A Review Paper On Big Data & It's Processing Using Hadoop

Mr. K. Omprakash, Mr.K.Mubarak Ali, Ms. M. Gowthami, Mr. S Abdul Basith, Mr. M Mohammed Sulthan, Mr. M. Nandakumar, Mr. M. Thillai Rajan, M
Al-Ameen Engineering College, Erode.

Abstract

The increased focus in encephalography to improve human-computer interaction (HCI) and build brain-computer interfaces (BCIs) for tracking and management applications necessitates fast information extraction from EEG devices. Due to its ability to train appropriate feature models from mixed data, DL has shown recently significant promise in working to make meaning of EEG data. The question of whether machine learning genuinely offers advantages over more standard EEG processing algorithms remains unanswered. Due to interference from external and internal artifacts, as well as physiological disruptions, focusing on the objective is challenging. Selecting elements that must be considered in future analysis can help to improve EEG-based emotion identification algorithms. As a result, the automatic selection of features of EEG signals is a crucial area of research. In deep learning methods, they present a multistep hybrid strategy combining the Reversed Correlation Method for automatic frequency band sensor combinations decision analysis. Our solution is easy to implement and decreases the number of sensors to just three to four channels. Tests on the image signal dataset were used to validate the suggested method of research. The validity of 2 factors and arousal-was tested using the effects produced. Our technique produced categorization findings that were 5.01–7.43 percent better than other scientific studies. Furthermore, because it belongs to unsupervised approaches, it can be viewed as a universal EEG signal categorization methodology.

Keywords: human-computer interaction, EEG signals, reversed correlation algorithm, brain-computer interfaces, deep learning.

1. Introduction

Emotion states are related to a wide range of human emotions, thoughts, and actions, and so influence our capacity to behave rationally in circumstances such as decision-making, vision, and human cognition. As a result, studies into emotion detection using emotional signals have improved the efficacy of BCI systems in medical applications and human interpersonal connections [1]. Biosensors are being used to monitor emotional situations while adding natural components of emotions in order to uncover therapy for psychological diseases such as ASD, ADHD, and chronic anxiety [2]. EEG-based emotion detection systems have been a popular research area in recent years.

The most difficult challenges in building an EEG-based emotion recognition system are efficient feature retrieval and appropriate classification. EEG signals are complicated, quasi-random, and stochastic [3]. They are buried in a variety of sound sources. As a result, in the creation of a reliable emotion detection system, processing and extracting relevant properties from EEG data is crucial. To quantify EEG data and as classifier properties, features extraction is used. Many features in the temporal domain [4], spectral domain [5], and combined time-frequency domain [6] have been extracted from EEG data in smart emotion recognition systems.

'db8', 'db4', 'coif5', and'sym8' were used by [7] to extract statistical information from EEG signals, including standard error, amplitude, and internal energy. Five distinct categories of emotions were classified using the KNN classifier. The maximum accuracy rate was around 90% on 59 stations, and it was 80.76 percent on 30 channels. [8] used the Gabor function and the wavelet technique to extract temporal, spectral, and spatial features from four EEG channels.

The ANN classifier, which had a 70.67 percent accuracy rate, was used to classify six different types of emotions. [9] extracted spectral information from 10 EEG channels, including wavelet transform energy and volatility. The maximum classification accuracy using KNN was 84 percent for arousal and 78 percent for valence. [10] employed the K-S test to choose the optimum channel for gathering sample entropy and passing it into an SVM as a feature. Jie's method is 91.2 percent accurate for arousal and 80 percent accurate for valence. [11], using three classifiers: QDA, SVM, and KNN, to describe emotion phases, integrated wavelet entropy, wavelet energy, updated energy, and statistical characteristics of EEG data. Alie's method has a 90 percent overall accuracy rate.

