrany, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian pediatrics*, 48(4), 277-287.

Huang, G. B., Zhu, Q. Y., and Siew, C. K., (2004). Extreme learning machine autoencoders. *IEEE Trans. Neural Networks*, 17(1), 145-149.
Muzaffer Aslan, received the B.Sc. degree in Electrics and Electronics from the Gazi University, Turkey, in 1993, and M.Sc. degree in Electronics Education, from the Firat University, Turkey, in 2004, and the Ph.D. degree in Electrical and Electronics Engineering, from the Firat University, Turkey, in 2016. From 1993 to 2018, He worked as a teacher in the Department of Electrical and Electronics, Technical Anatolian High Schools, Ministry of National Education. He is currently an Assistant Professor in the Engineering and Architecture Faculty of Bingol University. His research interests include pattern recognition, medical image processing, and computer vision.

Raja Majid Mehmood is an Assistant Professor in the School of Computing and Data Science, Information and Communication Technology Department, Xiamen University Malaysia, Sepang, Malaysia. Dr. Raja is an IEEE Senior Member. He served as a research professor in the Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea. He was Lecturer in the Software Engineering Department at King Saud University, Saudi Arabia. He obtained his Ph.D. in Computer Engineering from the Division of Computer Science and Engineering, Chonbuk National University, South Korea. He obtained his M.S. in Software Technology from Linnaeus University, Sweden. He received his B.S. in Computer Science from Gomal University, Pakistan. He has authored more than 50 research papers and is supervising or co-supervising several graduate (MS/PhD) students. His main research interests include affective computing, brain computer interfaces, information visualization, image processing, pattern recognition, multi-task scheduling, and software engineering. Dr. Raja has served as a reviewer of various international journals including IEEE Communications Magazine, IEEE Transactions On Affective Computing, Int. J. of Information Technology and Decision Making, IEEE Access, Multimedia Tools and Applications, ACM Transactions on Multimedia Computing Communications and Applications, Journal of King Saud University - Computer and Information Sciences, Biomedical Engineering (BME), IEEE Internet of Things Journal, IEEE Transactions on Systems, Man and Cybernetics and so on.