



GameEmo-CapsNet: Emotion Recognition from Single-Channel EEG Signals Using the 1D Capsule Networks

Suat Toraman¹, Ömer Osman Dursun^{2*}

¹ Air Traffic Control Department, School of Aviation, Firat University, Elazig 23119, Turkey

² Aircraft Electric-Electronics Department, School of Aviation, Firat University, Elazig 23119, Turkey

Corresponding Author Email: oodursun@firat.edu.tr

<https://doi.org/10.18280/ts.380612>

ABSTRACT

Received: 3 September 2021

Accepted: 28 November 2021

Keywords:

emotion estimation, EEG, fusion, deep learning, capsule networks

Human emotion recognition with machine learning methods through electroencephalographic (EEG) signals has become a highly interesting subject for researchers. Although it is simple to define emotions that can be expressed physically such as speech, facial expressions, and gestures, it is more difficult to define psychological emotions that are expressed internally. The most important stimuli in revealing inner emotions are aural and visual stimuli. In this study, EEG signals using both aural and visual stimuli were examined and emotions were evaluated in both binary and multi-class emotion recognitions models. A general emotion recognition model was proposed for non-subject-based classification. Unlike previous studies, a subject-based testing was carried out for the first time on the GAMEEMO dataset. Capsule Networks, a new neural network model, has been developed for binary and multi-class emotion recognition. In the proposed method, a novel fusion strategy was introduced for binary-class emotion recognition and the model was tested using the GAMEEMO dataset. Binary-class emotion recognition achieved a classification accuracy which was 10% better than the classification performance achieved in other studies in the literature. Based on these findings, we suggest that the proposed method will bring a different perspective to emotion recognition.

1. INTRODUCTION

Machine learning plays an important role in human life. Artificial intelligence (AI) applications based on machine learning are used in many engineering and medical fields such as disease treatment, healthcare, nuclear industry, and robotics [1-4]. Additionally, AI applications that imitate human thinking and behavior also contribute to people's various decision-making processes. More to the point, emotions, which are abstract concepts, can be interpreted more easily thanks to AI technologies. Emotions occur in two ways: physical and psychological. While emotions such as gestures and facial expressions that are reflected in the external environment are reflected physically, emotions such as happiness, sadness, boredom, and fear of the person's inner world emerge psychologically. Moreover, physical emotions such as speech, facial expression, and body language can be defined in a simple way, whereas it is more difficult to define emotions that arise psychologically. Electroencephalography (EEG) signals are widely used to describe emotions expressed internally or psychologically [5]. These signals are obtained from the brain in the form of electrical waves and are used to make sense of emotions with the help of machine learning techniques. There are three types of stimuli in emotion recognition applications using EEG signals, including aural, visual, and aural-visual stimuli. Sound in aural stimuli and various pictures in visual stimuli are used as external stimuli that allow the subject's senses to be revealed. Studies show that emotions are revealed more effectively when subjects are stimulated with aural-visual stimuli [6]. Emotion recognition

models in the literature are divided into two types: discrete-category and dimensional. Ekman et al. suggested many different categories of emotions, one of which included six basic emotions [7]. Plutchik's wheel of emotion suggestion and Russel's arousal-valence scale showed that emotion should be evaluated dimensionally, not discretely [8, 9]. Figure 1 shows the dimensional emotion model.

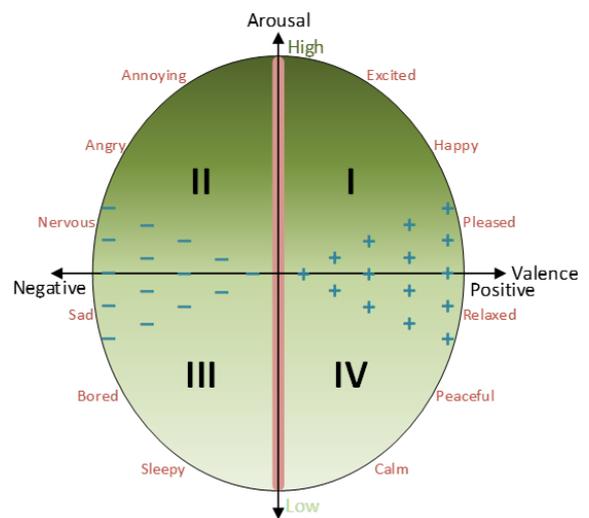


Figure 1. Dimensional emotion model

According to Figure 1, emotion can be expressed in two dimensions: valence (horizontal axis) and arousal (vertical

axis), whereby the arousal axis is expressed as low and high, while the valence axis is expressed as negative and positive. In this model, four zones are considered. The first zone includes emotions of high arousal positive valence (HAPV): pleased, happy, and excited. In the second zone, there are emotions with high arousal negative valence (HANV), such as nervous, angry, and annoying. In the third zone, there are emotions with low arousal and negative valence (LANV): sad, bored, and sleepy. Finally, the fourth zone consists of low arousal and positive valence (LAPV): relaxed, peaceful, and calm [10].

In the proposed study is considered emotions, both binary-class as negative and positive, and multi-class as LAPV, HAPV, LANV, and HANV.

1.1 Related work

The EEG signals used in emotion recognition have a non-linear structure. Non-linear feature extraction techniques are used quite frequently when processing EEG signals. Salankar et al. used the DEAP (a Database for Emotion Analysis using Physiological Signals) dataset, in which they divided the EEG signals with the Empirical Mode Decomposition (EMD) method and extracted the features during the pre-treatment process. The extracted features were evaluated in binary and multi-class classifications using support vector machine (SVM) and Multilayer Perceptron classifiers, of which Multilayer Perceptron achieved a classification performance of 100% [11]. Tan et al. performed emotion recognition based on spiking neural network (SNN) using spatial-temporal EEG signals. The authors used DEAP and MANHOB-HCI databases consisting of facial videos. Binary classification was made for arousal (high-low) and valence (negative-positive) and the classification of arousal provided an accuracy of 78.97% with DEAP and 79.39% with MANHOB-HCI, while the classification of valence provided an accuracy of 67.76% with DEAP and 72.12% with MANHOB-HCI [12]. Tuncer et al. [13] presented a new fractal pattern feature generation using GAMEEMO database for feature extraction by using the features extracted from EEG signals and classified the features with linear discriminant analysis (LDA), k-nearest neighbor (k-NN), and SVM. Of these, SVM showed the best classification accuracy (99.82%). Nawaz et al. [14] employed a three-dimensional (3D) emotion recognition model and recorded the EEG signals of 1-minute-long videos watched by the participants. Statistical properties such as power, entropy, and fractal dimension were obtained from EEG signals and then were classified with SVM, k-NN and Decision Tree (DT) using relief-based algorithms and principal component analysis (PCA). The accuracy of classification was 77.62%, 78.96%, and 77.60% for PCA, SVM, and DT, respectively. Additionally, the authors suggested that PCA is an effective feature selection technique.

