October 2018

Emotion Recognition Using Deep Convolutional Neural Network with Large Scale Physiological Data

Astha Sharma
University of South Florida, asthasharma017@gmail.com

Emotion Recognition Using Deep Convolutional Neural Network with Large Scale Physiological Data

by

Astha Sharma

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Computer Science
Department of Computer Science and Engineering
College of Engineering
University of South Florida

Major Professor: Shaun Canavan, Ph.D.
Paul A. Rosen, Ph.D.
Marvin Andujar, Ph.D.

Date of Approval:
October 18, 2018

Keywords: Affective Computing, Deep Learning, BP4D+, DEAP

DEDICATION

This thesis is dedicated to my wonderful parents who always encouraged me to learn and grow, supported me for my every step and loved me unconditionally.

ACKNOWLEDGMENTS

I want to thank my thesis advisor, Dr. Shaun Canavan for giving me this opportunity to research with him and trusting me when I knew nothing about research. This work would not be possible without his guidance and support. I would also thank Dr. Marvin Andujar for his thoughtful advice and research ideas, Dr. Paul A. Rosen for being a supportive member in my defense committee, Dr. Yu Sun for his valuable teaching which helped me a lot with my research and all the professors who taught me during my master's program.

This acknowledgement will not complete without thanking my friends and lab mates Sk Rahatul Jannat, Ghada Alzamzmi, Rahul Paul, Diego Fabiano Santalices, Palak Dave and Md Taufeeq Uddin who supported me and helped me whenever I needed. I would also like to thank my friends and roommates Astha Kakkad and Priya Narvekar who made my time during master's a wonderful and memorable journey.

At last, I want to thank the staff of the CSE main office and the entire department for their help and courtesy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Classification of emotions plays a very important role in affective computing and has real-world applications in fields as diverse as entertainment, medical, defense, retail and education. These applications include video games, virtual reality, pain recognition, lie detection, classification of Autistic Spectrum Disorder (ASD), analysis of stress levels, and determining attention levels. This vast range of applications motivated us to study automatic emotion recognition which can be done by using facial expression, speech and physiological data.

A person's physiological signals such are heart rate, and blood pressure are deeply linked with their emotional states and can be used to identify a variety of emotions; however, they are less frequently explored for emotion recognition compared to audiovisual signals such as facial expression and voice. In this thesis, we investigate a multimodal approach to emotion recognition using physiological signals by showing how these signals can be combined and used to accurately identify a wide range of emotions such as happiness, sadness, and pain. We use deep convolutional neural network for our experiments. We also detail comparisons between gender specific models of emotion. Our investigation makes use of deep convolutional neural networks, which are the latest state of art in supervised learning, on two publicly available databases, namely DEAP and BP4D+. We achieved an average emotion recognition accuracy of 98.89% on BP4D+ and on DEAP it is 86.09% for valence, 90.61% for arousal, 90.48% for liking and 90.95% for dominance. We also compare our results to current state of the art, showing the superior performance of our method.

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Statement

Emotional state is mental composition which shows our reaction on an experience. It is an integral part of human communication and behavior [3]. Recognizing emotions using machines has important roles in human-computer interaction. While recognizing other's emotions, we generally look at their facial expressions, speech or body language, however, these features can be misguiding. Facial expression, voice and body languages can be faked. Face can be occluded which can make feature identification difficult. On the top of it facial expressions can be contradictory, for example it has been observed that people smile during negative emotional experiences [4]. Considering this, physiological signals such as heart rate, blood pressure, respiratory signals, and Electroencephalogram (EEG) signals can be important traits for identifying emotions accurately. However, less attention is paid to physiological signals so far as opposed to audiovisual emotional channels [5]. One reason for this can be the relative difficulty in labeling the data with specific emotion, and in some cases there is a requirement of heavy machines to record the data, such as an Electroencephalogram, which are fine to be used in experimental setup but are very hard to use in daily life. This can make large-scale datasets more difficult to obtain. On the other hand, collecting and annotating audiovisual signals is comparatively easy and straightforward because they can be felt as opposed to physiological signals [6]. However, importance of recognizing emotions with physiological signals cannot be ignored, as the study by Lindh et al., to assess pain based on the heart rate variability during heel lancing procedure [7], says that there is a strong relation between pain intensity and increase in the heart rate. The experiments show that heel squeezing, which is a painful event, generated the highest heart rate.

In this thesis, We conduct experiments on BP4D+ [1] and DEAP [2] datasets which are the two largest databases having physiological data for emotions. For the BP4D+ dataset, we combine a total 8 physiological signals across 4 types, namely blood pressure (BP), heart rate (HR), EDA, and respiration. For the DEAP dataset, we combine multiple variations of the 32 channels of EEG signals along with 8 peripheral signals, listed in table 3.3 on page 13. We use this data to train a deep convolutional neural network to identify emotions. For our experiments, we first do smoothing and normalization for data preprocessing to remove noise and make the physiological signals consistent. We are using a 9 layer convolution neural network to process this data and recognize emotion. We also compare our results to the current state of the art showing superior performance. Figure 1.1 on page 3 gives an overview of the way our system works.

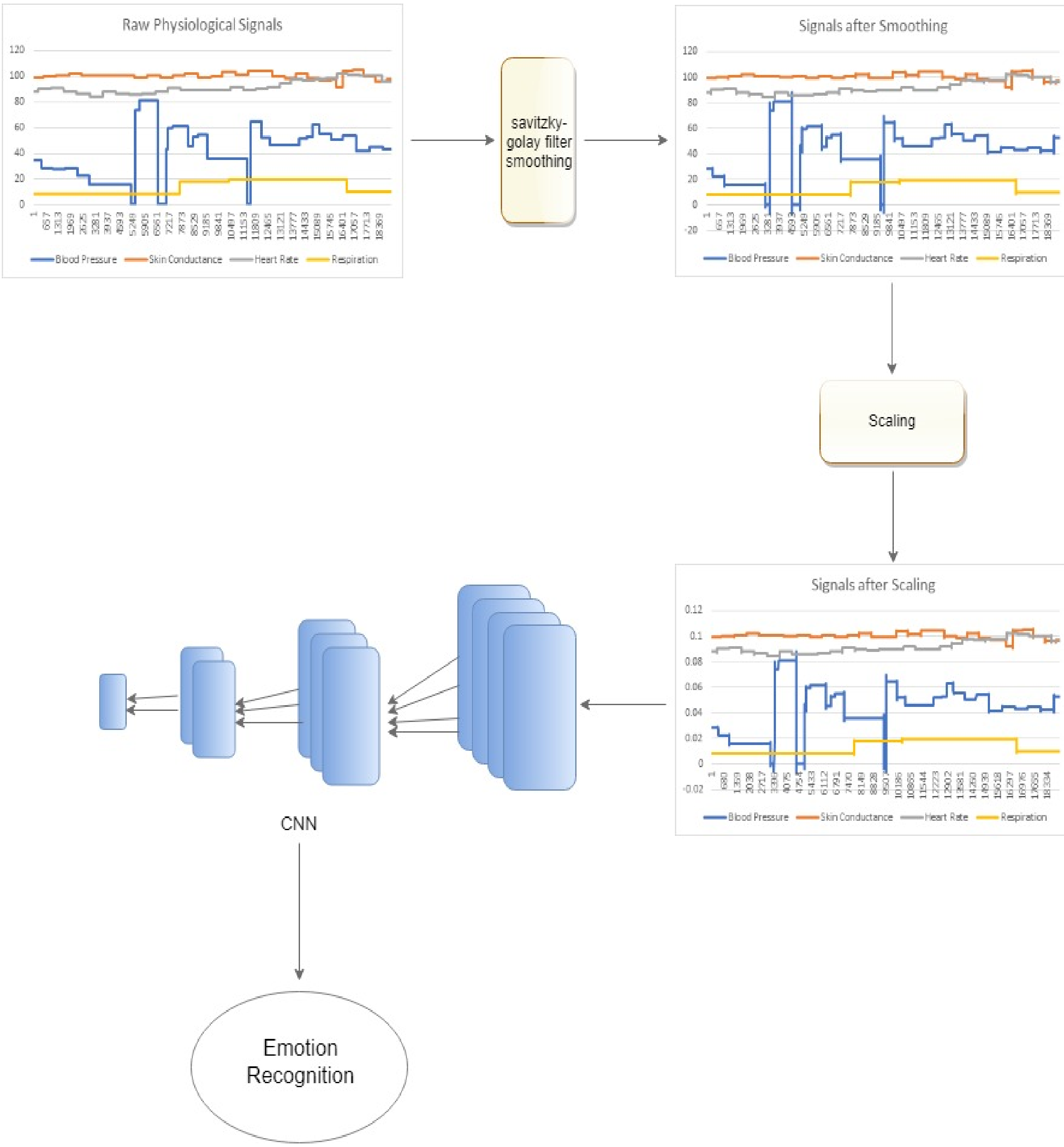

Figure 1.1 System overview.

1.2 Contributions

The main contribution of the work is a new method for recognizing emotion based on physiological signals by using deep convolutional neural network. The contributions of this thesis are 5-fold and are detailed below.

1. We develop a new method for recognizing emotion using physiological data and deep convolutional neural networks (CNN): We fed sequential physiological signals of multiple subjects, collected at different instances, to a deep convolutional neural network model and evaluated the results on the accuracy matrix. The model was designed using multiple convolutional, maxpooling and fully connected layers. We used regularization to avoid overfitting and evaluated the model against fully unseen data.

2. We used the new BP4D+ database: BP4D+ [1] is a large database that contains spontaneous emotional data sequences of 140 subjects and 10 emotion classes. To the best of our knowledge, we are the first to use the physiological signals from this database for emotion classification and achieved good classification results.

3. Comparison of our results to state-of-the-art methods: We did our experiments on two databases – BP4D+ [1] and DEAP [2]. We show state of the art performance on DEAP and to the best of our knowledge we are the first to report results using physiological data on BP4D+, contributing a base-line result to the community.

4. Analysis of gender-specific models of emotion, using physiological signals from BP4D+ [1]: To the best of our knowledge, we are the first to develop physiological signal-based gender-specific models for emotion, by training 2 separate CNNs. This part of our work is useful to study generalization of the CNN model, because in the real world, there may be instances when we do not have enough data for one gender. The first is trained on data only from female subjects and the second is trained on data only from male subjects. We report results of testing both genders on both models, showing some physiological differences between genders when modeling emotion.

5. One of first works to use deep learning with EEG data for emotion: To the best of our knowledge, deep convolutional neural network classifier has not been used before with EEG data for emotion recognition. This thesis can be the baseline for this kind of work.

CHAPTER 2

RELATED WORK

Deep learning with physiological signals for emotion recognition was first used by [8]. Skin conductance and blood pressure were combined to use for this method. The efficiency of this model was compared with standard feature extraction and feature selection methods. Results show that deep learning performed better than standard feature selection algorithms and the method shows more generalization. In another study [9], physiological data was used to measure stress. For 5 days, skin conductance for 18 participants was collected with wrist sensors, as well as their mobile phone usage like call, SMS, and location were monitored. A survey was done to know stress, mood, sleep, tiredness, general health, alcohol or caffeinated beverage intake and electronics usage. Correlation analysis was applied to find important features that shows stress and machine learning was used to classify whether the participant is stressed or not.

Electroencephalography (EEG) signals contain information about electrical activity of the brain, and can be useful in diagnosis of epilepsy. A computer-aided diagnosis (CAD) system was developed to automate the classification of EEG signals in three categories - normal, preictal, and seizure. This method achieved 88.67%, 90.00% and 95.00% accuracy for respective classes[10]. In another research, EEG signals were used to classify different emotions such as happiness, fear, and sadness. The experiments were performed on DEAP database. Shannon Entropy was used for feature extraction and a multi-class Support Vector Machine (MCSVM) was used for training. The accuracy obtained for classification was 94.097% [11].

