EEG-Based Analyzing Human Idiosyncrasies Within Unpredictable Circumstances Using 3D Convolutional Neural Networks

by

Tanzim Ahmed
15201040
Sabbir Al Fahad
16301191
Abu Sayed Bin Abdullah
15301053
Joy Paul
15301074
Rufaida Tasin
15201031

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science

Department of Computer Science and Engineering
BRAC University
October 2020

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Tanzim

Tanzim Ahmed
15201040

Sabbir

Sabbir Al Fahad
16301191

Abu Sayed

Abu Sayed Bin Abdullah
15301053

Rufaida

Rufaida Tasin
15201031

Joy Paul

Joy Paul
15301074

Approval

The thesis/project titled "EEG-Based Analyzing Human Idiosyncrasies Within Unpredictable Circumstances Using 3D Convolutional Neural Networks" submitted by

1. Tanzim Ahmed (15201040)
2. Sabbir Al Fahad (16301191)
3. Abu Sayed Bin Abdullah (15301053)
4. Rufaida Tasin (15201031)
5. Joy Paul (15301074)

Of Summer, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on October 5, 2020.

Examining Committee:

Supervisor:
(Member)

Muhammad Iqbal Hossain,PhD
Assistant Professor
Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam,PhD
Associate Professor
Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Prof. Mahbub Majumdar
Chairperson
Dept. of Computer Science & Engineering
Brac University

Mahbubul Alam Majumdar,PhD
Chairperson
Computer Science and Engineering
BRAC University

Abstract

Automatic recognition of emotions is a very challenging of task. To detect emotions from EEG signals, a sophisticated learning algorithm is needed that can reflect high-level abstraction. Using electroencephalogram (EEG) signals, in particular the learning of spatiotemporal characteristics, various methods have been used to improve the robustness of the emotion detection systems. The use of a model inspired by the Siamese Network in this approach, which deals with data by concatenation and 3D convolution. The Adam optimizer was used in this method for preparation. This model is a custom model which can deal with a stack for a single time instant, which gives better efficiency than 3D Convolutional Network. This type of model is usually used for Face Recognition, but use of such a model for this purpose opens a new horizon of opportunity for Automatic recognition of emotion. Highest accuracy of our project is around 70% has been achieved by Dense Net and lowest accuracy was 30%.

Keywords: 3D Convolutional Neural Network, Seed IV, Adam Optimizer, Siamese Network

Acknowledgement

First and foremost, we would like to thank God Almighty for giving us the ability and opportunity to undertake the thesis and to complete it. This achievement would not have been impossible without his blessings. After that, we want to thank and express our gratitude toward our supervisor Dr. Muhammad Iqbal Hossain for his constant guidance and support as well as providing us necessary information regarding our work. His valuable feedback helped us to improve ourselves and complete the thesis with our hard work. We are also very grateful to our family and friends who supported us directly and indirectly in our entire thesis period. Sometimes, we used to feel less confident because of facing difficulties in our work. Then, our family members and friends gave the mental support and encouraged us a lot to stick with the work and complete it in time. We would like to acknowledge the assistance that we found a huge number of online resources specially the work of our fellow researches. Finally, we are grateful to BRAC University to give us the opportunity to do such research that helped us to enhance our knowledge.

Contents

List of Figures

1 Introduction

Human emotions are isolated as the primary source of feelings from the external appearances. In any case, it is realized that a few people may conceal their genuine emotions and use outward appearances to deceive [1]. Analysts then keep fast to use specific data wellsprings which are accurate and not defenseless to misrepresentation [2]. The brain-computer interface (BCI) was used in biomedical engineering. For decades, brain-computer is a very interesting as well as important fields of research. It offers a promising technology that helps individuals, by modulating their brain waves, to manipulate outside machines. For non-invasive brain signal processing, most BCI innovations have been developed that are feasible to apply in real-world scenarios. Not only can BCI be used for instruments of mental function, but it can also be applied to recognize our states of mind. Emotion detection is one such program. Automatic algorithms for emotion recognition may theoretically cross the difference between human and computer experiences. This ranges from detrimental to positive. The Thrill From unconscious to active and neurological, it is a feeling of being awake or sensitive to stimuli. The dimensional emotion model of valence-arousal, defined in Figure 1, is commonly used in many research studies. The electroencephalogram (EEG) is a record of the oscillation of the brain's electrical potential likely to occur from the flow of ionic current through neurons of the brain. EEG signals are produced by calculating the behavior of the brain at the location of the scalp electrode [3][4].

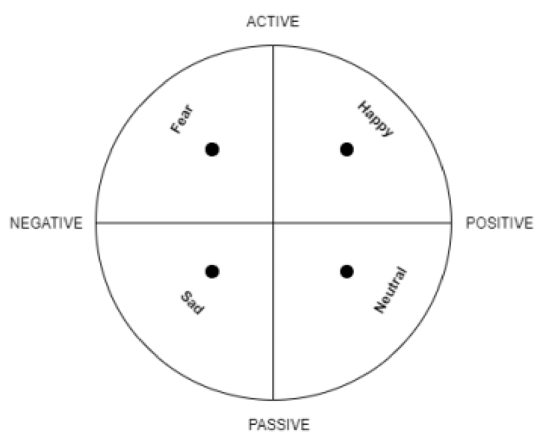

Figure 1: Dimensional Model

Moreover, with Gaussian noise signals n with 0 average and unit difference, the random sample size from the original EEG is developed with N elements to derive the noisy EEG signals. Ultimately, the noisy version of s will be obtained by integrating all samples of s and n signals. The Phase of Augmentation was implemented in the preparation phase only. And during test point, seamless representations of the signals have been used. The 3D-CNN can learn time and computational characteristics, as already described. This includes 3D visualization of the EEG signals. To that end, the research framework creates a 3D replication process.[7] EEG data are typically collected from each signal from separate channels. The data from each c channel is divided into small chunks (frames) using a linear kernel w,[8]and each chunk is labelled to boost working flows. Software apps that can detect the current emotional condition of the user are being desperately needed. Study has been directed for copying communication from the face as well as voice to understand feelings. We have therefore chosen to identify brain signal detection using the DNL algorithm, using the Valence-Arousal dimensional model, which defined their existing emotional situation and brain signal location mostly on base of an axis. With the growing importance of BCI, consumer EEGs (electroencephalograms) have also been studied BCI. If the EEG just indicates an unknown bodily response, or whether or not the feeling is mentally interpreted is also informative. Right EEG- recognition of artificially evoked emotion is currently only about 60 percent, but much research is hopefully showing the EEG's suitability for our future implementation [6]. EEG is one of the most common epilepsy screening measures. Usually, a therapeutic EEG routine, the session lasts 20-30 minutes (and as well as preparation time). This is a test that uses thin, metal discs (electrodes) connected to the scalp to identify electric brain activity. In clinical settings, EEG is commonly used to assess variations in brain function that could be effective in diagnosis and treatment brain activity. Disorders, seizures in particular, or some other seizure impairment. In order to diagnose or cure the following disorders, an EEG can also be helpful.[10]

- Brain tumor
- Brain damage due to head injury
- Brain dysfunction with specific factors (encephalopathy) (encephalopathy)
- Stroke
- Sleep disorders

It can also:

- Differentiate epileptic seizures from several other forms of symptoms, like the non-epileptic psychogenic seizures, fainting, prefrontal motion disorders, as well as different types of migraine.
- Distinguish between "organic" encephalopathy as well as delirium including major psychiatric disorders including catatonia
- Function as an alternative brain death screening for coma patients
- Prognosticate in coma patients (in some cases)
- Evaluate if anti-epileptic drugs need to be medicated.

A routine EEG is often not adequate to assess the diagnosis and/or the correct way to proceed in terms of therapy. For this case, initiatives can be taken when a seizure is ongoing to track an EEG. As compared to an inter-ictal tracking that relates to the EEG recording throughout seizures, it is classified as an ictal recording. A sustained EEG is usually done followed by a time-synchronized audio / video playback to achieve an ictal history. It can be achieved whether as an outpatient (at home) or after a hospitalization, generally with nurses as well as other staff skilled in the care of patients experiencing seizures in an Epilepsy Control Unit (EMU). Usually, outpatient ambulatory visual EEGs last 1 to 3 days[11]. Typically, an appointment to an epilepsy monitoring facility lasts most days, but can extend for 7 days or more. Seizure drugs are typically removed when in the hospital to improve the risk that a seizure may occur after admission. For medical concerns, drugs are not removed outside the facility while in an EEG. Ambulatory video EEGs, though, have the benefit of convenience are much less costly than hospital admission, but the downside is that a clinical incident is less likely to be recorded.[12] Usually, epilepsy testing is performed to differentiate epileptic seizures from many other forms of seizures, along with psychosomatic nonepileptic seizures, syncope (fainting), prefrontal motion disturbances as well as varieties of headache, to classify seizures for medical purposes, and also to pinpoint the region of the brain through which an epilepsy derives for future seizure surgery. In addition, we can use EEG as an indirect indicator of neuronal excitability during carotid endarterectomy to evaluate the extent of anesthesia, or to track the phenobarbital effect mostly during Wada test.[13] EEG can be used in brain activity tracking in hospitals and clinics to monitor non-convulsive seizures / non-convulsive epileptic state, to control the impact of anesthetic / sedation in medically induced coma patients (for the treatment of severe seizures or elevated respiratory rate) and to control secondary brain injury in cases such as subarachnoid hemorrhage (curing) That's because the electric charge of scalp EEG is reported by the cerebrospinal uid, skull and scalp smear. Usually, in these situations, neurosurgeons insert electrode strips and grids (or penetrating depth electrodes) underneath the dura mater, either by a laparoscopy or a burr opening. The analysis of these stimuli is related to as electrocardiography (ECoG), subdural EEG (sdEEG) or intracranial EEG (icEEG)all words for about the same issue. The signal reported from ECoG is on a distinct spectrum of activity than that of the brain function obtained from scalp EEG. Low voltage, wide frequency elements which can not be readily (or even at all) in scalp EEG can still be seen directly in ECoG. In comparison, tiny electrodes (which occupy a smaller portion of the surface of the brain) make it easier to see both lower voltages and faster elements of brain function. Some medical sites report breaching microelectrodes.[15] For the diagnosis of headache, EEG is not indicated. A typical pain concern is chronic headache, and this technique is often used in a diagnostic quest, although it has little benefit over routine clinical assessment.