Traditional feature extraction methods have various limitations. For instance, the experience of the expert is highly important in the feature extraction stages. In addition, the selected features or the effect of these features on the classification are also remarkably important. With the development of hardware technologies, deep learning, which is a different machine learning method, has emerged as a popular technique for automatic feature extraction from raw data. In this method, data that are not used in traditional feature extraction methods are used as part of the performance. The most widely applied architecture in deep learning is convolutional neural networks (CNN) architecture. Among

the studies using CNN architecture, Er et al. [15] performed two-channel multi-class discrete emotion recognition from EEG signals using pre-trained deep learning networks: AlexNet and VGG16. The authors obtained EEG signals from nine participants who regularly listened to Turkish Art, Turkish Folk, Turkish Pop, and Turkish Jazz Music. The results indicated that VGG16 showed the highest classification performance on the beta frequency channel, with an accuracy of 73.28%. In a study by Yin et al., a fusion model of long short-term memory (LSTM) and graph convolutional neural networks (GCNN) was used for emotion recognition. A binary classification of arousal-valence was made by using the DEAP dataset comparing SVM, DT, random forest, and GCNN. The subject-dependent average classification accuracy was found to be 90.54% for valence and 90.60% for arousal, while the subject-independent average classification accuracy was found to be 84.81% for valence and 85.27% for arousal [16]. Wang et al. applied electrode-frequency distribution maps (EFDMs) with short-time Fourier transform (STFT) to EEG signals in the SEED (SJTU Emotion EEG) and DEAP datasets. The results were used as an input to CNN and the average classification accuracy was found to be 90.59% with SEED and 82.84% with DEAP [17]. Chen et al. [3] transformed one-dimensional vector EEG vector sequences into two-dimensional mesh-like matrix sequences by the EEG signals in DEAP dataset for emotion recognition. The authors proposed a progressive hybrid convolution recurrent neural network and a parallel hybrid convolution recurrent neural network and obtained 93% accuracy using both hybrid convolutional recurrent neural networks. Cui et al. [18] proposed the regional-asymmetric convolutional neural network (RACNN) and used an asymmetric difference layer for feature extraction. The model was tested with the DEAP and DREAMER datasets and provided an accuracy of over 95% for both datasets. Wei et al. [19] applied the simple recurrent units network (SRU) and the ensemble learning method based on recurrent neural networks (RNN) using the SEED dataset for the EEG signals. The authors divided the EEG signals into five sub-bands and performed emotion recognition with 78.5% accuracy using SRU and lower EEG bands. Sabour et al. [20] proposed a new neural network model named capsule networks to overcome the drawback of CNN architectures in performing recognition without using the location and orientation information of objects.

Thus, in this study, a new capsule network architecture was proposed for GAMEEMO, a new audio-visual stimuli dataset. In the suggested method, EEG signals expressing positive and negative emotions were classified as binary and multi-class. Furthermore, a novel fusion strategy was introduced for the two-class emotion recognition problem. According to the results, it was shown that the presented method is an effective method for binary and multi-class emotion recognition. The flowchart of the proposed method is shown in Figure 2.

1.2 Motivation

Over the last decade, AI applications have gained the ability to examine much larger datasets through novel hardware architectures that have emerged in parallel with the developments in the game industry. As a result of these developments, deep learning architectures, which are algorithms motivated by AI and are used for analyzing raw information or information fusion, have become highly popular and remarkably effective in understanding,

recognizing, and analyzing emotions. Human emotions can be understood from physical movements such as facial expressions, speech, and gestures. In these physical movements, however, people can intentionally or unintentionally hide their true emotions. In psychological signals (i.e. EEG signals), human emotions can be defined as more objective and reliable and these signals are more sensitive and also can respond to emotional situations in real

time. Therefore, EEG signals can be more effective in determining people’s emotional states. Moreover, EEG-based emotion recognition has recently attracted the attention of many researchers, and many techniques have been proposed for this subject. In the present study, an emotion recognition method is proposed for determining the mood of gamers using a single channel EEG signal. The emotions of the individuals were examined in binary and multi-class emotion recognition.

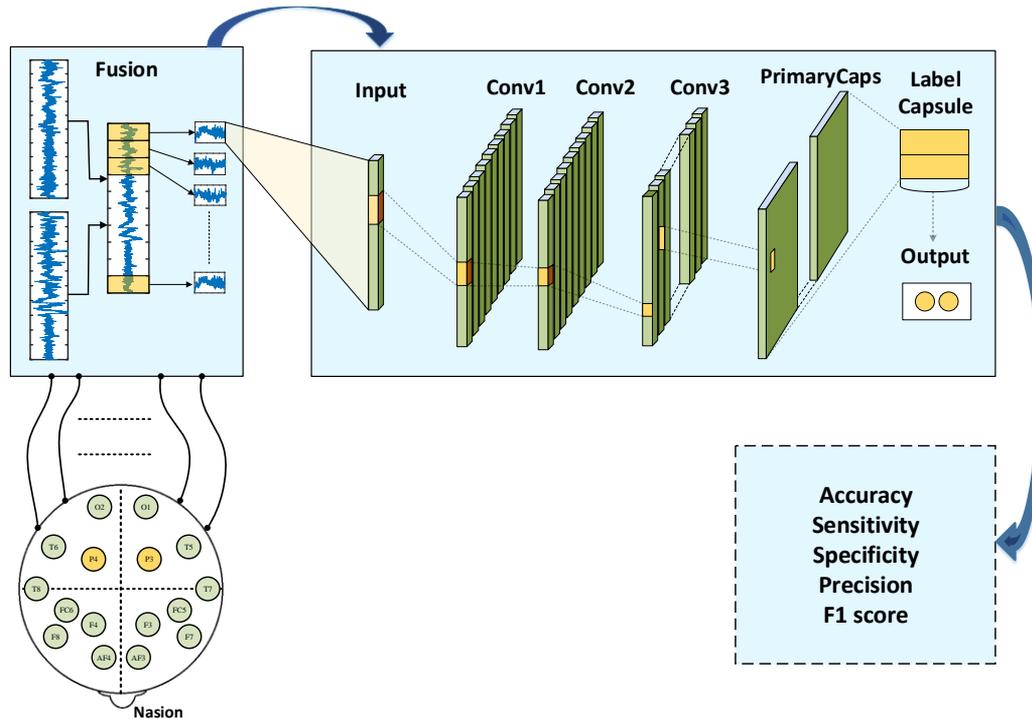


Figure 2. Block diagram representing the flowchart of the proposed method

2. MATERIALS AND METHODS

2.1 The dataset

EEG signals were obtained from 28 healthy people aged 20-27 years in Firat University Technology Faculty Software Engineering Department, Elazig, Turkey [10]. 16-channel EEG data was obtained with 128 Hz sampling frequency. Two channels (P3, P4) were used as reference. A wearable EMOTV EPOC + Mobile EEG device were used to obtain 14-channel EEG data (see Table 1). The signals were obtained while the subjects were playing four different games. 14-channel EEG data is shown in Figure 3.