Some works have combined multiple physiological signals to find the affective state. A musical induction method was used to ignite real emotion in subjects for data collection[12]. Electromyogram, electrocardiogram, skin conductivity and respiration signals were collected in the laboratory. Classification of four musical emotions (positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal) was performed by feature-based multiclass classification. The accuracies they obtained were 95% for subject-dependent and 70% for subject-independent classification.

Another study was done that shows variations in physiological signals on a daily basis. They claim that 'the features of different emotions on the same day tend to cluster more tightly than do the features of the same emotion on different days' [13]. The technique of seeding a Fisher Projection with the results of Sequential Floating was proposed to handle this variation. The accuracy obtained with this approach was 81% on 8 emotion classes.

There is another database DEAP, that presents a multimodal set of data with varieties of human emotional states. In this database, the EEG and peripheral physiological signals of 32 participants were recorded [2]. Using DEAP, an emotion classification method based on the Bayes classifier was developed [14]. A new function called weighted-log-posterior function for the Bayes classifier was proposed which is the weighted sum of likelihood function of each feature plus bias factor under the assumption of feature independence. Supervisory learning was used to calculate weights and bias. The accuracies they obtained for valence and arousal classification are 66.6% and 66.4%. Another study on DEAP was done which uses single-trial binary classification for each of four emotional dimensions - arousal, valence, dominance and liking. Average classification accuracies for each subject were reported - arousal (68.4%), valence (76.9%), dominance (73.9%) and liking (75.3%) [15]. A recent study by Liu et al. [16], using Bimodal Deep autoencoder (BDAE) for feature extraction and deep learning for training, got the results - arousal (80.5%), valence (85.2%), dominance (84.9%) and liking (82.4%).

These works motivated us to research further about techniques to recognize emotions using physiological data. The technique we are using for classification is deep convolutional neural networks, which has really improved state-of-the-are of classification methods in speech, face recognition and object recognition [17]. A recent study is done by Google, developing a system called FaceNet [18], in which they trained a deep convolutional network on the facial images in Wild (LFW) dataset and achieved 99.63% accuracy.

Deep neural networks has been proved to work well on multimodal data as well. In a study done using multiple modalities related to series of different tasks [19], researchers used cross modality feature learning using audio and video data, with training the optimizer on audio-only data and testing on vidio-only data and vice-versa. They used CUAVE and AVLetters datasets. Motivated by these studies, we propose a new method for emotion recognition by training a CNN on combined physiological signals, using BP4D+ and DEAP datasets. Our proposed method is detailed in the next chapter, followed by our results, related discussion, and thoughts on the future of this work.

CHAPTER 3

METHOD

3.1 Data

Our body generates various signals during the functioning of its physiological systems. These signals are called physiological signals and include but are not limited to heart rate, pulse rate, muscle movement, skin conductance and brain activities. Dynamic changes in these signals' effect perceptual, affective, and cognitive state [20]. Physiological signals can be a compelling approach in recognizing internal cognitive and emotional state. Because of the high temporal resolution of these signals, a large amount of authentic and relevant data can be gathered [21]. For our experiments, we are using two databases- BP4D+ and DEAP.

BP4D+ [1] is a multimodal spontaneous emotion corpus (MMSE) which includes physiological data sequences like blood pressure, EDA (skin conductance), heart rate and respiration as well as 2D, 3D and thermal images and videos. The huge versatility of gathered data makes BP4D+ the largest database of this kind. This data was gathered from 140 subjects (58 males and 82 females) from 18 to 66 years age range. Each participant went through 10 tasks as listed in table 3.1, to evoke a range of authentic emotions in a laboratory setting [1].

In the dataset BP4D+, physiological signals were captured under a sampling rate of 1000 points per seconds. Figure 3.1 on page 11 illustrates different physiological signals for various emotions. Refer Table 3.1 on page 10 for corresponding emotion for each task. The dataset contains 8 different sequences representing 8 different signals which are measured for each participant and every combination of signals at an instance corresponds to a specific

Table 3.1 Activities to elicit emotions, taken from BP4D+ [1].

Task	Activity	Target Emotion
1	Interview: Listen to a funny joke	Happiness or Smusement
2	Graphic show: Watch 3D avatar of participant	Surprise
3	Video clip: 911 emergency phone call	Sadness
4	Experince a sudden burst of sound	Startle or Surprise
5	Interview: True or false question	Skeptical
6	Improvise a silly song	Embarressment
7	Experience physical threat in dart game	Fear or Nervous
8	Cold pressor: Submerge hand into ice water	Physical pain
9	Interview: Complained for the poor performance	Anger or upset
10	Experience smelly odor	Disgust

emotion. These details are provided in separate text files for each participant which are listed below in table 3.2 [1].

Table 3.2 Details about physiological signals, taken from BP4D+ [1].

Feature	File Associated	Measurement
Blood Pressure	BP Dia_mmHg.txt LA Systolic BP_mmHg.txt LA Mean BP_mmHg.txt BP_mmHg.txt	[-25, 300 mmHg]
Respiration	Respiration Rate_BPM.txt Resp_Volts.txt	[0, 200 breaths/min Voltage]
Heart Rate	Pulse_Rate_BPM.txt	[30, 300 beats/min]
EDA	EDA_microsiemens.txt]	Micro Siemens

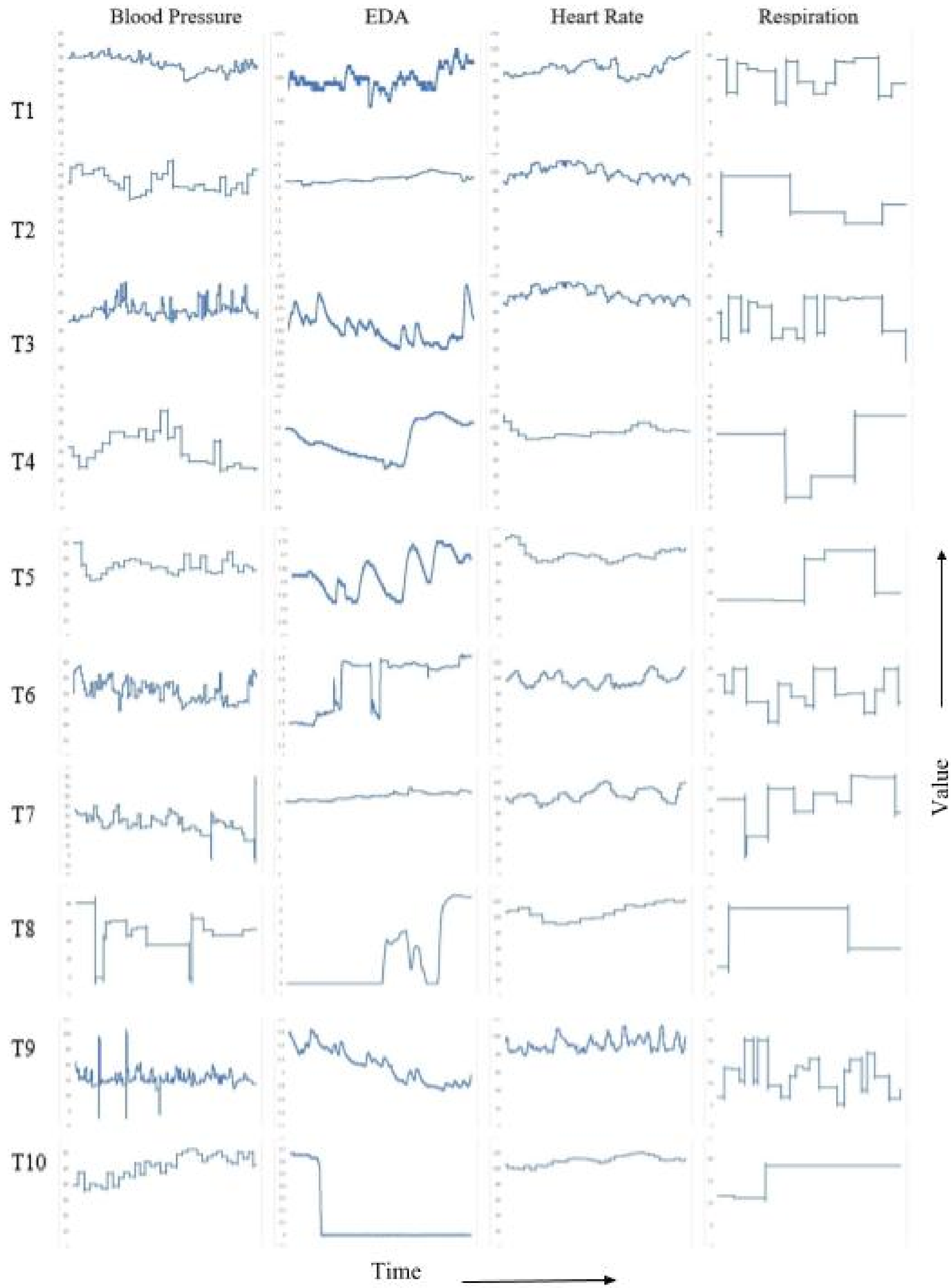

Figure 3.1 Signal vs task.

Another effective physiological signal that directly relates to emotions is brain activity which is measured in terms of flow of neurons as the large number of active neurons generate electrical activity at the surface of the scalp and recorded with EEG electrodes. EEG devices read these electrical brain activities from the scalp with high temporal resolution. DEAP [2] contains EEG data which is based on Valence-Arousal emotion model depicted in Figure 3.3. The database also has sequences of peripheral data which includes EOG (eye movements), EMG (muscle movement), GSR, respiration, blood pressure and temperature. The data was collected from 32 participants (19-37 years age, 50% male and 50% female) playing total 40 one-minute long music video to elicit emotions. 32 EEG channels based on the 10–20 system [22] for recording EEG data and 8 channels for peripheral physiological data were used. These channels are listed in table 3.3 on page 13. In this table, the first 32 channels are the EEG channels and last 8 channels are the peripheral channels used in DEAP database. These signals were recorded with a sampling rate of 512 Hz which was downsampled to 128 Hz after preprocessing. The data was labeled with arousal, valence, dominance, and liking values ranging from 1 to 9 showing intensity of each emotional state. Figure 3.2 shows effect of intensities of these emotions on affective state.

Table 3.3 40 Channels (32 EEG and 8 peripheral) used in DEAP dataset [2].

Channel #	**Channel Name**
1	Fp1
2	AF3
3	F7
4	F3
5	FC1
6	FC5
7	T7
8	C3
9	CP1
10	CP5
11	P7
12	P3
13	Pz
14	PO3
15	O1
16	Oz
17	O2
18	PO4
19	P4
20	P8
21	CP6
22	CP2
23	C4
24	T8
25	FC6
26	FC2
27	F4
28	F8
29	AF4
30	FP2
31	Fz
32	Cz
33	hEOG (horizontal EOG, hEOG1 - hEOG2)
34	vEOG (vertical EOG, vEOG1 - vEOG2)
35	zEMG (Zygomaticus Major EMG, zEMG1 - zEMG2)
36	tEMG (Trapezius EMG, tEMG1 - tEMG2)
37	GSR (Ohm)
38	Respiration belt
39	Plethysmograph
40	Temperature

Figure 3.2 Effect of intensity of affective state on emotions.

As can be seen in figure 3.2, when a user has high arousal they are alert and excited, however, low arousal implies disinterest or boredom. For valence, when it is high the person can generally be considered happy, on the other-hand they can be considered sad when it is low. High dominance results in feeling empowered, while low dominance implies a feeling of helplessness. Intuitively, a high liking level implies a feeling of love, while low is dislike (or even disgust). The level of valence and arousal together can determine the emotional states as detailed in figure 3.3 [23] on page 15.