2 Literature Review

Emotional states are associated with a large spectrum of human feelings , perceptions and behavior in situations such as decision-making, intuition, and abstract thinking, influencing our ability to behave logically. The brain-computer interface (BCI)[16] systems are also enhanced through studies on emotional awareness using emotional cues as an important subject for therapeutic applications and human social experiences. Physiological impulses are used to analyze mental states in order to elucidate therapeutics for psychiatric disorders such as autism spectrum disorder (ASD), concentration de cit hyperactivity disorder (ADHD), and anxiety disorder by taking into consideration usual emotional dimensions. The innovation of EEG-emotion recognition systems has become a common research topic among cognitive scientists in recent years.[18] In well-documented works, the capacity of EEG signals to recognize emotions has been thoroughly covered. Verma and Tiwari have documented the use of EEG signals for sentimental analysis using transmitted signal as features and the Support Vector Machines (SVM) along with k-Nearest Neighbors (KNN) as classifiers. [7]. Yoon and Chung[8] have previously used EEG signals to introduce a modern method of emotion detection. They took data from the EEG segments using Fast Fourier Transform (FFT) method using the coefficient Pearson correlation to extract the features. A probabilistic classifier is proposed on the basis of Baye 's theorem. In addition, it implements supervised learning with a convergence algorithm for perceptrons. For part elimination, Singular Value Decomposition (SVD), QR Pivoting Column Factoration (QRcp), and F-ratio are used. With the SVM, the classification process is finished. For sufferers from amyotrophic lateral sclerosis (als), an emotion recognizer has been developed by Choppin . In addition, Choppin studied in order to provide certain persons with the ability to communicate their emotions, only EEG signals understand sentiment. Furthermore, this study used neural networks to identify EEG signals online and, by using three emotion groups and restricted training data sets, obtained a right classification rate for new unseen samples of about 64 percent. Musha et al examined EEG signals to read the emotions of an individual. Cross-correlation has been extracted by them in order to convert these coefficients linearly into a 4 component vector that refers to four essential emotions, coefficients between the EEG behaviors from various places and calculated a 'emotion matrix.' The figures in the vector represent how powerful the EEG signals contain the particular emotion. The approximate subjective feeling across neural networks was analyzed in previous study to classify emotional responses on the basis of EEG characteristics. In addition, for each of the four emotional states, this analysis has indicated that average accuracy varies from 54.5 percent to 67.7 percent. Heraz et al.4 eventually arranged an operator for predicting human emotions during learning. An efficiency of 82.27 percent for identifying eight emotional responses was the highest distinction in the analysis, utilizing k-nearest neighbors as a classification model as well as the amplitudes of 4 EEG elements as characteristics. Moreover, while using EEG time-frequency data as characteristics and SVM as a classifier to classify EEG signals in 3 emotional responses, Chanel et al.5 achieved an average precision of 63 percent. A modern, feature-based emotional recognition framework was designed by Atkinson and Campos[10] in which statistical features like the average, standard deviation as well as coefficient curtosis were extracted from the EEG section. This current paradigm combines traditional knowledge with selection methods and kernel classifiers. The minimum-Redundancy-Maximum-Relevance (mRMR) is also used, including its data, to minimize redundancy. Alhagry[11] et al, to derive temporal features using RNN and EEG signals. Two fully connected LSTM layers, a dropout layer and a thick layer, form the RNN. Zhang[12] et al. suggested a deep learning method named the spatiotemporal recurrent neural network (STRNN) to integrate the learning of the features of spatiotemporal emotional recognition using the SJTU Emotion EEG Dataset (SEED-IV). While the accuracies obtained by the above study are comparatively strong, it is still important to further develop emotion detection. Normally, one or more different modalities may be used to conduct the automated process of identifying emotions: face, voice, body movements and EEG signals. Research focuses on the timely resolution of the problem of the association between emotions using EEG signals. By default, feelings endure over a short or lengthy time, not just a moment. Therefore, in improving identification accuracy, the interaction between segments of emotions in time is highly effective. Motivated by the recent popularity of deep learning approaches, the 3D-CNN approach suggests the EEG-signals spatiotemporal information model. Initially, the data expansion stage is implemented to increase the amount of EEG samples required to accomplish this goal.[19] The EEG segments then generate 3D input representation. And this system hopes that the proposed 3D-CNN model system can be designed with greater precision and will simplify the execution of work more than the preceding implementation precision. The DEAP (Dataset of emotion recognition using EEG and Physiological and Video Signals) will be used to validate the proposed approach and will provide a multimodal dataset called "SEED-IV" to analyze human axial state data. The system will run with an appropriate and effective DLN algorithm for problems where there are many unlabeled data sets and a couple of categorized ed data sets.

3 About Dataset

A tentative research has carefully selected 72 film clips that appear to trigger happy, sad, fear or neutral emotions.

Figure 2: Input Emotions

The experiment involved a combined number of 15 participants. Three sessions are conducted on alternating schedules for each respondent, and each session includes 24 trials. The participant watches one of the film clips in one session, while his EEG signals as well as eye motions are captured with the ESI NeuroScan Device 62-channel & SMI eye-tracking glasses. The following image is a time table diagram of the trial was used.

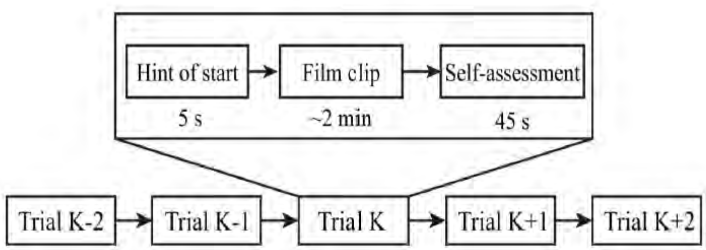

Figure 3: Schedule Diagram

3.1 EEG Signal Feature extraction

The raw EEG data is sampled first with an EEG signal processing sampling rate of 200 Hz. The EEG data is processed with a 1 Hz and 75 Hz bandpass filter and noise is applied and artifacts are removed. Afterwards, we extract power spectral density (PSD) at 5 frequency bands per section and Entropy Differential (DE):

$$PSD = E[x^2]$$
$$DE = -\int_{-\infty}^{\infty} P(x)\ln(P(x))dx$$

The EEG signals are supposed to be transmitted in Gaussian Distribution:

$$x \sim N(\mu, \sigma^2)$$

DE features can then be calculated in a simpler way:

$$DE = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} exp\frac{(x-\mu)^2}{2\sigma^2} ln(\frac{1}{\sqrt{2\pi\sigma}} exp\frac{(x-\mu)^2}{2\sigma^2})dx = \frac{1}{2} ln2\pi e\sigma^2$$

After pre-processing, the meaningful features are removed from the EEG data. The extraction of feature done using ad-hoc brain methods for the brain process that ranging from hand-crafted features to more sophisticated techniques like linear and non-linear spatial filtration. This ranges from nonexclusive techniques for example, principal component and independent component analysis to more EEG specific methods like CSPs and power feature variants and X-Dawn for temporal methodologies. Figure 3 demonstrates one of the easiest methods of feature extraction, which effectively subsamples the cleaned EEG signals directly into the temporal or frequency domain.

