IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 16 April 2023, accepted 8 May 2023, date of publication 11 May 2023, date of current version 17 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3275565

RESEARCH ARTICLE

Simplified 2D CNN Architecture With Channel Selection for Emotion Recognition Using EEG Spectrogram

LIA FAROKHAH [1,2], RIYANARTO SARNO [1], (Senior Member, IEEE), AND CHASTINE FATICHAH [1], (Member, IEEE)

[1]Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember (ITS), Sukolilo, Surabaya 60111, Indonesia

[2]Department of Informatics, Faculty of Technology and Design, Institut Teknologi dan Bisnis ASIA, Malang 65142, Indonesia

Corresponding author: Riyanarto Sarno (riyanarto@if.its.ac.id)

This work was supported in part by Asian Development Bank under Higher Education for Technology Innovation (ADB HETI) Project; the Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program; and Institut Teknologi Sepuluh Nopember (ITS) under project scheme of the Publication Writing and Intellectual Property Right (IPR) Incentive Program Pusat Pengembangan Hak Kekayaan Intelektual (PPHKI).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Rina Dewi Indahsari of Institut Teknologi dan Bisnis ASIA Malang under Application No. 5, and performed in line with the Emotion Analysis Data.

ABSTRACT Emotion Recognition through electroencephalography (EEG) is one of the prevailing emotion recognition techniques achieving higher accuracy rates. Nevertheless, one of the problems is the emotion recognition for inter-subjects where accuracy measures are lower. This happens because the EEG is a non-stationary signal which is resulting a domain shift across recordings of subjects, even under the same emotions, thus making emotional patterns difficult to identify. Another common observation is the emotion recognition of inter-subject and intra-subject by using all channels originating from the standard EEG recording mechanism. This requires higher computational resources and deep networks which are more complex including DenseNet, ResNet, etc. In this paper, we propose a novelty emotion recognition classification model that offers a simplified structure and employs only the selected channels from the 32 recorded channels. Residing on a standard approach of transforming EEG data – Database for Emotion Analysis of Physiological Signals (DEAP) – into 2D images using Short-Time Fourier Transform (STFT) EEG signals for training and analysis. Follows the channel selection approach and makes use of a simplified 2D CNN model. The channel selection adopts a search and retention approach using the selected samples of the data. The experimental results show that the performance of the proposed architecture improves the accuracy of the inter-subject emotion recognition using 32 channels by 9.73% and 11.7% on valence and arousal, respectively. While the use of the proposed selected channel method, with only 10 channels, performance increased by 3.53% and 7.2% on valence and arousal classes, respectively. Thus, by keeping a lower complexity level, the proposed architecture attains higher performance rates.

INDEX TERMS Classification, emotion recognition, inter-subject, spectrogram.

I. INTRODUCTION

Emotion recognition based on EEG signals has been one of the prevailing problems in the biomedical and signal-processing domain. There are two methods by which the task can be achieved in emotion recognition: using physiological and non-physiological signals. Recognition of emotions through physiological signals, EEG, renders an accurate analysis option because it is difficult to fake [1], [2], [3]. Nevertheless, one of the challenges in recognizing emotions

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Asif.

based on this method is their validation for inter-subject or cross-subject, or independent subjects [4], [5], [6]. Domain shifts in the EEG recordings across subjects make emotional pattern recognition even more difficult. The existing models make use of a large number of channels for the recognition problem. This often gets associated with highly complex deep learning models requiring large memory resources and irrelevant information [7], [8], [9], [10].

The existing studies of open-source DEAP datasets offer 32-channel EEG signals targeting different parts of brain activity [4], [11]. The sensitivity of EEG recording may hamper the accurate recognition problem and therefore a need for extracting the desired information from the recorded signals is essential. However, the problem reappears when we reduce the number of channels, and the performance also decreases [12], [13]. To cater to this issue, researchers have come along opting for highly deep networks including DenseNet, ResNets, etc [4], [11], [14] and use all the channels. This is associated with a heavy computational load and a highly complex network architecture making the learning carried out by the proposed network difficult to follow. However, in this study, we will still see the performance of the proposed architecture using 32 channels as well as the results of the channel selection method which uses 10 channels. This is done to carry out comprehensive and comparative learning.

The pre-processing methods for detecting emotions through EEG signals are numerous. Currently, research is focused on converting 1D EEG signals into 2D images and 2D frame sequences. This approach offers an easier analytical interpretation of changes in brain activation areas [15]. In the DEAP, research using 2D images is still minimal compared to the use of 2D frame sequences. In addition, the performance of emotion validation for the inter-subject is lower compared to emotion recognition for the intra-subject or dependent subject [16]. Performance differences can be around 5% [17] or even up to about 20% [18], [19].

Emotional recognition through EEG can be validated using two methods, namely intra-subject and inter-subject. Intra-subject refers to a classification model where training, testing, and validation datasets are generated from the same individual through a suitable split [18], [20], [21], [22]. Inter-subject is defined as a classification model trained on data from a set of subjects and tested on an unseen dataset [18] [23], [24], [25]. Accurate recognition of emotions for inter-subject is a research challenge, and there is a need to devise such models that render better performance measures. This contributes to broader EEG-based emotion recognition implementations. The study, therefore, involves both the intra-subject and inter-subject implementations and performance comparisons for the emotion recognition problem.

The motivations and contribution of the study are as follows:

1) The inter-subject performance is generally lower compared to the intra-subject performance. This needs to be improved.
2) Most of the previous studies looking for better emotion recognition performance on the DEAP dataset used 32 channels, thus a channel selection architecture shall be developed.
3) In reducing computation, the number of EEG channels needs to be reduced. Therefore, it is necessary to develop an appropriate method for selecting potential channels.
4) The use of highly complex and computationally intensive models demands high computational resources. Therefore, there is a need to explore simple architectures offering better or comparable performance.

Based on the gaps identified above, this work offers the following contributions:

1) tests and offers an improved version of the inter-subject emotion recognition in both arousal and valence classes with 32 channels and proposed selected channel method which uses 10 channels.
2) develops a channel selection architecture to select the channels with the significant information for classification.
3) provides performance outcomes based on a simplified CNN architecture having a performance level at par with the more complex architectural networks.

The paper is structured as follows: Section I describes the background of the problem. Section II describes research related to emotion recognition in 2D images and 2D frame sequences. However, we will add some non-2D-based research as additional knowledge. Section III describes the details of the dataset, the experimental scenarios, the proposed methodology, and the evaluation. Section IV describes the results of experimental research compared to previous studies on both intra-subject and inter-subject validation. Part 5 is the conclusion.

II. RELATED WORK

Research related to emotion recognition using the DEAP open-source repositories attempts to improve performance accuracy. Intra-subject research has been carried out in many previous studies and obtained accuracy rates above 90%, even up to 99%. According to the existing research, intra-subject accuracy measures are superior to inter-subject accuracy measures. In inter-subject research, the work is still very rare, resulting in poor performance, using a large number of channels, and often using a very deep architecture for classification. The underlying model for both of these modes remains the same, i.e., the use of 32 channels. Therefore, the algorithm offers architectural development for classification using the basic CNN architecture, and channel selection to obtain the most important information from the EEG recordings to improve accuracy performance.