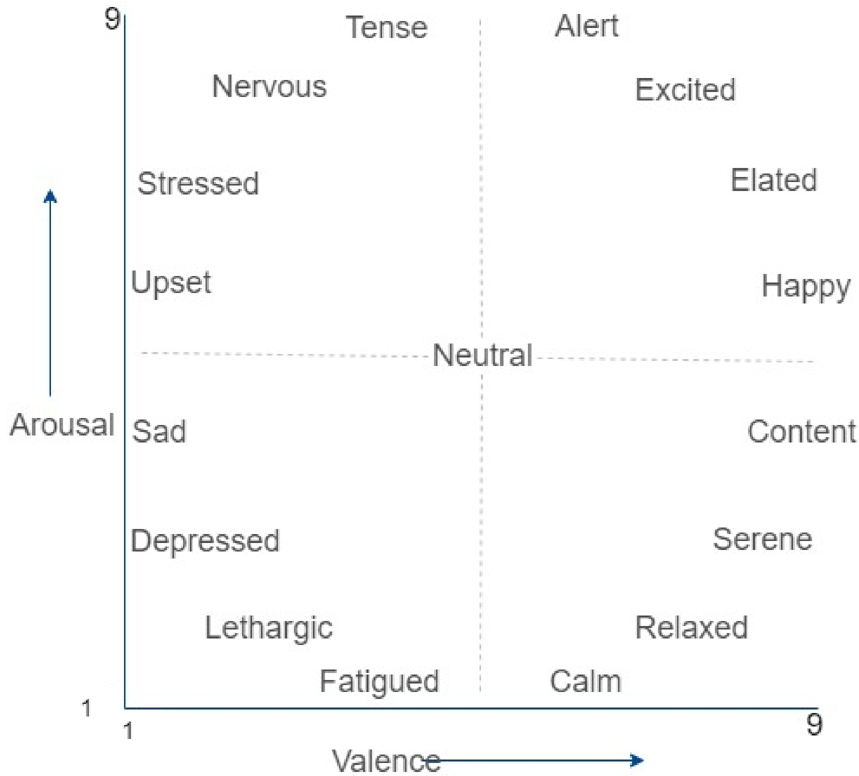

Figure 3.3 Effect of valence and arousal on emotions.

3.2 Preprocessing

Raw data can be noisy and inconsistent, and it can be difficult to obtain reliable results with such data. Data preprocessing can help to resolve such issues and prepares raw data sequences to be processed further as the input to a neural network. We used Smoothing and Scaling for preprocessing of BP4D+ signals. We detail both of these in the following subsections.

3.2.1 Smoothing

Smoothing is a technique to increase signal to noise ratio without deforming the signal a lot by removing rough, fast edging components and highlighting slow changes in values, which makes it easier to see trends in the data. This is done using various filters. These filters replace each data point with the best fit value with respect to its neighbors. Three types of filters are used in this technique: Moving average filter, Savitzky-golay filter and Median filter. In our experiments we are using savitzky-golay filter [24] as it is more efficient in handling delay alignment and the transient effect at the start and end of the sequence, compared to the other two [25]. This keeps the signal from getting deformed, making the filter better than the other two for our requirement.

Savitzky-golay filter, also known as digital smoothing polynomial filter or least square filter performs polynomial filtering to data frames which are the segments of whole data. The approach is to find a filter coefficients c, that preserve higher moments, and a polynomial of higher order within for each point in the sequence within a moving window of filter. The point at center position in the moving window is the target point and it is replaced by the result obtained by least-squares fitting done for the polynomial to all n L + n R + 1 points in the moving window. The next iteration is done by moving the window to the i+1 point and repeating this procedure for all points in the dataset [24].

$$g_i = \sum_{n=-n_L}^{n_R} c_n f_{i+n} \tag{3.1}$$

Here, i is the position in sequence for target data point, n_L is the number of points to the left of the data point within the moving window, n_R is the number of points to the right of the data point within the moving window, c_n is the filter coefficient and f_{i+n} is the polynomial function used for filtering. Figure 3.4 on page 17 shows the signals before and after applying savitzky-golay filter.

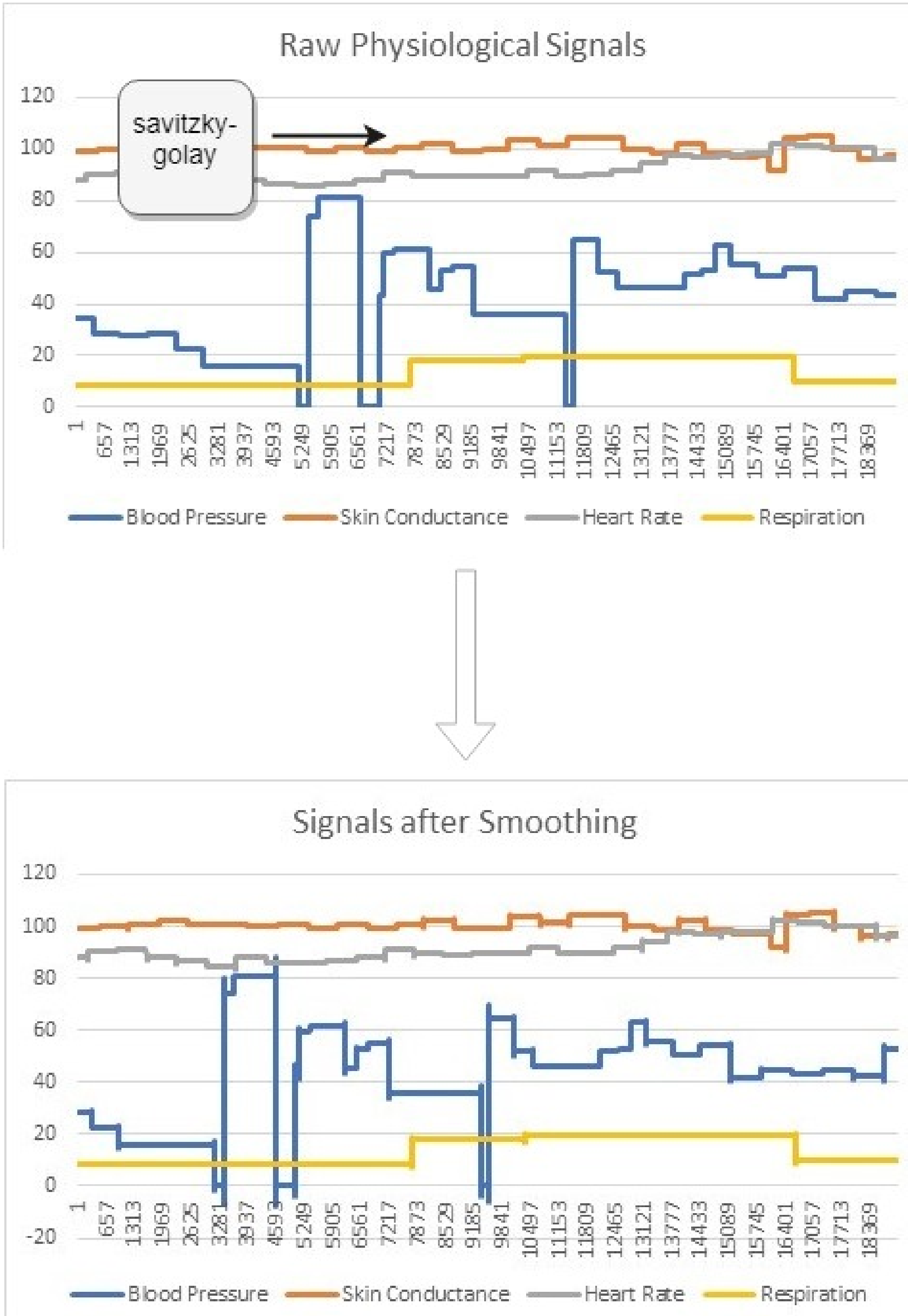

Figure 3.4 Physiological signals before and after applying savitzky-golay filter.

3.2.2 Scaling

The range of values of different features in the raw data may have a large variation. Some machine learning algorithms are sensitive to this variation. Normalization is a way to bring signals to identical levels which leads to faster optimization of machine learning models [26].

In our experiments we used scaling which is a technique to perform normalization by bringing the dataset between range 0 to 1. Figure 3.5 on page 18 shows the data sequences after scaling.

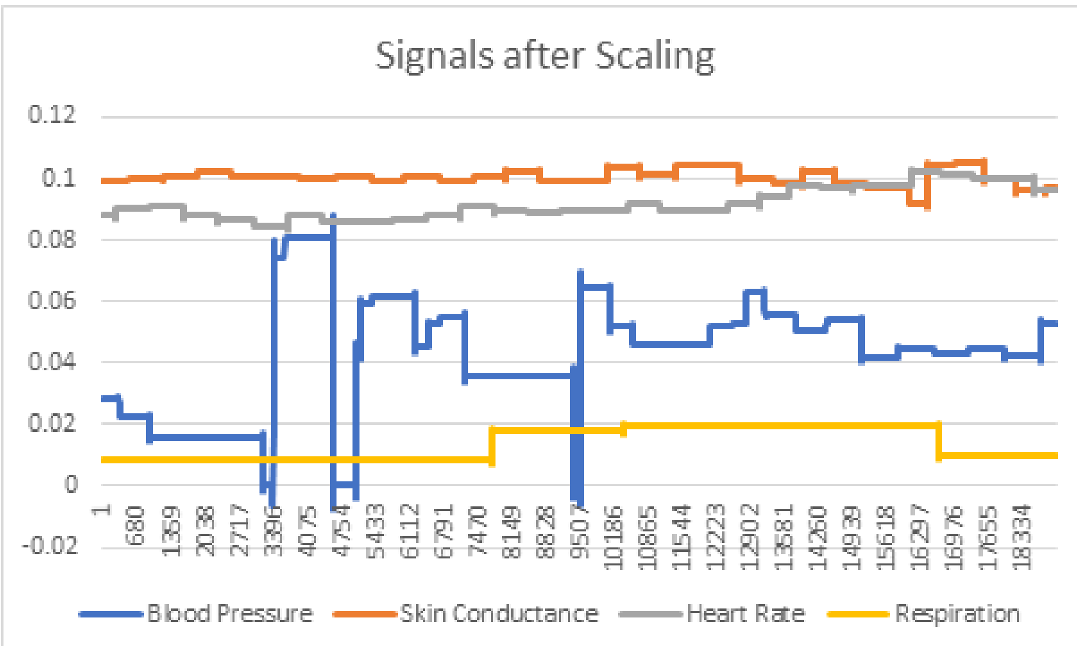

Figure 3.5 Data sequences after scaling.

We did not do any extra preprocessing on DEAP database the required preprocessing like normalization and artifact removal were already done on it [2].

3.3 Procedure

We used Deep Learning [17] as our supervised learning technique. Deep neural networks are the latest state of art in classification techniques. It has been proved highly efficient and has outperformed the ability of human in audio, images and text classification [27]. This machine learning approach is based on representation learning, i.e. it processed multiple layers of input data and automatically learns and finds the features required for classification by modifying training parameters recursively [28]. It is being used in wide verities of tasks in medical, gaming and entertainment fields like medical image classification [29], time series classification and prediction of future sales prices [30]. Considering this effectiveness of deep learning architectures, we chose to use it for our experiments.

Deep learning with neural networks can use hundreds of hidden layers and it requires large amounts of training data. Building a good neural network model always requires careful consideration of the architecture of the network as well as the input data format [31]. We created a 9-layer deep neural network [17] and used multiple layers of convolution, pooling, regularization, flattening and fully connected layers. We optimized the model using RMSProp optimizer [32].