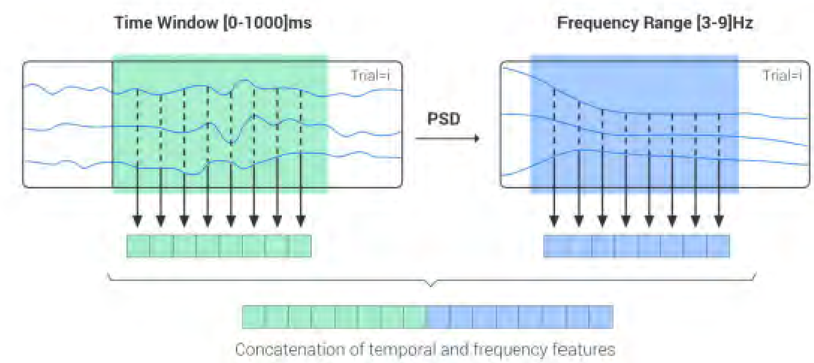

Figure 4: An example of Feature Extraction

The features extracted are usually tailored to the particular application: differences between experimental conditions (for example attention levels, responses to mismatched actions), the differentiation among the groups of predefined classes (for example, speller), behavior prediction (for example, anticipation of motion in neurorehabilitation), the finding of anomalies with respect to a normative database (e.g. Present cutting edge approaches include geometry classifiers, movie banks and adaptive classifiers in Riemann, used to solve EEG data problems at differing degrees of performance.[23]

4 Deep-Learning for EEG

Machine learning has radically changed data science in many ways (e.g. computer vision, speech, classification algorithms, etc.), by giving detailed functions and flexible systems that can operate with original information and perform the specialized training for an issue at hand. These designs could use massive quantities of data to concretely obtain characteristics as well as to effectively identify the design pattern, that can then be transformed and/or modified for different tasks. Such end-to-end learning ability matches well with the parameters of EEG research, although multiple interdependent techniques are necessary and have before lately been explicitly designed for each separate function.[34]

5 Flow Chart of Works Done

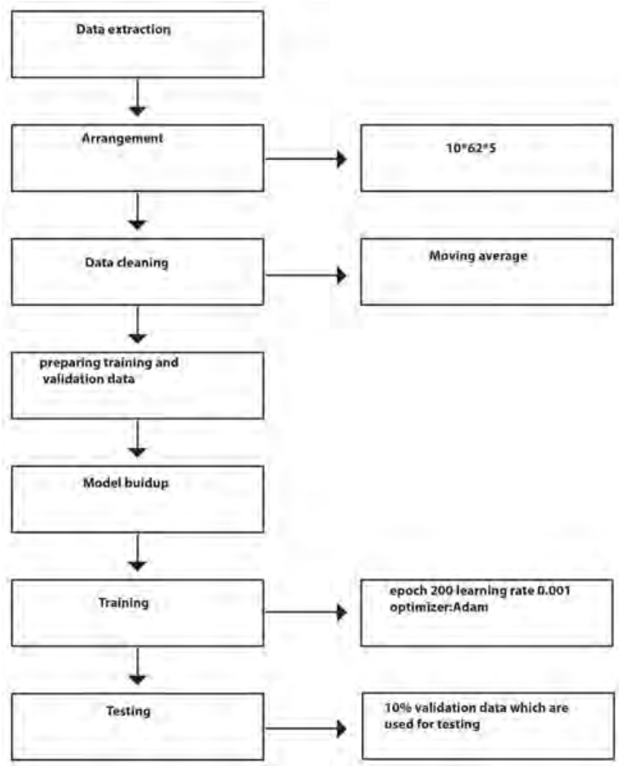

Figure 5: The Flowchart of the Proposed System.

5.1 Data Extraction

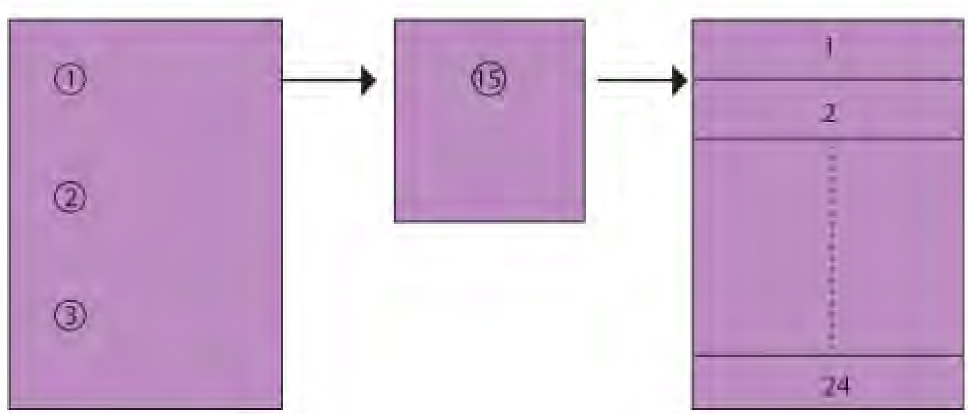

Figure 6: Data Extraction Flowchart

There was three sessions in our dataset,each session gives 15 .mat file of 15 individuals which gives 24 set of data, 24 different types of brain response due to watc hing 24 different kinds element. These signal are then denoised . Among various sorts of denoising Moving Average method was used.

5.2 Arrangement of data

There was 62 channels by which EEG data was taken in frequency of 5. This data creates a matrics of (10,62,5) Where , 10 is supposed time duration of data taken,62 is channel number and 5 is the frequency. Time parameter was taken vertically for personal convenience . In dataset time parameter was in horizontal axis. For convenience time parameter was taken in vertical axis and the channel component was taken horizontally. And illustration of it is being shown:

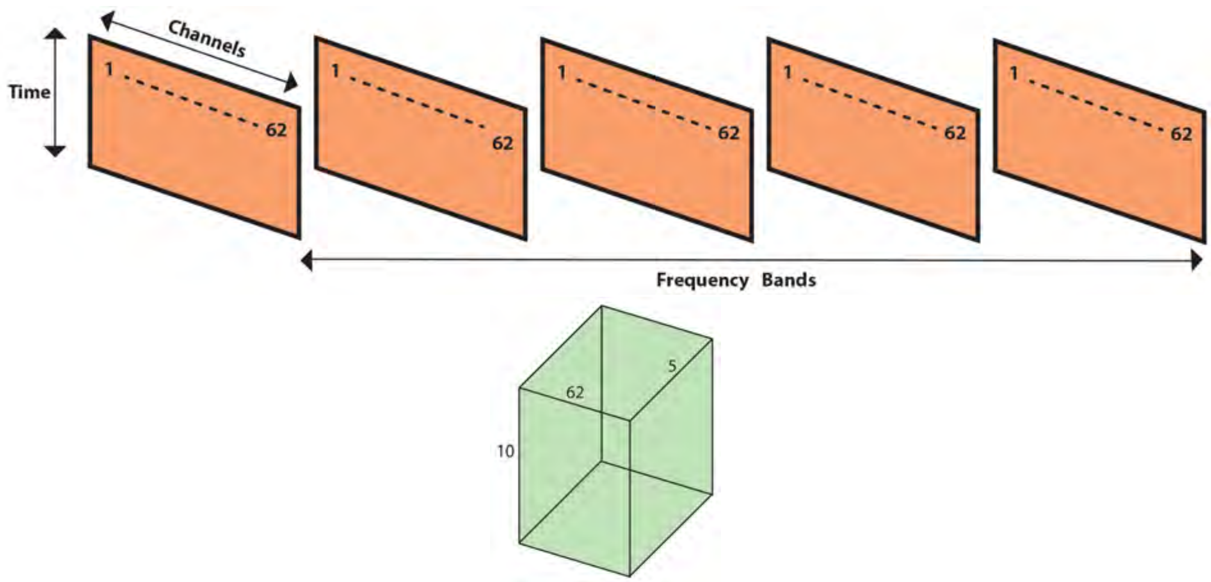

Figure 7: Arrangement of data, horizontally channel number, vertically time.

Two types of Brainwave component play vital role in analysis of data .

1. **PSD:** The frequency spectrum distribution (power spectrum) emits the 'frequency quality' of the transmitter or the spread of transmission power across frequency. Several EEG-based anesthesia depth assessment monitors are (at least in part) based on EEG spectral analysis

2. **DE:** Like most neural activity, delta waves are monitored with an electroencephalogram (EEG) and are strongly correlated utilizing deep Non - rapid eye movement sleep stage 3, also defined as slow-wave sleep (SWS), which aid in sleep intensity classification.

5.3 Data cleaning

The efficiency of artifact identification and methods of elimination can be evaluated in several ways. A visual examination of the EEG time series is also given as satisfactory confirmation of the effectiveness of the method of elimination. Less specific measures can also be used. Metrics may be defined as decreasing some of the EEG characteristics that are undesirable. It is needed to accomplish such cuts in such a manner that other important characteristics are retained. Intolerance and specification calculations, for instance, the accuracy of object tracking systems is quantified being used in channels considered to be infected by objects about the number of artifacts extracted and the effect of this elimination on the clean EEG. A rather noticeable topic is ignored by such a focus on the elimination of artifacts. What else does EEG literally look like, notably, tidy? The assumption, often unspoken, is that anything which does not include objects is included in clean EEG. This specification ignores the likelihood of unrecognized objects. An artifact removal technique designed for glances, for example, would assume the residual EEG was clean. Conversely, the EEG can also include other artifacts, such as electrode pops, electromyography of muscular noise (EMG), etc. There is also a need to do so EEG to be explained analyticallThis can take the form of allowing more robust testing of artifact removal approaches. Presently, artifact elimination experiments identify different approaches to accomplish their effectiveness.

