Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 7, No. 1, February, 2023

Continuous Capsule Network Method for Improving Electroencephalogram-Based Emotion Recognition

I Made Agus Wirawan [1,2], Retantyo Wardoyo [1*], Danang Lelono [1], Sri Kusrohmaniah [3]

[1] *Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia.*

[2] *Education of Informatics Engineering Department, Faculty of Engineering and Vocational, Universitas Pendidikan Ganesha, Singaraja 81116, Indonesia.*

[3] *Department of Psychology, Faculty of Psychology, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia.*

Abstract

The convolution process in the Capsule Network method can result in a loss of spatial data from the Electroencephalogram signal, despite its ability to characterize spatial information from Electroencephalogram signals. Therefore, this study applied the Continuous Capsule Network method to overcome problems associated with emotion recognition based on Electroencephalogram signals using the optimal architecture of the (1) 1st, 2nd, 3rd, and 4th Continuous Convolution layers with values of 64, 128, 256, and 64, respectively, and (2) kernel sizes of 2×2×4, 2×2×64, and 2×2×128 for the 1st, 2nd, and 3rd Continuous Convolution layers, and 1×1×256 for the 4th. Several methods were also used to support the Continuous Capsule Network process, such as the Differential Entropy and 3D Cube methods for the feature extraction and representation processes. These methods were chosen based on their ability to characterize spatial and low-frequency information from Electroencephalogram signals. By testing the DEAP dataset, these proposed methods achieved accuracies of 91.35, 93.67, and 92.82% for the four categories of emotions, two categories of arousal, and valence, respectively. Furthermore, on the DREAMER dataset, these proposed methods achieved accuracies of 94.23, 96.66, and 96.05% for the four categories of emotions, the two categories of arousal, and valence, respectively. Finally, on the AMIGOS dataset, these proposed methods achieved accuracies of 96.20, 97.96, and 97.32% for the four categories of emotions, the two categories of arousal, and valence, respectively.

Keywords:

Electroencephalogram;
Emotion Recognition;
Differential Entropy;
Baseline Reduction;
3D Cube; Capsule Network;
Continuous Convolution.

Article History:

Received:	09	June	2022
Revised:	17	August	2022
Accepted:	14	September	2022
Available online:	07	November	2022

1- Introduction

Emotions are psychological reactions to daily social interactions [1], which emerge due to certain conditions or problems encountered in achieving the desired target [2]. It is categorized into arousal and valence, with positive and negative values, respectively. Valence is an individual's reaction toward an event, while arousal is the excitement to behave accordingly or express the feeling [3]. A combination of arousal and valence labels is categorized into four quadrants: the 1st represents high arousal and positive valence (HAPV), 2nd depicts high arousal and negative valence (HANV), 3rd implies low arousal and negative valence (LANV), while the 4th describes low arousal and positive valence (LAPV) [4, 5]. The emotional reactions in each quadrant represent mental health and human performance [6]. It is essential to recognize the emotions in each quadrant to understand these individuals' mental state and performance.

* **CONTACT**: rw@ugm.ac.id

DOI: http://dx.doi.org/10.28991/ESJ-2023-07-01-09

Several preliminary studies have been carried out both externally and internally. External emotional reactions are usually hidden by individuals in social environments [7, 8]. In contrast, internal emotions are primarily recognized through Electroencephalogram (EEG) signals. These have several advantages, such as (i) EEG signals have spatial information that tends to represent human affective experiences [9], and (ii) they directly reflect changes in emotional reactions and responses in the human brain, particularly in the limbic system [10, 11].

This study's success was impacted by the EEG signals' characteristics, such as low-frequency and spatial data [12]. Several studies identified these attributes in feature extraction, feature representation, and classification processes [13, 14]. Regarding feature extraction, the Differential Entropy (DE) method is used to detect low amplitudes [15], as well as characterize the time series of EEG signals [16]. The most accurate and stable features for emotional recognition are generated by this method [17–19]. Although the feature extraction process executed using the DE method tends to consider the characteristics of EEG signals, different participant characteristics significantly affect the emotional responses [12, 20]. Yang et al. (2018) investigated the baseline reduction approach and its ability to overcome these problems. This approach can improve the accuracy of emotion recognition [18].

Due to the fact that the signals have characteristics related to spatial information, it is crucial to determine an appropriate representation method for these features before carrying out the classification process. Based on several preliminary studies, the representation features using the 3D Cube method can maintain spatial information between channels and frequency bands, such as Theta, Alpha, Beta, and Gamma [12, 18, 21]. In addition to the feature extraction and representation phases, the classification stage is essential for emotional recognition. Several studies stated that the adoption of the capsule network method is used to characterize the spatial relationships between different features [22]. This procedure is adequately trained on a smaller scale than a Convolutional Neural Network (CNN) [23]. However, the convolution process in the capsule network method can result in a loss of spatial information from the EEG signal. Therefore, using a continuous convolution approach in the capsule network method is crucial. A continuous convolution has been investigated in previous studies; however, this approach is applied to the CNN method. This procedure does not involve a pooling process, and zero-padding is a SAME in the convolution layer [12, 18]. A combination of the capsule network and continuous convolution approaches is called the continuous capsule network method.

This study contributes to the development of the Continuous Capsule Network method and its utilization to overcome the loss of spatial information from EEG signals. Several secondary datasets such as DEAP, DREAMER, and AMIGOS were used to test this contribution. The emotional outcomes were divided into four classes: high arousal and positive valence; high arousal and negative valence; low arousal and positive valence; and low arousal and negative valence. In addition, this method was used to classify two classes: high and low for arousal and high and low for valence.

2- Related Works

EEG signals with respect to emotional recognition have a few essential characteristics, such as low-frequency and spatial data. Previous studies distinguished these attributes based on extraction, representation features, and classification. Feature extraction is obtaining characters relevant to an EEG signal's emotional state. It is grouped into three features: time, frequency-domain, and time-frequency domain [12]. Based on a few studies, the extraction method executed using Differential Entropy (DE) emphasizes a few points of interest compared to other methods. The DE method tends to recognize low-frequency EEG signals [15]. In addition, it is also used to characterize the time series [16]. The most accurate and stable features are generated for emotional recognition [17–19]. Although the feature extraction process using the DE method considers the characteristics of EEG signals, that of diverse participants also significantly affects the emotional responses [12, 20]. These responses are also affected by a few characteristics, such as mental capacities, gender, and identity [14, 24]. Several studies examined the baseline reduction approach to represent emotional reactions based on the different characteristics of each participant [18, 25]. The reduction approach was implemented by reducing the DE feature of the experiment EEG signal using the average value of the baseline EEG signal [18]. The baseline EEG signals were recorded before the participants were given the stimulus medium. This procedure also increases the accuracy of emotional recognition [18, 21]. The classification process employed the DE feature of the reduced signals.

Due to the spatial information characteristics contained by the signal, it is essential to determine the appropriate representation method before the classification process. Chen et al. (2020) designed a feature representation using the 2D mesh sequence method. The study stated that the EEG signals data is directly represented in a $9 \times 9 \times 128$ matrix, whereby 9 represents the height and width of the matrix, while 128 denotes its depth. The feature mapping of 9×9 matrices is based on the International 10-20 System. However, this does not consider spatial information between the frequency bands [26]. Chao et al. (2019) proposed a representation method using the Multiband Feature Matrix (MFM) approach, which considers its frequency band. This procedure represents feature values as a 9×9 matrix, and each depicts theta, alpha, beta, and gamma frequencies. Feature representations are presented in a line to form a matrix measuring 18×18 [1]. This method produces a more extensive data size than the feature representation using the 3D Cube method. Like the MFM method, the 3D Cube was presented in a stack such that combining these four matrices produced the exact size of 9 × 9. The 3D Cube method keeps up spatial data between channels and frequencies, while its representation refers to an RGB image. Therefore, this study examines the use of the 3D Cube method for feature representation [12, 18, 21].

Several classification techniques are used to identify the four categories of emotions, such as the kNN method proposed by Liu et al. (2018). Based on the tests carried out on the DEAP dataset, this approach tends to produce an average accuracy of 86.05% [27]. Soroush et al. (2019) designed the multi-class support vector machine (MSVM) method used to identify four classes of emotions. However, by testing the DEAP dataset, the proposed method produces an accuracy of 81.67% [28]. The Graph regularized Extreme Learning Machine (GELM) method was designed by Zheng et al. (2016). In the DEAP dataset test, this approach produces an accuracy of 69.67% in recognizing the four emotional categories [29]. Mei and Xu (2017) proposed the Deep Learning approach using the CNN method. Based on the tests carried out on the DEAP dataset, this approach accurately recognizes four categories of emotions by 75% [30]. To improve its recognition accuracy, Zhao et al. (2020) designed the application of baseline reduction on the CNN method. In accordance with testing the two secondary datasets, this technique produces an accuracy of 93.53% and 95.86% for the DEAP and AMIGOS datasets, respectively [31].

In addition, emotions are classified into high and low categories, denoting arousal and valence, respectively. Liu et al. (2019) employed the kNN method to recognize these two categories of emotions, which produced an accuracy of 86.46% and 84.90% using the DEAP dataset [32]. He et al. (2017) used the SVM method to make the same classification and obtained an average accuracy of 67.9% and 70.9% [33]. Parui et al. (2019) used the XGboost method to perform this classification, and through the tests carried out with an average accuracy of 74.20% and 75.97% for arousal and valence, respectively [34]. Pan et al. (2020) employed a Logistic Regression method through the Variable Splitting and Augmented Lagrangian (LORSAL) for this classification and obtained an average accuracy of 77.17% and 77.03% for arousal and valence, respectively [35]. Garg and Verma (2020) proposed the CNN method with GoogleNet architecture, and through the tests carried out, an average accuracy of 61.23% and 92.19% was obtained for arousal and valence, respectively [36]. Song et al. (2018) designed the Dynamical Graph Convolution Neural Network (DGCNN) method for this categorization, and an average accuracy of 84.54% and 86.23% was obtained [10]. The use of the CNN method for classifying emotions into two categories was also proposed by Huang et al. (2019). This approach produced an average accuracy of 84.5% and 83.7% for valence and arousal [37].

Wardoyo et al. (2022) proposed the Radius SMOTE technique to improve the performance of the CNN method through the data oversampling process to realize an average accuracy of 82.11% and 78.99% for arousal and valence, respectively [20]. Another attempt was made by Yang et al. (2018), where they proposed a Continuous CNN method to identify the two categories. This approach does not employ pooling and utilizes the SAME padding in the convolution process. It aims to overcome the loss of feature information from the EEG signal, which consisted of average accuracy of 69.55% and 68.56% for arousal and valence, respectively. Yang et al. (2018) enhanced the performance of the Continuous CNN method by applying a baseline reduction approach. This led to achieving an average accuracy of 90.24% and 89.45% for arousal and valence, respectively [18].

