

Fine Grained Emotion Recognition from EEG Signal Using Fast Fourier Transformation and CNN

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

Abstract—The process of identifying human emotion is known as emotion recognition. Emotions are mental states originating in the human brain and this is closely related to the activities of the nervous system. Electroencephalogram (EEG) is a well-established approach to record neuron activities which is reliable for emotion recognition compared to the non-physiological clues. Emotion recognition is most often used for neuroscience, psychology, cognitive science, computer science, artificial intelligence, and brain-computer interfaces. So far, there have been reports of various searches for active patterns involving different emotions. However, with time, their stability is still not fully investigated. Hence, this paper presents convolutional neural network (CNN) models working on the DEAP dataset and it contains emotional states which are arousal, valence, dominance, and liking where each state is rated from 1-9. We are able to get better accuracy from what has been done about binary classification so far with the DEAP dataset for valence and arousal. Our binary models achieved 96.63% and 96.17% accuracy respectively for valence and arousal. Only four emotion spaces are found with binary classification whereas 8-class classification is worked more precisely with sixty-four emotion spaces. This paper attempts to work with 8-class classification also and get a promising result with the accuracy of 93.83 % and 93.79% respectively for valence and arousal. For both cases, Fast Fourier Transformation (FFT) has been used as the feature extraction method and all the four classification models are created under 1D-CNN using the same architecture. The design and application of CNN models depend on CNN's ability to automatically detect new emotion-related EEG features in future work. Furthermore, our study is a testament to the robust neural network classifiers for brain signals that will be even surpassing traditional theoretical learning techniques.

Index Terms—Emotion recognition, EEG, CNN, DEAP, FFT

I. INTRODUCTION

Emotions are a significant factor of human knowledge, behavior and communication. It is a neural impulse that moves an organism to action, prompting automatic reactive behavior that has been adapted through evolution as a survival mechanism to meet a survival need [1]. Emotion recognition is the process of identifying human emotions. Emotions can be divided

into two categories: Positive and Negative [2]. While positive emotions are needed for good health, negative emotions can cause mental health problems, such as depression, stress, and anxiety [3]. Emotions are known to arise from the central and peripheral nervous system and cause temporal movement due to the synchronized execution of neurons [4]. They are expressed by internal signals as well as external expressions like facial expression, speech, body posture, eye blinking, and skin response. If only external expressions are used for emotion measurement, incorrect results may be obtained, because in many cases external expressions can be controlled. That's why internal signals get priority. Electroencephalogram (EEG), Temperature (T), Electrocardiogram (ECG), Electromyogram (EMG), Galvanic Skin Response (GSR), Respiration (RSP) are examples of such internal signals. Excluding others, EEG is selected for its non-invasive, fast and low-cost characteristics, as compared with other physiological signals.

Neuropsychological measurement of electrical activity in the brain is known as Electroencephalography (EEG) and it is recorded by electrodes. These are normally placed on the scalp, or in special cases, subdurally and in the cerebral cortex. EEG estimates voltage fluctuations that result due to ionic flows inside the neurons of the cerebrum. EEG provides an excellent temporal resolution, even though it has a poor spatial resolution and requires many sensors placed on the scalp. Pure EEG signal is a composition of sub-bands: theta (3 - 7 Hz), alpha (8 - 13 Hz), beta (14 - 29 Hz) and gamma (30 - 47 Hz) [17]. Individual sub-bands are associated with individual relevant physical activities. For instance, Theta wave refers to REM sleep, deep and raw emotions, and cognitive processing. A drowsy state indicates the Alpha waves. It also becomes the cause of relaxation and calmness. Beta points to the conscious state during the thought process. The Gamma waves are available when trying to perceive two different senses at the same time as sound and sight [7].

Yet now, there comes out of extensive research using

machine learning to identify states of emotion with EEG. Machine learning-based theoretical methods are often effectively used for emotion classification, while the disadvantage of such methods is that researchers have to spend a lot of effort to detect and design different emotion-related features from the resulting noisy signals and these features are time-consuming calculations. Various methods of EEG feature extraction have been explored in recent years, although the Fast Fourier Transformation (FFT), Short-time Fourier Transform (STFT) and Discrete Wavelet Transformation (DWT) etc are most effectively used. FFT is used to decompose this work's EEG signal data. Then deep learning models are applied to the extracted features so that they are trained for recognizing emotions. Support Vector Machine (SVM), Linear and nonlinear regression, Decision trees and K-nearest neighbor (KNN) are examples of mostly used machine learning model architecture. Convolution Neural Networks (CNN), Long-Short Term Memory (LSTM), Convolution Long-Short Term Memory (CLSTM) are popular model architectures from deep learning for this field. There are so many EEG datasets that are publically available and some researchers use their own dataset. Some famous publically available datasets are DEAP (A Database for Emotion Analysis using Physiological Signals), SEED (SJTU Emotion EEG Dataset) and MAHNOB (MAHNOB-HCI-Tagging database) etc. Among these, the DEAP dataset has been used for this research work.