The output is then calculated as the number of artifacts detected fewer objects not identified, separated by the overall amount of artifacts, while quality is defined by measurements of sensitivity and accuracy about changes in artifact and – anti-artifact EEG materials. To calculate as clean an EEG epoch is, a metric will allow for rigorous correlations between such experiments to be made. For example, object removal approaches may be tested by highlighting the differences in the EEG cleanliness metric pre and post-use of the system. Attempts to assess EEG image strength focus mainly on the impedance and the estimated collection of instances discovered. In comparison, the identification precision or data transmission rate (ITR) can be used as a severe likelihood for transmission strength mostly in the context of BCI tests. These experiments are oblique and apply to a single study as well. A signal-to-noise ratio (SNR) approach tests between the EEG moments when a nod was instructed to the subject and the periods when no wink was found. The SNR (about just blinks) between both the EEG as well as the electrocorticogram (ECoG) was measured using this scale. However, this technique depends on the participants being able to generate blinks freely but does not account for other entity forms (e.g. objects of motion). Moreover, it does not allow a thorough estimation of measurement noise. Recent research uses cumulative EEG performance measures and classifier actions to make a decision when to switch between different control modes in a modified joystick-BCI regulated game. Each EEG performance is calculated by the amount of the EMG (as calculated by the minimum of magnitude) and the precision of the measuring performance of the BCI game. Therefore, the variable is special to this BCI request. Note that the increased interest in hybrid BCIs is also an area where EEG performance measures will be of great use. By contrast, in the dreaming adult mind, it offers superb descriptions of what a standard EEG would look like. Even so, the whole project is designed for healthcare professionals and related fields who examine and identify EEG visibly. Therefore, the translation of the EEG graphic representations into EEG accuracy empirical metrics remains to be done. First, we define several metrics in this review based on a clean EEG synopsis. We then aim to describe the spectrum and boundaries of these parameters in a clean EEG (identified by unbiased raters). Both clean and artifact-contact are the data sets used. EEG among stable patients, stroke survivors, and spinal cord injury (SCI) survivors. EEG is known to vary significantly between such groups Like clinicians allow EEG thresholds to be narrowly defined Data are learned by differential evolution (DE) on the optimal detection of clean epochs. Eventually, to define clean EEG epochs, the optimal level thresholds are supplied and an algorithm is answered. In our situation, we have used a moving average data filtering strategy, which provides us with improved results for noise eradication. A successful strategy for data cleaning is the Moving Average Approach.

In our case, we have used a moving average method for data cleaning which provides us better efficiency by eradicating noise. Moving Average Method is an effective way to clean data.

5.3.1 Moving Average Method

A moving average (tracking poll or running average) is a computation of the analysis of data points that produce a sequence of averages of different subsets of the real data. It is also known as a moving mean (MM)[1] or rolling mean and is a type filter of finite impulse. Differences include :easy and measurable, or weighted aspects (described below). The first component of the moving average is achieved by dividing the average of the number series' initial fixed sub-set, given the number series and the size of the fixed sub-set. And so the subgroup is modified by 'moving forward'; that is, the first number of the pattern is removed and the next value attached to the subset. A moving average is typically used to balance out short-term variations of time series analysis and explain longer-term averages or periods. A short-term to long-term threshold depends on the selection and the moving average variables are set accordingly. For example, it is often used in the technical study of financial records, such as share prices, profits, or quantities of trade. It is also used in economics to look at national income, salaries, or other economic and financial time series. Quantitatively, a moving average is a form of the convent, because it is being used as an instance of a low-pass filter used in data analysis. So if seen in non-time series data, a moving average filters greater range elements without any specific respect of time, even though typically some kind of purchasing is indicated. Which can be seen reductively as a layering of the results. An easy, analytical method is the moving average. Generally, moving averages are calculated to compute a stock's trend route or to evaluate its rate of assistance and opposition. Since it is based on past standards, it is a fad-following or missing predictor.

For the moving average, the longer the time interval, the greater the latency. A 200-day moving average will also have a significantly higher degree of lag than a 20-day MA, as it combines rates for the last 200 days. Shareholders and merchants are coupled with the increase by the 50-day and 200-day moving average representations for stocks and are regarded as major trading signals. An entirely flexible metric is moving averages, which implies that when evaluating an average, an entrepreneur can openly select any timeframe they want. 15, 20, 30, 50, 100, and 200 days are by far the most popular reason distances used in moving averages. The relatively short time frame used to generate the average, the more adaptable it would be to changes in prices. The longer the span of time, the average will be less receptive. Investors can choose various time cycles of various intensity to calculate moving averages premised on their exchange objectives. Although the potential movement of a given stock is difficult to forecast, using scientific research and review can help you make stronger forecasts. A rising moving average implies an uptrend in security, while a declining trend is implied by a declining moving average. Likewise, uptrend acceleration is validated with an optimistic crossover, which occurs whenever a short-term line chart reaches a longer-term moving average. Put differently, with a bullish spinoff, that also occurs whenever a short-term. The calculation can however form the foundation for many other technical indicators metrics, including the Moving Average Convergence Divergence (MACD) while calculating moving averages is valuable in its own right. Shareholders just use moving average convergence divergence (MACD) to track the relationship of all moving averages. By deducting it from a 12-day simple moving average, a 26-day moving averages average is usually calculated. The short-term mean is over the long-term average while the MACD is positive. It's an upward progress sign. If the short-term mean is under the long-term average, this is an indicator that the trend is decreasing. Several traders will also be watching a move up or down zero axes. A turn above zero is a trigger to buy, while a move under zero is a signal to offer.

5.3.2 Types of Moving Averages

Simple Moving Average

The precise nature of a moving average, defined as a simple moving average (SMA), is obtained by dividing the simple mathematical mean of a given set of values. In other words, a data packet or values are linked with each other in the case of financial products and then divided by the number of values in the set. The method of calculating the simple moving average of safety is as continues to follow:

$$SMA = \frac{A_1 + A_2 + A_n}{n}$$

where,

A=average in period n n=number of time periods

6 Training and Validation

A common task in machine learning is to research and construct algorithms that can learn from and predict data. Such algorithms operate by making data-driven predictions or choices by constructing a mathematical model from input data. Typically, the knowledge used to construct the final model comes from several datasets. In specific, three datasets are widely used in different phases of the development of the model. Initially, the model fits into a database of preparation, which is a list of examples used to align parameters of the model (e.g. relationship weights in convolutional neural networks between neurons). The model is trained using just a supervised learning algorithm (e.g. a deep net or a naive Bayes classifier) on the training dataset using, for example, optimization methods such as linear regression or stochastic gradient classification. In exercise, the test set often uses a pair of an input (or scalar) matrix and the expected outputs (or scalar) vector, where even the reaction key is usually called the aim (or label). The actual simulation is validated with the test set and outputs a response for each input image in the test set, which is then matched with the goal. Depend on the results of the correlation and the learning basics algorithm in use, the figure's variables are changed. In model fitting, both factor collection and variable approximation can be used. In a second incident, named the certain for the results, the fitted model is progressively used to simulate the answers. The evaluation dataset contains an objective method test that fits into another testing dataset when adjusting the model's hyperparameters (e.g. the number of hidden neurons (layers and layer widths) in a deep net). Verification databases could be used for normalization by reducing overfitting (stopping training when the failure on the validation dataset increases since this is an indication of overfitting the training dataset). In practice, this basic technique is complicated by the fact that during testing, the validation dataset error will fluctuate, creating several local minimums. This ambiguity has resulted in many ad-hoc guidelines being developed to decide when overfitting has truly begun. Eventually, the reference dataset is a collection that is used to include an unbiased indicator of a final design that fits into the database for practice. [5] The testing set is sometimes alluded to as a stint database if the information in the testing data has never been used in testing (for example, in cross-validation).

In this project 10% of the data was taken to be validation data. The rest 90% data was used for training.

7 Model Buildup

To detect emotion from EEG dataset two kinds of attempts took place.

1. CNN (Convolutional neural network).
2. 3D Convolutional neural network.
3. Model inspired by Siamese network.

7.1 Convolutional neural network

For Visual Imagery Ananlysis, now Covolutionary neural network is most used world wide. They are called shift invariance based on their shared-weight design and translation invariance features. They have picture and recognition software, image classification, medical images analysis and neural language processing] and succession of financial times. Regularized variations of multilayer perceptrons are CNNs. Generally, multilayer perceptrons mean fully connected networks, that is, every neuron in one layer in the following layer is connected with all neurons. The "full-connection" of these networks makes them weak against data overfitting. Loss function and calculation of weights are inclued to the typical ways of regularization. utilizing smaller and easier patterns, they have bit of advantage of the progressive patterns of data and collect more assemble intricate patterns. In this way, CNNs are on the lower end on the scale of connectedness and complexity. Biological cycles influenced Convolutionary networks in that the correspondence design between neurons takes after the association of the visual cortex of the animal. Just in a limited territory of the visual field known as the receptive field do individual cortical neurons react to stimuli. The responsive fields of different neurons halfway cover with the end goal that the entire visual field is filled. In contrast with other image classification algorithms, CNNs utilize next to no pre-processing. This infers that the network knows the filters that were hand-designed in conventional algorithms. A noteworthy advantage is this opportunity from past experience and human exertion in feature design. By applying significant filters, a ConvNet can effectively capture the Spatial and Temporal dependencies in a image. Attributable to the decrease in the number of parameters involved and the reusability of weights, the architecture is more qualified to the dataset. At the end of the day, to comprehend the complexity of the image better, the network can be taught. The CNN approach was carried out with the EEG dataset. One of the medicinal approaches is EEG or Electroencephalography. EEG is the most popular brain function investigation tool. EEG is utilized for the determination of cerebrum problems. The output of EEG signal signal is time-series data. Therefore, it is possible to equate performance data with the normal EEG signal to understand the illness or problem that exists. As of late, neural networks have assumed a critical part in the advancement of innovation. In several manufacturing sectors and classification schemes, the neural network has been used. Deep learning is one of the executions of neural networks. There is a rather complex layer of deep learning compared to a traditional neural network. Deep learning for the network has been used by many application frameworks. Deep Learning implementations are used to identify images. In specific manners, image classification plays a significant job. Deep learning has been developed lately. Deep learning creation begins with hand-written image classification, DenseNet, and GoogleNet. Here, we have established a new CNN approach to emotion recognition.