Several intra-subject studies based on 2D images originating from 32 EEG channels are available. Lin et al. focus on converting 1D EEG into a frequency band image and using AlexNet architecture, yielding valence and arousal class accuracy results of 85.50% and 87.30% [26]. Kwon et al. conducted a multimodal study using EEG and galvanic skin response (GSR) by transforming 1D EEG into a spectrogram using CNN architecture, resulting in valence and arousal class accuracy of 78.12% and 81.25%, respectively [27] by using a variety of subjects. Research on the combination of dimensionality reduction and architecture, namely CNN, SAE, and DNN, in the 2D form generated from the frequency band resulted valence and arousal accuracy of 89.49% and 92.86% [28], respectively. The research has used a variety of subjects in training. Wang et al. use the concept of concatenated images in treating the existing 32 channels. Two separate models have been trained (one for valence classes and other for arousal classes). The study resulted in validation accuracy for valence and arousal of 99.99% and 99.98%, respectively [14].

Several intra-subject studies based on 2D frame sequences using 32 EEG channels are available. Yang et al. use the CNN-LSTM model, with 8 subjects having an average accuracy of 90.6% and 91.03% on valence and arousal, respectively [29]. Another model makes use of 32 subjects using a single subject for training through a regional-asymmetric convolutional neural network (RACNN) to achieve valence and arousal accuracy of 96.65% and 97.11%, respectively [16]. A study using the BiDCNN model employed 16 subjects separately and achieved average accuracy on valence and arousal of 94.38% and 94.72%, respectively [30]. Research using graph CNNs and LSTM produces accuracy for valence and arousal of 90.45% and 90.60%, respectively [17]. The 2D frame sequence research that has used training with various subjects is Chao's research, which produces 68.28% and 66.73% valence and arousal accuracy performances, respectively.

Some inter-subject studies based on 2D images have lower performance measures. Pandey and Seeja researched 2D images of scalograms using 32 channels and frontal channels in the valence and arousal classes [31]. The accuracy performance obtained for valence and arousal is 59% and 58%, respectively, while with the frontal channel, the accuracy performance is 61% for the valence class and 58% for the arousal class. Another study used the Inception Resnet-V2 architecture for positive and negative emotion classes [32]. These emotions, transformed into the DEAP dataset, are called valence classes. The performance obtained is 72.81%.

Furthermore, the highest performance for inter-subjects research currently available is obtained by using a spectrogram with the DenseNet-121 and DCERNET architectures, which only provide performance details on the valence class [4]. The performance of the valence class using Densenet-121 is 85.57%, while the performance of the valence class using DCERNET and softmax is 88.57%, and DCERNET and SVM is 88.10%. In addition, the validation of inter-subject emotions using the 2D frame sequences resulted in 68% and 63% accuracy on valence and arousal, respectively [30]. Non-2D research, such as the use of 3D meta-learners, yields valence and arousal performance of 69.92% and 68.89%, respectively [22]. Research on 3D cubes with inter-subjects has been carried out with valence and arousal performances of 84% and 85% [17]. Based on previous studies, it is concluded that the highest accuracy performance in the valence class is 88.57% and the arousal class is 85%. Previous inter-subject research on the DEAP dataset used more than 32 channels to detect emotions.

Some of the channel selection methods in the DEAP EEG signal are also available, but limited to 2D-based. Peng et al. perform the channel selection method through a 2D frame sequence for intra-subject classification with accuracy of valence 95.18% and arousal 95.58%, which the accuracy got average from each subject [33]. Reference [12] makes use of channel selection using the ReliefF and NCA methods. Both of these methods are based on irradiation to form a computer-generated holographic feature map. It has been tested on intra-subject. We try to compare the channel selection method with research [12] because they use 10 channels, image-based and have the highest accuracy performance on several public datasets, such as DEAP, DREAMER, and AMIGOS, of around 80–90%. In addition, there are several channel selection methods in the DEAP dataset that use 10 channels in valence or arousal class or both of them [9], [31], [34], [10]. These studies are not all 2D-based and employ deep learning. In general, they have an accuracy performance below 90%. Bagzir et al. got accuracy performance results of 91.1% and 91.3% in valence and arousal classes, respectively [34]. They used frequency band and did not use deep learning. Based on the explanation above, this study will choose the 10 highest channel values from the performance metrics indicator using deep learning and image-based.

Several non-deep learning channel selection studies or not using 10 channels or 2D-based, have been carried out. Li et al. selected a single channel for emotion classification through brain rhythm sequencing using frequency subbands within 10 seconds [7]. The proposed method yields an intra-subject classification of about 70%. Msonda et al. propose a method based on searching for channel subsets with minimal errors by utilizing the frequency sub-band [35]. They perform inter-subject validation with a combination of several channels that produce a valence accuracy performance of around 50–60%. Reference [36] is one of the most recent studies in 2023, proposing a channel selection method with graph-based for intra-subject classification with accuracy around 90% and 70% for inter-subject. In addition, channel selection still uses the old method, which is based on the brain area. Yin et al. conduct channel selection for inter-subject classification by developing locally-robust feature selection (LRFS) with approximately 60% accuracy performance for valence and arousal classes [37]. Arjun et al. used the attention channel for inter-subject classification, collaborating LSTM and autoencoder, and achieved accuracy rates of 65.9% and 69.5% for valence and arousal classes, respectively [38]. Yang et al. used ReliefF and random forest to select the best 10 channels and 11 channels for intra-subject and inter-subject classification, with inter-subject performance results of 81.27% and 82.36% in valence and arousal classes [9]. The use of minimal channels (channel selection schemes) for inter-subject classification is yet to be extensively explored. In addition, performance can still be improved.

In the articles reviewed above, three problems can be identified. First, it improves accuracy performance when using 32 channels either intra-subject or inter-subject. Second, it is desired that the research should lead to the use of minimal channels but well-performing. Third, the architecture complexity is seen as a trade-off with the performance measures. Therefore, the proposed research targets all these gaps and investigates inter-subject performance using minimal channels and a simplified architecture, without the need for continuous re-trainings.

III. MATERIALS AND METHOD

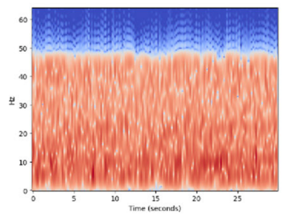
The architecture proposed in this study was validated on the DEAP, public dataset. The results of the emotional validation experiment using EEG are divided into two parts, namely intra-subject and inter-subject validation. The proposed method involves the processing of the raw data with fewer preprocessing steps. The Python development environment has been used to downsample data to 128 Hz, remove EOG artifacts, and perform bandpass frequency filtering from 4–45 Hz. The raw EEG data is transformed into a spectrogram image using the STFT method and stored as RGB images. This is done using the MATLAB (octave and oc2py) library in Python. After the spectrogram is generated, the next step is to classify emotions using a deep learning network. The architecture developed is a simplified CNN architecture comprising convolution, pooling, dropout, and fully connected layers. The combination of convolution followed by max pooling and dropout several times serves as a feature elimination and selection layer.

A. DEAP DATASET

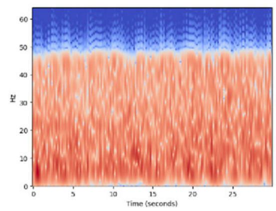
The DEAP dataset has 32 participants with a composition of 50% male and 50% female samples [39]. The stimulus used is a video with a duration of 60 seconds. The results of the EEG recording have a duration of 63 seconds, of which 3 seconds are the baseline. One participant has 40 recordings of the EEG trials. The results of the EEG trial recordings have been measured using the Self-Assessment Manikin (SAM). SAM has four emotional dimensions, namely valence, arousal, dominance, and liking, with ratings 0-9. For valence and arousal classes, emotion labels are divided into two categories: low and high. In this study, the low class has a rating below 4.5 and the high class is equal to or more than 4.5. The EEG tool used is the BioSemi ActiveTwo system, which has 32 channels and 8 peripheral channels. The datasets with 32 channels have been employed in this study.