From the EEG signal, using main data and only two emotion labels (arousal and valence), it is possible to recognize human emotion properly. Each emotion label is divided into two equal parts and a total of four emotions are created by classifying the binary-class using both emotion labels. It is possible to increase the accuracy here more than in the previous work. We have tried and succeeded in overcoming this deficit through our experiment. Although most of the research is done with binary-class classification, this is not possible to precisely recognize emotions with these four classes. To recognize emotion more precisely, we have worked with sixty-four emotion spaces where many real-life emotions exist.

Significant contributions of this paper:

- We have brought the best accuracy using the binary classification which is the conventional emotion recognition method and for this, we have used a very simple and light 1D-CNN model.
- Also we have proposed the eight-class emotion classification method which is able to recognize emotion much more accurately and we have also found a satisfactory classification accuracy for it.

The residual part of this study is highlighted as follow:

An overview of literature review in section II. A brief description of the DEAP dataset in section III. The research methodology of this paperwork is described including EEG data preprocessing, feature extraction, labeling and normalization as well as CNN model structure in section IV. Experiment and result comparison with previous DEAP dataset-related work in section V. An indication for our future work and the

conclusion of this paper are given in section VI.

II. LITERATURE REVIEW

There has been a lot of research on publicly available datasets (DEAP, SEED, MAHNOB, LUMED) for emotion recognition. But some have tried to recognize emotions on their own datasets. Since the DEAP dataset is used in this paper, it would be better to review the contributions and the recent studies using the DEAP dataset.

Rahul Sharmaa Et al. [4] achieved an accuracy of 82.01% with a ten-fold cross-validation technique using Long short-term memory (LSTM) by decomposing with DWT. They worked on only two dimensions, namely arousal and valence, and with four quadrants respectively LaLv (low arousal low valence), HaLv (high arousal low valence), LaHv (low arousal high valence) and HaHv (high arousal high valence). Kit Hwa Cheah Et al. Anubhav Et al. [7] earned a handsome accuracy when classifying emotions using valence and arousal labels, of 94.69% and 93.13% respectively. Although they tested KNN, SVM, Decision Tree, and Random Forest for classifying emotion, their best accuracy was achieved with the use of LSTM.

Zhongke Gao Et al. [9] proposed a model named Channel-fused dense convolutional network (CDCN), consisting of a 1D convolution layer and 1D dense layer. For pre-extracting, they used differential entropy (DE) and worked on four emotions. Their model applied the SEED dataset and DEAP dataset and obtained an accuracy of 90.6% and 92.58% respectively.

Yuling Luo Et al. [10] demonstrated their best performance with Spiking Neural Networks (SNNs) using three pre-extracting methods: DWT, Variance and FFT. They gained their best results with the use of variance, both on SEED and DEAP datasets. The emotion states of arousal, valence, dominance and liking were classified with accuracies of 74%, 78%, 80% and 86.27% for the DEAP dataset, as well as an overall accuracy of 96.67% for the SEED dataset.

Eman A. Abdel-Ghaffar Et al. [13] proposed a two-dimensional emotion model named Log-Euclidean Riemannian Metric (LERM) using Symmetric Positive Definite manifold (SPD). They received an accuracy of 88.3% for HVHA, 84.38% for LVHA, 79.3% for LVLA, and 78.4% for HVLA. The average accuracy for valence was $74.6\% \pm 3.9$, and $72.6\% \pm 6.7$ for arousal.

Fei Wang Et al. [5] used the Electrode-frequency distribution maps (EFDMS) model for classifying and short-time Fourier transform (STFT) for feature extraction. Gradient weighted class Activation mapping (Grad-CAM) is used in their research to obtain a better understanding of their selected features. When they applied their model on the SEED dataset, they obtained 90.59% for accuracy, and on the DEAP dataset, they obtained an accuracy of 82.84%. However, they only worked with valence labels with three classes: negative, neutral, positive.

Kit Hwa Cheah Et al. [8] used two types of CNN models: single-path CNN and two-path CNN model using 4 folds of