The application of a baseline reduction technique to the CNN method was also studied by Wirawan et al. (2021), and this approach uses the Relative Difference procedure. Based on the tests carried out on the DEAP dataset, the proposed CNN method produced an average accuracy of 82.10% and 81.47% for arousal and valence, respectively [21]. Of the several classification methods studied, CNN produces higher accuracy than others [18]. However, it cannot represent spatial information between parts of an object and its whole, which affects the accuracy of emotion recognition [12]. This issue is resolved with the Capsule Network method. It is based on a deep learning approach with several advantages, such as characterizing spatial relationships between objects [22] and being trained effectively on a much smaller data scale than the CNN method [23]. The Capsule Network technique is an improvement of the CNN method. Recognition of emotions based on EEG signals using this approach was first studied by Chao et al. (2019). The capsule network method only achieved an accuracy of 68.28% and 66.73% in recognizing arousal and valence emotions on the DEAP dataset [1]. Liu et al. (2020) also seek to improve its performance by applying Multi-Level Feature maps on the Primary Capsule. Regarding the DEAP dataset's testing, this method produces an average accuracy of 64.36% and 62.57% for arousal and valence, respectively [23].

Irrespective of the fact that it is an improvement of the CNN method, its accuracy is still low. This problem is caused by the loss of spatial information from the EEG signals during the convolution process [1, 12]. Therefore, the Continuous Convolution approach in the Capsule Network method is essential. This approach produces higher accuracy than the Capsule Network technique, irrespective of its application to the CNN method [18, 23]. The application of the Continuous Capsule Network method is expected to be able to represent spatial information and overcome its loss in the EEG signals during the classification process. This study proposes the Differential Entropy, 3D Cube, and Continuous Capsule Network methods for the feature extraction, representation, and classification, respectively.

3- The Proposed Model

This study developed a Continuous Capsule Network architecture based on EEG signals for emotion recognition, as shown in Figure 1.

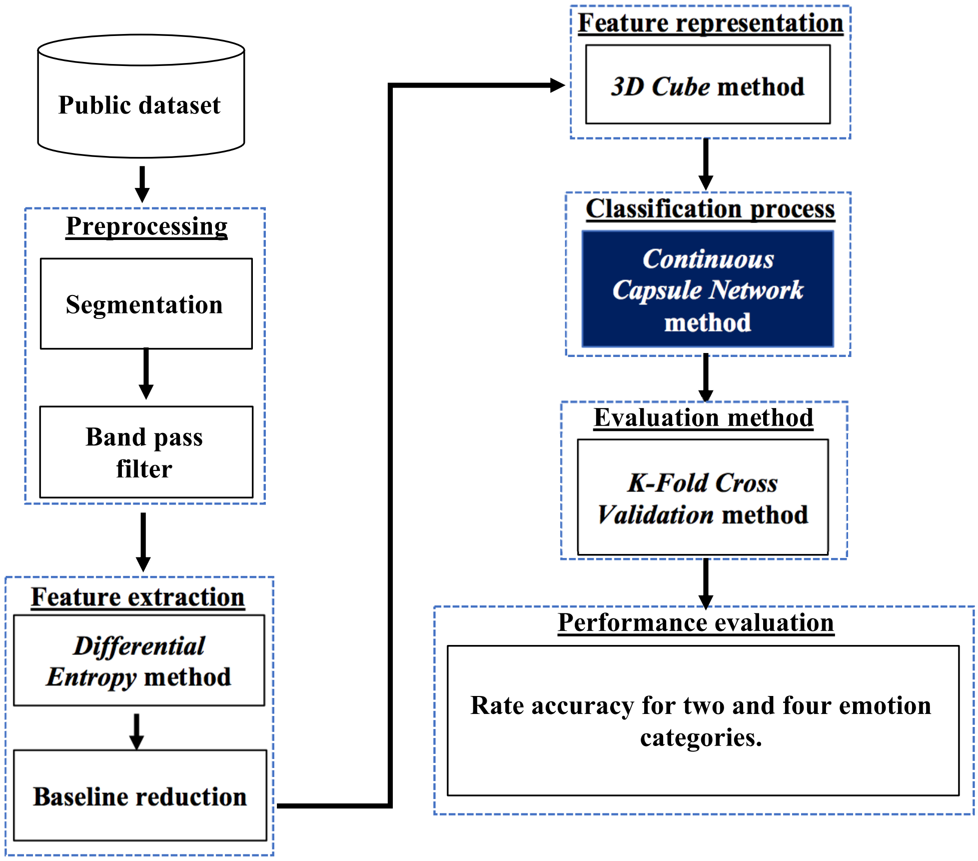

Figure 1. Steps of emotion recognition are based on EEG signals

In addition, three secondary datasets, DEAP, DREAMER, and AMIGOS, were selected. This was based on several considerations, such as the number of channels and participants, duration of the trial, and trial scenarios (individual and group scenarios). The DEAP dataset used 32 EEG channels, the duration of each trial was 63 s (one participant watched 40 trial videos), while the total number of participants was 32. This dataset carried out individual scenarios [38]. In the DREAMER dataset, 14 EEG channels were used, and the duration of each trial ranged from 67 s to 394 s. One participant of the 23 others watched 18 trial videos. The trial scenarios in this dataset are carried out individually [39]. The AMIGOS dataset used 14 EEG channels, the duration of each trial ranges from 56.61 s to 155.46 s for short videos and 851.15 s to 1420.42 s for the long ones (one participant watched 16 short trial videos and four lengthy ones). The total number of participants who can be used in this dataset is 31 people, and the trial were carried out both individually and in groups [40]. These three datasets were used to identify four and two emotional categories. Based on Figure 1, the emotion recognition process is divided into six steps, one of which is a blue rectangle.

3-1- Pre-processing

At this stage, the raw EEG signals from the secondary dataset is segmented every second and at each channel. It contains 128 data points in one second. In the DEAP data set, the first 3 seconds are the baseline segment, while the subsequent 4 seconds are the experiment segment. The DREAMER dataset comprises the base segment, which has a duration of 5 seconds, while the experiment segment is from 67 to 394 seconds. Lastly, in the AMIGOS data set, 5 seconds is the base segment, while the next 6 seconds is for the experiment segment.

After the segmentation, decomposition was performed to obtain theta, alpha, beta, and gamma bands. This process uses the bandpass filter method, however, the decomposition process for one second on the DREAMER dataset is shown in Figure 2. The bandpass filter method decomposes a one-second EEG signals generated from channel AF3, leading to the generation of four frequency bands.

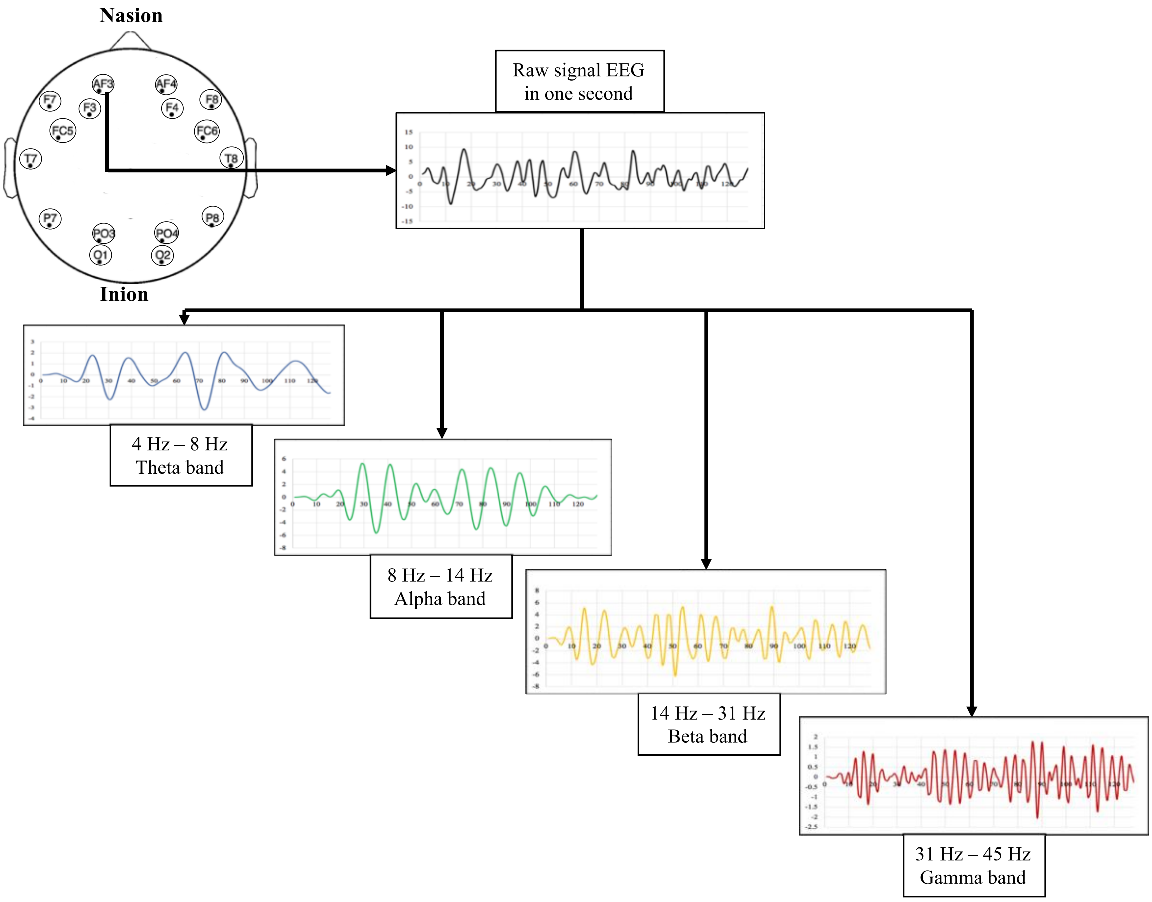

Figure 2. Illustration of the decomposition processes for AF3 channel

3-2- Feature Extraction

Feature extraction is performed for each second, frequency band, and channel, and the DE method is obtained using the following Equation 1 [18]:

$$h_i(X) = \frac{1}{2}\log(2\pi e\delta^2{}_i(X)) \quad (1)$$

where, *e* s Euler's constant, $\delta^2{}_i$ is the variance at the i^{th} seconds for the EEG signals segment, and $h_i(X)$ is the DE value at the i^{th} seconds for the EEG signals segment. Furthermore, the DE feature value from the experiment segment is reduced using its average from the baseline segment. It aims to represent emotional reactions according to the participants' characteristics. This approach tends to improve the accuracy of emotion recognition, and Equation 2 is used to obtain the baseline reduction.

$$Final_i(X) = Exper_i(X) - BaseMean(X) \quad (2)$$

where $Exper_i(X)$ is the DE feature of the X frequency band at the i^{th} seconds for the experiment segment, $BaseMean(X)$ is the DE feature of the X frequency band for the EEG baseline signals, $Final_i(X)$ is the DE feature after baseline reduction of the X frequency band at the i^{th} seconds for the experiment segment.

3-3- Feature Representation

Since the EEG signals have spatial information characteristics, it is crucial to determine an appropriate representation method for its features before performing the classification process. The feature representation was carried out on each frequency band and all channels. It aims to preserve spatial information between adjacent channels. Figure 3 illustrates the feature representation process, which lasts one second of the EEG signals on channel AF3. The 3D Cube method tends to maintain spatial information between all frequency bands and channels [12, 18, 21].