TABLE 1. Parameter of STFT.

No	*Parameter*	*Value*
1	HopSize	*25 samples*
2	Window	*Hanning*
3	FrameSize	50 *samples*
4	FFT Size	512



(a)



(b)

FIGURE 1. STFT images for individual's recording at channel 26. a. High valence. b. Low valence.

B. SPECTROGRAM GENERATION

Transformation of raw EEG to a 2D image spectrogram This transforms data into time-frequency information using the Short Time Fourier Transform (STFT). STFT provides a better description of localization in time and frequency using the color intensity level. The time used is the second half sample of each EEG signal, as also discussed in several previous studies [4], [40]. Some of the parameters used in STFT are described in Table 1.

STFT helps activate the spectral properties of the information contained in the individual channels. By using chunks of information (frames and overlaps), it identifies frequencies that exist in different parts of time. Its mathematical form is provided as Eq. 1:

$$STFTx\,[n] = X\,(n, \omega) = \sum_{m=-\infty}^{\infty} x\,[m]\,w\,[n-m]\,e^{-j\omega m} \quad (1)$$

where $w[m]$ is the time window that is applied to extract the framed signal, $e^{-j\omega m}$ infers the frequency response of the windowed signal. The formula above describes a standard Fourier transform applied to a framed signal, resulting in a 2D matrix with three channels. The values contained in each of these frequencies with time are depicted using a spectrogram. The windowing used is Hanning, which helps to make the frame smoother around the edges.

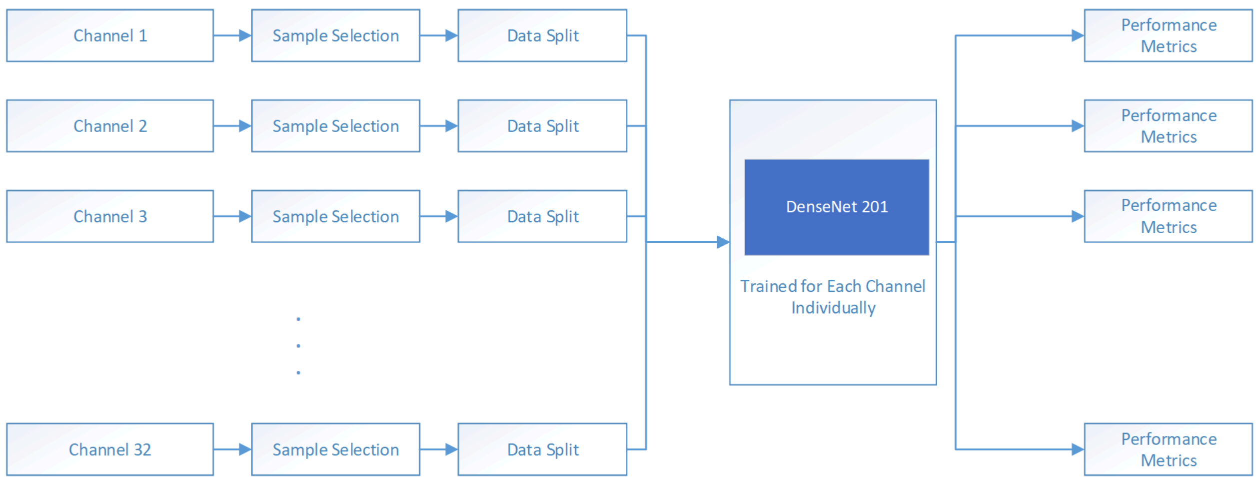


FIGURE 2. Proposed method of channel selection.

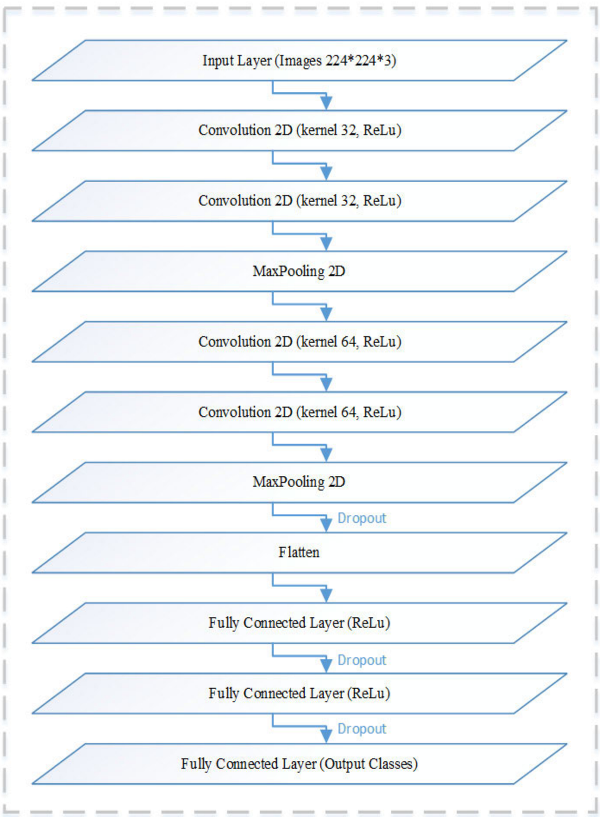


FIGURE 3. Proposed architecture.

The RGB images produced (having dimensions $224 \times 224 \times 3$ by resizing the spectrograms. The dimensions have been chosen based on the accepted input size of transfer learning models used in this study) are further normalized for the training operations. It helps to scale pixel values between 0 and 1 for the model to accurately infer hidden trends. All the channels are transformed into images and later subject to the channel selection and classification modes. A sample of STFT images produced against channel number 26 of an individual recording for high and low valence. They show in Fig. 1.

C. PROPOSED CHANNEL SELECTION

The use of 32 channels is the most common method used in previous studies for both intra-subject and inter-subject validation experiments. So, a gap is a wider, more logical computation that uses fewer channels for significant information. There is a need to have fewer EEG channels with good accuracy rates. In this research, channel selection is done by investigating individual channels and then selecting the ten best channels. Channel selection is based on the concept of ''garbage in, garbage out.'' The selection of channels using the brain area will certainly be less effective during the EEG sampling of the channel as there is a large enough amount of noise. This is the basis for channel selection, utilizing a single channel that has the best accuracy for emotion classification. The first step is to determine the architecture that will be used to detect the best channels. In this research, the DenseNet-201 architecture is used to see the performance on individual channels. We sampled each subject with 30% of the total 1280 samples of data for training. This was obtained from 40 trials x 32 subjects. The architecture works by taking individual channel data samples and dividing them into training, validation, and testing. The benchmark for channel selection is based on precision, recall, accuracy, and F1 score. The sequence of the proposed channel selection method is shown in Fig. 2.