7.2 3D Convolutional Neural Network

A 3d CNN stays regardless of what we express a CNN that looks similar to 2d CNN. Then again, actually it contrasts in these following points Firstly, the 2D convolutional layer is a entry per entry multiplications between the different filters and the input. Here, both input as well as filters are 2D matrices.

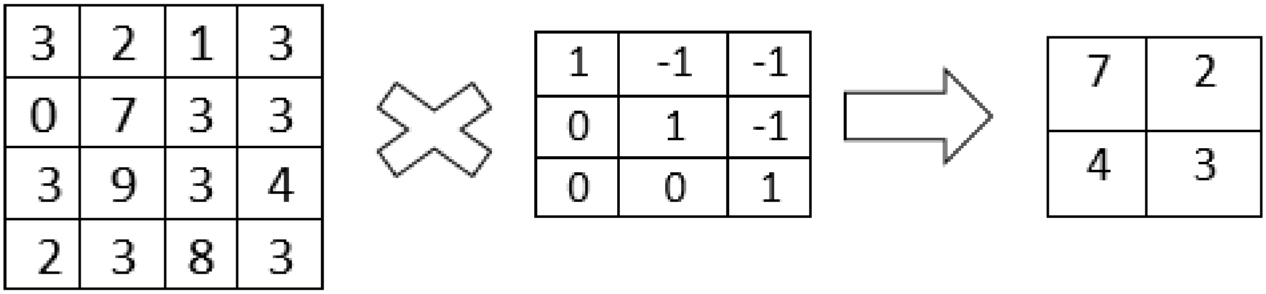

Figure 8: 3D Convolutional Matrice

We do same task in 3D Convolutional Network. We do these procedure on different sets of 2d lattices.

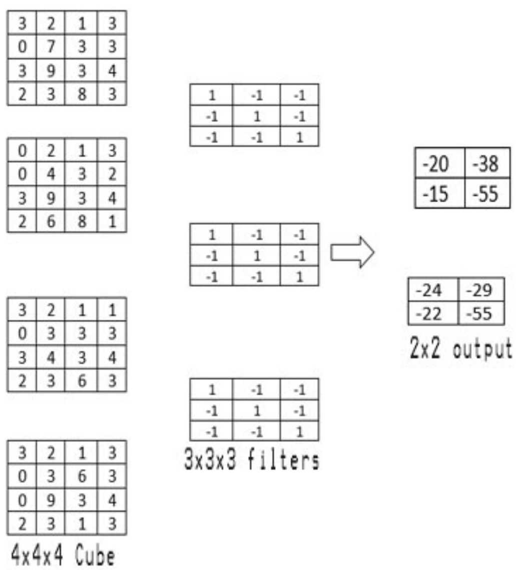

Figure 9: 3D Convolutional Matrices

Here, filter size is multiplied by a number of operations, and in addition , the size of the input itself multiplies the number of operations.

3D Maxpool

2d The maxpool layers (2x2 channel) are the maximum 2x2 square part that we can discern from the input.

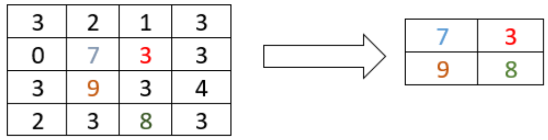

Figure 10: 2D Maxpool

We search for the maximum factor in a 2 cube width in a 3d Maxpool (2x2x2). This cube displays the area from the input, which is 2x2x2 area

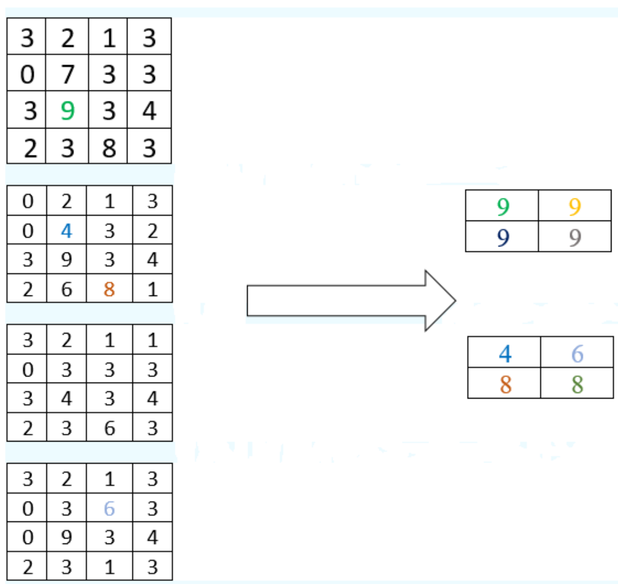

Figure 11: 3D Maxpool

The number of operations is compounded by the size of the filters used and the size of the input itself often compounded.

7.3 Attempt (1)

In the initial attempt two types of brainwave component goes through 3D convolution and resultant tensor goes through dense layer and then feeds to 4 neuron. Which determines which kind of emotion it was. Here, karnel size was determined to be (3,3,2). This model gives us 30% of accuracy.

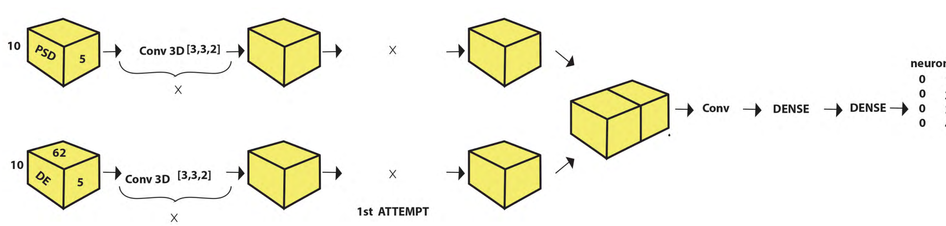

Figure 12: First Attempt of the model

Drawbacks of the first attempt

In this approach it was only possible to take a split amount of data as it was CNN 3D convolutional network.which may cause data loss or integrity of data reduction. As taken dataset deals with only limited chunk of data, the rest portion is unused. Which leaves the chance of missing spikes of brain activity in significant portion. As a result, this attempt did not meet expected result of successful emotion detection accuracy. This approach gave maximum of 30% accuracy which is not satisfactory. Thus this approach was not followed.

7.4 Attempt (2)

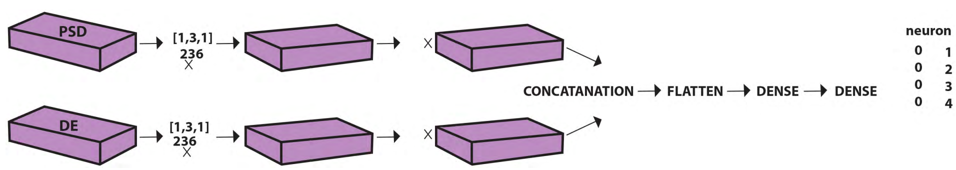

Figure 13: 2nd Attempt of the model

Unlike the the first attempt here data was taken for a single time instant. Both psd and de component are taken for single time instant and the gone through catenation and then through dense layer. As data was taken for single time instant here probability of missing spike is lower than before. Psd component is stored in variable Xpsd , De component is stored in Xde and result label or determined emotion is stored in Y variable in code.The karnel size was taken (1,3,1) and this model can work with 256 data at a time instant. Both component gone through catenation ,which gives resultant tensor. This model is inspired by Siamese network.