The 9 × 9 matrix is the DE feature value for each frequency band. A combination of these four matrices is called the 3D cube representation. Each participant engaged in all the experimental processes generated 2400, 3728, and 6192 3D Cube for the DEAP, DREAMER, and AMIGOS datasets, respectively.

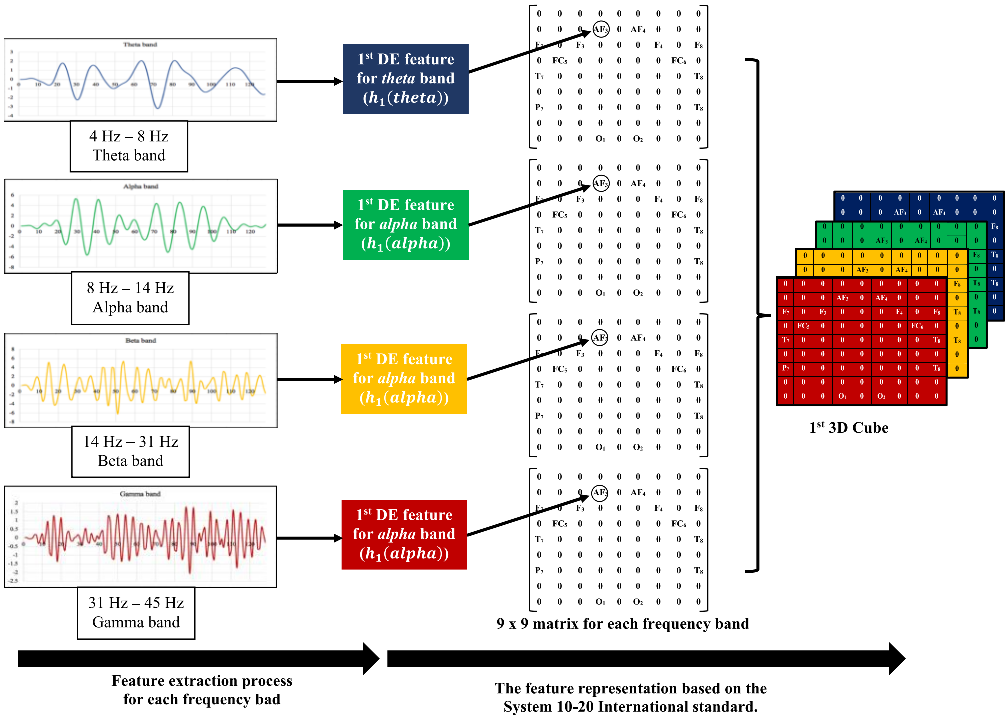

Figure 3. **Feature extraction and representation processes for AF3 in the DREAMER dataset**

3-4- Classification

The emotion classification process involves the use of a Continuous Capsule Network method. Yang et al. 2018, proposed a similar approach using the CNN method. The Continuous Convolution method is a procedure that uses the SAME padding and does not use pooling during the Convolution process. This approach aims to overcome the loss of feature information from the 3D Cube during the procedure [18]. It is essential to investigate this issue, considering that the 3D Cube as input data has a small size (9 × 9 × 4). Although, its use does not reduce the size of the 3D Cube data to a feature map on convolution. The Capsule Network applied in this study references its architecture designed by Liu et al. (2020) and Chao et al. (2019). Based on their proposal, its framework consists of three parts: the Convolution, Primary, and Emotion Capsules. The Capsule Network method in emotion recognition based on EEG signals can represent spatial information from EEG signals [1, 22, 23]. However, the Convolution process in this method can result in the loss of spatial data between channels of each frequency band. This study proposes a Continuous Convolution approach to the Capsule Network method to overcome these problems. Its architecture is a combination of Continuous Convolution with the Capsule Network architecture. It is designed for the recognition of the four categories of emotions, and its architecture is shown in Figure 4.

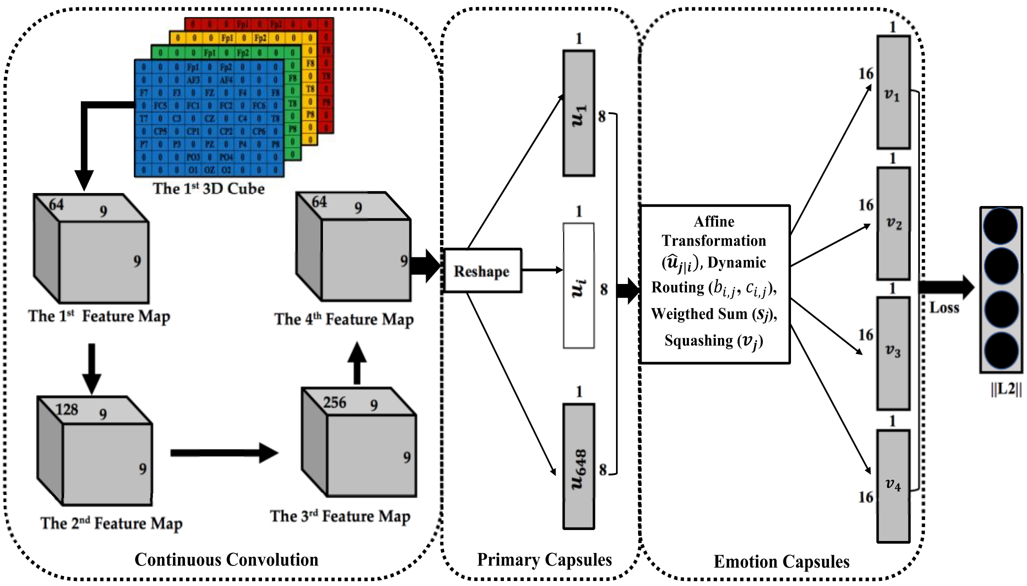

Figure 4. **Continuous Capsule Network architecture for four outputs**

As shown in Figure 4, the classification of emotions was divided into three stages, Continuous Convolution, Primary, and Emotion Capsules [1, 18, 23].

3-4-1- Continuous Convolution Stage

At this stage, the Continuous Convolution process is repeated four times, and its formula is represented by Equation 3:

$$FM[i]_{j,k} = \left(\sum_m \sum_n N_{[j-m,k-n]} F_{[m,n]}\right) + bF \tag{3}$$

where *FM[i]* is the i^{th} feature map matrix, *N* is the input data matrix, *F* is the filter Continuous Convolution matrix, *bF* is the bias value of the filter, *j* and *k* are pixel positions in the input data matrix, and *m* and *n* are the pixel positions in the input data matrix. In the first Continuous Convolution, the 3D cube data is convoluted using 64 filters with each having a size of 2 × 2 × 4 to produce the first feature map with a size of 9 × 9 × 64. The first feature map was then convoluted using 128 filters, with each having a size of 2 × 2 × 64. The second Continuous Convolution produces a feature map with a size of 9 × 9 × 128. The third procedure produced a feature map with a size of 9 × 9 × 256, using 256 filters, each has a size of 2 × 2 × 128. Finally, the third feature map is convoluted, using 64 filters with each having a size of 1 × 1 × 256 to produce one with a size of 9 × 9 × 64. This fourth Continuous Convolution process aims to combine feature maps and reduce computations. The four procedures used the SAME padding, no pooling, stride value of 1, and ReLU activation. The first Continuous Convolution process is shown in Figure 5.

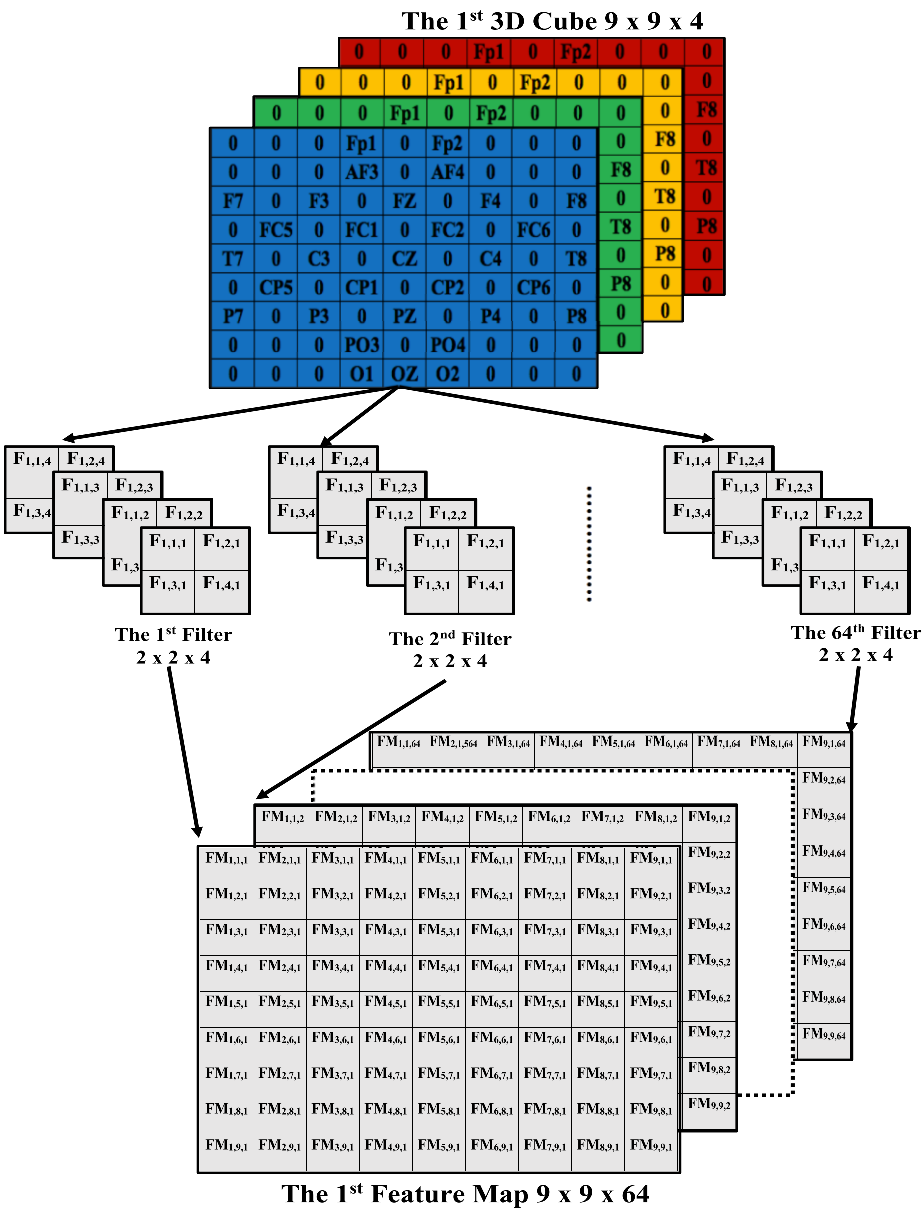

Figure 5. **The illustration of the first Continuous Convolution stage**

3-4-2- Primary Capsule Stage

At this stage, the feature map data obtained from the fourth Continuous Convolution process were divided into eight blocks, with each having a size of 9 × 9 × 8. After separation, each of them was reshaped to generate a vector u_i with a size of 8 × 648 (8 × 9 × 9). This vector is to encode brain areas by representing DE features into an 8 × 648 matrix [23]. Figure 6 shows the reshaping process of the fourth feature map.