TABLE I: Overview of literature review

No	Research	Year	Feature extraction	Modeling technique	No. of class	Working label	Performance
1	Rahul Sharmaa Et al. [4]	2020	DWT	LSTM	2-Class	Arousal, Valance	82.01%
2	Divya Acharya Et al. [3]	2020	FFT	LSTM	2-Class	Arousal, Valance	89.83%
3	Zhongke Gao Et al. [9]	2020	DE	CDCN	2-Class	Arousal, Valance	92.58%
4	Yulong Luo Et al. [10]	2020	Variance	SNN	2-Class	Arousal, Valance	Valence:78% Arousal:74%
5	Yucel Cimtay Et al. [11]	2020	Raw Data	CNN	2-Class	Arousal, Valance	72.81%
6	Eman A. Et al. [13]	2020	SPD	LERM	2-Class	Arousal, Valance	Valence: 74.6% ± 3.9, Arousal: 72.6% ± 6.7
7	Xiaolong Zhong Et al. [14]	2020	Down sampling	CNN	2-Class	Arousal, Valance	Valence:66.23% Arousal:68.50%
8	Yucel Cimtay Et al. [15]	2020	Raw Data	CNN	2-Class	Arousal, Valance	91.5%
9	Guolu Cao Et al. [12]	2019	PCA	CNN	2-Class	Arousal, Valance	Valence: 81.2±3.0% Arousal: 84.3±4.0%
10	Soheil Rayatdoost Et al. [1]	2018	HOC,PSD, DE, HOS	RF	2-Class	Arousal, Valance	Valence:60.86% Arousal:58.08%
11	Ningjie Liu Et al. [16]	2018	LFCC	KNN, ResNets	2-Class	Arousal, Valance	Valence:90.39% Arousal:89.06%
12	Abeer Al-Nafjan Et al. [22]	2017	PSD	DNN	2-Class	Arousal, Valance	Valence:82% Arousal:82%
13	Samarth Tripathi Et al. [24]	2017	GD	CNN	2-Class	Arousal, Valance	Valence:81.41% Arousal:73.36%
14	Xiang Li Et al. [21]	2016	CWT	C-RNN	2-Class	Arousal, Valance	Valence:72.06% Arousal:74.12%
15	Wei-Long Zheng Et al. [25]	2016	DE	GELM	2-Class	Valence, Arousal	69.67%
16	Wei Liu Et al. [23]	2016	PSD, DE	BDAE	2-Class	Valence, Arousal	Valence:85.20% Arousal:80.50%
17	Hyun Joong Yoon Et al. [18]	2013	FFT	Bayes classifier	2-Class	Valence, Arousal	Valence:70.9%, Arousal:70.1%
18	Viktor Rozgić Et al. [19]	2013	PCA	SVM	2-Class	Valence, Arousal	Valence:76.9% Arousal:68.4%
19	Xiaowei Zhang Et al. [20]	2013	Sliding 4-second windows with a 2-second overlap	Ontology Reasoning BIO-EMOTION	2-Class	Valence, Arousal	Valence:75.19% Arousal:81.74%

cross-validation. However, they did not use any manual pre-extraction methods, and instead, worked with only valence and arousal. Each emotional state was divided into three classes. Single-path CNN received an accuracy of 97.59% and 98.4% for 3-class valence and arousal, while two-path CNN received 98.75% and 97.58% for 3-class valence and arousal. Yucel Cimtay Et al. [11] did not use any manual pre-extraction methods, and instead, depended on the CNN Deep Learning method. Three datasets were used in their research work: DEAP, SEED and LUMED. In the SEED dataset when studying two classes of emotions, an accuracy of 86.56% was obtained, and 78.34% was obtained for three classes of emotions. When applied to the DEAP dataset, they received

72.81% accuracy for two emotion states: valence and arousal. Guolu Cao Et al. [12] created a CNN model, utilizing Principal Component Analysis (PCA) as the pre-extracting technique. They worked with two classes for arousal and valence with an accuracy of 84.3±4.0% and 81.2±3.0% respectively. Xiaolong Zhong Et al. [14] concentrated on the physiological forms of brain waves. Their method was efficient in recognizing emotions, especially in beta and gamma waves. 2D SE_CNN was applied to the DEAP and the MAHNOB-HCI datasets. In their study, the DEAP dataset received an accuracy of 66.23% for valence and 68.50% for arousal, while the MAHNOB-HCI dataset obtained 70.25% for valence and 73.27% for arousal. Yucel Cimtay Et al. [15] used the InceptionResnetV2 CNN

model, following a hybrid fusion strategy on the DEAP dataset as well as LUMED-2 datasets with facial expressions and galvanic skin response (GSR). They achieved maximum of 91.5% accuracy on the DEAP dataset with arousal and valence.

From the above discussion, we can see that almost everyone has worked on binary-classification. But the testing accuracy that has been found so far is not enough and it is possible to improve more. We have been able to achieve the best accuracy through our work. Only four emotional spaces can be found with a binary-classification which is unable to accurately recognize real-life emotions. To overcome this lack, we have introduced in our experiments an 8-class classification technique that can recognize a variety of emotions and has achieved satisfactory accuracy in this method as well.

III. DATASET

The DEAP dataset is a publicly available multimodal dataset [17] that includes electroencephalogram (EEG) signals and used for detecting human emotional states. A team of researchers at the Queen Mary University of London developed this dataset.

TABLE II: Synopsis of the DEAP dataset

Types of dataset	Multimodal dataset
No. of participant	32
No. of EEG channel	32
Data collection method	Showing one-minute long excerpts of music videos
No. of used data collection resource	40 music videos
Sampling rate	128Hz
Rating values	Continuous scale 1-9
Rating scales	Arousal and Valence