Experts in cognitive neuroscience have proposed a number of approaches and concepts for describing and separating non-cognitive and cognitive emotional states. (12, 13, 14) The valence-arousal paradigm [14] is a two-dimensional framework that includes HAHV, HALV, LAHV, and LALV emotional states. As a consequence, the valence-arousal model (Figure. 1) may be utilised to describe and evaluate each of the common emotional reactions.

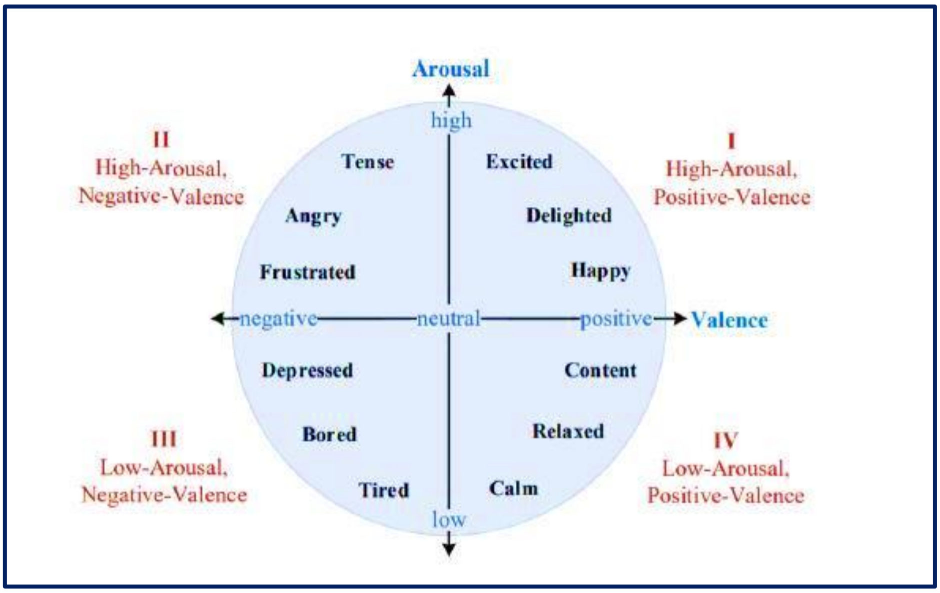

Figure. 1. The valence-arousal paradigm interprets various emotions.

Emotions have a critical role in human cognition, particularly in rational decision-making, vision, interpersonal touch, and intellect. Activity recognition appears to address the gap in emotion, particularly in HCI, by merging technology with emotions. HCI records emotional interactions between humans and machines in order to identify a user's emotional state. Computer systems that detect emotions are an area of computer science that aims to enhance human-machine interaction in a range of scenarios, including medical, commercial, military, and entertainment. [15]

Discrete and dimensional emotion models are the two types of emotion models. The dimensional framework portrays the persistence of an emotional state throughout time. Both valence and arousal are included in the majority of dimensional models. The discrete paradigm of emotions indicates that for a given number of feelings, there is a higher amount of emotion. Valence refers to the degree of pleasantness connected with a feeling. The valence spectrum depicts the shift from an unpleasant to a pleasant state. The intensity of an emotional experience is referred to as arousal. This arousal cycle goes from dormant to active in a continuous loop. The following are the definitions for dominance, valence, and arousal emotion types:

- Valence: in contrast to negative emotions, positive emotions lead to increased frontal coherence in alpha signals as well as greater right parietal beta networks and data.
- Arousal: activation of the parietal lobe with increased beta signal amplitude and coherence, but decreased alpha signal activity.
- Dominance: evaluated by an increase in the beta/alpha signal activity ratio in the prefrontal cortex and a reduction in beta activity in the frontal lobes on the EEG.

2. Methodology

We'll need to collect and analyse EEG data from a huge number of people in order to extract differentiating traits that may be utilised to detect distinct moods. The traits are then utilised to create data sets, which are then classified. This section will walk you through the data analysis process step by step.