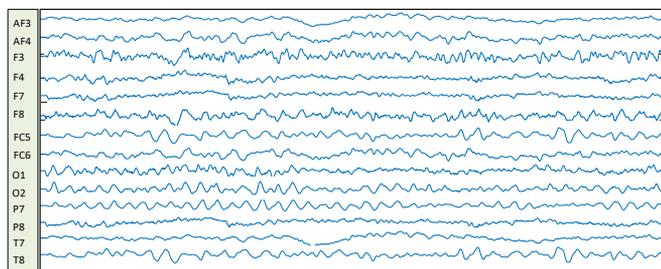


Figure 3. 14-channel EEG data obtained from the subjects

In addition, EEG signals for binary and multi-class emotion recognition were examined in both subject-based and non-

subject-based testing, both of which used a single EEG channel. In doing so, it was investigated as to whether an effective emotion recognition can be performed from a single EEG channel. The channel chosen for this study was AF4, mainly because the authors of the original article achieved the highest classification success with this channel [10]. All the experiments were carried out with a Linux Server (Ubuntu 16.04.4) with NVIDIA GTX 1080 GPU using Python Keras library.

2.2 Preprocessing

The GAMEEEMO dataset was examined for both binary and multi-class classification. In the first phase, the EEG signals in the LAPV and HAPV zones, which are known as positive emotion signals, were fused with the novel fusion strategy for binary emotion recognition. Likewise, negative emotion signals (LANV, HANV) were also fused (see Figure 4). With the new fusion method, a new EEG signal was obtained instead of the EEG signals, expressing two separate positive emotions. This process was carried out by taking the average of the EEG signals that defined the negative emotion.

In the second phase, EEG signals in LAPV, HAPV, LANV and HANV zones were used for multi-class emotion recognition. Before these processes, the data were normalized. Then, the *Z score* normalization method was used for normalization. Thus, *Z score* has facilitated the convergence of the classifier more quickly. While normalizing the data with

Z score, mean (μ) and standard deviation (σ) are considered. Z score is given in Eq. (1). Here, x represents the EEG signal.

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

EEG signals were segmented using the sliding window technique and each signal had a length of 38252 data-points. The signals were divided into segments of 100 units using the 10-unit sliding window method and a total of 3815 x 100 segments were obtained for each EEG signal. Thus, a total of 7630 x 100 segments, positive and negative EEG signals, were obtained in binary-class emotion recognition. This procedure was repeated for 28 subjects (for channel AF4). Positive and negative emotions were used for non-subject-based testing with binary-class emotion recognition to feed the capsule networks (2 x 106820) x 100 segments (28 x 3815 = 106820). The signal preprocessing steps that perform binary-class emotion recognition are shown in Figure 4. In the multi-class emotion recognition model (LAPV, HAPV, LANV, HANV) (4 x 3815) x 100 segments were used as input data for subject-based classification. For non-subject-based classification, 106820 x 100 segments were used as inputs for each class.

2.3 Capsule networks

Convolutional Neural Networks (CNN) architectures are highly successful in the field of image/signal processing thanks to the convolution layer structure added one after the other [21-23]. CNN models extract a feature map in each layer and transfer it directly to the next layer or reduce the data by

pooling to the next layer. However, performing pooling transfers important information in the data to the next layer, while the other data is viewed as unimportant and these data are not transferred to the next layer. As a result, the network is prevented from learning the small details [24]. In capsule networks, pooling is not used and this little information is used while training the network. In addition, capsule networks carry information to give the positions and orientations of the objects in the image thanks to the capsule structure. Another feature of the capsule networks is that it uses ‘squashing’ as the activation function, as shown in the following equation [20]:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} \quad (2)$$

where, v_j is the output of capsule j and s_j is the total entries of the capsule. v_j suppresses the long vectors towards one if there is an object in the image, while the image narrows the short vectors towards zero if there is no object [20, 24]. In capsule networks, a different margin loss is proposed to determine whether there are objects in a class (see Eq. (3)).

$$L_n = T_n \max(0, m^+ - \|v_n\|)^2 + \lambda(1 - T_n) \max(0, \|v_n\| - m^-)^2 \quad (3)$$

Here, $T_n = 1$, $m^+ = 0.9$ and $m^- = 0.1$ is selected. The direction of the vector depends on the object’s size, pose, and orientation [25, 26].

Table 1. Attributes of the GAMEEMO dataset

Data type	Information
Signal	EEG
Subject	Number of subjects: 28 subjects from students of Firat University Technology Faculty Software Engineering Department Age of subjects: 20-27
Device	EEG device type: 14-channel EMOTIV EPOC+ Mobile EEG device EEG electrodes location: 16 different scalp zones Connectivity: Wifi Sampling rate: 128 Hz Bandwidth: 0.16 Hz – 43 Hz
Game	Operating systems: Windows, Mac, IOS and Android Game type: 4 different games (funny, boring, horror, calm) Game recording time: 5 minutes (20 minutes in total for 4 games)