D. PROPOSED SIMPLIFIED CLASSIFICATION ARCHITECTURE

The simplified architecture has been developed using a combination of convolutional, pooling, dropout, and fully

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 222, 222, 32)      896

 conv2d_1 (Conv2D)           (None, 220, 220, 32)      9248

 max_pooling2d (MaxPooling2D  (None, 110, 110, 32)     0
 )

 conv2d_2 (Conv2D)           (None, 108, 108, 64)      18496

 conv2d_3 (Conv2D)           (None, 107, 107, 64)      16448

 max_pooling2d_1 (MaxPooling  (None, 53, 53, 64)       0
 2D)

 dropout (Dropout)           (None, 53, 53, 64)        0

 flatten (Flatten)           (None, 179776)            0

 dense (Dense)               (None, 256)               46022912

 dropout_1 (Dropout)         (None, 256)               0

 dense_1 (Dense)             (None, 128)               32896

 dropout_2 (Dropout)         (None, 128)               0

 dense_2 (Dense)             (None, 2)                 258

=================================================================
Total params: 46,101,154
Trainable params: 46,101,154
Non-trainable params: 0
```

FIGURE 4. Classification network stacked architecture.

connected layers. The layers are stacked together, formulating a sequential model as a combination of convolution, max-pooling, and dropout layers several times to infer the best features used for the emotion classification process. The proposed architecture is shown in Fig. 3. The architecture comprises a convolutional scheme where inference is made with the aid of fewer layers compared to DenseNet or other deep architectures. The architecture is hypothesized to offer decent accuracy while keeping the complexity levels lower. The architectural details and the number of parameters generated from the Python code running on the Google Colab engine are shown in Fig. 3 and Fig. 4.

E. EXPERIMENTATION

Fig. 5 depicts that there are two proposed scenarios, namely the use of the entire channel as in previous research and using channel selection to reduce computations and offer a logical implementation. The performance evaluation of the scenario is measured from the performance metrics such as accuracy, precision, recall, and F1 score. Accuracy will be used as the main indicator.

The parameters used in the proposed network are shown in Table 2.

The number of epochs chosen above are based on the experimental outcomes of training and validation outcomes (accuracy and loss figures) of the deep learning model. Two optimizers have been tested: RMSProp and Adam, where the latter offers better accuracy results. In addition, the categorical cross-entropy loss function has been tried, but the MSE loss function produces better results, and there are no NaN losses during the training and validation processes, which are generally the burst losses from the gradient. The learning rates have been chosen to be suitably smaller for the model to learn the training deeply as well as keep a suitable learning complexity. The overall evaluation of the proposed architecture is tested using a confusion matrix, namely accuracy, F1 score, precision, and recall. The formula used to find the accuracy is shown in Eq. 2.

TABLE 2. Hyper parameters for the experimentation framework.

Parameter	Value
Epoch	300 for intra-subject and 150 for inter-subject
Batch size	32 for intra-subject and 16 for inter-subject
Dropout	60%,70%, and 50%
Pooling Layer	Max Pooling
Learning rate	lr = 1e-4
Decay	decay = 1e-7
Optimizer	Adam
Loss	MSE
Activation function	ReLU

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives. The precision and recall are then evaluated, as shown in Eq. 3 and 4, respectively.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FP} \tag{4}$$

The next evaluation is the F-score, which is a combination of accuracy between recall and precision. The F-score formula is shown in Eq. 5.

$$F1score = \frac{2(precision)(recall)}{recall + precision} \tag{5}$$

In measuring the evaluation, eight scenarios will be used, as shown in Table 3. For inter-subject classification, it used leave one out subject cross validation (LOOCV). LOOCV is a standard evaluation for inter-subject classification. This can be seen from [4], [30], and [31].

IV. RESULT AND DISCUSSION

A. CHANNEL SELECTION

In the original dataset with 32 channels for each individual, 1280 images per subject have been generated. Initially, 49.920 STFT images have been generated and later reduced to 12.800 using the 10 selected channels under the channel selection approach. Following the approach presented in Fig. 2, the following results have been generated against each of the 32 channels as depicted in Table 4.

From the results presented in Table 4, the 10 best channels were selected to be processed because they had the best accuracy performance. These are channels 11, 12, 13, 14, 15, 23, 24, 27, 28, and 29. The analysis is carried out by taking

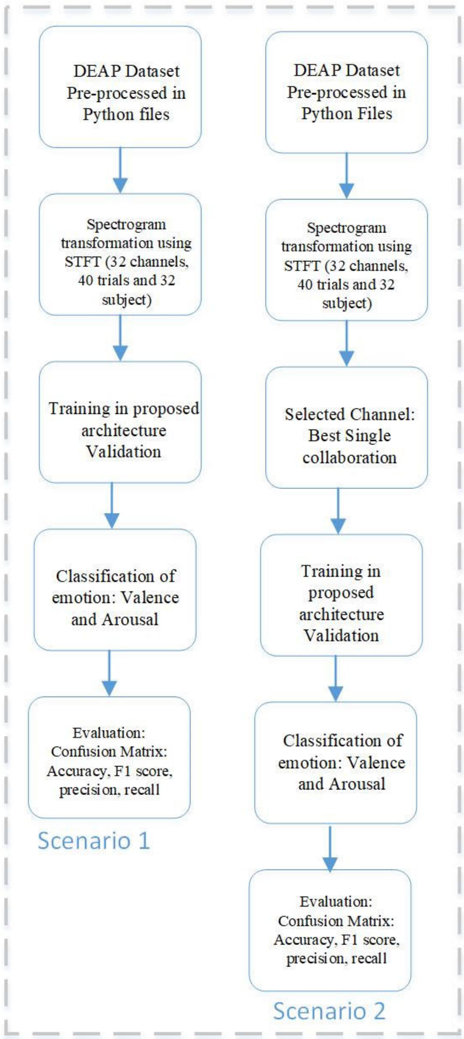


FIGURE 5. Experimentation framework for comparative analysis.

TABLE 3. Proposed experimentation reference.

Experiment	Test Data	Remark
Intra-1	90:10	All channels
Intra-2	90:10	10 channels with ReliefF method
Intra-3	90:10	Selected Channel with proposed method
Intra-4	90:10	10 channels with NCA method
Inter-1	Leave one out subject	32 channels
Inter-2	Leave one out subject	10 channels with ReliefF method
Inter-3	Leave one out subject	Selected Channel with proposed method
Inter-4	Leave one out subject	10 channels with NCA method

into account the performance of precision, recall, accuracy, and F1 score. The selection of the best 10 channels was made to compare with other methods that use 10 channels, namely the NCA and ReliefF methods [12].

TABLE 4. Performance outcomes against individual channels.

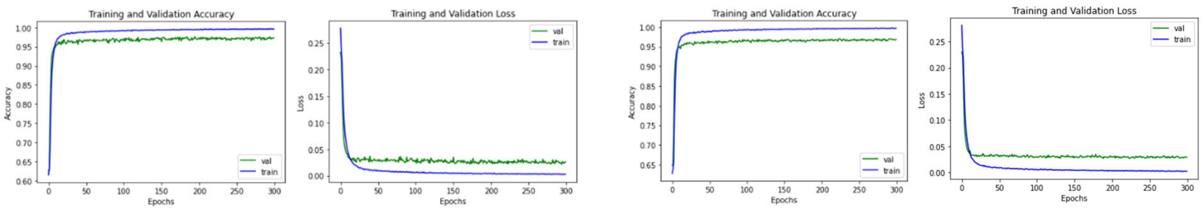
Channel No	Precision	Recall	Accuracy	F1 Score
1	0.353	0.500	0.485	0.414
2	0.318	0.159	0.594	0.212
3	0.529	0.281	0.637	0.367
4	0.551	0.286	0.672	0.376
5	0.543	0.6	0.664	0.57
6	0.582	0.549	0.702	0.565
7	0.578	0.721	0.711	0.642
8	0.834	0.524	0.782	0.644
9	0.744	0.527	0.754	0.617
10	0.754	0.714	0.812	0.733
11	0.862	0.678	0.838	0.759
12	0.847	0.758	0.854	0.8
13	0.895	0.644	0.836	0.749
14	0.9957	0.816	0.75	0.842
15	0.9961	0.81	0.807	0.86
16	0.999	0.235	0.091	0.586
17	1	0.574	0.376	0.672
18	0.9961	0.59	0.343	0.688
19	0.9971	0.617	0.607	0.705
20	0.9975	0.795	0.536	0.789
21	0.9957	0.737	0.665	0.793
22	0.9988	0.663	0.831	0.777
23	0.9934	0.816	0.654	0.817
24	0.9952	0.826	0.578	0.802
25	0.9935	0.783	0.69	0.813
26	0.9938	0.76	0.743	0.811
27	0.9972	0.872	0.789	0.872
28	0.9978	0.916	0.724	0.87
29	0.9982	0.914	0.749	0.883
30	0.4	0.478	0.555	0.436
31	0.545	0.437	0.695	0.485
32	0.57	0.421	0.706	0.484