The first layer of our network model is a convolution layer. It generates output in form of array of tensors as a convolution kernel convolved with the input of this layer. Input shape and dimensions must be determined if this layer is used as the first layer of network [33] [34]. The activation function used in this layer is Rectified Linear Unit (ReLU). ReLU has become very popular in recent years and it is proved to have 6 times improvement in convergence from Tanh function [35] as well as it gives better speed and performance than standard sigmoid function [35]. Mathematical formula of ReLU is:

$$R(x) = max(0, x) \tag{3.2}$$

$$R(x) = \begin{cases} 0, if x < 0 \\ x, if x >= 0 \end{cases} \tag{3.3}$$

The next layer is max pooling which does down-sampling of data to avoid overfitting by reducing the dimensionality. Another benefit of max pooling is that it reduces computational cost by decreasing number of parameters [36]. We are using two sets of convolutions, activation and max pooling layers in our CNN model.

We are using dropout for regularization, which helps the model to generalize better by reducing the risk of overfitting [35] [37]. Dropping out random nodes and their connections in the network makes nodes in the network insensitive to the weights of the other nodes, which makes the model robust [35] [37].

The next layer is Flatten which converts the 2-dimensional output of previous layer into a single long continuous linear vector. This is required before the flow goes to fully connected or dense layer. A dense layer is just an artificial neural network (ANN) classifier and requires individual features. Flattening is needed to convert the input array into a feature vector [38]. Next layer of the network is a fully connected or dense layer. In this layer, each input node is connected to each output node. It takes a flattened vector as input and gives n-dimension tensor as output. Using dense layer has several advantages. “It solves the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters” [39]. There are two fully connected layers used in the model. Activation function used in the last fully connected layer is softmax. It converges the output between 0 and 1 which makes it preferable to use it in the last layer of network for classification problems. [27]. It is used in categorial probabilistic distribution. Mathematical formula of softmax function is:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{3.4}$$

The model is compiled with the RMSprop optimizer which works by dividing the learning rate by an exponentially decaying average of squared gradients [32]. Default value of learning rate for RMSprop is 0.001 [40]. The objective function used to compile the model is categorical_crossentropy, which is commonly used when number of classes on the dataset is more than two [41]. Experiments were done with keeping number of epochs 150 and batch size 32. Learning is evaluated on accuracy metrics. Architecture of our CNN is depicted in figure 3.6.

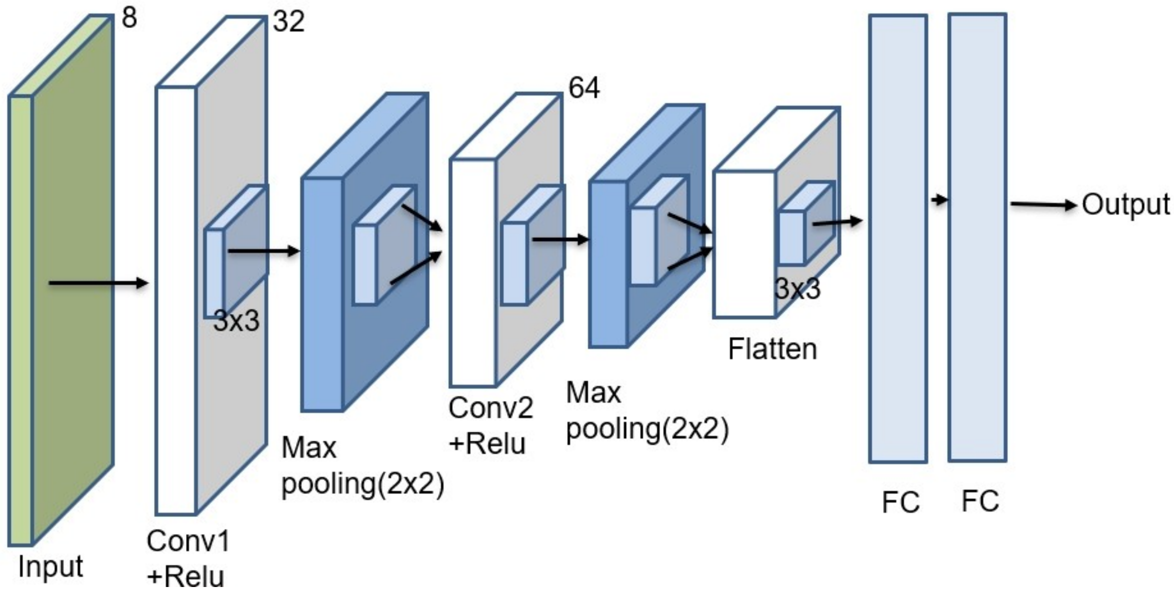

Figure 3.6 CNN architecture.

3.3.1 Experiment 1 : Training on BP4D+

BP4D+ database has 140 subjects consisting 58 males and 82 females. There is data for 10 emotions for each subject and 8 different physiological sequences for each emotion. Each set of features is labeled as a specific emotion as detailed in table 3.1. Figure 3.7 shows the feature vector structure of data used for training.

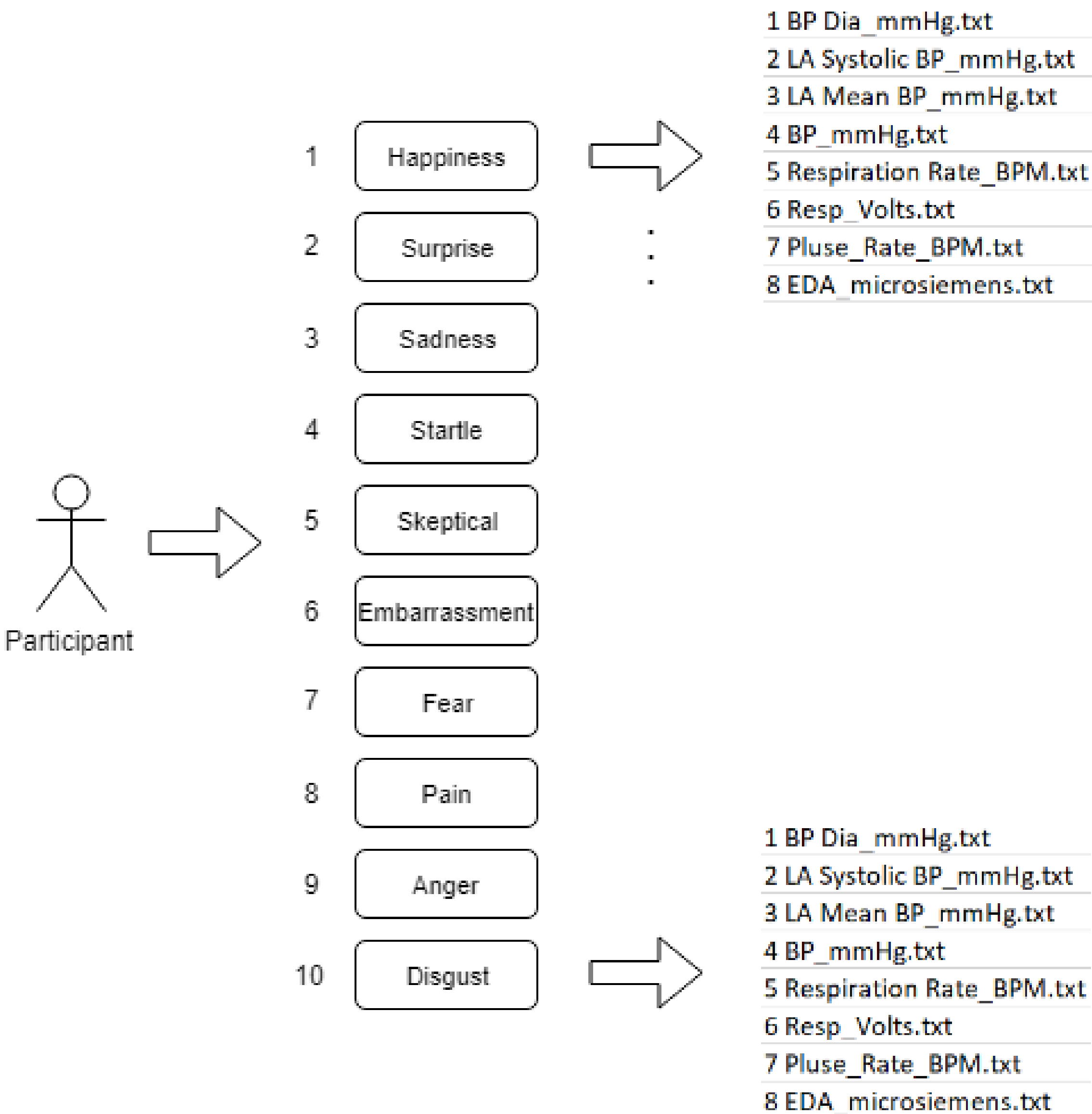

Figure 3.7 BP4D+ feature vector structure.

The following 6 iterations were performed on this data, with a ratio of 4:1 for training and testing for all experiments:

1. The first set of experiments was done on all subjects individually, which creates 140 different models for all different subjects and 140 sets of results. Mean and standard deviation of the results are calculated to evaluate the learning.

2. The second set of experiments was done on all female subjects individually, which creates 82 different models for all different female subjects and 82 sets of results. Mean and standard deviation of the results are calculated to evaluate the learning.

3. The third set of experiments was done on all male subjects individually, which creates 58 different models for all different male subjects and 58 sets of results. Mean and standard deviation of the results are calculated to evaluate the learning.

4. Fourth experiment was done using whole data at once. In this experiment 80% subjects were used for training and remaining 20% for testing the model.

5. Fifth experiment was done taking only females to train the model and testing was done for both males and females.

6. Sixth experiment was done taking only males to train the model and testing was done for both males and females.

3.3.2 Experiment 2 : Training on DEAP

DEAP database has 32 subjects participating in data collection. 40 trials were done for each participant and in every trial a one minute long video was shown to the participant to elicit different emotions. 32 EEG and 8 peripheral channels were recorded, which constitutes total 40 channels of data, listed in table 3.3 on page 13. Each channel has a sequence of 8064 data points. Each set of features is labeled positive(1) and negative(0) for all four emotions - valence, arousal, dominance and liking. We consider it positive if the emotion rating >= 5, otherwise it is negative. Figure 3.8 shows the feature vector structure of data used for training.

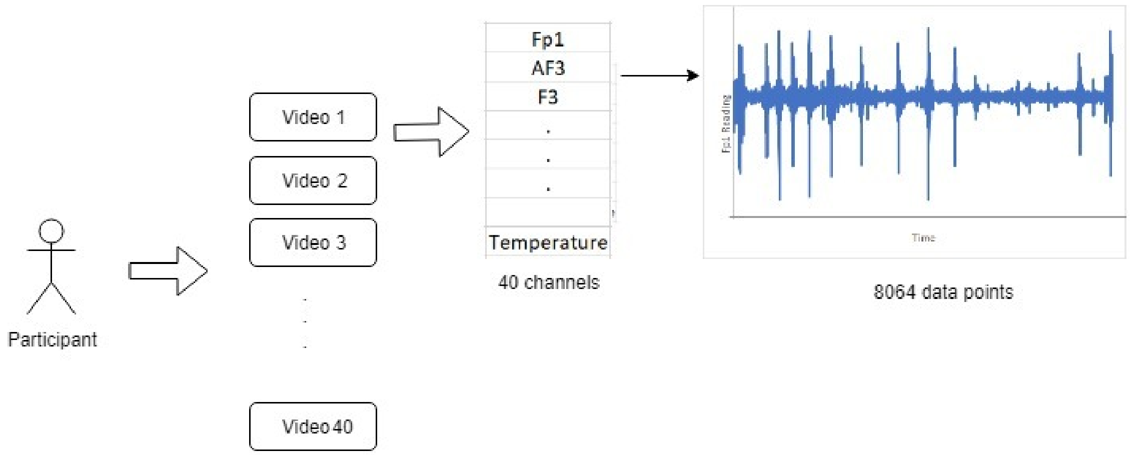

Figure 3.8 DEAP feature vector structure.