Layer (type)	Output Shape	Param #	Connected to
input_10 (InputLayer)	[(None, 1, 62, 5, 1)	0	
input_11 (InputLayer)	[(None, 1, 62, 5, 1)	0	
conv3d_11 (Conv3D)	(None, 1, 61, 4, 28)	140	input_10[0][0]
conv3d_14 (Conv3D)	(None, 1, 61, 4, 28)	140	input_11[0][0]
conv3d_12 (Conv3D)	(None, 1, 60, 4, 64)	3648	conv3d_11[0][0]
conv3d_15 (Conv3D)	(None, 1, 60, 4, 64)	3648	conv3d_14[0][0]
conv3d_13 (Conv3D)	(None, 1, 59, 4, 64)	8256	conv3d_12[0][0]
conv3d_16 (Conv3D)	(None, 1, 59, 4, 64)	8256	conv3d_15[0][0]
concatenate_1 (Concatenate)	(None, 1, 59, 4, 128	0	conv3d_13[0][0] conv3d_16[0][0]
batch_normalization_3 (BatchNor	(None, 1, 59, 4, 128	512	concatenate_1[0][0]
flatten_1 (Flatten)	(None, 30208)	0	batch_normalization_3[0][0]
dense_3 (Dense)	(None, 4000)	120836000	flatten_1[0][0]
batch_normalization_4 (BatchNor	(None, 4000)	16000	dense_3[0][0]
dense_4 (Dense)	(None, 500)	2000500	batch_normalization_4[0][0]
batch_normalization_5 (BatchNor	(None, 500)	2000	dense_4[0][0]
dense_5 (Dense)	(None, 4)	2004	batch_normalization_5[0][0]

Total params: 122,881,104
Trainable params: 122,871,848
Non-trainable params: 9,256

Figure 14: Summary of the model

7.5 Siamese Neural network

Same weights in order to calculate comparable output vectors when operating in tandem on two separate input vectors. One of the output vectors is always pre-computed, thus creating a baseline that contrasts the other output vector with it. This is similar to fingerprint comparison, but can be more formally defined as a distance function for locality-sensitive hashing. It is possible to construct a kind of structure similar to a Siamese network that is functional, but that implements a slightly different feature. This is usually used in various typesets to compare related instances. The use of similarity steps where a twin network can be used is the identification of hand-written checks , can be automated facial identification in camera pictures and comparing queries with indexed documents. Face-recognition may be the most common application of double networks, where well-known photographs of individuals are pre-compared to a tower style or otherwise. It's not clear at first, but two slightly different problems are involved.An person is identified among many other problems with face recognition and Deep face is a notable example.[4] It requires an individual in the extreme form of a single train station or airport. Other hand, it is face authentication in addition to seeing if the picture on a pass is like the person who claims to be the same person he/she is. The same double network can be used but can be applied more precisely. It is critical that the architecture not only of the sub-networks is identical, but that the weights also be shared among them in order to make the network known as "siamese." The key concept behind Siamese networks is that they can obtain useful knowledge descriptors which can be used to compare the inputs of the individual sub networks. The inputs can include numerical (usually complete connected layers, subnetwork), imagery (with CNNs as subnetworks), or even sequential (with RNNs as subnetworks), including sentences or time signals. Inputs may consist of any form of data. Siamese networks typically classify the output binary, classifying whether or not the inputs are all of the same class. Different loss functions can be used here during planning. Here the most common binary cross-entropy loss and it has common roles in loss function. The loss can be calculated as:

$$L = -y \log p + (1 - y) \log(1 - p)$$

The class mark (0 or 1) is y and the estimate is p when L's loss function. In order to train a network to recognise related and separate subjects and add losses we can feed it on a negative and positive occasion:

$$L = L_+ + L_-$$

The threefold loss can be used another way, D is a distance (e.g. loss of L2), a data set function is a sample, p is a positive random, and n a negative one. d is a distance function. m is an arbitrary margin used to differentiate between positive and negative scores. Siamese networks have wide applications. Some of them are here:

$$L = max(d(a, p) - d(a, n) + m, 0)$$

Siamese networks have wide applications. Some of them are here:

- **One-shot learning:** Learning with one shot, the trained (classification) network in this learning scenario is adding a new training dataset, with only one example per class. The classification accuracy of this new dataset will then be tested on another study dataset. In addition, Siamese networks may use these to generally understand entirely new classes and distributions, if they learn the distinctive features of a wide-range dataset. The authors use this capability to make one-shot learning on the MNIST dataset through a network that is trained in the Omniglot data set.
- Video video surveillance for pedestrians. This research incorporates a Siamese CNN network with the image patch size and location features to track several people in the camera's field of vision by identifying the position of each video frame, the associations of several frames and the trajectory computations.
- Jobs can be matched. In this adventurous programme, the network seeks to find matching jobs for candidates. A Siamese CNN network, a specialist in the field, will therefore extract and measure the syntactic and semantic similitude of deep contextual knowledge from the posts . The assumption is that matching resume pairs will be more similar in size than matching pairs.

7.5.1 Architecture of Siamese Network

In the circumstance of a CNN model, there is a sequence of convolution and pooling layers, followed by dense layers and a softmax output layer. As the convolutional layers here are responsible for extracting features from the picture, while the softmax layer provides a range of variables for each class. Siamese networks have the same composition as convolutional and pooling levels unless there is a softmax layer. We are then evaluating the image class with neurons with the highest probability value. We will end up with two layers, as described above because the network has two pictures as inputs.

The difference of these two layers is now determined and the effect is output in a single neuron with sigmoid activation (0 to 1). There must, therefore, be a list of two images and one variable, either 0 or 1, with the training data for this network.

- It includes a variety of convolutional and pooling layers in the CNN model, followed by several thick layers and an output layer with a softmax function.
- Convolutional layers are responsible for removing the feature from the picture and supplying a spectrum of probabilities with the highest probability value for each neuron type.

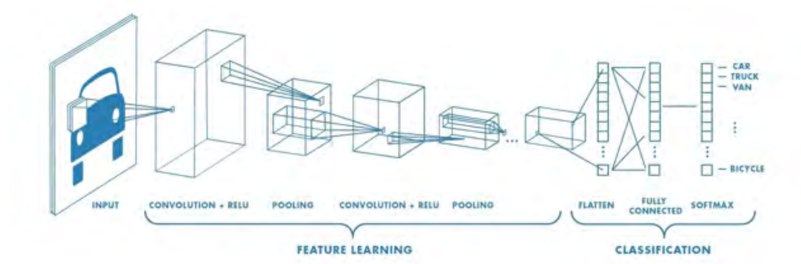

Figure 15: Siamese Network Architecture.

- Convolutional layers are responsible for removing the feature from the picture and supplying a spectrum of probabilities with the highest probability value for each neuron type.
- The network has two photos as references for the Siamese network, so there are two thick layers. The discrepancy between these two layers and the output outcome is bound to a single neuron with a feature of softmax activation(0 to 1).

7.5.2 Reasons for using This model inspired by Siamese Network

The reasons behind why we have chosen the Siamese network in our proposed model:

1. This model can deal with data for a single time instant which gives flexibility.
2. This model dealing with data for a single time instant makes it efficient as it eradicates any missing spike.
3. This model can give better efficienciency with comparatively less data. As we had less data so this model did enable us to work with as much as we had. Usually these works require a Terabyte amount of data, but this model did let us work with comparatively very low amounts of data than that.

In this research, Siamese Networks can provide three distinct advantages over the conventional CLASSIFICATION, which are true of any data, and not just pictures. It was designed for 3 Secret layer classification (200–100–2 size) & loss of Softmax. The output of its third hidden layer is presented in test images as embeddings. Siamese embedding is obviously suitable for long distance calculation not just to linearly separable. The network Siamese focuses on learning embedding (deeper) and brings together equivalent classes / ideas.

,Semantic similarity can be learned. Thus, This is different from the classification loss (e.g. logistic loss) that is only specifically rewarded for linear separation of groups. Moreover, One-shot Learning with Siamese Neural Networks can strongly solve the issues of having less data as we could not collect more data due to pandemic situations and used SEED-IV dataset around six gigabytes which is less in the era of Machine Learning. One-shot Learning with Siamese Neural Networks is the cursed breaker for small datasets in Deep Learning or Machine Learning.

Figure 16: Siamese Neural Networks Workflow

The activation function used in this research is ReLU and softmax. It produces unbounded network unit activation values which can be processed by Local Response Standardization. Normalization increases the sensitivity of an excited neuron to the neighborhood and damps a neuronal area when the activation in this neighborhood is generally high.