B. EXPERIMENT 1: EMOTION RECOGNITION FOR INTRA-SUBJECT

The results of the emotion recognition experiment on the DEAP dataset, which is used in the proposed architecture, are divided into two parts: intra-subject validation and inter-subject validation. The first part is for intra-subject, which is divided into 4 sub-sections: intra-1 (Exp. 1), where validation is carried out using 32 channels, intra-2 (Exp. 2), using 10 channels based on the ReliefF method; and intra-3 (Exp. 3), using the best single-channel collaboration selection

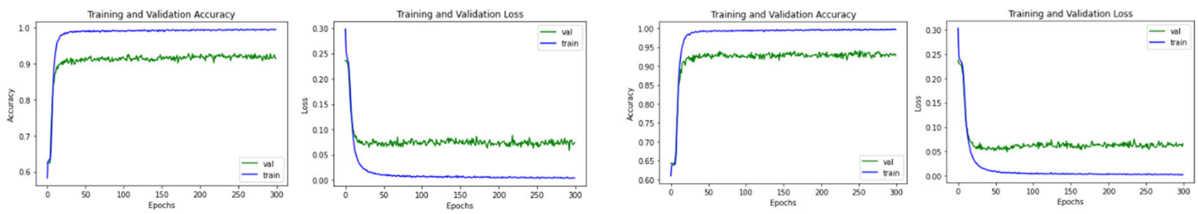
TABLE 5. Testing performance results for intra-subject.

Exp	Valence (%)		Arousal (%)		Testing Performance							
					Valence (%)				Arousal (%)			
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
1	99.62	97.20	99.73	97.17	97.4	98.5	94.3	96.4	97	96.9	94.7	95.8
2	99.59	94.97	99.67	92.97	90.9	93.7	81	86.9	92.7	92	87.3	89.6
3	99.69	92.88	99.70	94.44	92.4	93.6	85.7	89.5	93.4	91.1	89.9	90.5
4	99.61	94.79	99.75	94.53	92.2	91.4	87.2	89.3	92.3	92.9	85.1	88.8



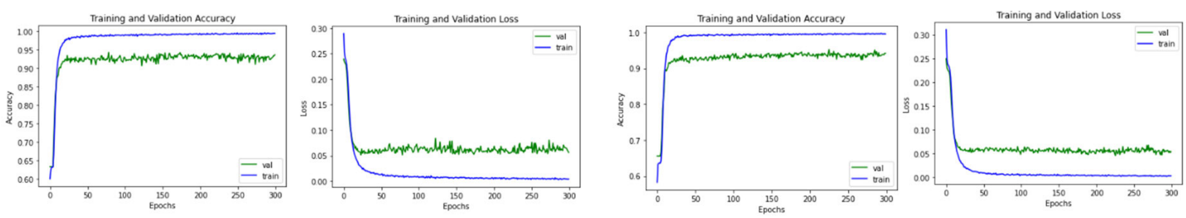
(a) (b)

FIGURE 6. (a). All channel accuracy and loss on valence (b) All channel accuracy and loss on arousal.



(a) (b)

FIGURE 7. (a). ReliefF accuracy and loss on valence (b) ReliefF accuracy and loss on Arousal.



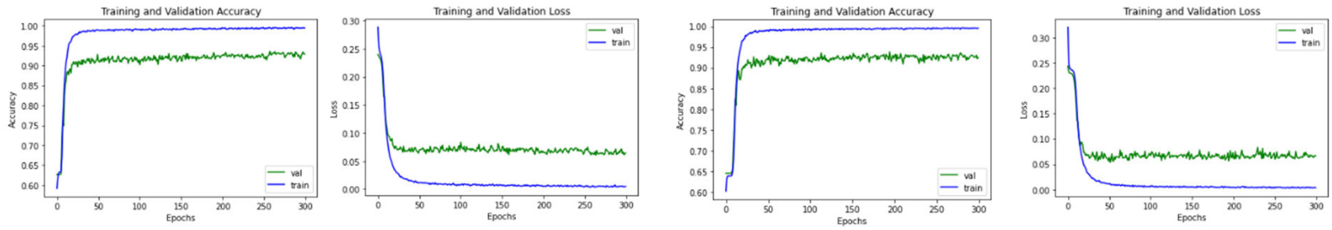
(a) (b)

FIGURE 8. (a). Proposed method accuracy and loss on valence (b) Proposed method accuracy and loss on arousal.

based on the proposed method. The fourth sub-section is intra-4 (Exp. 4), which uses 10 channels based on the NCA method. Scenarios for intra-subject validation can be seen in Table 3.

Table 5 displays the performance results for the intra-subject validation of emotions

As evident in Table 5, the proposed architecture produces a good accuracy performance of above 90%. On Exp. 1, which uses 32 channels, it produces good accuracy performance values. In the Exp. 2, experiment using ReliefF and Exp. 4 using NCA methods [12], Exp. 3 used the proposed selected channel. When compared with research [12], the proposed method has a performance comparison that is quite thin. The thing that needs to be underlined is the experiment in Table 5 using the proposed architecture for classification. In addition, compared to the paper classification [12] on the DEAP dataset, which uses the NCA and ReliefF methods based on holographic feature maps, they use a different CNN architecture. When using 10 channels with their architecture, the accuracy performance of the NCA method (N-HOLO-FM)



(a) (b)

FIGURE 9. (a). NCA accuracy and loss on valence (b) NCA accuracy and loss on arousal.

TABLE 6. Testing performance results for inter-subject.

Exp	Performance							
	Valence				Arousal			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
2	91.75	91.5	85.35	88.2	89.4	84.9	90.7	87.1
3	92.42	92.37	86.43	89.2	92.24	92.15	85.52	88.6
4	92.53	93.78	85.35	89.3	91.18	90.68	83.61	86.8

TABLE 7. Inter-subject performance of the proposed method.

Exp	Performance							
	Valence				Arousal			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
1	98.3	98.3	94.2	96.2	96.7	96.5	94.2	95.4
3	92.1	92.42	85.94	89.04	92.2	92.27	87.22	89.6

is 81.88% and 82.45% in the valence and arousal classes, respectively. In the ReliefF method (R-HOLO-FM), it is 83.26% and 83.85% in the valence and arousal classes. This is different when using the proposed architecture, which results in increased performance accuracy on the NCA: 10.32% and 9.85% in the valence and arousal and ReliefF method classes, respectively, with a performance increase of 7.64% and 8.85% in the valence and arousal classes. The proposed channel selection method outperforms the other selection methods in the valence and arousal classes [12]. The input for selecting the best channel for the proposed method uses EEG data to be classified, namely spectrogram data. The accuracy and loss plot results in Table 5 are shown in Figures 6, 7, 8, and 9. From the plot of the image, we can infer that there is no overfitting.

C. EXPERIMENT 2: EMOTION RECOGNITION FOR INTER-SUBJECT

The second part is inter-subject validation. The inter-subject classification was performed using the proposed architecture. In this study, inter-subject validation will be divided into two parts, which use 32 channels and use 10 channels resulting from the proposed reduction channel method. The first is inter-subject with the reduced channel; this section is divided into three sub-sections, namely, the use of 10 channels based on the ReliefF method (Exp. 2), 10 channels from the proposed method (Exp. 3), and 10 channels based on the NCA method (Exp. 4). Leave-One-Out validation approach has been used for inter-subject validation. At a given instance of training, one of the subjects S_n (in turn) is excluded and the model is exposed to no samples from it. Training is completed on the remaining 32 subjects $S_{1to32-n}$. S_n data is used as a validation data passes to the model. The same procedure is used with each subject to get the average classification accuracy. In this section, specifically for the use of 10 channels based on the ReliefF method, NCA, and the proposed method, we use 32 subjects but only average the first 16 subjects. This is because of the efficiency of the experimental process. The results can be seen clearly with an average indicator of half of the existing participants. The results of inter-subject performance are shown in Table 6. The type of experimental order is adapted from Table 3.

According to Table 6, the proposed selected channel provides the best performance (Exp. 3). The best detection results were analyzed from the sum and average accuracy in the valence and arousal classes using the best single-channel collaboration method compared to the other methods. After finding the best accuracy performance among the three experiments using 10 channels, the average of all subjects was sought. One thing that must be emphasized is that the paper [12] only performs intra-subject classification when reducing channels using NCA and ReliefF, and the results in Table 6 use the proposed architecture. Using the proposed architecture for the 10 channels, changes the performance results compared to the previous intra-subject classification. After that, an inter-subject classification was carried out for the use of 32 channels (Exp. 1). In the inter-subject evaluation using 32 or 10 channels, the average training accuracy performance for the valence and arousal classes is around 99%. The average of inter-subject testing performance for 32 channels (Exp. 1) and 10 channels (Exp. 3) is shown in Table 7.