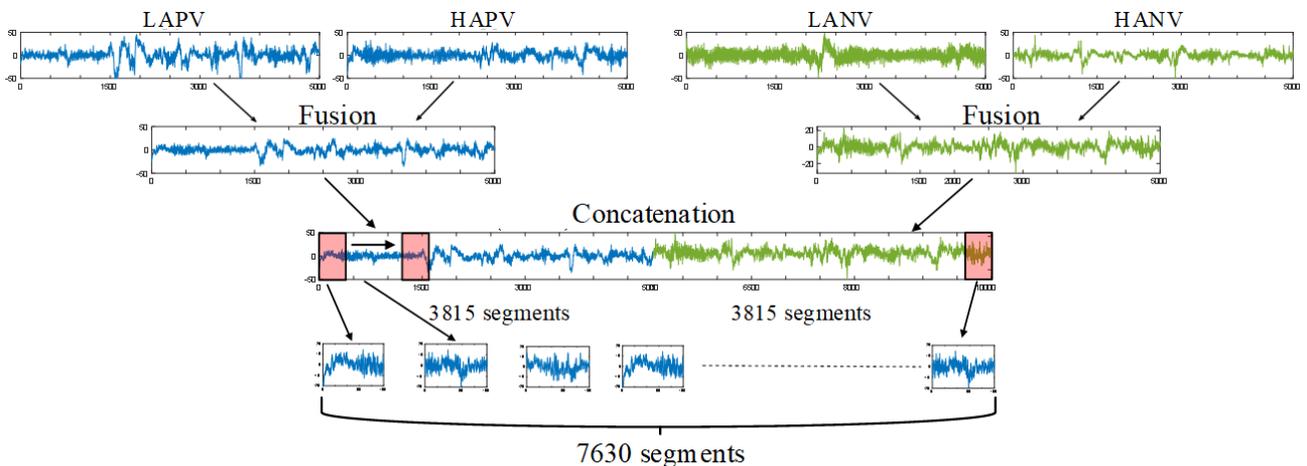


Figure 4. EEG signal preprocessing steps for binary-class emotion recognition

2.4 The proposed network architecture

A new network model with four convolution layers is proposed for 100×1 signal classification. Capsule networks have more computational load than traditional CNN architectures and require more powerful hardware. In addition, the size of the data to be given as input to the capsule network directly affects the system's speed. Therefore, a smaller and more efficient feature map was given as input to the primary layer by adding four convolution layers before the primary capsule layer. Figure 5 shows the structure of the proposed network architecture.

Details of the original capsule network and proposed network architecture are given in Table 2 and 3. The first two layers contain 16 and 32 kernels with size 5×5 and a stride of 1. Maximum pooling with a stride of 2 is applied to the second layer output. The third layer contains 128 filters with size 9×1 and a stride of 1. The fourth layer is the primary capsule and contains 32 different capsules, with each capsule applied a filter with size 9×1 and a stride of 1. The label capsule layer contains two size 16 capsules representing the positive and negative EEG signal classes in the binary-class model. Capsule networks can carry various features in signals in a single vector, thanks to their high dimensionality. Capsule networks also perform better on relatively small data than image data, such as the EEG signal [27].

ReLU activation function is used for all layers. The ReLU function takes zero value for neurons that produce negative values. Therefore, the ReLU function works faster and more efficiently than Tangent, Sigmoid, etc. [28].

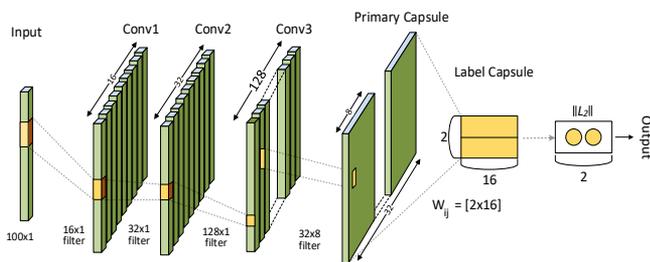


Figure 5. Capsule network structure proposed for automated emotion recognition

Table 2. Original capsule network architecture

Layers	Filter	Kernel size	Stride	Output
Input	-	-	-	28, 28
Conv1	256	9	1	20, 20
PrimaryCaps	32x8	9	2	32, 8, 6, 6
Digit capsule	-	-	-	16, 10
Output	-	-	-	10

Table 3. Details of layers and parameters of capsule network architecture for binary-class emotion recognition

Layers	Filter	Kernel size	Stride	Output
Input	-	-	-	100, 1
Conv1	16	5	1	100, 16
Conv2	32	5	1	100, 32
Maxpooling	-	-	2	50, 32
Conv3	128	9	1	50, 128
PrimaryCaps	256	9	1	1600, 8
Label capsule	-	-	-	16, 2
Output	-	-	-	2

Table 4. Details of layers and parameters of capsule network architecture for multi-class emotion recognition

Layers	Filter	Kernel size	Stride	Output
Input	-	-	-	100, 1
Conv1	16	5	1	100, 16
Conv2	32	5	1	100, 32
Maxpooling	-	-	2	50, 32
Conv3	64	5	1	50, 64
Maxpooling	-	-	2	25, 64
Conv4	128	9	1	25, 128
PrimaryCaps	256	9	1	800, 8
Label capsule	-	-	-	16, 4
Output	-	-	-	4

One convolution (128 kernels) and max pooling (2 of stride) layers were added to the binary-class capsule network architecture proposed for multi-class emotion recognition (Table 4). The reason for adding these layers was that the binary-class architecture performs poorly in recognizing multi-class datasets. In other words, while an efficient feature map was created with 3 convolution layers in a binary-class architecture, a multi-class architecture was also insufficient. That is, the increase in the number of classes increased the sensitivity. Therefore, deeper feature extraction was performed by adding a new convolution layer. The performance was increased with the addition of the layers. Comparative results are given in section 3.2.

2.5 Performance evaluation

The performance of the method was evaluated according to the most widely known indicators in the literature, including Accuracy, Sensitivity, Specificity, Precision, and F1 score. Additionally, True positive (TP), True negative (TN), False positive (FP), and False negative (FN) were used for the calculation, whereby TP indicated the positive emotions that were described correctly by the method, TN indicated the negative emotions that were described correctly by the method, FN indicated the positive emotions that were described incorrectly by the method, and FP indicated the negative emotions that were described incorrectly by the method.

$$Acc = (TP + TN)/(TP + FP + TN + FN) \quad (4)$$

$$Sen = TP/(TP + FN) \quad (5)$$

$$Spe = TN/(TN + FP) \quad (6)$$

$$Pre = TP/(TP + FP) \quad (7)$$

$$F1 = 2TP/(2TP + FP + FN) \quad (8)$$

3. EXPERIMENTAL RESULTS

An emotion recognition model was applied using the EEG signals of subjects while playing four different computer games. Information about the environment and equipment used for the experiment was delineated in the study by Alakus et al. [10]. The emotions of the subjects were classified into two and four classes and the emotions were identified in binary-class (positive, negative) or multi-class (LAPV, HAPV, LANV and HANV) comparisons [10].

3.1 Binary-class emotion recognition

Positive and negative emotions were analyzed in two ways. In the original article, the authors performed a general classification using all 14-channel EEG signals [12]. In the method proposed in the present study, a single channel was used to examine how successful a single channel would be in emotion recognition. In addition, a subject-based examination was also carried out in the present study, unlike in the original article. Considering the person-to-person nature of EEG signals, it is considered that it would be more efficient to compare the results of this examination with those of general signal analysis. In the proposed model, ReLU is used as the activation function. The learning rate (lr) was examined in the range of 10^{-2} , ..., 10^{-4} and the best performance was obtained with binary-class recognition (0.001). Additionally, 5-fold cross validation was performed for subject-based assessment (Figure 6). Table 5 presents the parameter results for all 28 subjects. Accordingly, Subject #26 had the lowest (90.22%) and Subject #01 had the highest value (99.99%) and the average accuracy rate for 28 subjects was 98.11%.