7.6 DenseNet

DenseNet is a kind of convolution neural network that uses dense layer connections via dense blocks in which we connect every layer directly with one another (with the matching function-map-sizes). Each layer obtains more inputs from all previous levels to maintain the feed-forward nature and transfers its own function mappings on to all subsequent layers. The next phase on the road to continuing to increase the depth of deep convolutional networks is Densely Linked Convolutional Networks, DenseNet. DenseNet requires less parameters than an analog CNN when connecting this way, as redundant maps are unnecessary to learn. In comparison, some ResNets variants have shown that several layers rarely participate and can be withdrawn. In fact, since each layer has its weights to learn, the number of parameters of ResNets are high. Instead, the layers of DenseNets (e.g. 12 filters) are very thin and they also provide a limited collection of new feature maps. Owing to the above flow of knowledge and gradients, another concern with very deep networks was the problems of preparation. It's because any layer has direct access to loss function gradients and original images and DenseNets can solve this problem. The CNN network can typically minimize the map size by Poing, or step¿1, and a dense connection mechanism requires a consistent map size in DenseNet's dense connection method. The DenseBlock+Transition structure in the DenseNet Network is used to address this problem. The DenseNet network layout, which comprises four DenseBlocks, and is connected by transformation by each DenseBlock. DenseNet has one significant benefit over traditional, deep CNNs: as it hits the end of the Network, information transmitted across multiple layers is not swept or lost. A basic connectivity pattern is required to accomplish this. Dense blocks and transformation layers are the principal components of a DenseNet. The first determines the combination of inputs and outputs, and the second restricts the number of channels, so that they are no longer too large. The data set is applied to the DenseNet architecture described in the research paper, SEED-IV. For preparation, the stochastic gradient descent optimizer was used by all architectures. For SEED-IV, the training size was 200 and the epoch size 30. The initial level of learning was set to 0.001 and remained constant. Why DenseNet is chosen and what benefits DenseNet has to bring in our model proposed model:

- Solid Gradient Flow: Easier to relay the error signal to previous layers. This is like implicit deep supervision, since earlier layers can be supervised directly from the final grading layer.
- Parameter and Computational Efficiency: The numbers of parameters in ResNet for each layer are directly proportional to C FunctionC, and the number of parameters in DenseNet is directly commensurate with l Functional. Moreover, DenseNets size is far smaller than ResNet's since k ¡ C.
- Further Diversification: As all previous layers are received as input by each layer in DenseNet, more diversifying characteristics and patterns appear to be richer.
- Classification uses the most complex features in Regular ConvNet. Classifier uses features of all complexities in DenseNet. It helps to make decisions more smoothly.It therefore explains because DenseNet is good for insufficient training information.

7.7 Categorical Cross-entropy loss or log loss

The function for evaluating a candidate solution is known as the objective function in connection with an optimization algorithm. We should try to make the objective work more efficient or less efficient, which means that you are looking for a solution with the highest or lowest score. We typically try to minimize the error with neural networks. This is also referred to as an expense or a loss function, with the estimated value of the loss function simply referred to as a loss function. Cross-entropy loss is almost always used in machine learning classification problems. In this study, we thought the theory and reasoning behind its extensive use would be interesting. Not as much was written on the subject as we expected, but we learned a few interesting things from what little that could be found. Though, We were frustrated with the various names and formulations that people were writing in their journals, and also with the loss layer names of the deep learning systems, such as Caffe, Pytorch, etc. TensorFlow. Calculates the efficiency of a classification model with a probability value of 0 to 1 production. It is legally programmed to measure the difference between the two probability distributions. This is the Categorical gorical cross entropy loss function measures the loss of an example by computing an example with amount of the following:

$$Loss = \sum_{i=1}^{outputsize} y_i.log\hat{y}_i$$

Here, i is the scalar value in the output of the model and Yi is the corresponding target and output value. In the model output, the size In the model output, the size This loss is a really good measure of how two discrete distributions of probability are distinguishable. These functions are modifications that we add to vectors leading up to loss computation coming from CNNs. A vector in the range (0, 1) is squashed by Sigmoid and applied to each individual element independently.This is also known as the logistic function here:

- Caffe: Sigmoid Cross-Entropy Loss Layer

- Pytorch: BCE With Logits Loss
- TensorFlow: sigmoid cross entropy.

Softmax it's a function, not a loss. Moreover, it depends on all elements of s and it cannot be applied independently to each s. Here,

- Caffe: SoftmaxWithLoss Layer is limited to multi-class classification.
- Pytorch: CrossEntropyLos s is limited to multi-class classification.
- TensorFlow: softmax cross entropy is limited to multi-class classification.

Usually, before the CE failure, the function activation is added (Sigmoid / Softmax). Estimate, Logistic loss and multinomial logistic loss are also terms for Cross Entropy loss. Loss of Sigmoid Cross Entropy is often called Cross Entropy Loss, and Softmax loss is also called Categorical Entropy Loss of Cross Entropy. With cross-entropy, the error between two probability distributions is determined by the maximum probability process. If the modelete has a classification problem where the input variables are converted to a class name, the likelihood of an example of each class can be predicted. There will be two classes with a binary classification problem, so we can predict the likelihood of the first class example. When classifying several groups, a likelihood for the example belonging to each class can be expected. The chances of a given class example in the training data set is 1 or 0 as each example in the training data set is a common example. To summarize the previous section and explicitly recommend the loss functions of the output layer to frame your prediction problem and to pick the loss function as the way for a given model frame to measure the error. Depending on the maximum probability system, an error between two probability distributions is determined by cross-entropy. The choice of the cost function is closely connected to the choice of the production unit. We use cross-entropy much of the time between the distribution of data and the model. In addition, the output is chosen as the transversal central entropy function.

7.8 Concatenation

Concatenation is the process of linking diverse strings together in the sense of programming. Concatenation, in addition to strings, can be generalized to every other data structure, including objects. Until the version of the concatenation string is modified, concatenation can be extended to a number of data types such as binary, integer, floating point, character, and boolean.By using one of the above operators, concatenation can then be effectively implemented. For objects, concatenation ensures that the data from inside object is concatenated.

7.9 Adam Optimizer

Implements the Algorithm of Adam optimization. Adam is a kind of stochastic gradient descent which figures the individual adaptive learning rates for various boundaries, in light of assessments of first and second order gradient moments. Implements the Algorithm of Adam optimization. Adam is a form of stochastic gradient descent that calculates individual adaptive learning rates from estimates of first and second-order moments of the gradients for different parameters.The optimization algorithm (or optimiser) is the key technique used today to train a machine learning model in order to decrease its error rate. In order to determine the efficiency of the optimiser, there are two metrics: the speed of convergence (the method of achieving a global optimum for gradient descent) and the generalization (the performance of the new data model). One or the other metric may be covered with algorithms like Adaptive Moment Estimation(Adam) or the Stochastic Gradient Descent (SGD), but both can not be covered by researchers. We used Adam optimiser instead of Stochastic Gradient Descent as when the learning rate is low, Stochastic Gradient Descent produces the same efficiency as standard gradient descent. Despite Adam's widespread popularity, recent research papers have noted that, under particular circumstances, an optimal solution can not be achieved. The article Improving Generalization Efficiency Moving from Adam to Stochastic Gradient Descent shows that , compared to Stochastic Gradient Descent, adaptive optimization methods such as Adam are poorly generalized. These had led the researchers to investigate innovative forms which could reinforce Adam 's conduct. Moreover, Adam is powerful and quick, among other things, the optimiser. We focused on the effective side of Adam Optimizer and fulfilled our hopes. Adam is characterized as combining the advantages of two other extensions of stochastic gradients. Particularly:

- Adaptive Gradient Algorithm (AdaGrad), which preserves a learning rate per parameter that increases output on sparse gradient issues (e.g. problems with natural language and computer vision).
- Root Mean Square Propagation (RMSProp), which often retains learning rates per parameter that are modified based on the average of recent weight gradient magnitudes (e.g. how rapidly it changes). This means that the algorithm performs well (e.g. noisy) on online and non-stationary issues.

Here, Adam learns the benefits of both AdaGrad and RMSProp. With in developed framework, Adam performs efficiently as Adam eliminates the limitation of limited data sets in artificial intelligence. As SEED IV dataset has a comparatively limited amount of data, which was very limited in Machine Learning. Learning with less data: Even with a single example, people have the potential to learn and are always able to identify new objects with very high accuracy. From the other hand, a huge proportion of classified information is necessary to practice and generalize deep learning models. This is a big drawback, and one-shot learning attempts to train neural networks with even small data sets. There really are multiple options we can really do. We need to change our loss function so that minute variations can be detected and better data representation can be learned. However one method which is frequently also used confirm photos is the Siamese platform.

8 Results

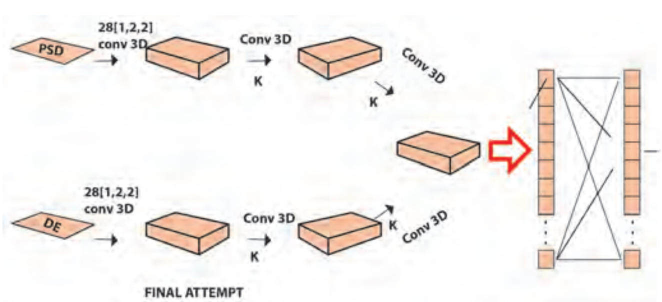

Figure 17: Final Model of the proposed system

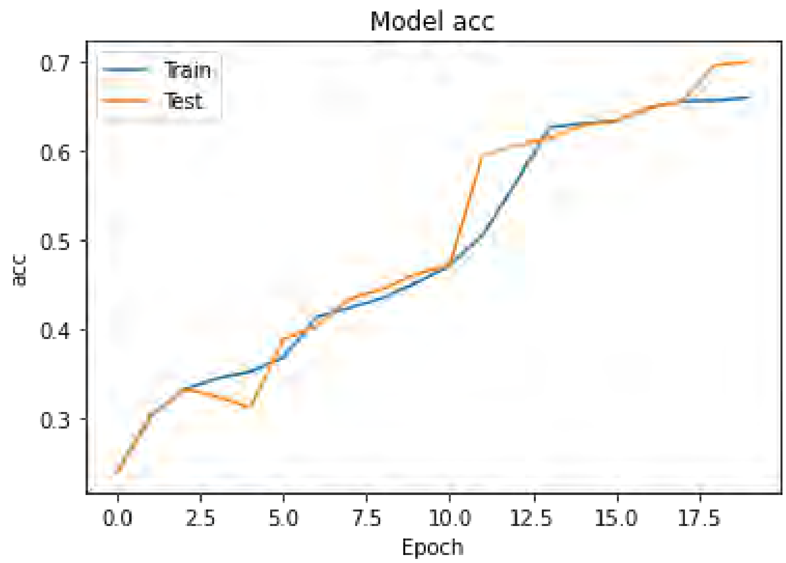

Figure 18: Training and Testing Accuracy of the model

In our experiment, using 90% data as training data and 10% data as validation data, we have found 69.9% validation accuracy. We performed training for 300 epochs and monitored the performance and saving weights at every epoch at the same time and found that, after 20 epochs, the model started to overfit significantly. As per the figure, this accuracy was the best among the 300 epochs. The categories are happy, sad, fear and neutral emotion for each input. As we have used softmax activation function at the last layer of our model, the ideal final output is like [1,0,0,0] for happy, [0,1,0,0] for sad and so on. But in practical test case, we get one highest value among the four neurons and that is the predicted result.