All experiments are done individually on each subject and evaluation is done by calculating mean and standard deviation for every set of experiments. 12 sets of experiments are performed for all 3 sets of channels (EEG, peripheral, both) against all 4 emotions (valence, arousal, liking, dominance) and each set contains 32 experiments for all 32 participants. Data was divided into training and testing sets with the ratio 4:1 before feeding it to the neural network.

CHAPTER 4

RESULTS

4.1 BP4D+

The first set of experiments was done on all subjects individually creating 140 different models and 140 sets of results. The second set of experiments was done on all female subjects individually creating 82 different models and 82 sets of results. The third set of experiments was done on all male subjects individually creating 58 different models and 58 sets of results. Table 4.1 shows these results in terms of mean and standard deviations for every set.

Table 4.1 BP4D+ results for training individually.

Data	**Mean**	**Standard Deviation**
All participants	98.89 %	1.647
Female	99.09 %	1.212
Male	98.88 %	1.916

As can be seen in Table 4.1, both male and female data had a relatively high mean accuracy, with a low standard deviation. This shows that the majority of the subjects (both male and female), where recognized with high accuracy. The rest of the experiments were done combining all the data which consist of 1- all participants, 2- all females and 3- all males. 80% subjects are picked randomly for training and testing was done on rest of the 20% of subjects. It is important to note that the same subjects did not appear in both the training and testing datasets. A gender specific comparison is done by training a model of female data and testing on male data and vice versa. Table 4.2 shows these results in terms of mean and standard deviations for every set.

Table 4.2 BP4D+ results for combining whole data.

Data	Test Data	Accuracy
All participants	All	94 %
Female	Female	96.77 %
Male	Male	93.60 %
Female	Male	15.35 %
Male	Female	15.08 %

These results give some insight into gender-specific models of emotion. When tested on the opposite gender, both models performed poorly at approximately 15% accuracy. Similar to the results in Table 4.1, the female model outperformed the male model in terms of overall accuracy. It is unclear why the female models perform better on this dataset. One possible explanation is the data has more variance resulting in an overall better model. This is an interesting and open question for future works. Figure 4.1 shows a distribution of results among all models trained on data for individual participants.

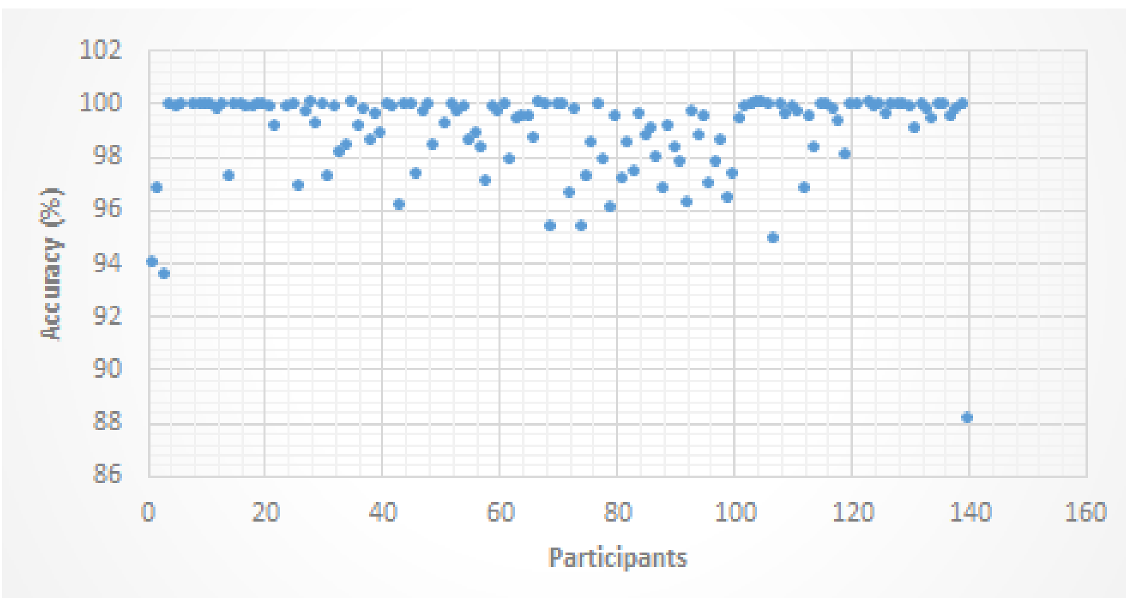

Figure 4.1 BP4D+ result.

As can be seen in Figure 4.1, many of the subjects were very consistent in their accuracy with a lot at or near 100% accuracy. There are some interesting outliers as subjects 1 and 3 are close to 94% accuracy, and even more so with subject 140 at approximately 88% accuracy.

More detailed analysis of this data, may lead to better models of emotion, thus increasing the overall accuracy of the proposed method.

4.2 DEAP

For DEAP, all experiments are done on each subject individually. There were 12 sets of experiments performed for 3 sets of channels (EEG, peripheral, both) against 4 emotions (valence, arousal, liking, dominance) and each set contains 32 experiments. For every experiment, 80% data was used for training and rest of the 20% for testing. Table 4.3 shows mean and standard deviations for all sets of experiments. We show the distribution of results among all models trained on all 12 sets of data in following figures from 4.2 - 4.13.

Table 4.3 DEAP dataset results.

Emotion Category	**EEG**	**Peripheral**	**Both**
Valence	Mean: 60.21% SD: 6.306	Mean: 86.31% SD: 6.186	Mean: 86.09% SD: 5.367
Arousal	Mean: 65.03% SD: 9.486	Mean: 88.83% SD: 4.455	Mean: 90.61% SD: 3.579
Liking	Mean: 67.59% SD: 11.295	Mean: 88.38% SD: 6.416	Mean: 90.48% SD: 4.954
Dominance	Mean: 66.22% SD: 12.528	Mean: 89.12% SD: 5.484	Mean: 90.95 % SD: 4.667

As can be seen above, EEG data performs the worst, and has some of the overall highest stand deviations. In 3 out of the 4 emotions combining both EEG and peripheral signals resulted in a higher accuracy compared to using just EEG or just peripheral signals. This result in intuitive, as it has been shown that a multimodal approach to classification can lead to higher accuracy rate compared to a single modality [42].

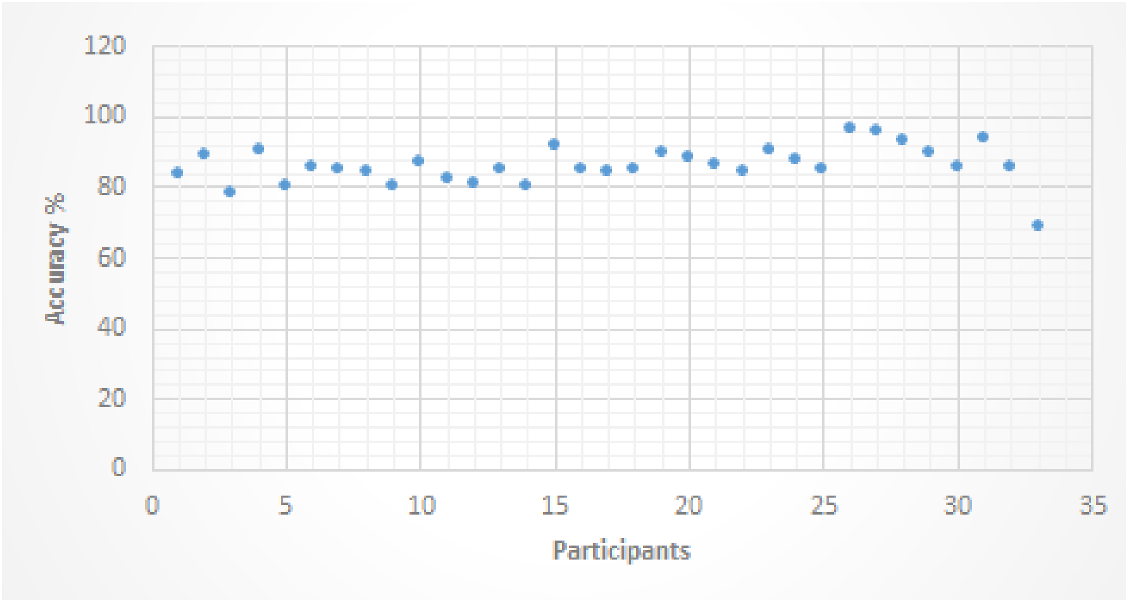

Figure 4.2 Valence results (EEG+peripheral).

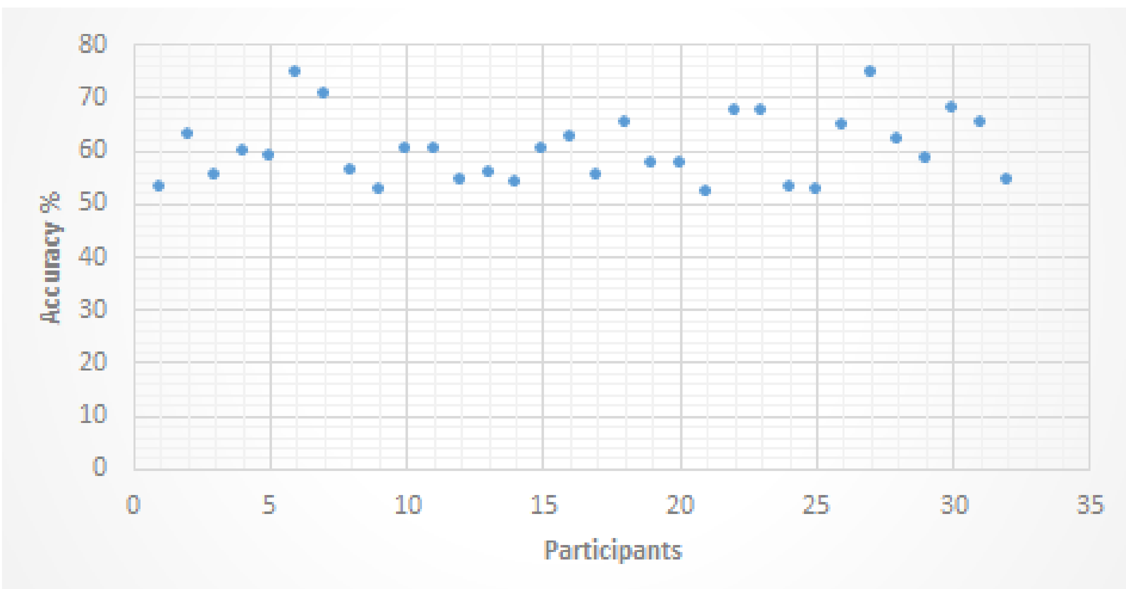

Figure 4.3 Valence results (EEG).