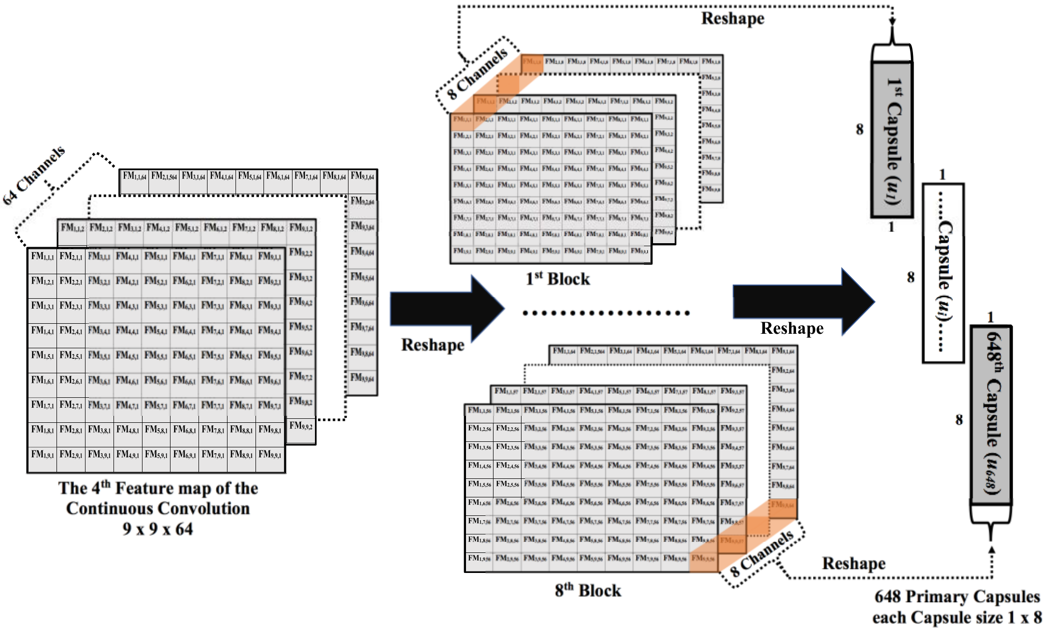

Figure 6. Illustration of the Primary Capsules stage

3-4-3- Emotion Capsule Stage

In this stage, the affine transformation, Weighted Sum, Dynamic Routing, and Squashing were performed [22, 41] as shown in Figure 7.

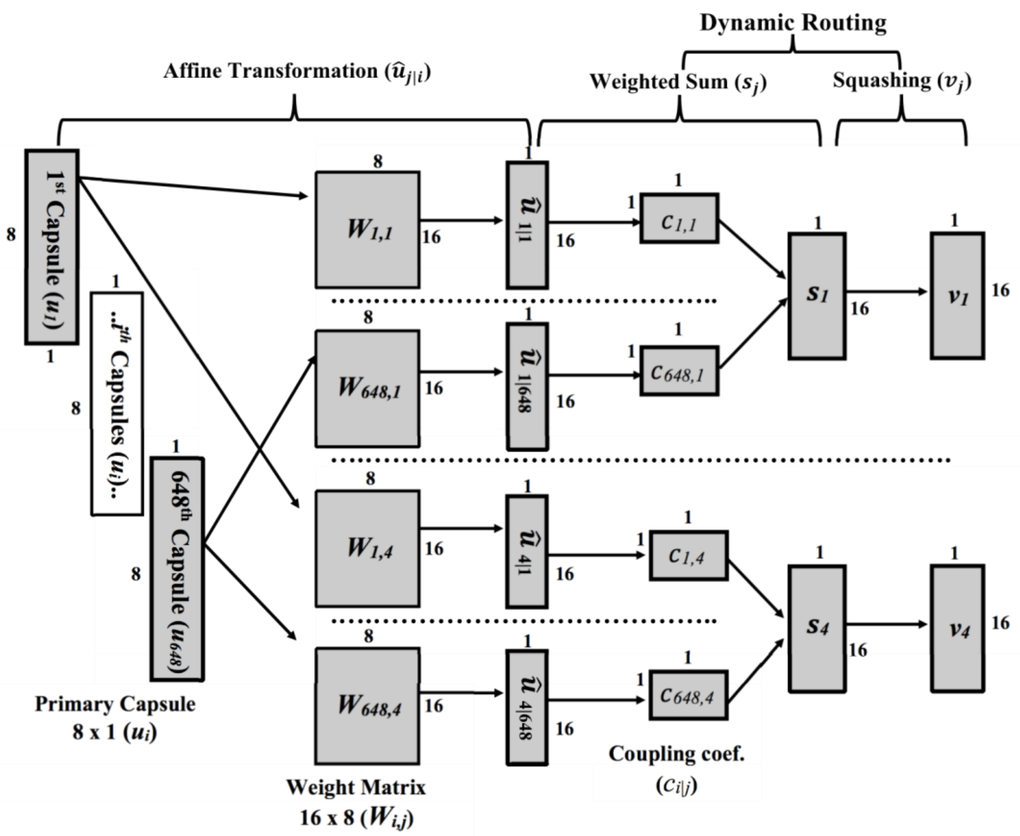

Figure 7. Illustration of the Emotion Capsules stage

Based on Figure 7, the value at the input node is a vector, and it emanates from u_1,..., u_{648} (each of the u_i is 1 × 8). The Capsule Network method does not use bias values as input, rather it is included in the next phase of the affine transformation process with the $W_{i|j}$ matrix. In the affine transformation process, the vector u_i (1×8) i is multiplied by the matrix $W_{i|j}$ (8 × 16) to produce a vector $\hat{u}_{j|i}$ measuring 1 × 16. This process was repeated 648 times to generate 648 vectors for one emotion class, where *j* represents each class, and *i* represents each vector's index. The value of the affine transformation process is obtained using Equation 4.

$$\hat{u}_{j|i} = W_{i|j} u_i \tag{4}$$

where u_i is the input vector for lower-level capsule (i), $W_{i|j}$ is the weight matrix, and $\hat{u}_{j|i}$ is the prediction vector. The transformation process represents the spatial relationship between the sub-objects and all others in a higher layer. Therefore, it is possible to predict whether these sub-objects correlate with higher-level ones. This process is followed by the Weighted Sum process between c_{ij} and the input vector $\hat{u}_{j|i}$. It aims to produce a Weighted Sum (s_j) by projecting several prediction vectors ($\hat{u}_{j|i}$) using the coupling coefficients (c_{ij}). According to Sabour et al. (2017), a dynamic routing process was used to determine the c_{ij} value. as shown in Table 1 [22].

Table 1. Dynamic routing algorithm

Routing dynamic algorithm
procedure ROUTING ($\hat{\boldsymbol{u}}_{j\|i}$, *r*, *l*)
for all capsule *i* in layer *l* and capsule *j* in layer (*l* + 1): $b_{ij} \leftarrow 0$
for *r* iterations do
for all capsule *i* in layer *l*: $c_i \leftarrow$ SoftMax function (b_i)
for all capsule *j* in layer (*l* + 1): $\boldsymbol{s_j} \leftarrow \sum_i c_{ij}\hat{\boldsymbol{u}}_{\boldsymbol{j\|i}}$
for all capsule *j* in layer (*l*+1): $\boldsymbol{v_j} \leftarrow$ squash function ($\boldsymbol{s_j}$)
for all capsule *i* in layer *l* and capsule *j in layer* (*l*+1): $b_{ij} \leftarrow b_{ij} + \hat{\boldsymbol{u}}_{\boldsymbol{j\|i}} . \boldsymbol{v_j}$
return $\boldsymbol{v_j}$

Furthermore, the $\boldsymbol{s_j}$ vector is activated using the Squashing activation function to obtain the probability values ($\boldsymbol{v_j}$) of the four emotional states (HAPV, HANV, LANV, and LAPV). Finally, the loss value was calculated based on the output and target using the L2 regularization method, as stated in Equation 5.

$$L_e = T_e \max(0, m^+ - \|\boldsymbol{v_e}\|)^2 + \lambda (1 - T_e) \max(0, \|\boldsymbol{v_e}\| - m^-)^2 \tag{5}$$

where T_e assuming the emotion class (e) matches the target, m^- is 0.1, m^+ is 0.9, v_e represents the output vector of class e, and λ is the down-weighting of the loss function. However, by default, the λ value was set at 0.5, and this study used a batch size of two for the optimization process by employing the Adam optimizer method and a routing coefficient value of three.

3-5- Evaluation Method

The task or classification model was assessed using K-fold cross-validation at this stage. 10-fold cross-validation (K=10) is the most commonly used method, with an illustration shown in Figure 8.

Total dataset per respondent

	1st fold	**2nd fold**	**3rd fold**	**4th fold**	**5th fold**	**6th fold**	**7th fold**	**8th fold**	**9th fold**	**10th fold**
Split 1	Valid	Train	Train	Train	Train	Train	Train	Train	Train	Train
Split 2	Train	Valid	Train	Train	Train	Train	Train	Train	Train	Train
Split 3	Train	Train	Valid	Train	Train	Train	Train	Train	Train	Train
Split 4	Train	Train	Train	Valid	Train	Train	Train	Train	Train	Train
Split 5	Train	Train	Train	Train	Valid	Train	Train	Train	Train	Train
Split 6	Train	Train	Train	Train	Train	Valid	Train	Train	Train	Train
Split 7	Train	Train	Train	Train	Train	Train	Valid	Train	Train	Train
Split 8	Train	Train	Train	Train	Train	Train	Train	Valid	Train	Train
Split 9	Train	Train	Train	Train	Train	Train	Train	Train	Valid	Train
Split 10	Train	Train	Train	Train	Train	Train	Train	Train	Train	Valid

Figure 8. Illustration of k-fold cross validation

In K-Fold cross-validation, the initial sampling was divided into K subsamples. One subsample (K) was used as the test set (T), while the K sample was employed for training (S). Each subsample was validated once, and the mean K value was used as the final result, moreover, cross-validation was repeated K times.

3-6- Performance Evaluation

In this study, the performance of the classification process was evaluated using the accuracy rate value. It is a proportion of correctly classified samples to their total, calculated using Equation 6.

$$Accuracy\ Rate\ (R) = \frac{True\ Labels}{True\ Labels + False\ Labels} \times 100\% \tag{6}$$

True label is the number of samples that can recognize two or four emotion labels, while the False one is the number of samples that cannot identify these labels. The four emotional labels were high arousal and positive valence, high arousal and negative valence, low arousal and positive valence, and low arousal and negative valence. Furthermore, two emotional labels were high and low for arousal and valence.