The range of being an input to happy, sad, fear or neutral is always in between 0 to 1. In this experiment we used the dataset where the main focus is to classify the inputs among four classes. The score can be easily achieved by printing the output matrix which is a floating-point number greater than 0 and less than 1. Change in this value does not affect the final result so much as the accuracy is calculated considering the correct and incorrect prediction, not the output score of softmax function.

9 Conclusion

The proposal gives a better insight into the emotional recognition by using EEG signals. Since 3D-CNN requires 3D inputs, a new method has been developed, which represents EEG signals in a 3D format from multi-channel signals. The time dimension is joined together to form the 3D input volume through a concatenation of consecutive frames. The SEED-IV dataset is used to show the efficiency of the method proposed. We have Enhanced EEG performance for signal acquisition and moving average method for signal enhancement are obtained by the EEG technique. Experimental outcomes indicate that our model understands emotions correctly. We also demonstrate the critical technique for an emotional recognition network in 3-D architecture, batch normalization and dense prediction. In addition , the idea of automatic emotional detection provides a viable way to recognize emotions in clinical and practical applications.

9.1 Drawbacks

We did not have sufficient data as we had only 10% data for testing, Moreover, we did not have better hardware support. We had to train in Google collaboration to overcome this issue partially and sometimes our system was stuck. And the final outrages that we have faced during our thesis project that the virus Covid-19 was spread out through the all parts of the world. So, Therefore, all earthly things come to an end. All over the there was a lockdown situation in which we have done our work through online process. We cant just meet with our group practically during this situation. Even though we couldn't make any physical contact with our supervisor.

9.2 Future Improvement

If we could collect more data, the model is used might give better accuracy. This will give better outcomes if we can implement inter-layer-loss sharing. Adding more data on the dataset to get more accuracy in our work. As we know that machine, learning gives us more accuracy if we have more data on the dataset.In addition, we want to detect more emotions in future. Using Siamese Network, the model can be less complex than the present one. In short, the proposed model can be output better accuracy if we can train it with more amount of data and can make it less complex by upgrading it.

Bibliography

[1] N.N. Khatri, Z.H. Shah, and S.A. Patel, “Facial expression recognition: A survey,” International Journal of Computer Science and Information Technologies, vol. 5, no. 1, pp. 149–152, 2014.

[2] Q. Yao, “Multi-sensory Emotion recognition with speech and facial expression,” Ph.D. dissertation, Computing. University of Southern Mississippi, 2014.

[3] W. Heller and J. B. Nitscke, ”Regional brain activity in emotion: A framework for understanding cognition in depresion,” Cognition & Emotion, vol. 11, no. 5-6, pp. 637–661, 1997.

[4] K. Nashiro, M. Sakaki, and M. Mather, “Age differences in brain activity during emotion processing: Reflections of age-related decline or increased emotion regulation,” Gerontology, vol. 58, no. 2, pp. 156–163, 2012.

[5] Anagnostopoulos, C.-N.; Iliou, T.; Giannoukos, I. Features and classifiers for emotion recognition fromspeech: A survey from 2000 to 2011.Artif. Intell. Rev.2015,43, 155–177. [CrossRef]

[6] Tzirakis, P.; Trigeorgis, G.; Nicolaou, M.A.; Schuller, B.W.; Zafeiriou, S. End-to-End Multimodal EmotionRecognition Using Deep Neural Networks.IEEE J. Sel. Top. Signal Process.2017,11, 1301–1309. [CrossRef]

[7] H. Greenspan, B. Van Ginneken, and R. M. Summers, editorial deep learning in medical imaging: Overview and future promise of an exciting new technique”, IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 11531159, 2016.

[8]H. J. Yoon and S. Y. Chung. “EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm,” Computers in Biology and Medicine, vol 43, no. 12, pp. 2230–2237. 2013.

[9]D. S. Naser and G. Saha. “Recognition of emotions induced by music videos using DT-CWPT,” 2013 Indian Conference on Medical Informatics and Telemedicine, India, 2013, pp. 53–57.

[10]J. Atkinson and D. Campos. “Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers,” Expert Systems with Applications, vol. 47, pp. 35–6, 2015.

[11]S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, “Emotion recognition based on the EEG using LSTM recurrent neural network,” International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, pp. 355-358, 2017.

[12]T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial-temporal recurrent neural network for emotion recognition,” Journal of Latex Class Files, vol. 13, no. 9, pp. 1–8, 2017

[13] Zheng, W.-L.; Lu, B.-L. Investigating Critical Frequency Bands and Channels for EEG-based Emo-

tionRecognition with Deep Neural Networks.IEEE Trans. Auton. Ment. Dev. (IEEE TAMD)2015,7, 162–175

[14] Koelstra, S.; Muehl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I.DEAP: A Database for Emotion Analysis using Physiological Signals.IEEE Trans. Affect. Comput.2012,3,18–31.

[15]S. Koelstra, C. Muhl, and M. Soleymani, "DEAP: A database for emotion analysis using physiological signals," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18–13, 2012.

[16]K. Li, et al., "Emotion recognition from multi-channel the EEG data through convolutional recurrent neural network,"inProceedings of IEEE International Conference onBioinformatics and Biomedicine, Shenzhen,China, 2017, pp. 352–359.

[17]K. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNetclassification with deep convolutional neural networks," Neural Information Processing Systems, pp. 1106–1114, 2012. [17]K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proceedings of Computer Vision and Pattern Recognition, 2014, pp. 1409-1556

[18] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proceedings of Computer Vision and Pattern Recognition, 2014, pp. 1409-1556.

[19]K. Li, et al., "Emotion recognition from multi-channel the EEG data through convolutional recurrent neural network,"inProceedings ofIEEEInternational Conference onBioinformatics and Biomedicine, Shenzhen,China, 2017, pp. 352–359.

[20]D. Maturana, and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, Hamburg, Germany, 2015.

[21] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," Expert Systems with Applications, vol. 47, pp. 35–41, 2016.

[22] R. Gupta and T. H. Falk, "Affective state characterization based on electroencephalography graph-theoretic features," in 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, 2015, pp. 577–580.

[23] D. S. Naser and G. Saha, "Recognition of emotions induced by music videos using DT-CWPT," in 2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT). IEEE, 2013, pp. 53–57.

[24] H. Candra, M. Yuwono, R. Chai, A. Handojoseno, I. Elamvazuthi, H. T. Nguyen, and S. Su, "Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine," in 2015 37th Annual International Conference of the IEEE on Engineering in Medicine and Biol-

ogy Society (EMBC). IEEE, 2015, pp

[25] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-time EEG-based human emotion recognition and visualization," in 2010 International Conference on Cyberworlds (CW). IEEE, 2010, pp. 262–269.

[26] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," The Scientific World Journal, vol. 2014, 2014.

[27] S. Blanco, H. Garcia, R. Q. Quiroga, L. Romanelli, and O. Rosso, "Stationarity of the EEG series," IEEE Engineering in medicine and biology Magazine, vol. 14, no. 4, pp. 395–399, 1995.

[28] A. I. Klistorner and S. L. Graham, "Electroencephalogram-based scaling of multifocal visual evoked potentials: effect on intersubject amplitude variability," Investigative ophthalmology & visual science, vol. 42, no. 9, pp. 2145–2152, 2001.

[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International conference on machine learning, 2015, pp. 448–456.

[30] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," Image and Vision Computing, vol. 31, no. 2, pp. 120–136, 2013.

[31] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[32] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," Journal of behavior therapy and experimental psychiatry, vol. 25, no. 1, pp. 49–59, 1994.

[33] J. A. Russell, "A circumplex model of affect." Journal of personality and social psychology, vol. 39, no. 6, p. 1161, 1980.

[34] S. Y. Chung and H. J. Yoon, "Affective classification using bayesian classifier and supervised learning," in 2012 12th International Conference on Control, Automation and Systems (ICCAS). IEEE, 2012, pp. 1768–1771.