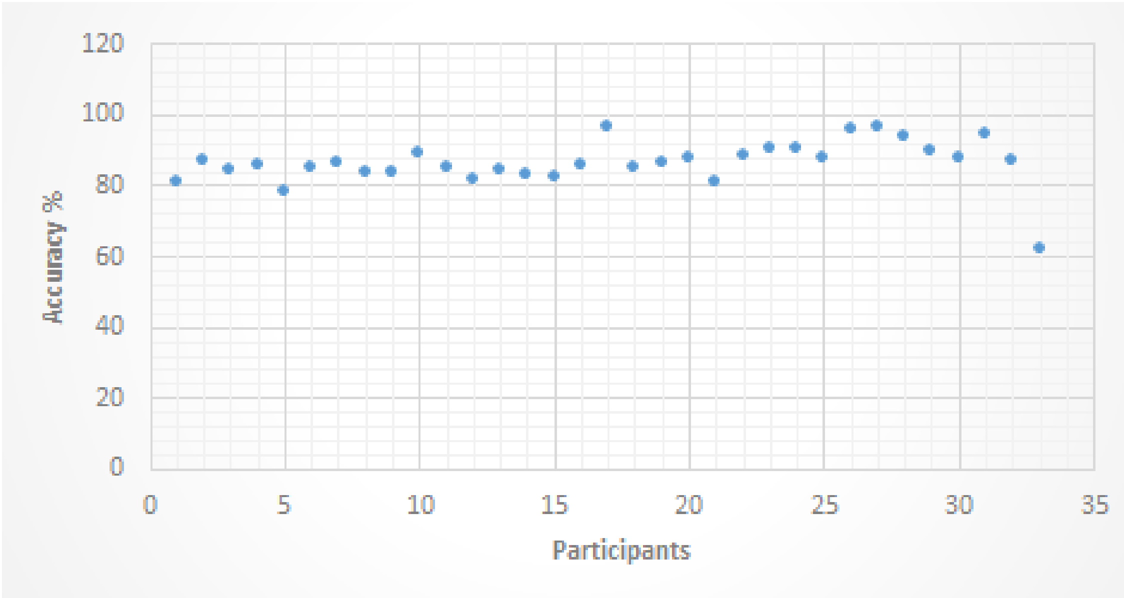

Figure 4.4 Valence results (peripheral).

While the EEG accuracy was lowest for valence accuracy compared to the other 3 (arousal, dominance, and liking), the standard deviations are more consistent for the three different experiments (EEG, peripheral, and EEG + peripheral). A more detailed study into how EEG data correlates to the peripheral data could yield interesting results for future fusion of these signals.

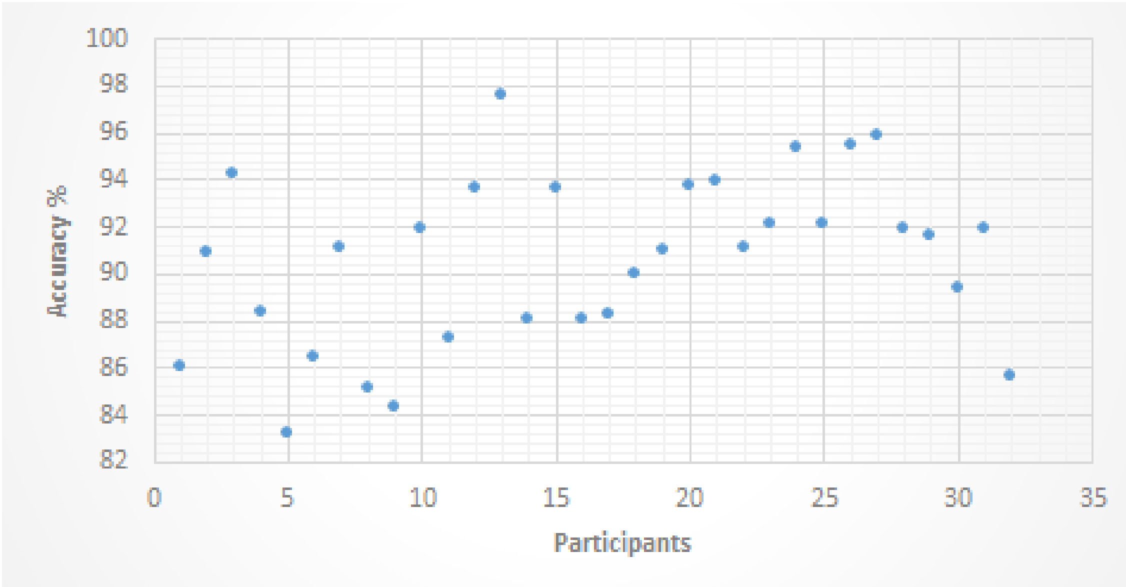

Figure 4.5 Arousal results (EEG+peripheral).

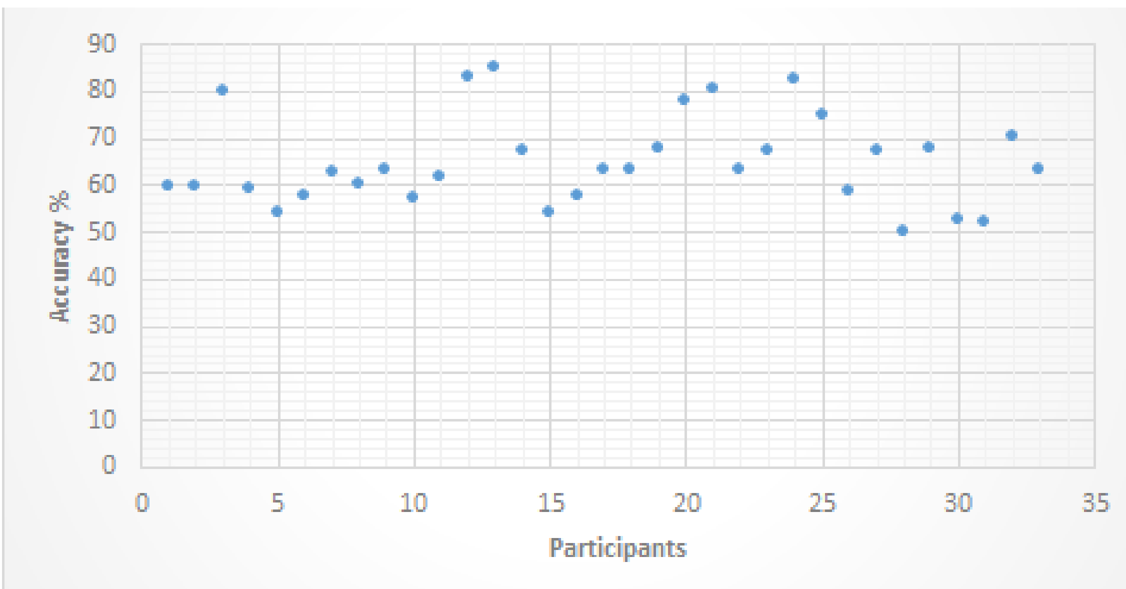

Figure 4.6 Arousal results (EEG).

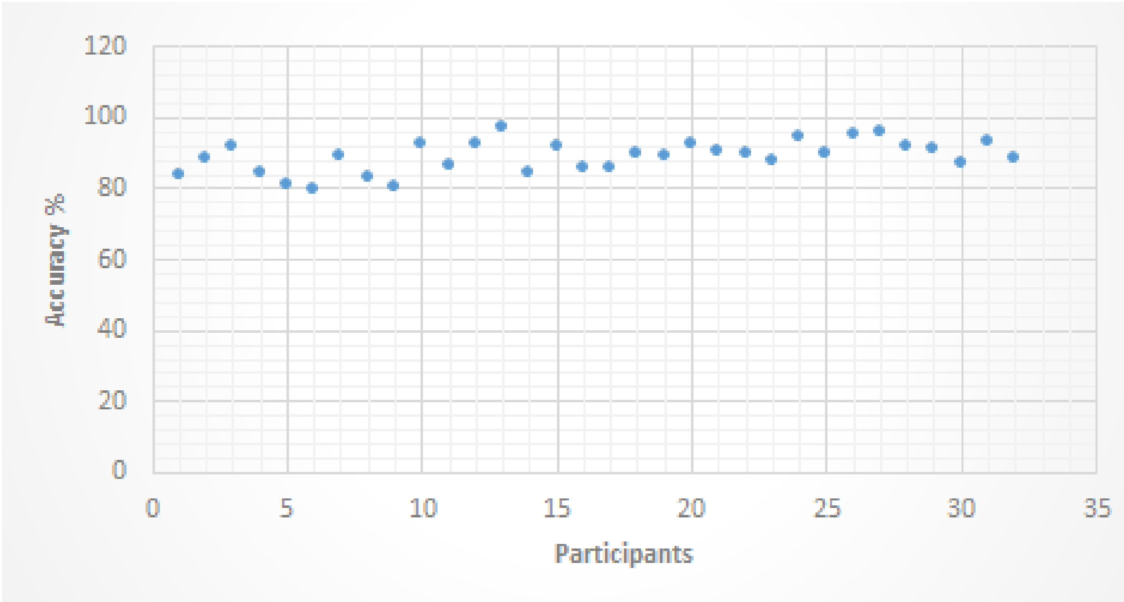

Figure 4.7 Arousal results (peripheral).

Arousal results had the lowest standard deviation for peripheral and EEG + peripheral experiments, although it is higher, compared to valence, for EEG data alone. This could be partially explained by EEG data not being able to easily generalize across subjects [43]

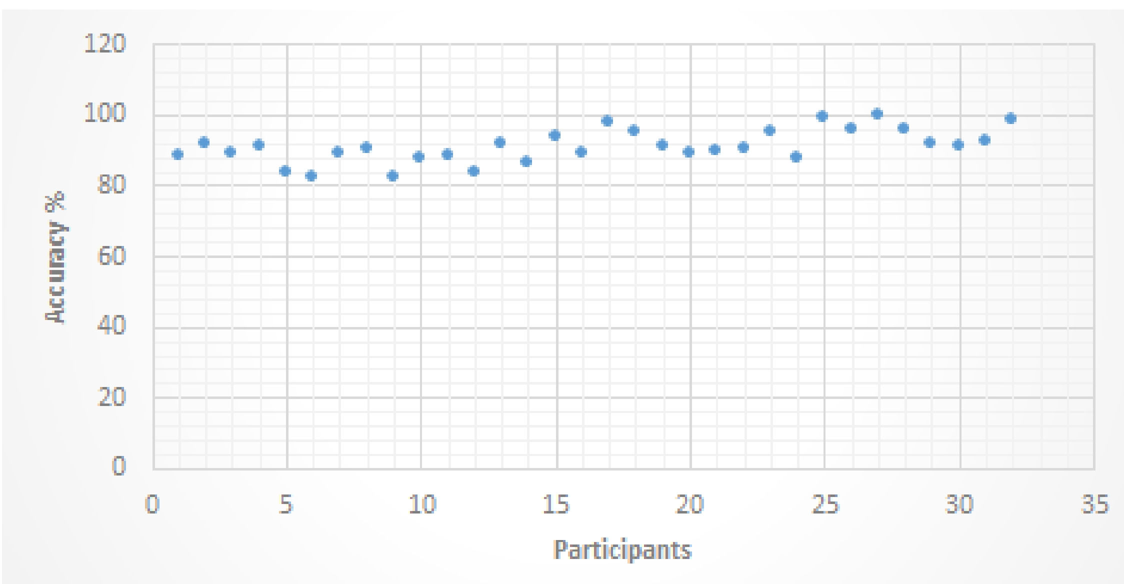

Figure 4.8 Dominance results (EEG+peripheral).