4- Results and Discussion

A Continuous Capsule Network is used to represent spatial information and overcome its loss in the EEG signals during the classification process. In addition, the feature extraction and representation process use Differential Entropy and 3D Cube methods to support the classification procedure. The DE is used to recognize low-frequency and characterize the time series [15, 16]. The features generated from this method are the most accurate and stable for emotion recognition compared to the others [17-19]. The DE features were represented using the 3D cube method. This method represents the spatial information between all channels and frequency bands. Two experimental stages were carried out to test the proposed Continuous Capsule Network method, namely:

4-1- First Experiment

This experiment aimed to determine the optimal Continuous Capsule Network architecture based on the accuracy and the number of parameters using the hyperparameter process [1, 23]. This procedure was carried out in two stages. In the first one, the hyperparameter process is based on the number of Continuous Convolution layers, while in the second, it is dependent on the kernel size [1]. Four test scenarios were carried out in the first stage and three in the second to test the Continuous Capsule Network architecture. Furthermore, the architecture generated from each scenario selects the best approach, namely the Continuous Capsule Network method, based on the highest accuracy value and least number of parameters [1, 23]. Accuracy and the number of parameters is the benefit and cost criteria. Their values were determined using the normalization techniques such as in Equation 7 [42].

$$r_{ij} = \begin{cases} \frac{x_{ij}}{Max_i(x_{ij})}, & If\ x_{ij}\ is\ a\ benefit\ criterion \\ \frac{Min_i(x_{ij})}{x_{ij}}, & If\ x_{ij}\ is\ a\ cost\ criterion \end{cases} \tag{7}$$

where x_{ij} is a cost or benefit criterion, r_{ij} is the normalized criterion, *i* is an alternative, and *j* is an criteria. Subsequently, the weighted sum process is carried out using Equation 8 for all criteria in each alternative.

$$V_i = \sum_{j=1}^{n} W_j r_{ij} \tag{8}$$

where V_i is the weighted sum value for each alternative, W_j is the weighted value for each criteria, and the weight used in this experiment is 1 for all criteria.

1. In the first stage, the hyperparameter process is used to determine the optimal Continuous Capsule Network architecture based on the number of Continuous Convolution layers. Four scenarios were implemented to ascertain the optimal Continuous Capsule Network architecture:

   - The first scenario involves the design of the first Continuous Capsule Network architecture, where the number of first, second, and third Continuous Convolution layers were 64, 128, and 256, respectively.
   - The second scenario comprises designing the second Continuous Capsule Network architecture, where the number of 1st, 2nd, 3rd, and 4th Continuous Convolution layers were 64, 128, 256, and 64, respectively.
   - In the third scenario, the third Continuous Capsule Network architecture was designed, and the number of the first, second, third, and fourth Continuous Convolution layers were 64, 128, 256, and 128, respectively.
   - The fourth scenario involves the design of the fourth Continuous Capsule Network architecture, where the number of 1st, 2nd, 3rd, and 4th Continuous Convolution layers are 64, 128, 256, and 512, respectively.

By default, the kernel size used in these four architectures is 4 × 4 × 4, 4 × 4 × 64, 4 × 4 × 128, and 4 × 4 × 256. In addition, four scenarios were tested using three secondary data sets to identify four classes of emotions. Tables 2 to 4 show the normalization results based on the accuracy and number of parameters for the DEAP, DREAMER, and AMIGOS datasets. Based on Table 2, Architecture 2 was ranked first with a value of 1.977297, although 4 produced the highest accuracy, and the number of parameters used was greater than the other three architectures. The DREAMER dataset also selected Architecture 2, which had a value of 1.955176. However, 4 produced the highest accuracy, and the number of parameters used was greater than the other three architectures. Based on the tests carried out on the AMIGOS dataset, Architecture 2 was ranked first with a value of 1.977297. On the other hand, 4 produced the highest accuracy, and the number of parameters used was greater than the other three. Architecture 2 of the Continuous Capsule Network was selected and retested in the second stage.

Table 2. Normalization results for the first stage of the DEAP dataset

Architectures	Average accuracy values (x_{i1}) %	Number of parameters (x_{i2})	Normalization (r_{i1})	Normalization (r_{i2})	Summation (V_i)	Ranking
Architecture 1	58.8815	1987008	0.992908	0.631043	1.623951	3
Architecture 2	57.9557	1253888	0.977297	1	1.977297	1
Architecture 3	58.6797	1847872	0.989505	0.678558	1.668062	2
Architecture 4	59.3021	5411776	1	0.231696	1.231696	4

Table 3. Normalization results for the first stage of the DREAMER dataset

Architectures	Average accuracy values (x_{i1}) %	Number of parameters (x_{i2})	Normalization (r_{i1})	Normalization (r_{i2})	Summation (V_i)	Ranking
Architecture 1	47.756	1987008	0.975761	0.631043	1.606804	3
Architecture 2	46.7485	1253888	0.955176	1	1.955176	1
Architecture 3	47.9675	1847872	0.980084	0.678558	1.658641	2
Architecture 4	48.9423	5411776	1	0.231696	1.231696	4

Table 4. Normalization results for the first stage of the AMIGOS dataset

Architectures	Average accuracy values (x_{i1}) %	Number of parameters (x_{i2})	Normalization (r_{i1})	Normalization (r_{i2})	Summation (V_i)	Ranking
Architecture 1	47.4117	1987008	0.980695	0.631043	1.611738	3
Architecture 2	47.8737	1253888	0.990252	1	1.990252	1
Architecture 3	47.8815	1847872	0.990412	0.678558	1.66897	2
Architecture 4	48.345	5411776	1	0.231696	1.231696	4

2. In the second stage, the hyperparameter process was used to determine the optimal Continuous Capsule Network architecture based on the kernel size. This was designed using three scenarios, and the architecture tested was selected from stage 1 (Architecture 2).
    - The first scenario used a kernel size of 2 × 2 × 4, 2 × 2 × 64, and 2 × 2 × 128 for the 1st, 2nd, and 3rd convolution layers, and in contrast, the fourth one has a size of 1 × 1 × 256.
    - In the second scenario, used a kernel size of 3 × 3 × 4, 3 × 3 × 64, and 3 × 3 × 128 for the 1st, 2nd, and 3rd convolution layers, and in contrast, the fourth one has a size of 1 × 1 × 256.
    - In the third scenario, the kernel size used for the 1st, 2nd, and 3rd convolution layers is 4 × 4 × 4, 4 × 4 × 64, and 4 × 4 × 128, while the fourth one is 1 × 1 × 256.

The normalization results of the three test processes carried out on the three secondary datasets are shown in Tables 5 to 7. According to Table 5, Architecture 2A was ranked first with a value of 1.959773. However, Architecture 2C produced higher average accuracy than the other two. The addition of several parameters did not significantly affect the average accuracy. Architecture 2A was also ranked first with a value of 1.96987 in the DREAMER dataset test, as shown in Table 6. The value of the parameter used in this architecture is 513536. Regarding Table 7, Architecture 2A was selected based on the AMIGOS dataset, and it yielded a value of 1.968684, while the number of parameters used was 513536.

Table 5. Normalization results for the second stage of the DEAP dataset

Architectures	Average accuracy values (x_{i1}) %	Number of parameters (x_{i2})	Normalization (r_{i1})	Normalization (r_{i2})	Summation (V_i)	Ranking
Architecture 2A	56.0742	513536	0.959773	1	1.959773	1
Architecture 2B	57.5326	719616	0.984734	0.713625	1.698359	2
Architecture 2C	58.4245	1008128	1	0.509396	1.509396	3

Table 6. Normalization results for the second stage of the DREAMER dataset

Architectures	Average accuracy values (x_{i1}) %	Number of parameters (x_{i2})	Normalization (r_{i1})	Normalization (r_{i2})	Summation (V_i)	Ranking
Architecture 2A	45.3343	513536	0.96987	1	1.96987	1
Architecture 2B	46.2623	719616	0.989723	0.713625	1.703348	2
Architecture 2C	46.7426	1008128	1	0.509396	1.509396	3

Table 7. Normalization results for the second stage of the AMIGOS dataset

Architectures	Average accuracy values (x_{i1}) %	Number of parameters (x_{i2})	Normalization (r_{i1})	Normalization (r_{i2})	Summation (V_i)	Ranking
Architecture 2A	46.3221	513536	0.968684	1	1.968684	1
Architecture 2B	46.6836	719616	0.976243	0.713625	1.689868	2
Architecture 2C	47.8196	1008128	1	0.509396	1.509396	3

The two hyperparameter stages in this experiment were used to design the optimal Continuous Capsule Network architectures where the number of the 1^{st}, 2^{nd}, 3^{rd}, and 4^{th} Continuous Convolution layers are 64, 128, 512, and 64, respectively. Furthermore, the kernel size used was $2 \times 2 \times 4$, $2 \times 2 \times 64$, and $2 \times 2 \times 128$ for the 1^{st}, 2^{nd}, and 3^{rd} layers, whereas the 4^{th} one is $1 \times 1 \times 256$. The stride value used in this architecture is one, and it does not use pooling rather it involves zero padding. An illustration of the selected architecture is shown in Figure 4 to identify four classes of emotions. Considering that the Continuous Capsule Network is an improvement of the Continuous Convolution Neural Network and Capsule Network methods, its architecture obtained from this experiment was compared with several related studies, specifically in terms of recognizing the two classes of emotions. A comparison of its accuracy with related classification processes is shown in Table 8.

Based on Table 8, the average accuracy of the Continuous Capsule Network in this study is higher than in other investigations. Irrespective of the extraction and representation features applied to the DE and 3D Cube methods, and the implemented Continuous Capsule Network approach produced higher accuracy than the Continuous CNN technique proposed by Yang et al. (2018) [18]. Considering that the Capsule Network is an improvement of the CNN method, it tends to represent spatial data from the EEG signals and can be trained with less information [22, 23]. The application of the Continuous Capsule Network also produces higher accuracy than the Capsule Network procedure proposed by Liu et al. (2020) and Chao et al. (2019). Although Liu et al. (2020) and Chao et al. (2019) applied it during the classification process, the loss of spatial information from the EEG signals causes the resulting accuracy to be less than the Continuous Capsule Network method [1, 23]. This usually occurs during the Convolution process using the Continuous procedure. The spatial information contained in the input data (3D Cube) is maintained, as evidenced by the size of the feature map generated for each convolution that does not decrease.

Table 8. Comparison of the accuracy of the Continuous Capsule Network method with several related studies in recognizing two classes of emotions

No.	Researchers	Methods	DEAP		DREAMER		AMIGOS	
			Arousal	Valence	Arousal	Valence	Arousal	Valence
1	Yang et al. (2018) [18]	Feature Extraction using DE; Feature representation using 3D Cube; Classification using Continuous CNN	69.55%	68.56%				
2	Chao et al. (2019) [1]	Feature Extraction using PSD, Feature representation using MFM Classification using *Capsule Network*	68.28%	66.73%				
3	Liu et al. (2020) [23]	Representation features using MLF; Extraction features & Classification using *Capsule Network*	64.36%	62.57%				
4	Present study (2022)	Feature Extraction using DE; Feature representation using 3D Cube; Classification using *Continuous Capsule Network*	72.63%	71.21%	74.40%	61.35%	69.50%	63.07%

Besides, the number of parameters used in the Continuous Capsule Network method is less than those applied in the Continuous CNN procedure. Yang et al. (2018) stated that the Continuous CNN method used 5989892 parameters [18]. The Capsule Network approach proposed by Liu et al. (2020) utilized 110850000 parameters [23]. Compared to the two procedures proposed in previous studies, the parameters used in the Continuous Capsule Network were relatively 513536. Based on this experiment, it was stated that the application of this approach represents and maintains spatial information from the EEG signals.