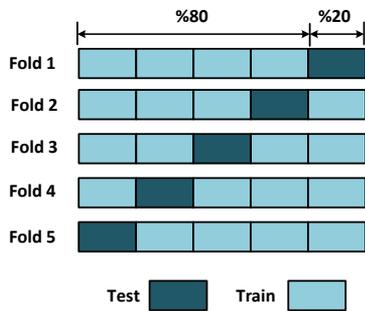


Figure 6. Graphical representation of training and test data

Table 5. Classification results for subject-based testing with binary-class emotion recognition (channel AF4)

Subject	Acc (%)	Pre (%)	Sen (%)	Spe (%)	F1 score (%)
s01	99.99 ± 0.03	99.97 ± 0.05	100.0 ± 0.00	99.97 ± 0.05	99.99 ± 0.03
s02	99.96 ± 0.08	99.92 ± 0.16	100.0 ± 0.00	99.92 ± 0.16	99.96 ± 0.08
s03	99.91 ± 0.15	99.82 ± 0.30	100.0 ± 0.00	99.82 ± 0.31	99.91 ± 0.15
s04	99.97 ± 0.05	99.95 ± 0.10	100.0 ± 0.00	99.95 ± 0.10	99.97 ± 0.05
s05	98.07 ± 1.05	97.48 ± 1.55	98.72 ± 0.52	97.43 ± 1.62	98.09 ± 1.03
s06	98.82 ± 0.40	99.16 ± 0.62	98.48 ± 0.69	99.16 ± 0.63	98.82 ± 0.40
s07	96.26 ± 1.67	96.20 ± 3.27	96.51 ± 3.05	96.02 ± 3.61	96.28 ± 1.65
s08	99.48 ± 0.51	99.63 ± 0.33	99.32 ± 0.75	99.63 ± 0.33	99.47 ± 0.52
s09	93.54 ± 1.50	94.05 ± 2.36	93.05 ± 2.69	94.02 ± 2.67	93.50 ± 1.51
s10	97.27 ± 1.65	98.02 ± 1.34	96.51 ± 3.04	98.03 ± 1.38	97.23 ± 1.72
s11	98.81 ± 1.36	99.25 ± 0.42	98.35 ± 2.47	99.27 ± 0.40	98.79 ± 1.40
s12	94.88 ± 0.90	94.69 ± 0.56	95.10 ± 2.19	94.65 ± 0.68	94.88 ± 0.96
s13	99.83 ± 0.10	99.79 ± 0.20	99.87 ± 0.20	99.79 ± 0.20	99.83 ± 0.10

s14	99.38 ± 0.49	99.25 ± 0.79	99.53 ± 0.20	99.24 ± 0.80	99.39 ± 0.49
s15	99.96 ± 0.05	99.92 ± 0.10	100.0 ± 0.00	99.92 ± 0.10	99.96 ± 0.05
s16	99.84 ± 0.18	99.84 ± 0.19	99.84 ± 0.31	99.84 ± 0.19	99.84 ± 0.18
s17	99.83 ± 0.12	99.82 ± 0.13	99.84 ± 0.15	99.82 ± 0.13	99.83 ± 0.12
s18	99.95 ± 0.08	99.90 ± 0.15	100.0 ± 0.00	99.99 ± 0.15	99.95 ± 0.08
s19	98.03 ± 1.38	97.67 ± 1.43	98.43 ± 1.66	97.64 ± 1.45	98.04 ± 1.39
s20	97.39 ± 0.86	98.50 ± 0.77	96.25 ± 1.09	98.53 ± 0.76	97.36 ± 0.87
s21	97.47 ± 0.68	98.00 ± 2.02	96.99 ± 1.38	97.96 ± 2.12	97.46 ± 0.66
s22	99.76 ± 0.18	99.84 ± 0.10	99.69 ± 0.32	99.84 ± 0.10	99.76 ± 0.18
s23	91.40 ± 1.40	91.68 ± 2.95	91.22 ± 1.97	91.59 ± 3.32	91.40 ± 1.31
s24	98.39 ± 0.88	97.50 ± 1.51	99.34 ± 0.52	97.43 ± 1.57	98.41 ± 0.86
s25	99.53 ± 0.25	99.22 ± 0.43	99.84 ± 0.13	99.21 ± 0.44	99.53 ± 0.25
s26	90.22 ± 1.76	88.75 ± 3.10	92.32 ± 2.53	88.13 ± 4.02	90.44 ± 1.62
s27	99.66 ± 0.21	99.56 ± 0.46	99.76 ± 0.24	99.55 ± 0.47	99.66 ± 0.21
s28	99.58 ± 0.13	99.56 ± 0.21	99.61 ± 0.19	99.55 ± 0.21	99.58 ± 0.13
Mean ±	98.11 ± 2.58	98.11 ± 2.70	98.16 ± 2.48	98.07 ± 2.80	98.12 ± 2.56
SD					

Acc: Accuracy, Pre: Precision, Sen: Sensitivity, Spe: Specificity, SD: Standard deviation

After performing subject-based testing, the AF4 channel data of 28 subjects were combined for a general recognition model to perform a non-subject-based classification and to examine whether a general emotion recognition model could be obtained. Table 6 presents the classification results of non-subject based data following the administration of 5-fold cross validation. Accordingly, the AF4 channel differentiated positive and negative emotions with 98.89% accuracy, 98.58% sensitivity, and 99.19% specificity.

Table 6. Classification results for non-subject-based testing with binary-class emotion recognition (channel AF4)

Fold	Acc (%)	Pre (%)	Sen (%)	Spe (%)	F1 score (%)
Fold 1	98.99	98.89	99.09	98.89	98.99
Fold 2	98.88	98.84	98.92	98.84	98.88
Fold 3	98.43	99.35	97.50	99.36	98.41
Fold 4	98.94	99.54	98.33	99.54	98.93
Fold 5	99.19	99.30	99.09	99.30	99.19
Mean ±	98.89 ± 0.25	99.18 ± 0.27	98.58 ± 0.61	99.19 ± 0.27	98.88 ± 0.26
SD					

Acc: Accuracy, Pre: Precision, Sen: Sensitivity, Spe: Specificity, SD: Standard deviation

3.2 Multi-class emotion recognition

Initially, the capsule network structure used for binary-class emotion recognition (see Table 3) was applied for multi-class emotion recognition, which provided a low accuracy (Figure 7a). Subsequently, the network structure given in Table 3 was applied for lr values (0.001 and 0.0001). Thus, an improvement in the results was observed (Figure 7b). The values 88.07% and 91.93% respectively were obtained for

Subject #01. Finally, as a result of parameter determination performed using the brute force method, the training-loss graphs shown in Figure 8 were reached (see Table 4). Table 7 shows the parameters of the capsule network architecture used for binary and multi-class emotion recognition and Table 8 presents the accuracy rate achieved for multi-class emotion recognition for all 28 subjects.