The DEAP dataset is available in two parts, with the first part containing an online self-assessment of 14-16 volunteers based on arousal, valence, and dominance for 120 one-minute music videos. The second part contains the participant ratings, physiological recordings and face video of an experiment where 32 volunteers watched 40 music videos which are the subset of the previously mentioned 120 music videos. Physiological signals and EEG signals were recorded where each participant also rated the videos following the above procedure. Facial expression at the time of watching videos was also recorded from 22 participants. Individual online self-assessment ratings, a list of the used YouTube music videos, the ratings that the participants gave for the videos, all the answers for the questionnaire of the participants before the experiment and participant's frontal face video recordings as well as raw physiological data recordings in BioSemi .bdf format are in the official dataset. Forty experiments for each of the 32 participants are found in the dataset. For each of the 40 experiments, the label array for each participant contained ratings of arousal, valence, dominance, and liking. For each participant, 8064 physiological / EEG signals data were collected with 40 different channels for each experiment and put into the data array. Among 40 channels 32 are EEG channels. Brain data is collected using electrode caps, EEG signal is collected

via 512 Hz sampling frequency. After watching the videos all participants rated those according to a 1-9 scale based on valence, arousal and dominance. The duration of every sampled data is 63s. For experiments normally pre-processed data is used where 128Hz downsampling, electrooculogram (EOG) removal, filtering, segmentation and so on have been used. Two versions of preprocessed data are found on the official website. One of them is the data_preprocessed_matlab folder processed with Matlab where files are in .mat format and another is the data_preprocessed_python folder processed with Python (numpy) where files are in .dat format. The preprocessed data folder contains 32 files and each file holds the individual data of each of the 32 subjects. The data format is in table III.

TABLE III: Data orientation of each subject

Name of array	Shape of array	Contents of array
data	40 x 40 x 8064	video/trial x channel x data
labels	40 x 4	video/trial x label (valence, arousal, dominance, liking)

Both .dat files and .mat files contain data field with shape 40*40*8064 and label field with shape 40*4, where data field shape 40*40*8064 stands for 40 trials, 40 channels and 8064 refers to 63*128. Here sampling time is 63 seconds and the sampling frequency is 128Hz. In label field shape 40*4 indicates 40 experiments and 4 dimensions respectively for valence, arousal, dominance and liking. Signals were recorded according to the international 10-20 system. From the data_preprocessed_python folder, some parts of a file out of 32 preprocessed files are plotted and shown in figure 1.

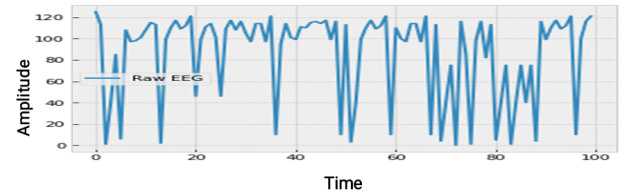


Fig. 1: Raw EEG signal for one subject

IV. METHODOLOGY

The procedure for EEG Data Analysis is shown below through figure 2. Firstly, the dataset is collected then raw data is cleaned using the various pre-processing techniques. Next, this cleaned raw data is segmented in the feature extraction step. Then those extracted features are used to train a model for getting a better classification result.

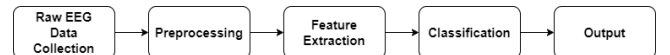


Fig. 2: Workflow of EEG data analysis method

A. Preprocessing

The DEAP EEG signal has been recorded with a good instrument and as a result, the chance of artifacts are minimized. The DEAP dataset's EEG signals are downsampled to 128 Hz first, so that the data content is collected in a good way between 0-48 Hz. Then the electromyogram (EMG) and electrooculogram (EOG) has been removed from the downsampled data. A bandpass filter has been applied to separate the delta waves from the analysis process. A blind source separation technique has been used removing eye artifacts. Using Common Average Reference (CAR) the data has been averaged. Each recording data has been partitioned into 60 seconds segments and a pre-trial baseline of 3 seconds has been removed.

B. Feature Extraction

In the research field, emotion classification potentiality depends on two factors: feature extraction and classification. Feature extraction reduces the initial dataset by identifying key features of data and later these features are used for classification. Distinguishing property, recognizable measurement, and functional components obtained from a section of a pattern are represented by features. A better classification accuracy comes if extracting features from a dataset are used instead of the original dataset. By Feature extraction, various advantages could be found like Minimizing the loss of important signal, decreasing the risk of over-fitting, improving the visualization of data, and reducing the implementation complexity.

Three types of features were found [7]:

- 1) Time-domain features: used for statistical features.
- 2) Frequency-domain features: decomposition of preprocessed signal data into sub-bands.
- 3) Time-Frequency domain features: used for non-stationary waveform signals [16].

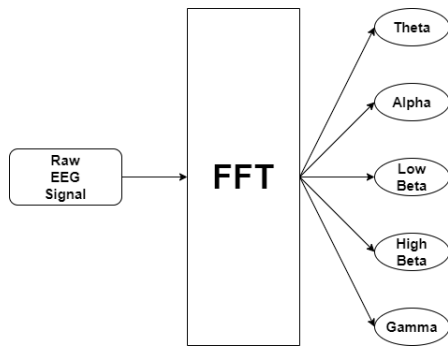


Fig. 3: Feature extraction method

Frequency-domain features are used for this work. Different methods are used for EEG feature extraction, including FFT, Wavelet Transform (WT), Time-Frequency Distribution (TFD), Equivocator methods (EM), Auto-Regressive methods (ARM), etc. From these methods, FFT is ultimately applied to the preprocessed EEG datasets by using a python module named PyEEG, which is formatted in NumPy.