4-2- Second Experiment

Irrespective of the fact that a hyperparameter was used to determine the optimal Continuous Capsule Network architecture, its accuracy was less than 75%. Therefore, applying a baseline reduction approach is crucial for boosting emotional recognition accuracy. This experiment implemented a baseline reduction procedure after the feature extraction process was performed, as illustrated in Fig. 1. A comparison of the accuracies of Continuous Capsule Network methods with and without baseline reduction for the four emotion categories in the three secondary datasets is shown in Figures 9 to 11.

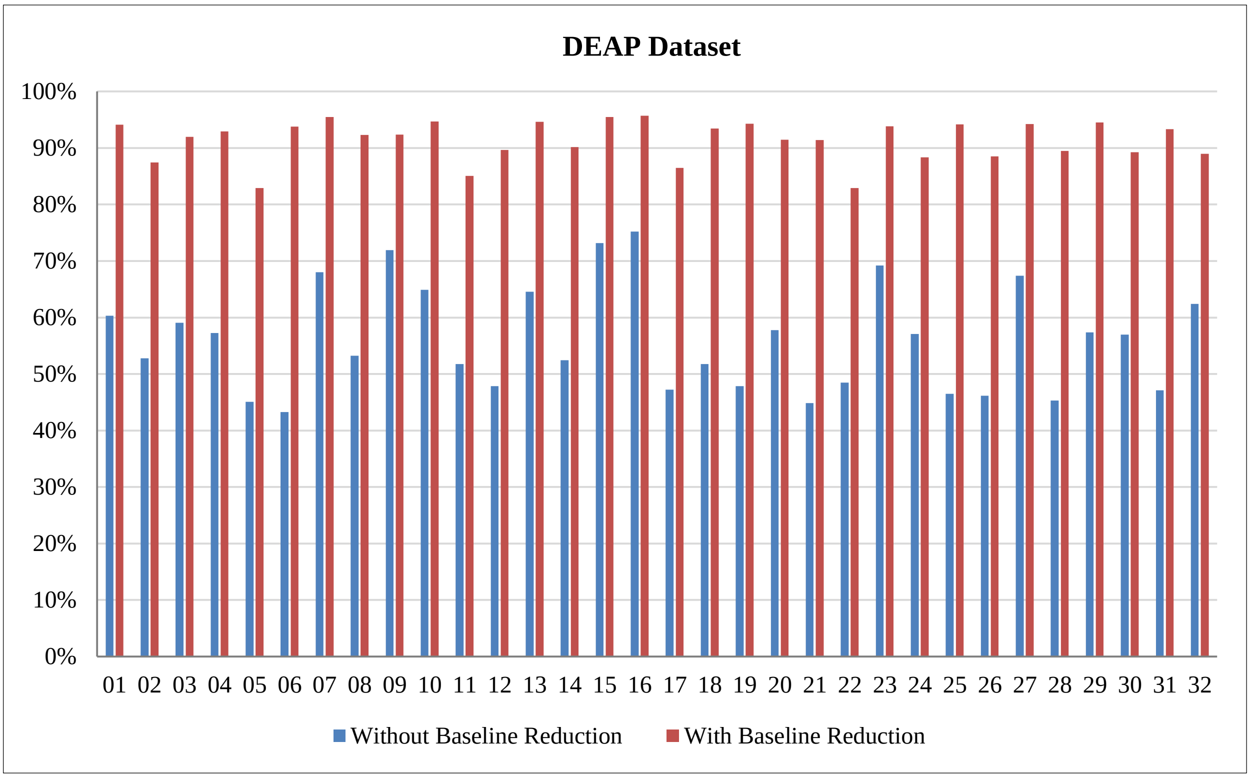

Figure 9. Comparison of accuracy with and without baseline reduction approach on the DEAP dataset

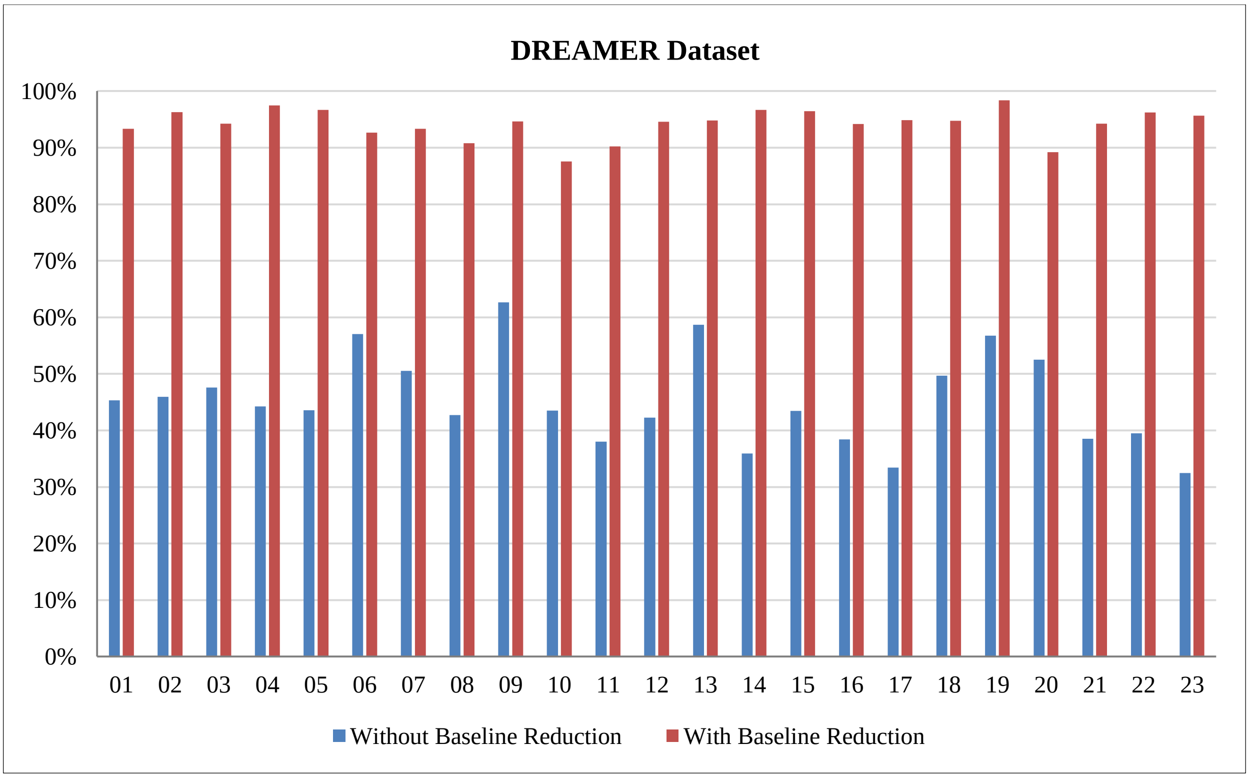

Figure 10. Comparison accuracy with and without baseline reduction approach on the DREAMER dataset

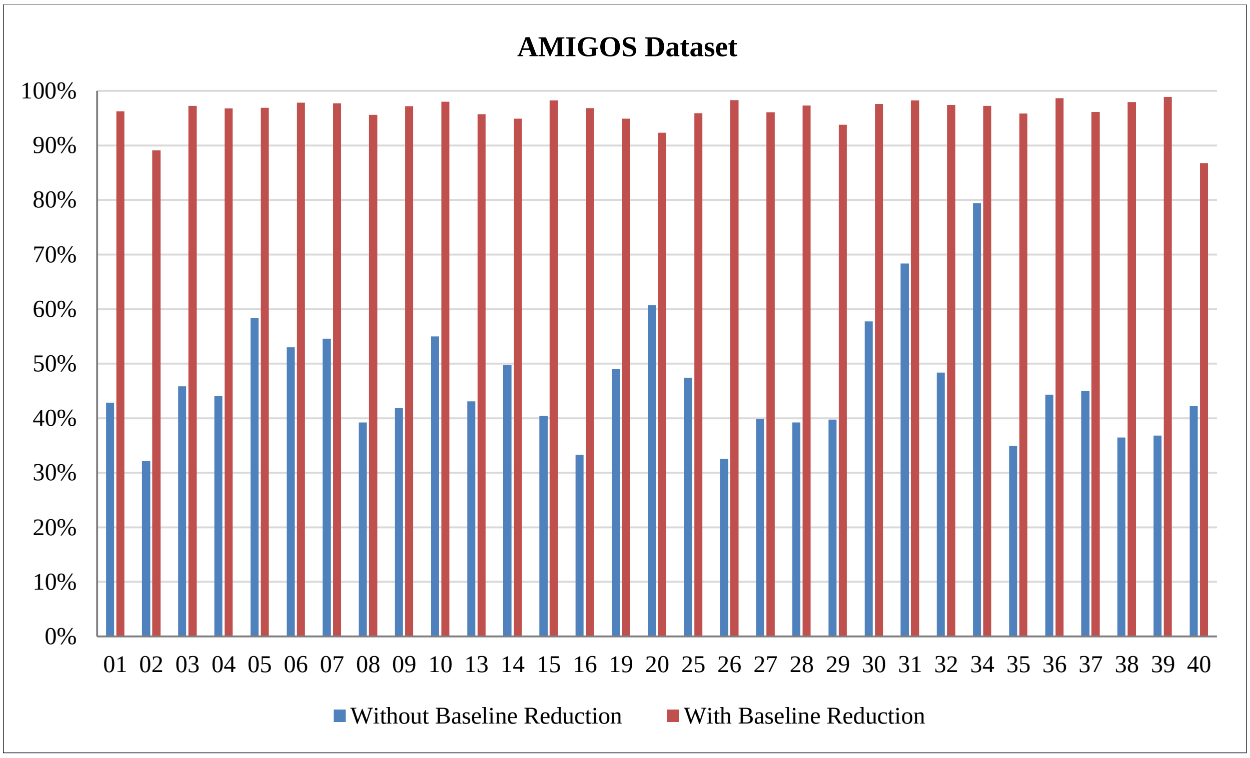

Figure 11. Comparison accuracy with and without baseline reduction approach on the AMIGOS dataset

Figure 9 shows that the implemented baseline reduction in the Continuous Capsule Network tends to boost the accuracy of emotion recognition in the DEAP dataset by an average increase of 35.28%. In addition, its standard deviation using the baseline reduction is 3.58.

Interestingly, certain improvements were also observed in the DREAMER dataset, as shown in Figure 10. The application of baseline reduction in the Continuous Capsule Network method boosts the accuracy of all subjects, with an average increase of 48.90 %. Its standard deviation using the baseline reduction approach was 2.68.