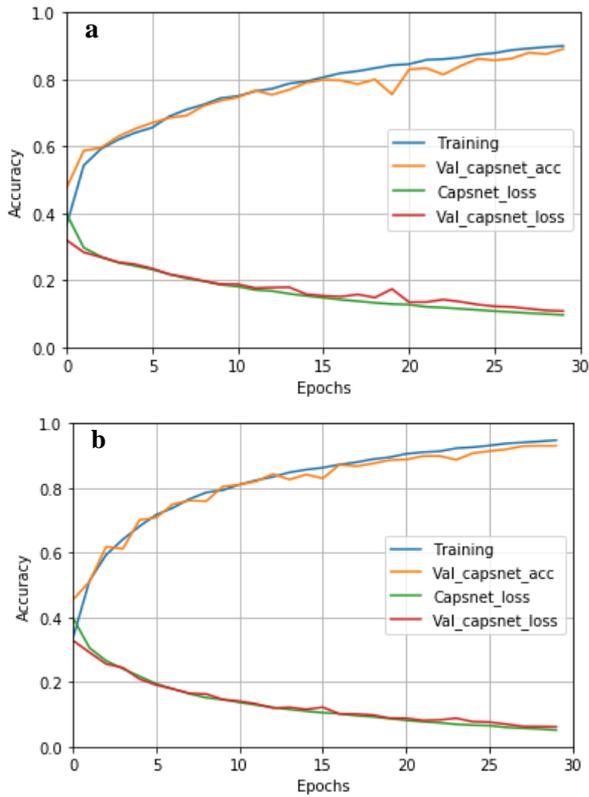


Figure 7. Application of the two-class model structure to the multi-class dataset according to (a) lr = 0.001, (b) lr = 0.0001

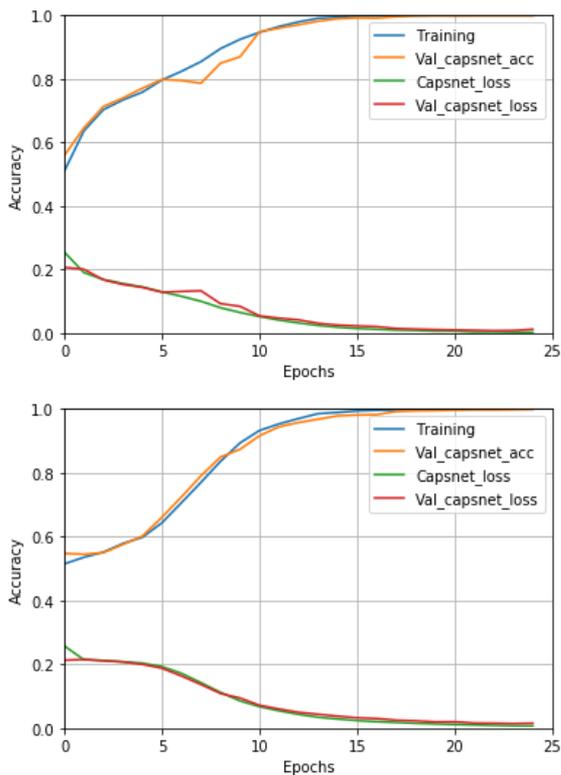


Figure 8. Sample folds of training and loss graphics (top: binary-class, bottom: multi-class)

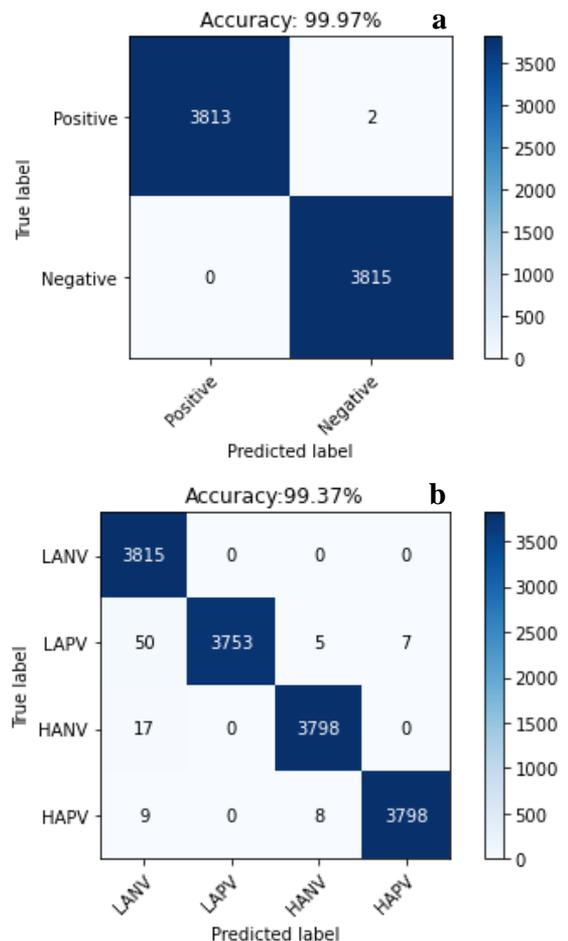


Figure 9. Confusion matrices of testing, a) binary-class, b) multi-class

Table 7. Hyper parameters of capsule network architectures

	Routing	Optimizer	<i>lr</i>	Loss weight	Batch size	Epoch	Time (s) (per epoch)
2 class	3	Adam	0.001	0.392	16	25	16
4 class	3	Adam	0.0001	0.392	32	10	23

lr: Learning rate**Table 8.** Capsule network results for the multi-class subject-based testing

Subject	Accuracy (%)
s01	99.76 ± 0.18
s02	99.99 ± 0.01
s03	99.99 ± 0.01
s04	99.37 ± 1.24
s05	98.49 ± 0.63
s06	98.36 ± 0.82
s07	98.77 ± 0.68
s08	96.81 ± 1.21
s09	92.17 ± 2.75
s10	97.53 ± 0.81
s11	98.22 ± 0.58
s12	94.33 ± 2.10
s13	97.49 ± 3.27
s14	99.12 ± 0.29
s15	99.61 ± 0.44
s16	99.79 ± 0.31
s17	99.98 ± 0.03
s18	99.98 ± 0.03
s19	99.81 ± 0.11
s20	99.52 ± 0.47
s21	98.58 ± 0.65
s22	96.24 ± 2.06
s23	99.26 ± 0.31
s24	98.05 ± 0.47
s25	98.41 ± 1.40
s26	97.64 ± 1.09
s27	98.96 ± 0.40
s28	98.59 ± 0.81
Mean ± SD	98.38 ± 1.75

Table 9. Capsule network results for the multi-class the non-subject based testing

Fold	Accuracy (%)
Fold1	90.31
Fold2	90.31
Fold3	90.53
Fold4	89.81
Fold5	90.73
Mean ± SD	90.37 ± 0.31

In non-subject-based recognition, an average accuracy rate of 90.37% was achieved with 5-fold cross validation results (Table 9).