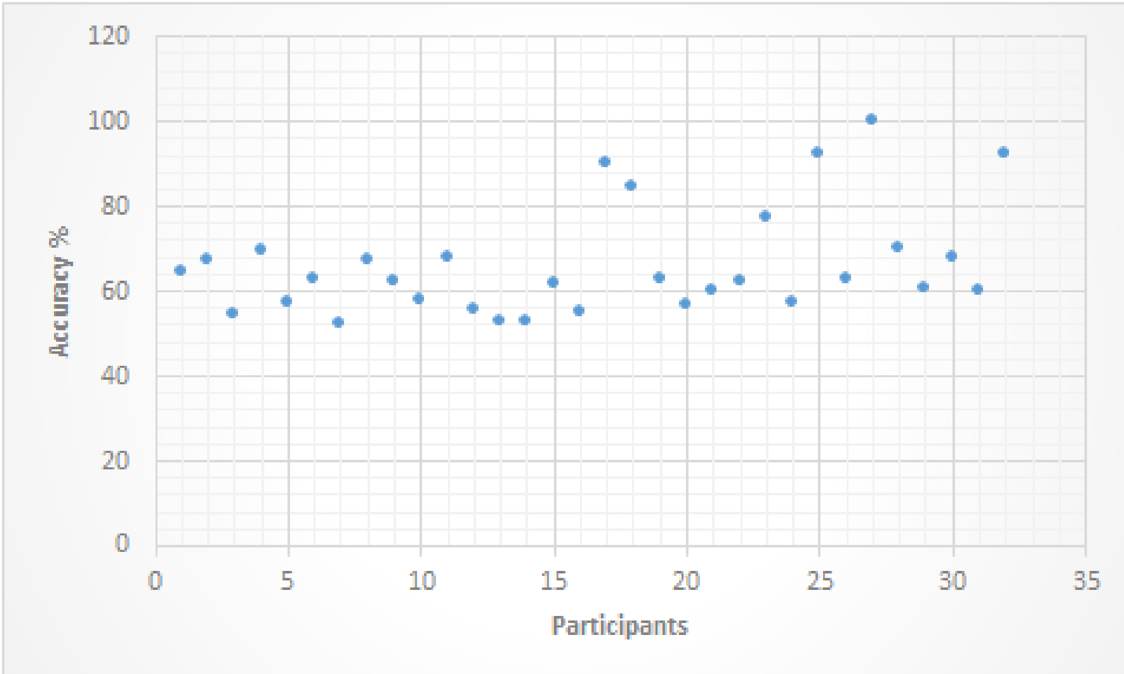

Figure 4.9 Dominance results (EEG).

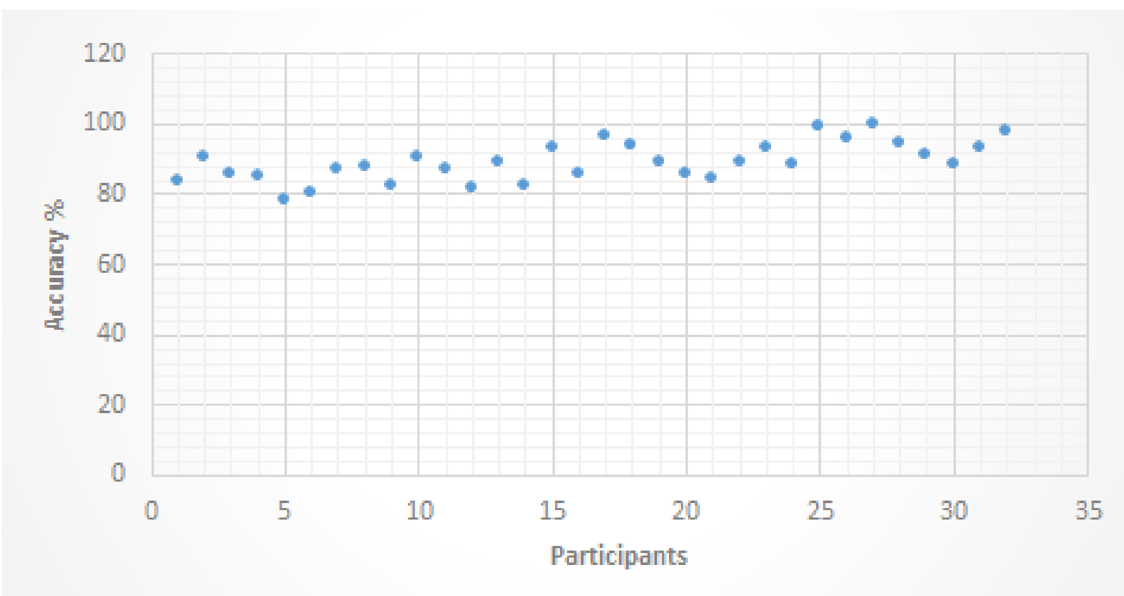

Figure 4.10 Dominance results (peripheral).

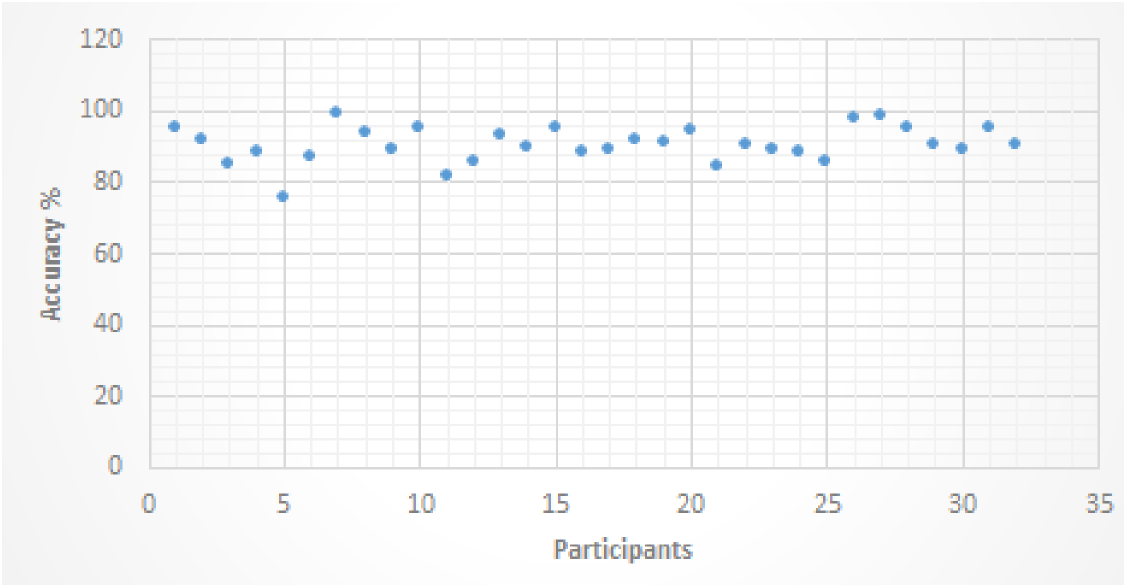

Figure 4.11 Liking results (EEG+peripheral).

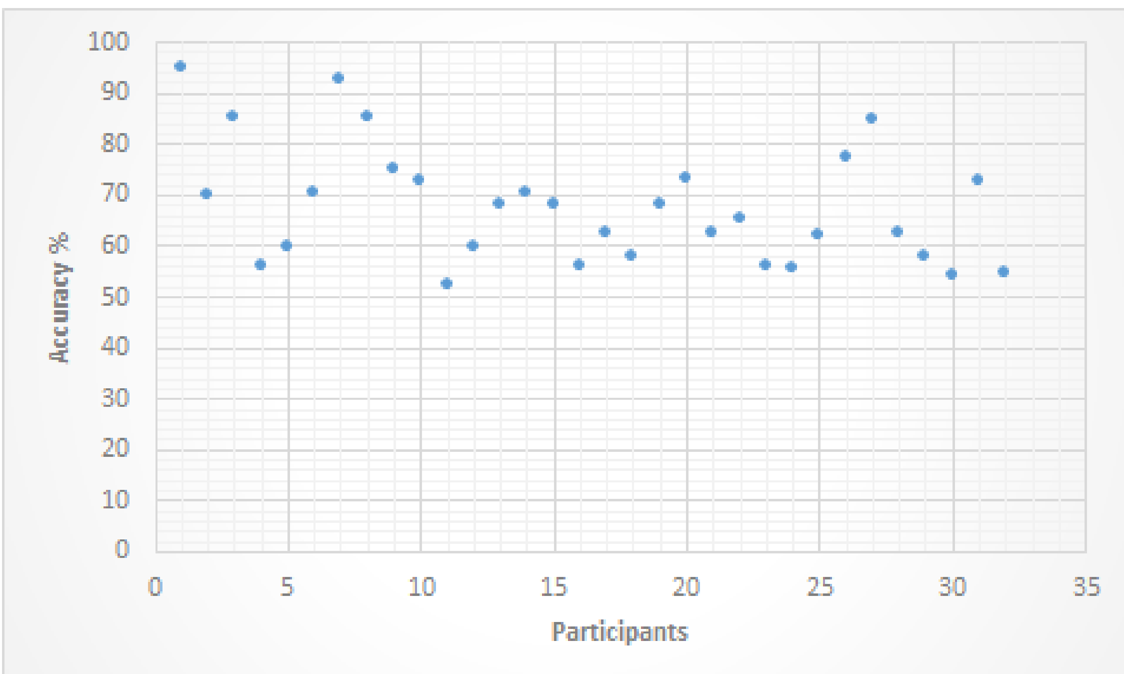

Figure 4.12 Liking results (EEG).

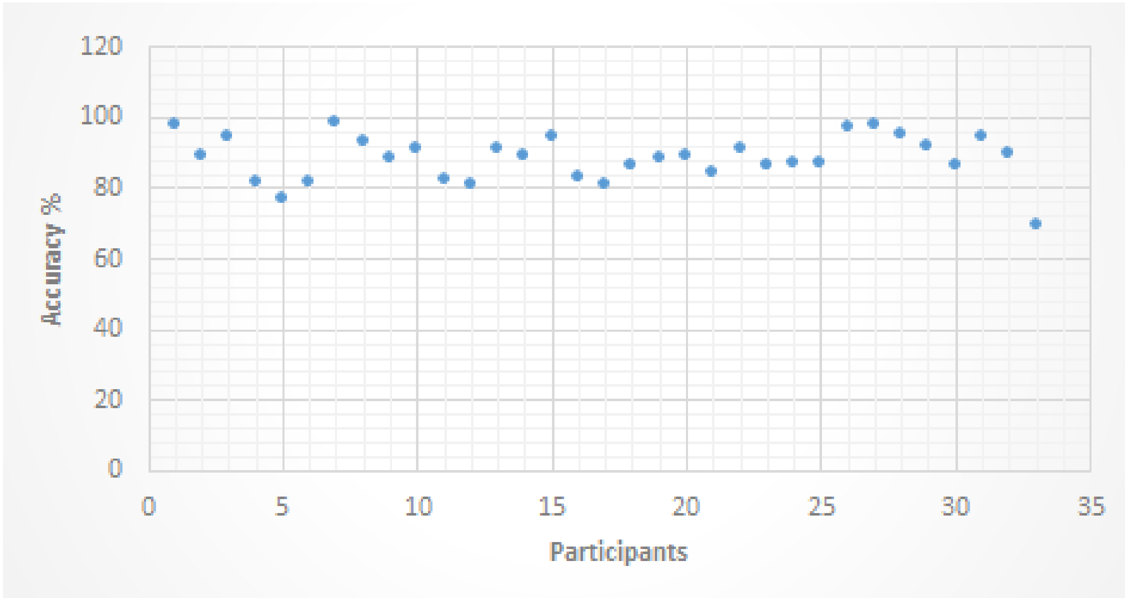

Figure 4.13 Liking results (peripheral).

For both dominance and liking, the standard deviation was highest for EEG data alone with liking over 11% and dominance over 12%. Similar to arousal, this can partially be attributed to EEG data's inability to easily generalize across subjects. A question that arises from this work is "Is it possible to generalize across subjects with EEG data alone?" With valence having the lowest standard deviation, that can be a good starting point for future experiments to answer this question.

CHAPTER 5

DISCUSSION

In our research, we worked upon utilizing deep convolutional neural network for classification of emotions based on physiological signals generated in human body upon any emotional elicitation. We used two different databases, BP4D+ and DEAP for our experiments. The proposed method generated encouraging results on both datasets. We also showed how our results outperforms previous works done using DEAP. We can see in the figure 4.1 on page 26 that the results using BP4D+ are very consistent among the subjects and have low standard deviation as shown in table 4.1 on page 25.