Fourier analysis is commonly used for signal processing to convert time-domain signals into frequency domain signals. Thus, it was used to decompose the EEG signal into frequency components. To compute the Fourier transformation, the most popular algorithm is FFT, which computes the Discrete Fourier Transform (DFT) of a sequence [3].

$$X_k = \sum_{i=0}^{N-1} x_i(n) e^{-j2\pi i k / N} \quad (1)$$

In equation (1), $k = 0, 1, 2, \dots, N-1$ and X_k is the coefficient of discrete Fourier, length available data is N and $x_i(n)$ is the time domain input signal.

As the purpose is to extract key features from pre-processed data and use them to classify emotions through a CNN-based deep learning model. The Fast Fourier Transform method is used when employing mathematical tools to extract the EEG features. By using power spectral density (PSD) estimation, the characteristics of the EEG signal are computed. The EEG spectrum of wave characteristics is then divided into four frequency bands. Through the approximate auto-correlation sequence of the Fourier converter, PSD can be counted accurately.

When extracting features by using FFT for EEG, there are mainly two types of techniques: the Periodogram Method and Welch's method.

Calculating PSD with a periodogram is the easiest way. Frequency decomposition is included by Periodogram, and the modulus squared of the Fourier transform of the signal is expressed as:

$$\tilde{P}_{xx}(f) = \frac{\Delta t}{N} \left| \sum_{i=0}^{N-1} x_i(n) e^{-j2\pi i k / N} \right|^2 \quad (2)$$

In equation (2), Δt refers to space between the samples, available data length is ' N ', $x_i(n)$ is the time domain input signal and $\tilde{P}_{xx}(f)$ denotes PSD for $x_i(n)$.

Welch's method is another PSD assessment strategy that is used to improve the modified periodogram accuracy. This method is established through the use of signals in overlapping windows where for each window a periodogram is calculated and then to calculate the PSD those periodograms are constructed.

Supposing signals $x(n)$ have finite length, then the relationship with power spectral density is estimated as:

$$x_i(n) = x(n + iD), \quad n = 0, 1, 2, \dots, M-1 \quad (3)$$

While $i = 0, 1, 2, \dots, L-1$

In equation (3), ' iD ' represents the start of the i^{th} sequence and ' L ' is the length of the formatted data segment.

The subsequent outcome periodograms give:

$$\tilde{\tilde{P}}_{xx}(f) = \frac{1}{MU} \left| \sum_{n=0}^{N-1} x_i(n) w(n) e^{-j2\pi f n} \right|^2 \quad (4)$$

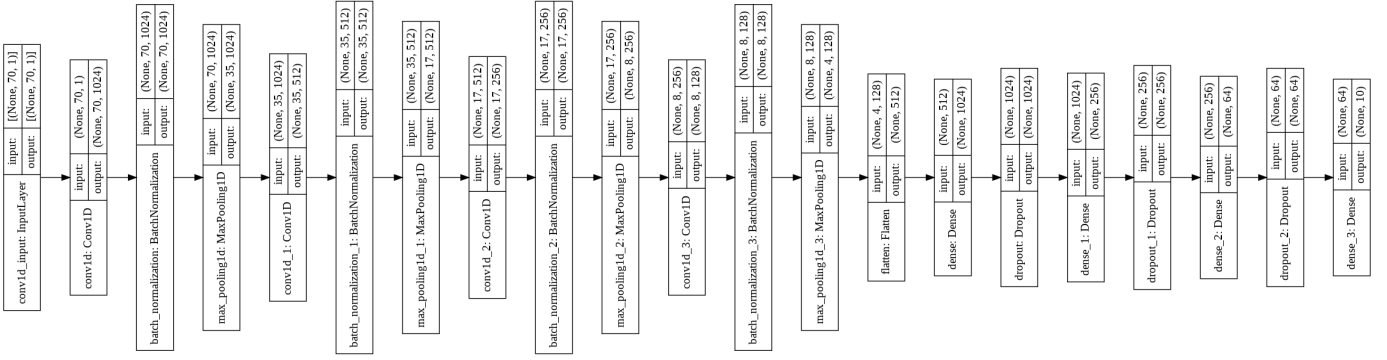


Fig. 4: CNN model architecture

In equation (4), ' U ' represents the normalization factor of the power in the window function and is expressed such that

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n) \quad (5)$$

In equation (5), $w(n)$ indicates the window function. The mean of these modified periodograms presents Welch's power spectrum which is considered as:

$$P_{xx}^W = \frac{1}{L} \sum_{i=0}^{L-1} \tilde{P}_{xx}^{(i)}(f) \quad (6)$$

In this paper 14 channels, 5 bands, window_size = 256, step size = 16, sample rate = 128 are used for per subject's data in feature extraction.