2.1. Data collection of EEG

We were given EEG signals that had been recorded before. We'd like to offer a quick description of the approach, even if we didn't do it ourselves. Two Actiview active systems and BioSemi software were used to record EEG signals. The data was collected using the 20-30 electrode implantation approach, which is widely used across the world. The 20-30 system, also known as the International 20–30 system, is a generally established method for defining and applying scalp electrode positions in an EEG research or experiment.

BDF was utilised to store all of the generated signals. Each participant was told to sit attentively in front of the lab computer while wearing a BioSemi head cover that was correctly fitted to fit over their heads. Electrode holders and pin-type active electrodes were required per the experiment design. Electrode gel was poured into electrode holders, and pin-type active electrodes were inserted using a needle into the electrode carriers. To decrease the danger of losing vital information, a high sampling frequency of 2048Hz was chosen.

The purpose of flashing a blank screen is to reset the participants' emotions by allowing them to relax and unwind. Each photograph had been processed in the same way. The average length of the signal captured for each participant was about 12 minutes. Figure 2 depicts the BioSemi data acquisition patterning. On EDFbrowser, the electrodes F3, Fp1, Fp2 C4, C3, and F4 were used to duplicate the channels A4, A1, A23, A8, A30, and A27. The initial letter of each electrode specifies the exact lobe of the brain from which the signal is gathered. Below is the letter that correlates to each of the brain lobes: F is for frontal lobe and C stands for central lobe.

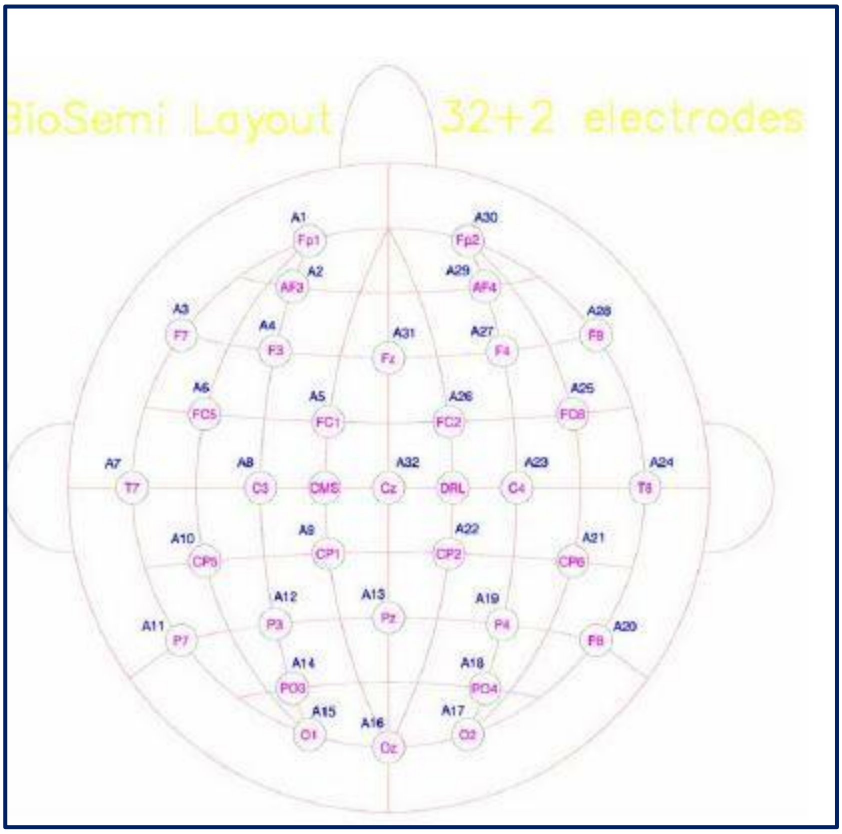

Figure. 2 – Positioning of electrodes in a BioSemi arrangement

2.2 Data processing of EEG

The obtained signals from the previous part must be processed further to reduce noise. The signals received were utilised to differentiate the signals produced on each subject when the IAPS picture was shown for the whole length of transmission. Finally, we extract a variety of features from those signals. As a result, all of these procedures will be described in the phases that follow under various titles.