An increase in the accuracy was also observed in the AMIGOS dataset, as shown in Figure 11. In addition, there were 40 participants, and only 31 were included in the test. The others had missing values in their EEG signals data and emotion labels. Based on the data acquired from the 31 participants, the implementation of baseline reduction in the Continuous Capsule Network method triggered the accuracy of the emotion recognition, with an average increase of 49.87% and a standard deviation of 2.65. The Wilcoxon statistical test was carried out when the baseline reduction approach was applied to the Continuous Capsule Network method for the four emotion categories to prove the significant increase in accuracy. This was aimed to compare the accuracy of the Continuous Capsule Network that uses the baseline reduction approach to the one that does not use the procedure. Two hypotheses were developed to support the Wilcoxon test:

- *Ho*: There was an insignificant difference in the accuracy of the results obtained with and without baseline reduction when applied to the Continuous Capsule Network. This hypothesis is true if the 2-tailed value is greater than or equal to the degree of significance ($\alpha = 0.05$), expressed as 2-tailed >= α.
- *Ha*: When applied to the Continuous Capsule Network, there was a significant difference in the accuracy of the results obtained with and without baseline reduction. This hypothesis is true when the 2-tailed value is greater than or equal to the degree of significance ($\alpha = 0.05$), expressed as 2-tailed < α.

The Wilcoxon test results realized from the application of baseline reduction in the Continuous Capsule Network method for the DEAP, DREAMER, and AMIGOS datasets are shown in Table 9.

Table 9. Wilcoxon test results for applying the baseline reduction in the Continuous Capsule Network method for four emotion categories

No.	Secondary dataset	Average Accuracy		Wilcoxon Test				
		Without baseline reduction	With baseline reduction	Positive Ranks	Negative Ranks	Ties	2-tailed	Hypotheses
1	DEAP	56.07%	91.35%	32	0	0	0.000	*Ha* accepted
2	DREAMER	45.33%	94.23%	23	0	0	0.000	*Ha* accepted
3	AMIGOS	46.32%	96.20%	31	0	0	0.000	*Ha* accepted

In accordance with Table 9, a combination of baseline reduction and the Continuous Capsule Network method used to identify the four emotion categories produced average accuracies of 91.35%, 94.23%, and 96.20% for the DEAP, DREAMER, and AMIGOS datasets, respectively. Its accuracy is higher than the approach without baseline reduction. Furthermore, several values are generated from the Wilcoxon test, such as Positive and Negative Ranks, Ties, and 2-tailed. The Positive Ranks are the data samples obtained from the Continuous Capsule Network method, with the baseline reduction having higher accuracy than the one without this procedure. The Negative Ranks were the data samples from the Continuous Capsule Network method with baseline reduction have lower accuracy than those without this approach. Ties are the data sample from the Continuous Capsule Network method with baseline reduction, which has the same accuracy as those realized without the procedure. The 2-tailed approach (Asymp. sig.) is the significant value between the Continuous Capsule Network method with and without baseline reduction. It was used to determine the hypothesis. Ha was accepted if the 2-tailed value was < 0.05; otherwise, Ho was accepted. Based on the Wilcoxon test carried out on the three secondary datasets, using the baseline reduction and the Continuous Capsule Network method for the emotion classification process significantly improves the emotion recognition accuracy based on EEG signals (2-tailed = 0.000). The average accuracy realized with the Continuous Capsule Network method with baseline reduction was compared with previous studies, particularly in recognizing the four emotional classes, as shown in Table 10.

Table 10. Comparison of accuracy results between the proposed method and other methods for the four categories of emotions

No.	Researchers	Methods	DEAP	DREAMER	AMIGOS
1	Liu et al. (2018) [27]	Extraction feature using ResNets dan LFCC; Classification using kNN.	86.05%	-	-
2	Soroush et al. (2019) [28]	Extraction feature using Angle Space; Classification using MSVM.	81.67%.	-	-
3	Zheng et al. (2016) [29]	Extraction feature using DE; Classification using GELM.	69.67%	-	-
4	Mei and Xu. (2017) [30]	Extraction feature using Pearson Correlation Coefficient; Classification using CNN.	75%	-	-
5	Zhao et al. (2020) [31]	Baseline reduction using Difference; Extraction feature and Classification using CNN.	93.53%	-	95.95%
6	Present study (2022)	Extraction feature using DE; Baseline reduction using Difference; Representation features using 3D Cube; Classification using Continuous Capsule Network.	91.35%	94.23%	96.20%

Considering that the recognition of emotions for the two categories was absolutely investigated in previous studies, the resulting accuracy compared to preliminary analyses is shown in Table 11.

Table 11. Comparison of accuracy results between the proposed method and other methods for the two categories of emotions

No.	Researchers	Methods	DEAP		DREAMER		AMIGOS	
			Arousal	Valence	Arousal	Valence	Arousal	Valence
1	Liu et al. [32]	Extraction feature using EMD; Classification using kNN.	86.46%	84.90%				
2	He et al. [33]	Extraction feature using MEMD; Classification using SVM.	67.90%	70.90%				
3	Parui et al. [34]	Extraction feature using Hjorth Parameters, Autoregressive Parameters, dan Wavelet Domain Features; Classification using XGboost.	74.20%	75.97%				
4	Pan et al. [35]	Extraction feature using DE; Classification using LORSAL.	77.17%	77.03%				
5	Garg & Verma [36]	Extraction feature and Classification using CNN with GoogleNet architecture.	61.23%	92.19%				
6	Song et al. [10]	Extraction feature using DE; Classification using DGCNN.			84.54%	86.23%		
7	Huang et al. [37]	Extraction feature using STFT; Classification using CNN.	84.50%	83.70%				
8	Yang et al. [18]	Extraction feature using DE; Baseline reduction using Difference; Representation features using 3D Cube; Classification using Continuous CNN.	90.24%	89.45%				
9	Wirawan et al. [21]	Extraction feature using DE; Baseline reduction using Relative Difference; Representation features using 3D Cube; Classification using Continuous CNN.	82.10%	81.47%				
10	Wardoyo et al. [20]	Extraction feature using DE; Oversampling data using Radius SMOTE; Representation features using 3D Cube; Classification using Continuous CNN.	82.11%	78.99%				
11	Present study	Extraction feature using DE; Baseline reduction using Difference; Representation features using 3D Cube; Classification using Continuous Capsule Network.	93.69%	92.85%	96.66%	96.05%	97.96%	97.32%

Several machine learning-based emotional classification methods have been evaluated based on EEG signals, including the KNN, SVM, XGboost, LORSAL, and GELM approaches [27, 28, 32-35]. However, this procedure is not highly accurate (less than 90%). This is because machine learning-based classification methods are unable to characterize EEG signal information in depth [12, 43]. The process is essential, considering that the signal has low-frequency characteristics and contains many noises [12]. To overcome this problem, a deep learning-based approach using the CNN method was investigated in previous studies [10, 30, 36, 37]. Although, it is unable to characterize the spatial information between channels and frequency bands of the EEG signals. The CNN method further requires much training data to improve the classification performance. The availability of secondary datasets on emotion recognition, specifically based on EEG signals, is still minute. Several publicly available datasets are in an unbalanced condition, and efforts to overcome this issue using the Radius SMOTE approach were employed by Wardoyo et al. (2022). This method was used to curb data limitations and imbalances [20]. Interestingly, these efforts are not absolutely accurate. The Capsule Network method is also used to overcome this problem, while the application of the Continuous Capsule Network technique proposed in this study can characterize the spatial information of the EEG signals as well as to overcome its loss and maintain it when the Convolution process is performed.

In addition, the challenge of recognizing emotions is also strongly influenced by the differences in the characteristics of each participant, such as gender, age, education, and personality traits. The baseline reduction approach was employed to overcome this problem [12, 18]. Its implementation in the Capsule Network process can produce significantly increased accuracy compared to without its usage. Irrespective of the fact that this approach was also reviewed by Yang et al. (2018), the CNN method employed in this study was unable to characterize the spatial information of the EEG signals, and the resulting accuracy is less than the proposed technique [18]. The baseline reduction approach was also investigated by Zhao et al. (2020), which used the CNN procedure to produce higher accuracy compared to the proposed method, involving the use of the DEAP test dataset and not the AMIGOS [31]. Another effort to optimize the baseline reduction approach was also reviewed by Wirawan et al. (2021) [21]. It was not maximized because the resulting accuracy was still less than the proposed method [20]. Based on the results and findings of this study, it was stated that the application of baseline reduction and the Continuous Capsule Network methods significantly increased the accuracy of emotion recognition based on EEG signals. Although the baseline reduction approach tends to improve emotion recognition accuracy, its EEG signals may be impaired [12]. This disturbance causes the resulting accuracy not to be maximal in the three secondary datasets. Therefore, the future study challenge is optimizing the baseline reduction approach to increase emotion recognition accuracy based on EEG signals.

5- Conclusion

Despite the use of the capsule network method can characterize spatial information from EEG signals, its loss reduces the performance of emotion recognition, resulting in poor accuracy. Therefore, this study applied the continuous capsule network as a solution to overcome these problems. Based on the two experiments carried out, the optimal architecture of this method for emotion recognition has (1) 1^{st}, 2^{nd}, 3^{rd}, and 4^{th} Continuous Convolution layers of 64, 128, 256, and 64, respectively, and (2) a kernel sizes of $2\times2\times4$, $2\times2\times64$, and $2\times2\times128$ for the 1^{st}, 2^{nd}, and 3^{rd} layers, while the 4^{th} is $1\times1\times256$. In addition, several methods are used to support the Continuous Capsule Network in the classification process, such as the DE approach for the feature extraction process, which is represented using the 3D Cube procedure. The choice of DE and 3D Cube methods is based on their ability to characterize spatial and low-frequency information. Based on the tests on the three secondary datasets, the Continuous Capsule Network produces higher accuracy than some of the proposed methods in previous studies, both for recognizing four and two classes of emotions. The number of parameters used in the continuous capsule network is less than those employed in previous studies.

Furthermore, applying the baseline reduction to the Continuous Capsule Network method increases emotion recognition accuracy. Although it also enhances emotion recognition accuracy, the baseline EEG signals used in this approach may be impaired. This disturbance causes the resulting accuracy not to be maximal in the three secondary datasets. The future study challenge is optimizing the baseline reduction approach to increase emotion recognition accuracy.

6- Declarations

6-1- Author Contributions

Conceptualization, R.W. and I.M.A.W.; preparation of introduction, I.M.A.W.; preparation of related work, I.M.A.W., R.W., D.L. and S.K.; data collection and analysis, I.M.A.W.; preparation of the preprocess methodology, I.M.A.W., R.W., D.L. and S.K; compilation of methods and theory of feature extraction, I.M.A.W. and R.W.; compilation of theory and methodology Baseline Reduction, I.M.A.W. and D.L.; compilation of the theory and methodology of feature representation, I.M.A.W. and R.W.; compilation of the theory and methods of classification process, I.M.A.W., R.W., D.L. and S.K; formulation of validation theory and methodology, I.M.A.W. and R.W.; compilation of theory and methods of accuracy measurement, I.M.A.W. and R.W.; development of feature extraction method, I.M.A.W.; development of representation feature extraction, I.M.A.W.; development of the classification method, I.M.A.W.; preparation of results and discussion, I.M.A.W., R.W., D.L., and S.K.; drawing conclusions, I.M.A.W., R.W., D.L., and S.K. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

This study used three secondary datasets, including the DEAP, DREAMER, and AMIGOS datasets. Where the DEAP dataset can be freely accessed at "*https://anaxagoras.eecs.qmul.ac.uk/request.php?dataset=DEAP*", the DREAMER dataset is freely accessible at "*https://zenodo.org/record/546113#.YoB2n5MzaRs*", and the dataset AMIGOS is freely accessible at "*https://anaxagoras.eecs.qmul.ac.uk/request.php?dataset=AMIGOS*".