Another component used to evaluate system performance is confusion matrix. Figure 9 presents the binary- and multi-class confusion matrices of a subject (#04). As seen in Figure 9a, only 2 out of the 3815 signal parts belonging to positive emotions were classified incorrectly, while all of the signal parts belonging to negative emotions were recognized correctly. In multi-class classification, however, a highly effective definition was obtained.

4. DISCUSSION

To date, various applications for emotion recognition from EEG have been performed using traditional feature extraction

methods. In these applications, traditional features of EEG signals such as EMD, entropy, and gray-level co-occurrence matrix (GLCM) have been extracted [11, 29-32]. However, with the emergence of deep learning architectures, automatic feature extraction methods have begun to replace the traditional feature extraction methods. With deep learning, in particular, more effective feature extraction can be made from raw signals [5]. Many datasets on emotion recognition using deep learning architectures are available, with the best-known ones including DEAP, SEED, MANHOB-HCI, and DREAMER. Additionally, CNN, RNN, LSTM, and deep belief networks are some of the deep learning architectures used for these datasets [16, 17, 19, 33]. However, to our knowledge, no deep learning architecture has yet been tried on the GAMEEMO dataset created for physical and spiritual emotion recognition. The applications performed on this dataset are given in Tables 10 and 11. As seen in Table 10, the authors initially implemented binary-class emotion recognition for positive and negative emotion prediction from the dataset and then performed multi-class emotion recognition for four-zone prediction (LANV, HANV, HAPV, and LAPV) (see Table 11).

Alakus et al. performed statistical, chaotic, and time-frequency analysis from these sub-signals and classified the extracted features with three different classifiers (k-NN, SVM, MLPNN) [12].

In another study, Turker et al. performed feature extraction from each EEG channel with the help of Tunable Q wavelet transform. The EEG signal was split into 30 subbands for feature extraction, 1024 features were extracted from each subband, and a total of 31744 feature vectors were obtained for one EEG channel. Since the feature vector was remarkably large, features were reduced with the IChi2 feature selector. The reduced features were then classified with k-NN, SVM, and LDA [13].

In both of the studies conducted on the GAMEEMO dataset, the authors divided the EEG signals into subbands and extracted features from these subbands. The extracted features were classified with different classifiers. Moreover, the authors applied traditional feature extraction methods which are both time-consuming and require serious processing load. In our study, however, no feature extraction was made from EEG signals and the features were automatically acquired and classified by the deep learning architecture. In doing so, a direct result was obtained from the raw data with an end-to-end method, thus eliminating the deficiencies encountered in traditional feature extraction methods or errors arising from the expert's experience.

Significant contributions of this study are as follows:

- Capsule network, a new neural network model, was used instead of traditional feature extraction methods.
- In the proposed study, feature extraction and feature selection by separating EEG signals into sub-bands were not needed.
- Raw EEG signals were used in the proposed method, which led to less processing load and also saved time.
- High-accuracy emotion recognition was performed from a single EEG channel (AF4).

Table 10. Binary-class classification comparison results for GAMEEMO dataset

Channel	Alakus et al.			Proposed method (PM)	
	kNN	SVM	MLPNN	PM1	PM2
AF4	75%	88%	87%	98.11%	98.89%

Table 11. Multi-class classification comparison results for GAMEEMO dataset

Channel	Alakus et al.			Tuncer et al.			Proposed method (PM)	
	kNN	SVM	MLPNN	kNN	SVM	LDA	PM1	PM2
AF4	55%	50%	75%	98.39%	99.64%	95.54%	98.38%	90.37%

PM1: Average accuracy of subject-based AF4 channel

PM2: Average accuracy of non subject-based AF4 channel

•The proposed capsule network architecture was tested for 28 subjects.

•The GAMEEMO dataset was examined using subject-based testing for the first time.

•The proposed method achieved a classification accuracy which was 10% better than the classification performance achieved in other studies in the literature.

In previous studies, the authors focused on general emotion recognition from the EEG signal. Using all channels both increases the processing load and the total duration of the procedure. Additionally, separating all channels into their subbands further increases the processing load. In the proposed method, the only preprocessing to classify raw EEG signals with capsule networks is Z score normalization. The advantages of the proposed capsule network are given above. The limitation of the study was that it examined a single EEG channel. Further studies may examine other EEG channels with capsule networks.

5. CONCLUSIONS

Emotion recognition was carried out using two different scenarios. In the first, the emotions were examined for binary and multi-class emotion recognition. In the second, subject-based recognition was performed, which, to our knowledge, had never been applied on this dataset in the literature. The results indicated that subject-based recognition was more successful than non-subject-based recognition, i.e. a general EEG model. Considering that each person's characteristics such as personal, psychological, or physical features are different, we suggest that creating personalized EEG datasets will play a more effective role in increasing emotion recognition performance.

REFERENCES

[1] Suman, S. (2021). Artificial intelligence in nuclear industry: Chimera or solution? *Journal of Cleaner Production*, 278: 124022. <https://doi.org/10.1016/j.jclepro.2020.124022>

[2] Górriz, J.M., Ramírez, J., Ortíz, A., et al. (2020). Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications. *Neurocomputing*, 410: 237-270. <https://doi.org/10.1016/j.neucom.2020.05.078>

[3] Chen, J., Jiang, D., Zhang, Y., Zhang, P. (2020). Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset. *Computer*

Communications, 154: 58-65. <https://doi.org/10.1016/j.comcom.2020.02.051>

[4] Turkoglu, M. (2021). COVID-19 detection system using chest CT images and multiple kernels-extreme learning machine based on deep neural network. *IRBM*, 42(4): 207-214. <https://doi.org/10.1016/j.irbm.2021.01.004>

[5] Sharma, R., Pachori, R.B., Sircar, P. (2020). Automated emotion recognition based on higher order statistics and deep learning algorithm. *Biomed Signal Process Control*, 58: 101867. <https://doi.org/10.1016/j.bspc.2020.101867>

[6] Lang, P.J. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report.