Here,

- **Channel:** The difference between the electrode and the weighted mean of contiguous electrodes is considered as an EEG channel.
- **Band:** A fixed range of wave frequencies and amplitudes over a time scale is called Frequency Bands in EEG.
- **Windows Size:** The length of a cutout (sliding) of a time sequence of data is known as the window size.
- **Step Size:** During the training period amounts of weights are needed to update, that is called step Size.
- **Sample Rate:** From a non-digital or continuous signal to create a digital or discrete signal how many samples are taken per second, this number is known as the sampling rate

C. Labeling and Normalization

The DEAP dataset contains raw data but after some pre-processing procedures, it becomes suitable to feed the classification model. To get the frequency domain of these data FFT has been applied. Then all the data are split into training and testing segments following the 7:1 ratio. Encoding is subsequently applied for labeling to avoid over-fitting. This dataset contains four label columns: arousal, valence, dominance and liking. But only arousal and valence labels are used for categorizing with the help of categorical function. Then normalization has been used to bring the different ranges of data between 0 and

1. Sometimes normalization helps to increase the accuracy of the models. Standard Scalar is one of the techniques for normalization. Initially, 2D arrays are found in the DEAP dataset but used models need 3D data as input. That's why 2D data is converted into 3D using reshaping.

D. CNN model structure

Initially, all preprocessing parts are completed before a model is created, so that models can learn various complex features correctly. A 1D-CNN was used whose hidden layer can be changed to increase the accuracy. A segment is created with a 1D Convolution layer, a batch normalization layer as well as a 1D max-pooling layer and this segment has been found a total of four times at the starting of the model architecture. The fourth segment's result is converted into a 1D array with the help of flattening. Then three connected layers (dense layers) are applied and after every connected layer a drop-out layer is also given to avoid the overfitting problem. Then the output layer has been constructed including the number of classes and the softmax activation function. The entire model architecture has been demonstrated in figure 4.

V. EXPERIMENT AND RESULT ANALYSIS

The extracted data contains the main data as well as the emotion state label. This research strives to better identify emotions with two different approaches. One of them is the conventional binary classification method and another is our proposed eight-class classification method. But in both classification methods, the same 1D-CNN model architecture has been used that is designed by us. This architecture is much more simple and light. Total four models are used for the entire experiment. Two models for the binary classification and the other two models for the eight-class classification, here all the model works on arousal and valence emotional states. The label array contains floating values from 1 to 9 for both valence and arousal. But among all the values, the amount of 9 is very poor. For the convenience of reducing calculation and space-complexity, we converted the 9 into 8.99. In this case, the difference between 9 and 8.99 is very negligible and it does not create any impact.

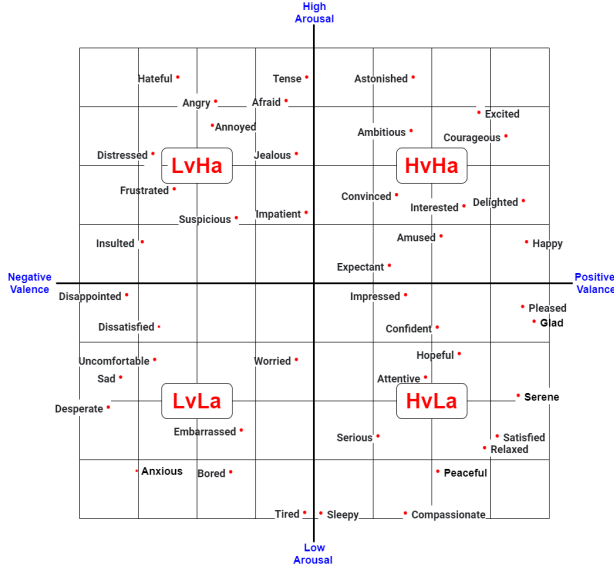


Fig. 5: Circumplex model for emotions

1) *Binary-Class*: In a two-dimensional emotion recognition system binary class classification is found as a conventional method. All the values of the label array are divided into two classes where 1-4.99 for one class and 5-8.99 for other classes. Then the binary arousal classification model and the binary valence classification model have been trained on all the data. Binary classification divides the entire emotion space into four classes and each of the four classes is the combination of more than one real-life known emotion. Four emotions are expressed by HaHv, LaHv, LaLv and HaLv. These four class ideas are demonstrated in figure 5.

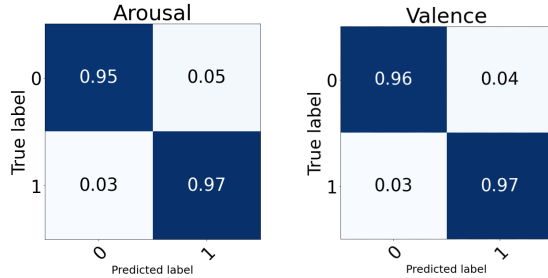


Fig. 6: Binary-class confusion matrix

The binary classifiers get 96.63% accuracy for arousal and 96.17% accuracy for valence. The binary arousal classifier achieved 99.65% training accuracy and 96.17% test accuracy in 131 epochs and at epoch 126, it provides the best accuracy. And the binary valence classifier provides 96.63% test accuracy when the training accuracy is 99.73% after 163 epochs but it shows the best test accuracy at 150th epoch.