6-3- Funding

The Ministry of Education Culture, Research, and Technology, Republic Indonesia, funded the publication of this article through the 2022 Penelitian Disertasi Doktor fund with Master Contract Number 089/E5/PG.02.00.PT/2022 and Derivative Contract Number 1902/UN1/DITLIT/Dit-Lit/PT.01.03/2022.

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Not applicable.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Chao, H., Dong, L., Liu, Y., & Lu, B. (2019). Emotion recognition from multiband EEG signals using CAPSNET. Sensors (Switzerland), 19(9), 2212. doi:10.3390/s19092212.

[2] Murali Krishna, N., Sekaran, K., Vamsi, A. V. N., Pradeep Ghantasala, G. S., Chandana, P., Kadry, S., Blazauskas, T., & Damasevicius, R. (2019). An Efficient Mixture Model Approach in Brain-Machine Interface Systems for Extracting the Psychological Status of Mentally Impaired Persons Using EEG Signals. IEEE Access, 7, 77905–77914. doi:10.1109/ACCESS.2019.2922047.

[3] Scherer, K. R., Schorr, A., & Johnstone, T. (2001). Appraisal processes in emotion: Theory, methods, research. Oxford University Press, Oxford, United Kingdom.

[4] Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. Proceedings IEEE International Conference on Advanced Learning Technologies. doi:10.1109/icalt.2001.943850.

[5] Gottardo, E., Pimentel, A.R. (2019). Inferring Students' Emotions Using a Hybrid Approach that Combine Cognitive and Physical Data. Enterprise Information Systems, ICEIS 2018. Lecture Notes in Business Information Processing, 363, Springer, Cham, Switzerland. doi:10.1007/978-3-030-26169-6_14.

[6] Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., & Yang, X. (2018). A review of emotion recognition using physiological signals. Sensors (Switzerland), 18(7). doi:10.3390/s18072074.

[7] Ekman, P., Friesen, W. V., & Simons, R. C. (1985). Is the startle reaction an emotion?. Journal of personality and social psychology, 49(5), 1416, doi:10.1037/0022-3514.49.5.1416.

[8] Li, Y., Huang, J., Zhou, H., & Zhong, N. (2017). Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. Applied Sciences (Switzerland), 7(10). doi:10.3390/app7101060.

[9] Hu, X., Chen, J., Wang, F., & Zhang, D. (2019). Ten challenges for EEG-based affective computing. Brain Science Advances, 5(1), 1–20. doi:10.1177/2096595819896200.

[10] Song, T., Zheng, W., Song, P., & Cui, Z. (2020). EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. IEEE Transactions on Affective Computing, 11(3), 532–541. doi:10.1109/TAFFC.2018.2817622.

[11] Kawintiranon, K., Buatong, Y., & Vateekul, P. (2016). Online music emotion prediction on multiple sessions of EEG data using SVM. 13th International Joint Conference on Computer Science and Software Engineering. doi:10.1109/JCSSE.2016.7748921.

[12] Wirawan, I.M.A., Wardoyo, R., & Lelono, D. (2022). The challenges of emotion recognition methods based on electroencephalogram signals: A literature review. International Journal of Electrical and Computer Engineering, 12(2), 1508–1519. doi:10.11591/ijece.v12i2.pp1508-1519.

[13] Bhandari, N. K., & Jain, M. (2020). Emotion recognition and classification using EEG: A review. International Journal of Scientific and Technology Research, 9(2), 1827–1836.

[14] Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion, 59(January), 103–126. doi:10.1016/j.inffus.2020.01.011.

[15] Pane, E. S., Wibawa, A. D., & Pumomo, M. H. (2018, November). Channel selection of EEG emotion recognition using stepwise discriminant analysis. 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM). doi:10.1109/CENIM.2018.8711196.

[16] Chen, D. W., Miao, R., Yang, W. Q., Liang, Y., Chen, H. H., Huang, L., Deng, C. J., & Han, N. (2019). A feature extraction method based on differential entropy and linear discriminant analysis for emotion recognition. Sensors (Switzerland), 19(7), 1631. doi:10.3390/s19071631.

[17] Li, J., Zhang, Z., & He, H. (2018). Hierarchical Convolutional Neural Networks for EEG-Based Emotion Recognition. Cognitive Computation, 10(2), 368–380. doi:10.1007/s12559-017-9533-x.

[18] Yang, Y., Wu, Q., Fu, Y., & Chen, X. (2018). Continuous convolutional neural network with 3D input for EEG-based emotion recognition. International Conference on Neural Information Processing. doi:10.1007/978-3-030-04239-4_39.

[19] Jiang, H., & Jia, J. (2019, November). Research on EEG emotional recognition based on LSTM. International Conference on Bio-Inspired Computing: Theories and Applications. doi:10.1007/978-981-15-3415-7_34.

[20] Wardoyo, R., Wirawan, I. M. A., & Pradipta, I. G. A. (2022). Oversampling Approach Using Radius-SMOTE for Imbalance Electroencephalography Datasets. Emerging Science Journal, 6(2), 382–398. doi:10.28991/ESJ-2022-06-02-013.

[21] Wirawan, I. M. A., Wardoyo, R., Lelono, D., Kusrohmaniah, S., & Asrori, S. (2021). Comparison of Baseline Reduction Methods for Emotion Recognition Based On Electroencephalogram Signals. 2021 6th International Conference on Informatics and Computing (ICIC). doi:10.1109/ICIC54025.2021.9632948.

[22] Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. Advances in Neural Information Processing Systems 30, NeurIPS Proceedings, 4-9 December, CA, United States.

[23] Liu, Y., Ding, Y., Li, C., Cheng, J., Song, R., Wan, F., & Chen, X. (2020). Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network. Computers in Biology and Medicine, 123(March), 103927. doi:10.1016/j.compbiomed.2020.103927.

[24] Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. Frontiers in Psychology, 8(AUG). doi:10.3389/fpsyg.2017.01454.

[25] Cheng, J., Chen, M., Li, C., Liu, Y., Song, R., Liu, A., & Chen, X. (2021). Emotion Recognition from Multi-Channel EEG via Deep Forest. IEEE Journal of Biomedical and Health Informatics, 25(2), 453–464. doi:10.1109/JBHI.2020.2995767.

[26] 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(6), 58–65. doi:10.1016/j.comcom.2020.02.051.

[27] Liu, N., Fang, Y., Li, L., Hou, L., Yang, F., & Guo, Y. (2018, April). Multiple feature fusion for automatic emotion recognition using EEG signals. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/ICASSP.2018.8462518.

[28] Zangeneh Soroush, M., Maghooli, K., Setarehdan, S. K., & Nasrabadi, A. M. (2019). A novel EEG-based approach to classify emotions through phase space dynamics. Signal, Image and Video Processing. doi:10.1007/s11760-019-01455-y.

[29] Zheng, W. L., Zhu, J. Y., & Lu, B. L. (2019). Identifying stable patterns over time for emotion recognition from EEG. IEEE Transactions on Affective Computing, 10(3), 417–429. doi:10.1109/TAFFC.2017.2712143.

[30] Mei, H., & Xu, X. (2017, December). EEG-based emotion classification using convolutional neural network. In 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC). Doi:10.1109/SPAC.2017.8304263.

[31] Zhao, Y., Yang, J., Lin, J., Yu, D., & Cao, X. (2020). A 3D Convolutional Neural Network for Emotion Recognition based on EEG Signals. Proceedings of the International Joint Conference on Neural Networks. doi:10.1109/IJCNN48605.2020.9207420.

[32] Liu, Z. T., Xie, Q., Wu, M., Cao, W. H., Li, D. Y., & Li, S. H. (2019). Electroencephalogram Emotion Recognition Based on Empirical Mode Decomposition and Optimal Feature Selection. IEEE Transactions on Cognitive and Developmental Systems, 11(4), 517–526. doi:10.1109/TCDS.2018.2868121.

[33] He, Y., Ai, Q., & Chen, K. (2017). A MEMD method of human emotion recognition based on valence-Arousal model. Proceedings - 9th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2017, 2, 399–402. doi:10.1109/IHMSC.2017.201.

[34] Parui, S., Bajiya, A. K. R., Samanta, D., & Chakravorty, N. (2019). Emotion recognition from EEG signal using XGBoost algorithm. 2019 IEEE 16th India Council International Conference (INDICON). doi:10.1109/INDICON47234.2019.9028978.

[35] Pan, C., Shi, C., Mu, H., Li, J., & Gao, X. (2020). EEG-based emotion recognition using logistic regression with gaussian kernel and laplacian prior and investigation of critical frequency bands. Applied Sciences (Switzerland), 10(5), 1–24. doi:10.3390/app10051619.

[36] Garg, D., & Verma, G. K. (2020). Emotion Recognition in Valence-Arousal Space from Multi-channel EEG data and Wavelet based Deep Learning Framework. Procedia Computer Science, 171, 857–867. doi:10.1016/j.procs.2020.04.093.

[37] Huang, Y. D., Wang, K. Y., Ho, Y. L., He, C. Y., & Fang, W. C. (2019). An edge AI system-on-chip design with customized convolutional-neural-network architecture for real-time EEG-based affective computing system. 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS). doi:10.1109/BIOCAS.2019.8919038.

[38] Koelstra, S., Mühl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing, 3(1), 18–31. doi:10.1109/T-AFFC.2011.15.

[39] Katsigiannis, S., & Ramzan, N. (2018). 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, 22(1), 98–107. doi:10.1109/JBHI.2017.2688239.

[40] Miranda-Correa, J. A., Abadi, M. K., Sebe, N., & Patras, I. (2021). AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. IEEE Transactions on Affective Computing, 12(2), 479–493. doi:10.1109/TAFFC.2018.2884461.

[41] Eldor, T. (2018). Capsule Neural Networks - Part2: What is a Capsule?. Towards Data Science. Available online: https://towardsdatascience.com/capsule-neural-networks-part-2-what-is-a-capsule-846d5418929f (accessed on August 2022).

[42] Wirawan, I. M. A., & Darmawiguna, I. G. M. (2017). New Concept of Learning Outcomes Assessment in Adaptive Mobile Learning. 2nd International Conference on Innovative Research across Disciplines (ICIRAD 2017). doi:10.2991/icirad-17.2017.49.

[43] Al-Nafjan, A., Hosny, M., Al-Ohali, Y., & Al-Wabil, A. (2017). Review and classification of emotion recognition based on EEG brain-computer interface system research: A systematic review. Applied Sciences (Switzerland), 7(12), 1239. doi:10.3390/app7121239.