[7] Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3-4): 169-200. <https://doi.org/10.1080/02699939208411068>

[8] Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4): 344-350.

[9] Russell, J.A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1): 145-172. <https://doi.org/10.1037/0033-295X.110.1.145>

[10] Alakus, T.B., Gonen, M., Turkoglu, I. (2020). Database for an emotion recognition system based on eeg signals and various computer games-GAMEEMO. *Biomedical Signal Processing and Control*, 60: 101951. <https://doi.org/10.1016/j.bspc.2020.101951>

[11] Salankar, N., Mishra, P., Garg, L. (2021). Emotion recognition from EEG signals using empirical mode decomposition and second-order difference plot. *Biomedical Signal Processing and Control*, 65: 102389. <https://doi.org/10.1016/j.bspc.2020.102389>

[12] Tan, C., Šarlija, M., Kasabov, N. (2021). NeuroSense: Short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns. *Neurocomputing*, 434: 137-148. <https://doi.org/10.1016/j.neucom.2020.12.098>

[13] 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. <https://doi.org/10.1016/j.chaos.2021.110671>

[14] Nawaz, R., Cheah, K.H., Nisar, H., Yap, V.V. (2020). Comparison of different feature extraction methods for EEG-based emotion recognition. *Biocybernetics and Biomedical Engineering*, 40(3): 910-926. <https://doi.org/10.1016/j.bbe.2020.04.005>

[15] Er, M.B., Çiğ, H., Aydılek, İ.B. (2021). A new approach to recognition of human emotions using brain signals and music stimuli. *Applied Acoustics*, 175: 107840.

- <https://doi.org/10.1016/j.apacoust.2020.107840>
- [16] Yin, Y., Zheng, X., Hu, B., Zhang, Y., Cui, X. (2021). EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Applied Soft Computing*, 100: 106954. <https://doi.org/10.1016/j.asoc.2020.106954>
- [17] Wang, F., Wu, S., Zhang, W., Xu, Z., Zhang, Y., Wu, C., Coleman, S. (2020). Emotion recognition with convolutional neural network and EEG-based EFDMS. *Neuropsychologia*, 146: 107506. <https://doi.org/10.1016/j.neuropsychologia.2020.107506>
- [18] Cui, H., Liu, A., Zhang, X., Chen, X., Wang, K., Chen, X. (2020). EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowledge-Based Systems*, 205: 106243. <https://doi.org/10.1016/j.knosys.2020.106243>
- [19] Wei, C., Chen, L.L., Song, Z.Z., Lou, X.G., Li, D.D. (2020). EEG-based emotion recognition using simple recurrent units network and ensemble learning. *Biomedical Signal Processing and Control*, 58: 101756. <https://doi.org/10.1016/j.bspc.2019.101756>
- [20] Sabour, S., Frosst, N., Hinton, G.E. (2017). Dynamic routing between capsules. *arXiv preprint arXiv:1710.09829*.
- [21] Toraman, S. (2020). Preictal and interictal recognition for epileptic seizure prediction using pre-trained 2D-CNN models. *Traitement du Signal*, 37(6): 1045-1054. <https://doi.org/10.18280/ts.370617>
- [22] Ismael, A.M., Alçin, Ö.F., Abdalla, K.H., Şengür, A. (2020). Two-stepped majority voting for efficient EEG-based emotion classification. *Brain Informatics*, 7(1): 9. <https://doi.org/10.1186/s40708-020-00111-3>
- [23] Chao, H., Dong, L., Liu, Y., Lu, B. (2019). Emotion recognition from multiband EEG signals using CapsNet. *Sensors*, 19(9): 2212. <https://doi.org/10.3390/s19092212>
- [24] Toraman, S., Alakus, T.B., Turkoglu, I. (2020). Convolutional CapsNet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. *Chaos, Solitons & Fractals*, 140: 110122. <https://doi.org/10.1016/j.chaos.2020.110122>
- [25] Beşer, F., Kizrak, M.A., Bolat, B., Yildirim, T. (2018). Recognition of sign language using capsule networks. In 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, pp. 1-4. <https://doi.org/10.1109/SIU.2018.8404385>
- [26] Zhang, X., Zhao, S.G. (2019). Cervical image classification based on image segmentation preprocessing and a CapsNet network model. *International Journal of Imaging Systems and Technology*, 29(1): 19-28. <https://doi.org/10.1002/ima.22291>
- [27] Butun, E., Yildirim, O., Talo, M., Tan, R.S., Acharya, U.R. (2020). 1D-CADCapsNet: One dimensional deep capsule networks for coronary artery disease detection using ECG signals. *Physica Medica*, 70: 39-48. <https://doi.org/10.1016/j.ejmp.2020.01.007>
- [28] Szandała, T. (2021). Review and comparison of commonly used activation functions for deep neural networks. *Bio-inspired Neurocomputing*, pp. 203-224. <http://dx.doi.org/10.1007/978-981-15-5495-7>
- [29] Lu, Y., Wang, M., Wu, W., Han, Y., Zhang, Q., Chen, S. (2020). Dynamic entropy-based pattern learning to identify emotions from EEG signals across individuals. *Measurement*, 150: 107003. <https://doi.org/10.1016/j.measurement.2019.107003>
- [30] Gao, Y., Wang, X., Potter, T., Zhang, J., Zhang, Y. (2020). Single-trial EEG Emotion recognition using granger causality/transfer entropy analysis. *Journal of Neuroscience Methods*, 346: 108904. <https://doi.org/10.1016/j.jneumeth.2020.108904>
- [31] Liu, J., Lughofer, E., Zeng, X. (2015). Could linear model bridge the gap between low-level statistical features and aesthetic emotions of visual textures? *Neurocomputing*, 168: 947-960. <https://doi.org/10.1016/j.neucom.2015.05.030>
- [32] Sbagoud, F., Djeha, M., Guiatni, M., Ababou, N. (2019). WPT-ANN and belief theory based EEG/EMG data fusion for movement identification. *Traitement du Signal*, 36(5): 383-391. <https://doi.org/10.18280/ts.360502>
- [33] Hassan, M.M., Alam, M.G.R., Uddin, M.Z., Huda, S., Almogren, A., Fortino, G. (2019). Human emotion recognition using deep belief network architecture. *Information Fusion*, 51: 10-18. <https://doi.org/10.1016/j.inffus.2018.10.009>