2) *8-Class*: This technique works precisely to recognize emotion properly using the DEAP datasets values. Many real-life emotions can be recognized using the eight-class classification technique where the binary class classification

is able to find out only four compound emotions and each of these compound emotions consists of multiple real-life emotions. Using eight-class classifications those emotions are possible to recognize, a small number of them are mentioned in figure 5.

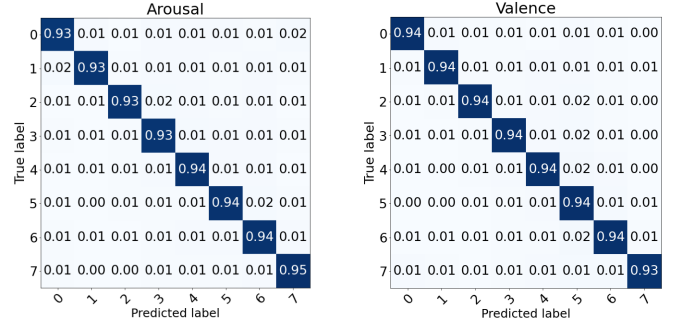


Fig. 7: 8-class confusion matrix

To be successful in this classification technique, arousal and valence states label array values are needed to divide into eight segments. Eight label segmentation have been created maintaining 1 - 1.99, 2 - 2.99, 3 - 3.99, 4 - 4.99, 5 - 5.99, 6 - 6.99, 7 - 7.99, 8 - 8.99 procedure. Then the eight-class arousal classification model and the eight-class valence classification model have been applied to all the data for training. The eight-class arousal classifier gets 98.83% training accuracy after running 151 epochs and gets 93.79% best test accuracy at 126th epoch. The eight-class valence classifier shows the best test accuracy of 93.83% at 120th epoch where it takes 160 epochs to get 98.92% train accuracy.

TABLE IV: Result summary

Type	Arousal	Valence
2-Class	96.18%	96.63%
8-Class	93.79%	93.83%

TABLE V: Result comparison for binary-class

No.	Modeling technique	No. of class	Accuracy
1	CNN [15]	2-Class	91.5%
2	LSTM [3]	2-Class	89.83%
3	CDCN [9]	2-Class	92.58%
4	KNN, ResNets [16]	2-Class	Valence:90.39% Arousal:89.06%
5	Our Model: 1D-CNN	2-Class	Valence:96.07% Arousal:96.33%

Using the DEAP dataset those model architectures are on the top list for better emotion recognition accuracy, the best accuracy models from each of them are demonstrated in table V. Yucel Cimtay Et al. [15] have worked with raw data for binary classification and get the highest accuracy among the reviewed CNN models with an accuracy of 91.5% . Many researchers have worked with the famous LSTM model for binary classification and also got good results. According to the literature review, Divya Acharya Et al. [3] have got 89.83% accuracy and this is the best accuracy for LSTM. To get

this best LSTM binary classification accuracy have used FFT as feature extraction. Analyzing the results of our reviewed papers, CDCN is also found in the top category for accuracy of binary class emotion recognition and Zhongke Gao Et al. [9] have worked with this model. They have got 92.58% accuracy where DE is the extraction method. Ningjie Liu Et al. [16] have applied their KNN, ResNets on DEAP dataset to get valence: 90.39% and arousal: 89.06% accuracy. But in the same field using FFT with the help of two 1D-CNN models we got 96.07% accuracy for valence and 96.33% accuracy for arousal which is the highest compared to all other models of table V.

VI. CONCLUSION AND FUTURE WORK

The models that have been used are very simple and light in terms of architecture. Our proposed CNN models are more effective for emotion recognition and outperform previous research in terms of accuracy. These are able to effectively classify preprocessed EEG data. For arousal-valence binary classification accuracy exceeds the same benchmark activities that shows a noticeable difference and introduces a much more precise eight-class classification approach which provides a satisfactory result also.

For future work, we would like to work with our methodology on real-time data so that the emotions of mentally challenged and autistic people can be expressed easily. We will also focus on how to make our recognizing models more efficient and portable.