2.2.1.Bandpass filtering for EEG

With the application of a proper bandpass filter, noise in recordings, such as superimposed artefacts from diverse sources, might be effectively eliminated. Blinking of the eyes has the most impact below 5 Hz, whereas heartbeat causes irregularities around 3 Hz and muscle artefacts distort the EEG spectrum about 25 Hz. At a frequency of 40-55 Hz, power lines cause non-physiological irregularities. 13 Another reason to use bandpass filtering is because the frequency range of the EEG is quite important.

To fulfil the requirement for both eliminating artefacts and retaining signals within the precise frequency range of interest, i.e. frequencies inside the Alpha (9-15 Hz) and Beta (15-28 Hz) bands, we employ the 10th order "Butterworth bandpass filter." As a consequence, we ensure that non-physiological and physiological artefacts are eliminated from the acquired EEG signals by extracting just the and B bandwidths. High-order filters may be necessary to achieve the requisite levels of stopband retardation or cutoff accuracy because they have larger roll-off rates between the pass and stop bands.

1. The amplitude response in the pass-band is as flat as feasible.
2. Outstanding overall performance
3. Chebyshev's pulse velocity is better.
4. Bessel's absorption rate is higher.

We may apply the filter we want to our signal in the browser by selecting it from the menu bar, as shown in figure 3.

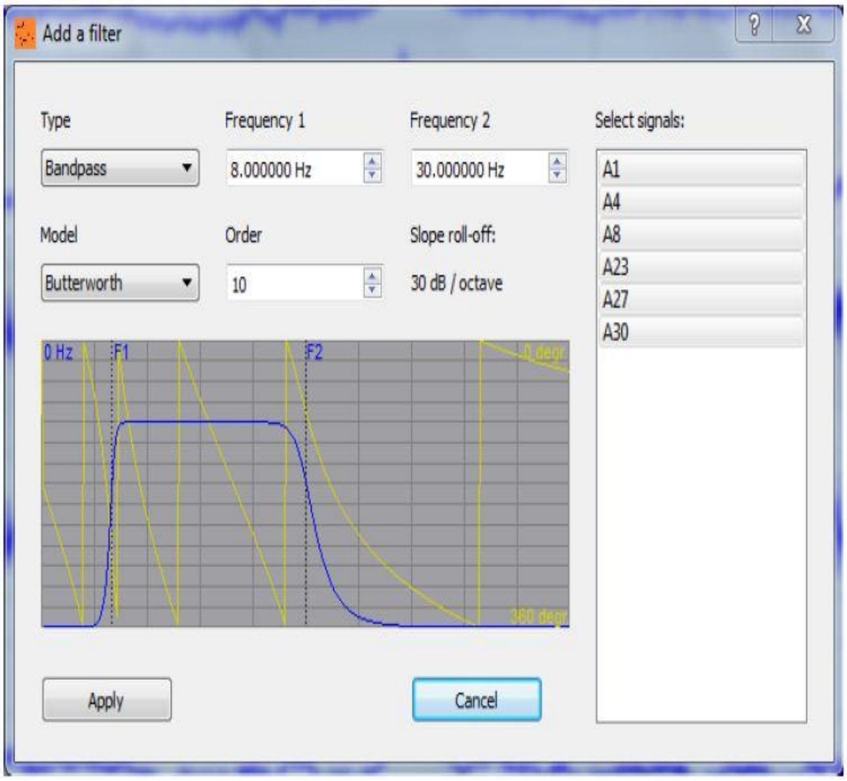

Figure. 3 – Applying a filter

As a result of this filter, new messages with frequencies ranging from 9 to 40 Hz will emerge, as shown in Figure 4.