D. COMPARISON WITH PREVIOUS RESEARCH ON INTRA-SUBJECT EMOTION RECOGNITION

A comparison of performance with previous research is carried out to validate the results obtained from this research experimentation. This section is validation for intra-subjects. Intra-subject validation is emotional validation using the same subject in the training process. The results of the comparison of research on intra-subjects are shown in Table 8.

In Table 8, the use of 32 channels to improve performance on the DEAP dataset is more dominant. Several studies conducted training on each subject and achieved average accuracy. This is more common in 2D frame sequences. This is a very minimal generalization of the model. In this study, we used a training model for all subjects so that the model results were rich in variance. The proposed research is better than the method that uses 2D frame sequences. In research based on 2D images when using 32 channels, this study is indeed 2% smaller in valence and arousal accuracy compared to Wang's study in 2022. However, the proposed architecture is simple and computationally less intensive. Although Wang's research did not mention the number of parameters used in their proposed architecture, this can be seen from the depth of the proposed layer in [14]. In the proposed study, compared to Wang's, they both use CNN architecture development. The difference is that the architecture proposed in the architectural research is not very deep. One of the most significant indicators is that the number of convolution layers in this study is only 4, while Wang's research has 8 convolution layers. This is a minor computational contribution to intra-subject classification. According to several studies, lighter computing can be called a contribution or a new focus for more implementable artificial intelligence [4], [8], [42], [43]. On the other hand, the intra-subject classification for reduced channels is also described. We compared with a 2D-based study [12], which used 10 channels. The accuracy of the proposed research is much better. We also compare with 2D frame sequence research, which has a higher accuracy performance than ours, which is 95.18% and 95.58% for valence and arousal class, respectively [33]. However, this study produced accuracy through the average of each subject. This means that there is no variation in the model. We treat intra-subject classification by using all data subjects, namely 32 people. We only have a few previous research comparisons when using reduced channels. This is because it is still very limited, especially the use of deep learning. To persuade our research, we compare it to the reduced channel method, which does not use 10 channels or 2D-based or deep learning. We discuss this in sub-II-related work [7], [33], [34], [36]. In discussing related work, we know that our performance accuracy is still higher than all of them.

TABLE 8. Comparison of intra-subject performance results with previous studies.

Author	Model	Accuracy		Remark
		Valence	Arousal	
Chao et al.[24]	CapsNet	68.28 %	66.73 %	- Varian subject training - 2D EEG Frame Matrix based - All channels
Islam and Ahmad[41]	CNN	81.51 %	70.42	- Varian subject training - Images-based - All channels
Yang et al [29]	CNN+RNN	90.80 %	91.03	- Single subject and average accuracy - 2D EEG Frame Matrix based - All channels
Cui et al[16]	RACNN	96.65 %	97.11 %	- Single subject and average accuracy - 2D EEG Frame Matrix based - All channels
Lin et al[26]	AlexNet	85.50 %	87.30	- Single subject and average accuracy - Images-based - All channels
Kwon et al[27]	CNN	78.12 %	81.25 %	- Varian subject training - Images-based - All channels

TABLE 8. *(Continued.)* **Comparison of intra-subject performance results with previous studies.**

TABLE 8. *(Continued.)* **Comparison of intra-subject performance results with previous studies.**

Liu et al [28]	CNN+SAE+DNN	89.49 %	92.86 %	- Varian subject training - Images-based - All channels
Liu et al[20]	MLF-CapsNet	97.97	98.31	- Single subject and average accuracy - 2D EEG Frame Matrix - All channels
Huang et al [30]	BiDCNN	94.38 %	94.72 %	- Single subject and average accuracy - 2D EEG Frame Matrix based - All channels
Topic et al[12]	CNN + SVM (R-HOLO-FM)	83.26 %	83.85 %	- Varian subject training - Images based - 10 channels
Topic et al[12]	CNN + SVM (N-HOLO-FM)	81.88 %	82.45 %	- Varian subject training - Images based - 10 channels
Wang et al[14]	CNN	99.99 %	99.98	- Varian subject training - Images based - All channels
Peng et al [33]	TR-CA	95.18 %	95.58 %	- Single subject and average accuracy - 2D Frame - Attention channel
Proposed method	CNN	97.4	97	- Varian subject training - Images based - All channels
Proposed method	Simplified CNN	92.4	93.4	- Varian subject training - Images based - 10 channels from the best single collaboration

E. COMPARISON WITH PREVIOUS RESEARCH ON INTER-SUBJECT EMOTION RECOGNITION

In order to compare the performance of this study with previous research for inter-subject classification, we will compare using 32 channels and less than 32 channels. The results of the accuracy performance of the proposed architecture have the best performance using 32 channels. A comparison of emotional validation performance for inter-subjects with previous studies is shown in Table 9.

Based on Table 9 regarding the emotional validation of inter-subjects, we know that the proposed method, both architecture and channel selection methods produce better accuracy performance than previous studies. In addition, the increase in the percentage of emotion classification in valence class for inter-subjects with 10 channels increased by 3.53% compared to the results of the previous highest research, namely 2D-based and 3D-based [4]. For arousal class, it compares with the previous highest research, namely 3D-based [17]. It experienced an increase in accuracy of 7.2%. In the arousal class, if only compared with similar studies in the form of 2D EEG, the accuracy performance increased by 23.14% [22]. With the use of 32 channels, the proposed architecture improves valence and arousal class accuracy from another study by

TABLE 9. Comparison of inter-subject performance results with previous studies.

Author	Model	Accuracy		Remark
		Valence	Arousal	
Yin et al[17]	3D Cube - ECLGCNN	84%	85%	- Graph-based dan Deep Learning - 3D Cube
Pandey and Seeja[31]	CNN	59%	58%	- Scalogram images -10 frontal channels: a. Valence: 61% b. Arousal:58%
Cimtay et al [32]	Inception Resnet-V2	72,81 %	-	- Images based - All channels
Huang et al [30]	2D BiDCNN	68%	63%	- 2D EEG Frame Matrix based - All channels
Bhosale et al [22]	Raw EEG + CNN (Meta-learner)	69,92 %	68,89 %	- 2D EEG Frame Matrix based - All channels