REFERENCES

- [1] Rayatdoost, Soheil, and Mohammad Soleymani. "Cross-corpus EEG-based emotion recognition." In 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1-6. IEEE, 2018.
- [2] Acharya, Divya, Shivani Goel, Rishi Asthana, and Arpit Bhardwaj. "A novel fitness function in genetic programming to handle unbalanced emotion recognition data." *Pattern Recognition Letters* 133 (2020): 272-279.
- [3] Acharya, Divya, Shivani Goel, Harshit Bhardwaj, Aditi Sakalle, and Arpit Bhardwaj. "A long short term memory deep learning network for the classification of negative emotions using EEG signals." In 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2020.
- [4] Sharma, Rahul, Ram Bilas Pachori, and Pradip Sircar. "Automated emotion recognition based on higher order statistics and deep learning algorithm." *Biomedical Signal Processing and Control* 58 (2020): 101867.
- [5] Wang, Fei, Shichao Wu, Weiwei Zhang, Zongfeng Xu, Yahui Zhang, Chengdong Wu, and Sonya Coleman. "Emotion recognition with convolutional neural network and EEG-based EFDMS." *Neuropsychologia* 146 (2020): 107506.
- [6] Hassan, Reshad, Sakib Hasan, Md Jubaer Hasan, Md Rafat Jamader, David Eisenberg, and Tanmoy Pias. "Human Attention Recognition with Machine Learning from Brain-EEG Signals." In 2020 IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), pp. 16-19. IEEE, 2020.
- [7] Nath, Debarshi, Mrigank Singh, Divyashikha Sethia, Diksha Kalra, and S. Indu. "An Efficient Approach to EEG-Based Emotion Recognition using LSTM Network." In 2020 16th IEEE International Colloquium on Signal Processing Its Applications (CSPA), pp. 88-92. IEEE, 2020.
- [8] Cheah, Kit Hwa, Humaira Nisar, Vooi Voon Yap, and Chen-Yi Lee. "Short-time-span EEG-based personalized emotion recognition with deep convolutional neural network." In 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 78-83. IEEE, 2019.
- [9] Gao, Zhongke, Xinmin Wang, Yuxuan Yang, Yanli Li, Kai Ma, and Guanrong Chen. "A channel-fused dense convolutional network for EEG-based emotion recognition." *IEEE Transactions on Cognitive and Developmental Systems* (2020).
- [10] Luo, Yuling, Qiang Fu, Juntao Xie, Yunbai Qin, Guopei Wu, Junxiu Liu, Frank Jiang, Yi Cao, and Xuemei Ding. "EEG-based emotion classification using spiking neural networks." *IEEE Access* 8 (2020): 46007-46016.
- [11] Cimtay, Yucel, and Erhan Ekmekcioglu. "Investigating the use of pre-trained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition." *Sensors* 20, no. 7 (2020): 2034.
- [12] Cao, Guolu, Yuliang Ma, Xiaofei Meng, Yunyuan Gao, and Ming Meng. "Emotion recognition based on CNN." In 2019 Chinese Control Conference (CCC), pp. 8627-8630. IEEE, 2019.
- [13] Abdel-Ghaffar, Eman A., and Mohamed Daoudi. "Emotion Recognition from Multidimensional Electroencephalographic Signals on the Manifold of Symmetric Positive Definite Matrices." In 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 354-359. IEEE, 2020.
- [14] Zhong, Xiaolong, Zhong Yin, and Jianhua Zhang. "Cross-Subject emotion recognition from EEG using Convolutional Neural Networks." In 2020 39th Chinese Control Conference (CCC), pp. 7516-7521. IEEE, 2020.
- [15] Cimtay, Yucel, Erhan Ekmekcioglu, and Seyma Caglar-Ozhan. "Cross-subject multimodal emotion recognition based on hybrid fusion." *IEEE Access* 8 (2020): 168865-168878.
- [16] Liu, Ningjie, Yuchun Fang, Ling Li, Limin Hou, Fenglei Yang, and Yike Guo. "Multiple feature fusion for automatic emotion recognition using EEG signals." In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 896-900. IEEE, 2018.
- [17] Koelstra, Sander, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. "Deap: A database for emotion analysis; using physiological signals." *IEEE transactions on affective computing* 3, no. 1 (2011): 18-31.
- [18] Yoon, Hyun Joong, and Seong Youb Chung. "EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm." *Computers in biology and medicine* 43, no. 12 (2013): 2230-2237.
- [19] Rozgić, Viktor, Shiv N. Vitaladevuni, and Rohit Prasad. "Robust EEG emotion classification using segment level decision fusion." In 2013 IEEE international conference on acoustics, speech and signal processing, pp. 1286-1290. IEEE, 2013.
- [20] Zhang, Xiaowei, Bin Hu, Jing Chen, and Philip Moore. "Ontology-based context modeling for emotion recognition in an intelligent web." *World Wide Web* 16, no. 4 (2013): 497-513.
- [21] Li, Xiang, Dawei Song, Peng Zhang, Guangliang Yu, Yuexian Hou, and Bin Hu. "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network." In 2016 IEEE international conference on bioinformatics and biomedicine (BIBM), pp. 352-359. IEEE, 2016.
- [22] Al-Nafjan, Abeer, Manar Hosny, Areej Al-Wabil, and Yousef Al-Ohali. "Classification of human emotions from electroencephalogram (EEG) signal using deep neural network." *Int. J. Adv. Comput. Sci. Appl* 8, no. 9 (2017): 419-425.
- [23] Liu, Wei, Wei-Long Zheng, and Bao-Liang Lu. "Emotion recognition using multimodal deep learning." In International conference on neural information processing, pp. 521-529. Springer, Cham, 2016.
- [24] Tripathi, Samarth, Shrinivas Acharya, Ranti Dev Sharma, Sudhanshi Mittal, and Samit Bhattacharya. "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset." In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 4746-4752. 2017.
- [25] Zheng, Wei-Long, Jia-Yi Zhu, and Bao-Liang Lu. "Identifying stable patterns over time for emotion recognition from EEG." *IEEE Transactions on Affective Computing* 10, no. 3 (2017): 417-429.