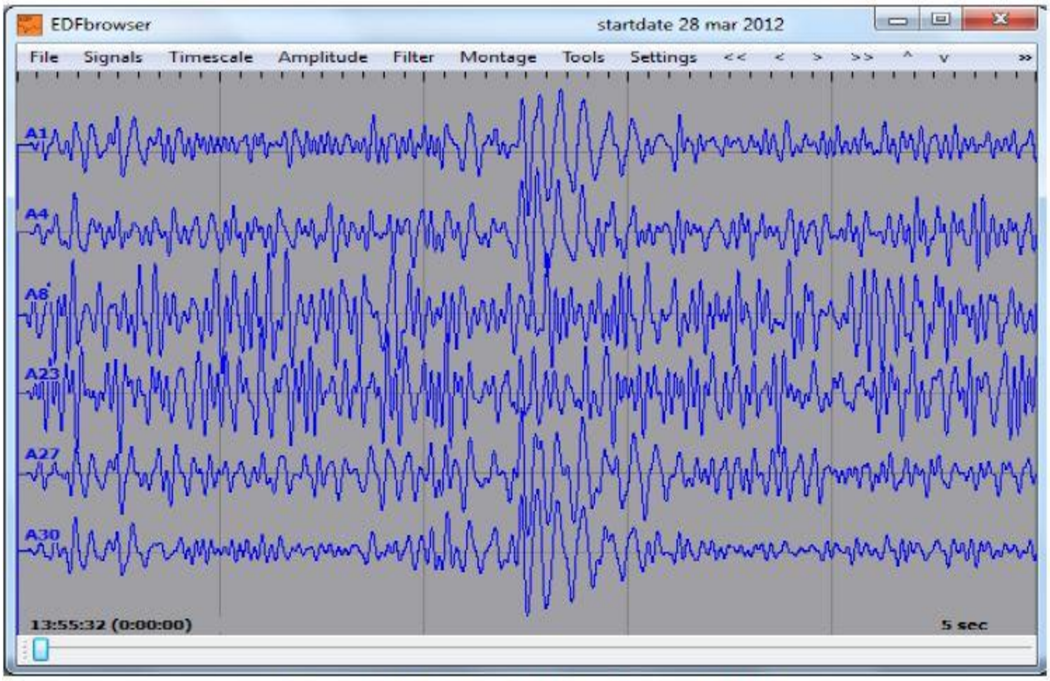

Figure. 4 –filtered signal sample

2.2.2 Data segmentation of EEG

The filtered data from the preceding phase must be processed further to extract the signals generated when the subject sees stimuli during the transmission. That is, we are instructed to save 30 segments of data, each lasting 5 seconds, for each topic for which IAPS photographs are supplied. For signal separation, we only require signals from six sensors. As a result, here's how to go about it: The EDF browser loads the BDF file, and 6 channels A4, A1, A30, A8, A23, and A6 are selected based on the wires mapping on the BioSemi design, as shown in figure 5.