To investigate generalization, we compared gender specific models. Upon observing these results in table 4.2 on page 26, we can see that model trained on female physiological data worked good when tested on female data, and did not gave good results when tested on male data. Similarly model trained on male physiological data worked good when tested on male data, and did not gave good results when tested on female data. Interestingly, a previous study notes that physiological signals are similar during similar emotions in male and female [44], however, our results show that they are different from a machine learning standpoint. These results are supported by another study that notes that the neurons flows in different parts of brain in men and women during emotion elicitation. For women, these neurons connect the parts of brain that regulates internal areas of body that impacts blood pressure, respiration and hormones, while in men, these neurons connects to the areas of brain that controls vision and movement [45] [46]. This can cause the difference in physiological signals in men and women. This contradiction requires more detailed analysis.

We can also observe in table 4.1 and 4.2 that the results for data for females are a little better than the results for data for males, this can be because females have more intense expression of emotions than males [47] [44]. The articles [47][44] talks about difference in perception of emotions in men and women saying that women feel the same emotion in a richer and more intense way than men.

Moving forward, we can observe that for DEAP database, we got good results for peripheral signals and combining EEG with peripheral signals and our results for all the modalities outperforms the current state of art. We can say that for this data, deep convolutional network performed better than previous studies which used other methods like single-trial binary classification by Rozgic et al. [15] for each of four emotional dimensions, or recent work using DEAP which was done by Liu et al. [16] using Bimodal Deep autoencoder (BDAE) for feature extraction and deep learning for training. Table 5.1 shows a comparison of results with previous researches on DEAP database.

Table 5.1 Comparison with current state of art for DEAP database.

	Valence	**Arousal**	**Dominance**	**Liking**
Proposed	**86.31%**	**90.61%**	**90.95%**	**90.48%**
Liu et al. (arXiv, 2016) [16]	85.2%	80.5%	84.9%	82.4%
Rozgic et al. (ICASSP, 2013) [15]	76.9%	69.1%	73.9%	75.3%
Li et al. (SIGIR, 2015) [48]	58.4%	64.3%	65.8%	66.9%
Koelstra et al (IEEE Trans. Affective Computing, 2012) [2]	65.2%	63.1%	N/A	64.2%

Observing table 4.3 on page 27 in results section, we can see that results for peripheral data and combination of EEG with peripheral data are good, but results for EEG alone are not as good as other set of modalities. Figures 4.2 - 4.13 also shows in similar way that the results obtained with EEG data are less consistent and distributed over a large scale from 50% to 100% accuracy, while the results of peripheral data and the combination of EEG with peripheral data are much more consistent. The reason may be because EEG data may have several artifacts while getting recorded, for example there are chances that heart beats can be misinterpreted as a brain activity which can introduce noise while recording data. These results with EEG data may be improved by trying some different methods for preprocessing or post processing and noise removal. Further analysis is needed for improving results with EEG data.

Another observation can be made upon comparing results of BP4D+ and DEAP databases for physiological data that results for BP4D+ are better than results for DEAP for multi-modal experiments with physiological data. BP4D+ has the combination of blood pressure, heart rate, respiration and EDA, while DEAP database has a different combination of physiological data i.e. EOG, EMG, GSR, respiration, blood pressure and body temperature. Based on our proposed method and experimental design, it can be said that the first combination of physiological data is better than the second combination. However, more detailed analysis of these data types if required to confirm or deny this. Due to this being the first work to report on BP4D+ physiological data, in this manner, our work can be used as a baseline for future experiments.

CHAPTER 6

CONCLUSION AND FUTURE WORK

Affective computing is a highly promising field of study due to advantageous applications that can make people's lives better. Various techniques are being tried for emotion recognition. In our experiments we study the capability of physiological signals to determine human emotion. For classification of emotions, we used deep learning which is the latest state of art in classification techniques. The goal of our study is to determine if the combination of multimodal data along with deep learning can be useful for classifying emotions with high accuracy.

The results we have from our experiments are very exciting and enlightens some untouched aspects like the gender specific comparison using physiological signals. We have developed a new method of emotion recognition which is a multimodal approach that combines different physiological signals of a person together to classify their emotional state. We used two databases – BP4D+ and DEAP for our experiments and deep learning as the classification technique. We also compared our results with the current state of art on DEAP, which is outperformed by our method, and we detail a baseline for future experiments on BP4D+.

The work we did opens doors for very exciting studies for classification of emotions using physiological signals especially brain data. We are doing more analysis of EEG signals and working on training classifiers with data of channels from particular parts of brain that are responsible for emotions specifically, for example frontal and temporal lobes [49]. We are also planning to try building multiple classification models using a single EEG channel, to compare the contribution of different EEG channels for classification of emotions.

Further studies can be done using multimodal fusion i.e. how different EGG channels can be combined to improve results with EEG data. Combining EEG and with other physiological signals can be tried for further studies. We are also planning to try heterogeneous data correlations i.e. combining images, audio, thermal and physiological data. We are going to do detailed investigation into age, gender, and ethnicity for emotion recognition. We are also working on improving results for EEG signals and making it more consistent by trying more preprocessing techniques for noise removal and planning to try a different deep net architectures (e.g. LSTM) for further experiments.

LIST OF REFERENCES

[1] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, *et al.*, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3438–3446, 2016.

[2] 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 Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[3] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[4] P. Ekman, "The argument and evidence about universals in facial expressions," *Handbook of social psychophysiology*, pp. 143–164, 1989.

[5] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.

[6] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.

[7] V. Lindh, U. Wiklund, and S. Håkansson, "Heel lancing in term new-born infants: an evaluation of pain by frequency domain analysis of heart rate variability," *Pain*, vol. 80, no. 1-2, pp. 143–148, 1999.

[8] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *IEEE Computational Intelligence Magazine*, vol. 8, no. 2, pp. 20–33, 2013.

[9] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pp. 671–676, IEEE, 2013.

[10] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in biology and medicine*, 2017.

[11] A. E. Vijayan, D. Sen, and A. Sudheer, "Eeg-based emotion recognition using statistical measures and auto-regressive modeling," in *Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on*, pp. 587–591, IEEE, 2015.

[12] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pp. 940–943, IEEE, 2005.

[13] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[14] S. Y. Chung and H. J. Yoon, "Affective classification using bayesian classifier and supervised learning," *Proceedings of the 12th International Conference on Control, Automation and Systems (ICCAS), Jeju, Republic of Korea*, pp. 1768–1771, 2012.

[15] V. Rozgić, S. N. Vitaladevuni, and R. Prasad, "Robust eeg emotion classification using segment level decision fusion," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 1286–1290, IEEE, 2013.

[16] W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using multimodal deep learning," *arXiv preprint arXiv:1602.08225*, 2016.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[18] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[19] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 689–696, 2011.

[20] H. D. Critchley and S. N. Garfinkel, "The influence of physiological signals on cognition," *Current Opinion in Behavioral Sciences*, vol. 19, pp. 13–18, 2018.

[21] X. Li, P. Zhang, D. Song, G. Yu, Y. Hou, and B. Hu, "Eeg based emotion identification using unsupervised deep feature learning," 2015.

[22] G. H. Klem, H. O. Lüders, H. Jasper, C. Elger, *et al.*, "The ten-twenty electrode system of the international federation," *Electroencephalogr Clin Neurophysiol*, vol. 52, no. 3, pp. 3–6, 1999.

[23] D. Peterson, A. Hotere-Barnes, and C. Duncan, *Fighting shadows: Self-stigma and mental illness*. Mental Health Foundation of New Zealand, 2008.

[24] W. H. Press and S. A. Teukolsky, "Savitzky-golay smoothing filters," *Computers in Physics*, vol. 4, no. 6, pp. 669–672, 1990.

[25] H. Azami, K. Mohammadi, and B. Bozorgtabar, "An improved signal segmentation using moving average and savitzky-golay filter," *Journal of Signal and Information Processing*, vol. 3, no. 01, p. 39, 2012.

[26] P. Trebuňa, J. Halčinová, M. Fil'o, and J. Markovič, "The importance of normalization and standardization in the process of clustering," in *2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 381–385, Jan 2014.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[28] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[29] A. Ortiz, J. Munilla, J. M. Gorriz, and J. Ramirez, "Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease," *International journal of neural systems*, vol. 26, no. 07, p. 1650025, 2016.

[30] M. H. Rafiei and H. Adeli, "A novel machine learning model for estimation of sale prices of real estate units," *Journal of Construction Engineering and Management*, vol. 142, no. 2, p. 04015066, 2015.

[31] X.-W. Chen and X. Lin, "Big data deep learning: challenges and perspectives," *IEEE access*, vol. 2, pp. 514–525, 2014.

[32] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," *Cited on*, p. 14, 2012.

[33] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv preprint arXiv:1603.07285*, 2016.

[34] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," 2010.

[35] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for lvcsr using rectified linear units and dropout," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 8609–8613, IEEE, 2013.

[36] A. Giusti, D. C. Ciresan, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Fast image scanning with deep max-pooling convolutional neural networks," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pp. 4034–4038, IEEE, 2013.

[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[38] J. Jin, A. Dundar, and E. Culurciello, "Flattened convolutional neural networks for feedforward acceleration," *arXiv preprint arXiv:1412.5474*, 2014.

[39] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks.," in *CVPR*, vol. 1, p. 3, 2017.

[40] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[41] R. Gomez, "Understanding categorical cross-entropy loss, binary cross-entropy loss, softmax loss, logistic loss, focal loss and all those confusing names," 2018.

[42] G. Zamzmi, C.-Y. Pai, D. Goldgof, R. Kasturi, T. Ashmeade, and Y. Sun, "An approach for automated multimodal analysis of infants' pain," *International Conference on Pattern Recognition*, 2016.

[43] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring eeg features in cross-subject emotion recognition," *Frontiers in Neuroscience*, 2018.

[44] D. Cummins, "Are males and females equally emotional?," in *Psychology Today*, 2014.

[45] M. Bradley, "Emotions — differences between men and women," in *HealthGuidance for better health*, 2014.

[46] S. Whittle, M. Yücel, M. B. Yap, and N. B. Allen, "Sex differences in the neural correlates of emotion: evidence from neuroimaging," *Biological psychology*, vol. 87, no. 3, pp. 319–333, 2011.

[47] B. Goldman, "Two minds - the cognitive differences between men and women," in *Stanford Medicine*, 2017.

[48] X. Li, P. Zhang, D. Song, and Y. Hou, "Recognizing emotions based on multimodal neurophysiological signals," *Advances in Computational Psychophysiology*, pp. 28–30, 2015.

[49] H. J. Rosen, K. Pace-Savitsky, R. J. Perry, J. H. Kramer, B. L. Miller, and R. W. Levenson, "Recognition of emotion in the frontal and temporal variants of frontotemporal dementia," *Dementia and geriatric cognitive disorders*, vol. 17, no. 4, pp. 277–281, 2004.

ABOUT THE AUTHOR

Astha's journey with Computer Science started in year 2008 when she took admission in bachelor's degree in Computer Science back in her home town, Bhopal, India, knowing nothing about software engineering. Her passion for programming started the day she wrote her first program and motivated her to peruse a carrier as a Software Engineer. She started her first job in year 2012 working on mobile application development and worked upon many mobile and web projects for the next four years.

She always wanted to learn and expand her horizons. This was the thrust along with wanderlust which made her come to the USA, far away from home and peruse a master's degree in Computer Science and Engineering. While studying in University of South Florida, one of the best research-oriented schools in the United States, she got great opportunity for working upon cutting edge technologies, which got her academic interest started shifting towards research and she started working with Dr. Shaun Canavan, who is her thesis advisor. She started working as a Teaching Assistant here and also wrote her first academic paper and discovered her interest in writing.

Apart from her work, she enjoys singing, reading, traveling and exploring all the new places and new hobbies. She have this deep desire to utilize her knowledge for environment welfare while working on cool projects in industry or by her own which she would like to do after her graduation from USF.