Pusarla et al [4]	DenseNet121	85,57	-	- Spectrogram Images based - All channels
Pusarla et al [4]	DCERNet+softmax	88,57	-	-Spectrogram Images based - All channels
Pusarla et al [4]	DCERNet+SVM	88.10	-	- Spectrogram Images based - All channels
Proposed Method	Simplified CNN	98.3	96.7	- Spectrogram Images based - All channels
Proposed Method	Simplified CNN	92.1	92.2	- Spectrogram Images based - 10 Channels based best single collaboration

9.73 and 11.7%, respectively. However, we are having difficulty finding research comparisons for inter-subject classification in 2D-based when using 10 channels. Hence, we also present non-2D-based, non-10-channel, or non-deep learning research in subsection II Related Work. It to ensure the results of our reduced channels research [9], [35], [36], [37], [38]. The comparison shows that the results we obtained in this study were better than those of previous research, even across domains. The indicator used is accuracy performance.

TABLE 10. Comparison of parameters on architectures.

Author	Parameters
Cimtay et al[32]	55,970,000
Pusarla et al[4]	51,760,644
Proposed Architecture	46,101,156

F. COMPARISON WITH PREVIOUS RESEARCH ON INTER-SUBJECT NUMBER OF ARCHITECTURAL PARAMETERS

A comparison of the number of architectural parameters in the proposed inter-subject emotion validation with previous research is shown in Table 10. We compare computations to previous papers that state the number of parameters. In addition, our comparative focus is on the study by Pusarla et al [4] because their research has the highest accuracy.

In table 9 and Table 10, we know that in addition to having a better performance on inter-subject emotion classification, we also offer a lower level of complexity on the proposed architecture.

V. CONCLUSION

The work presented in this article focused on novelty and a simplified architecture for the classification of EEG data taken from DEAP open-source repositories. Three aspects have been focused on: improving accuracy performance when using 32 channels, both intra-subject and inter-subject. Second, use the channel selection method to find the channel with the most significant emotional information but good performance. Third, the architecture complexity is seen as a trade-off with the performance measures. It has been found that this architecture contributes to an increase in the performance of emotion recognition for intra-subjects when using 10 channels. In addition, the increase in accuracy performance also occurs in emotion recognition for inter-subjects in the DEAP dataset using both 32 channels and 10 channels. By collaborating on the developed CNN architecture, the proposed method outperforms the ReliefF and NCA methods with images-based or non-images-based contributions in related work, sub II. Additionally, the results generated for intra-subject and inter-subject classification using the simplified network architecture reveal performance on par with the more complex networks like DeepNets and Resnet. For future work, the channel optimization or selection method must be further developed so that the most optimal number of channels can be obtained. In addition, it will be compared with other datasets, so this will ensure that the channel selection method can be used between different datasets.

REFERENCES

[1] B. Farrow and S. Jayarathna, "Technological advancements in post-traumatic stress disorder detection: A survey," in *Proc. IEEE 20th Int. Conf. Inf. Reuse Integr. Data Sci. (IRI)*, Jul. 2019, pp. 223–228, doi: 10.1109/IRI.2019.00044.

[2] R. Sarno, M. N. Munawar, and B. T. Nugraha, "Real-time electroencephalography-based emotion recognition system," *Int. Rev. Comput. Softw.*, vol. 11, no. 5, pp. 456–465, 2016, doi: 10.15866/irecos.v11i5.9334.

[3] L. J. Zheng, J. Mountstephens, and J. Teo, "Four-class emotion classification in virtual reality using pupillometry," *J. Big Data*, vol. 7, no. 1, pp. 1–9, Dec. 2020, doi: 10.1186/s40537-020-00322-9.

[4] N. Pusarla, A. Singh, and S. Tripathi, "Learning DenseNet features from EEG based spectrograms for subject independent emotion recognition," *Biomed. Signal Process. Control*, vol. 74, Apr. 2022, Art. no. 103485, doi: 10.1016/j.bspc.2022.103485.

[5] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Oct. 2019, Art. no. 051001.

[6] E. Lashgari, D. Liang, and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," *J. Neurosci. Methods*, vol. 346, Dec. 2020, Art. no. 108885, doi: 10.1016/j.jneumeth.2020.108885.

[7] J. W. Li, S. Barma, P. U. Mak, F. Chen, C. Li, M. T. Li, M. I. Vai, and S. H. Pun, "Single-channel selection for EEG-based emotion recognition using brain rhythm sequencing," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 6, pp. 2493–2503, Jun. 2022, doi: 10.1109/JBHI.2022.3148109.

[8] N. Pusarla, A. Singh, and S. Tripathi, "Normal inverse Gaussian features for EEG-based automatic emotion recognition," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3205894.

[9] L. Yang, S. Chao, Q. Zhang, P. Ni, and D. Liu, "A grouped dynamic EEG channel selection method for emotion recognition," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2021, pp. 3689–3696, doi: 10.1109/BIBM52615.2021.9669889.

[10] Z. Wang, S. Hu, and H. Song, "Channel selection method for EEG emotion recognition using normalized mutual information," *IEEE Access*, vol. 7, pp. 143303–143311, 2019, doi: 10.1109/ACCESS.2019.2944273.

[11] J. Cho and H. Hwang, "Spatio-temporal representation of an electoencephalogram for emotion recognition using a three-dimensional convolutional neural network," *Sensors*, vol. 20, no. 12, p. 3491, Jun. 2020, doi: 10.3390/s20123491.

[12] A. Topic, M. Russo, M. Stella, and M. Saric, "Emotion recognition using a reduced set of EEG channels based on holographic feature maps," *Sensors*, vol. 22, no. 9, p. 3248, Apr. 2022, doi: 10.3390/s22093248.

[13] M. A. Asghar, "EEG-based multi-modal emotion recognition using bag of deep features: An optimal feature selection approach," *Sensors*, vol. 19, no. 23, pp. 1–16, 2019, doi: 10.3390/s19235218.

[14] Y. Wang, L. Zhang, P. Xia, P. Wang, X. Chen, L. Du, Z. Fang, and M. Du, "EEG-based emotion recognition using a 2D CNN with different kernels," *Bioengineering*, vol. 9, no. 6, p. 231, May 2022, doi: 10.3390/bioengineering9060231.

[15] D. Shah, G. Gopan K., and N. Sinha, "An investigation of the multi-dimensional (1D vs. 2D vs. 3D) analyses of EEG signals using traditional methods and deep learning-based methods," *Frontiers Signal Process.*, vol. 2, pp. 1–15, Jul. 2022, doi: 10.3389/frsip.2022.936790.

[16] H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, and X. Chen, "EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106243, doi: 10.1016/j.knosys.2020.106243.

[17] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106954, doi: 10.1016/j.asoc.2020.106954.

[18] G. C. Jana, A. Sabath, and A. Agrawal, "Capsule neural networks on spatio-temporal EEG frames for cross-subject emotion recognition," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103361, doi: 10.1016/j.bspc.2021.103361.

[19] J. Zhu, W. Zheng, and B. Lu, "Cross-subject and cross-gender emotion classification from EEG," in *Proc. IFMBE*, vol. 51, 2015, pp. 1188–1191, doi: 10.1007/978-3-319-19387-8.

[20] Y. Liu, Y. Ding, C. Li, J. Cheng, R. Song, F. Wan, and X. Chen, "Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103927, doi: 10.1016/j.compbiomed.2020.103927.

[21] J. Guo, F. Fang, W. Wang, and F. Ren, "EEG emotion recognition based on Granger causality and CapsNet neural network," in *Proc. 5th IEEE Int. Conf. Cloud Comput. Intell. Syst. (CCIS)*, Nov. 2018, pp. 47–52, doi: 10.1109/CCIS.2018.8691230.

[22] S. Bhosale, R. Chakraborty, and S. K. Kopparapu, "Calibration free meta learning based approach for subject independent EEG emotion recognition," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103289, doi: 10.1016/j.bspc.2021.103289.

[23] H. Li, Y. M. Jin, W. L. Zheng, and B. L. Lu, "Cross-subject emotion recognition using deep adaptation networks," in *Proc. Int. Conf. Neural Inf. Process.*, vol. 11305, 2018, pp. 403–413, doi: 10.1007/978-3-030-04221-9_36.

[24] H. Chao, L. Dong, Y. Liu, and B. Lu, "Emotion recognition from multiband EEG signals using CapsNet," *Sensors*, vol. 19, no. 9, p. 2212, May 2019, doi: 10.3390/s19092212.

[25] J. Li, S. Qiu, Y. Shen, C. Liu, and H. He, "Multisource transfer learning for cross-subject EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3281–3293, Jul. 2020.

[26] W. Lin, C. Li, and S. Sun, "Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal," in *Proc. Int. Conf. Image Graph.*, vol. 10667, 2017, pp. 385–394, doi: 10.1007/978-3-319-71589-6_33.

[27] Y.-H. Kwon, S.-B. Shin, and S.-D. Kim, "Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system," *Sensors*, vol. 18, no. 5, p. 1383, Apr. 2018, doi: 10.3390/s18051383.

[28] J. Liu, G. Wu, Y. Luo, S. Qiu, S. Yang, W. Li, and Y. Bi, "EEG-based emotion classification using a deep neural network and sparse autoencoder," *Frontiers Syst. Neurosci.*, vol. 14, pp. 1–14, Sep. 2020, doi: 10.3389/fnsys.2020.00043.

[29] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7, doi: 10.1109/IJCNN.2018.8489331.

[30] D. Huang, S. Chen, C. Liu, L. Zheng, Z. Tian, and D. Jiang, "Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for EEG emotion recognition," *Neurocomputing*, vol. 448, pp. 140–151, Aug. 2021, doi: 10.1016/j.neucom.2021.03.105.

[31] P. Pandey and K. R. Seeja, "Subject independent emotion recognition system for people with facial deformity: An EEG based approach," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 2, pp. 2311–2320, Feb. 2021, doi: 10.1007/s12652-020-02338-8.

[32] Y. Cimtay and E. Ekmekcioglu, "Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition," *Sensors*, vol. 20, no. 7, pp. 1–20, 2020, doi: 10.3390/s20072034.

[33] G. Peng, K. Zhao, H. Zhang, D. Xu, and X. Kong, "Temporal relative transformer encoding cooperating with channel attention for EEG emotion analysis," *Comput. Biol. Med.*, vol. 154, Mar. 2023, Art. no. 106537, doi: 10.1016/j.compbiomed.2023.106537.

[34] O. Bazgir, Z. Mohammadi, and S. A. H. Habibi, "Emotion recognition with machine learning using EEG signals," in *Proc. 25th Nat. and 3rd Int. Iranian Conf. Biomed. Eng. (ICBME)*, Nov. 2018, pp. 1–5, doi: 10.1109/ICBME.2018.8703559.

[35] J. R. Msonda, Z. He, and C. Lu, "Feature reconstruction based channel selection for emotion recognition using EEG," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2021, pp. 1–7, doi: 10.1109/SPMB52430.2021.9672258.

[36] X. Lin, J. Chen, W. Ma, W. Tang, and Y. Wang, "EEG emotion recognition using improved graph neural network with channel selection," *Comput. Methods Programs Biomed.*, vol. 231, Apr. 2023, Art. no. 107380, doi: 10.1016/j.cmpb.2023.107380.

[37] Z. Yin, L. Liu, J. Chen, B. Zhao, and Y. Wang, "Locally robust EEG feature selection for individual-independent emotion recognition," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113768, doi: 10.1016/j.eswa.2020.113768.

[38] Arjun, A. S. Rajpoot, and M. R. Panicker, "Subject independent emotion recognition using EEG signals employing attention driven neural networks," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103547, doi: 10.1016/j.bspc.2022.103547.

[39] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[40] C. Qing, R. Qiao, X. Xu, and Y. Cheng, "Interpretable emotion recognition using EEG signals," *IEEE Access*, vol. 7, pp. 94160–94170, 2019, doi: 10.1109/ACCESS.2019.2928691.

[41] M. R. Islam, M. M. Islam, M. M. Rahman, C. Mondal, S. K. Singha, M. Ahmad, A. Awal, M. S. Islam, and M. A. Moni, "EEG channel correlation based model for emotion recognition," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104757, doi: 10.1016/j.compbiomed.2021.104757.
[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.
[43] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The computational limits of deep learning," 2020, *arXiv:2007.05558*.

LIA FAROKHAH was born in Jombang, in August 1990. She is currently pursuing the Ph.D. degree in computer science with Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. She is also a Lecturer and a Researcher with the Department of Computer Science, Institut Teknologi dan Bisnis ASIA, Malang.

RIYANARTO SARNO (Senior Member, IEEE) received the Doctoral (Ph.D.) degree, in 1992.

He is currently a Professor with the Informatics Department, Institut Teknologi Sepuluh Nopember (ITS). He is interested in research projects about machine learning, the Internet of Things, and knowledge engineering. He has researched electronic nose, image processing, and signal processing. Being an author of more than five books and over 300 scientific articles led him incorporated in the top 2% world ranking scientist, in 2020, by Standford University.

CHASTINE FATICHAH (Member, IEEE) was born in Pasuruan, in December 1975. She received the bachelor's degree in informatics, in 2000, the master's degree in computer science, in 2008, and the Ph.D. degree in computational intelligence, in 2012. She is currently a Doctor in intelligent computation and vision with the Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember (ITS). She was awarded the Young Researcher Award from Advanced Computational Intelligence and Intelligent Informatics (JACIII), Fuji Press, Tokyo, Japan, in 2014.