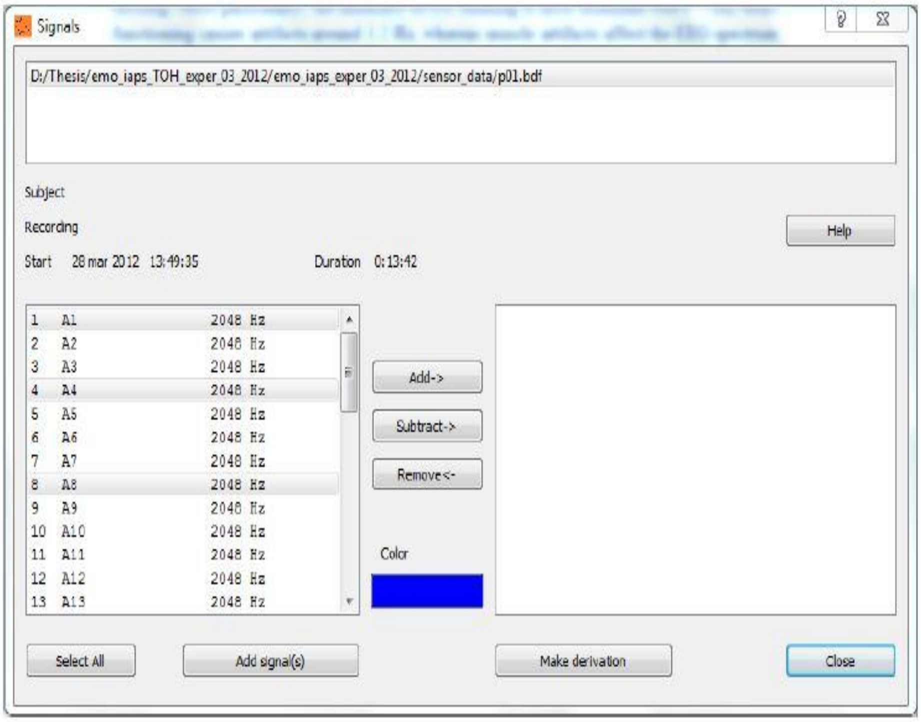

Figure. 5 – EDF browser, Signal dialog

Then, for more in-depth examination, specific channels are added. By putting the glass on the display, the collected signals may be seen. The original signal from one of the participants' trials, which lasted 14 minutes and 43 seconds in total, is depicted in thefollowing image, with each page presenting 10 seconds of data.

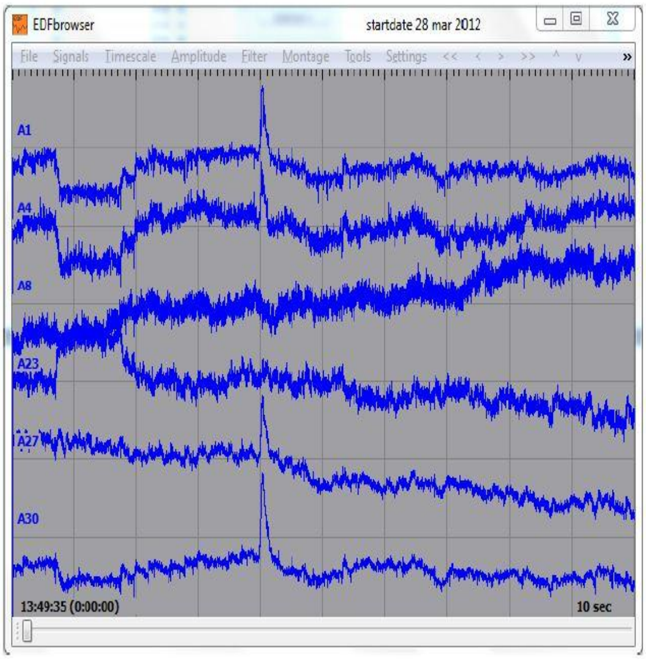

Figure. 6 – signals sample

We know the timing of each second of the experiment since each subject's recording time slot is available and has been logged on a text file. For example, the baseline for this experiment began at 13:54:22 and ended at 13:55:21 for a total of 59 seconds (when we talk about a baseline, in essence, we are talking about the duration of signal for which subjects are not shown any picture before actually starting the experiment). Then, at 13:55:29, a cross shape projection is started to attract the subject's attention, and at 13:55:32, the first IAPS picture is displayed for the subject for 5 seconds, meaning the picture ends at 13:55:37, and the process is repeated until 30 pictures have been completed and displayed for the subject.

The data from the first 5 seconds of the image is what we're seeking for, and it's required for additional processing and feature extraction in order to create the data set. So, according to the time of showing the photographs, we must now remove the 5 seconds of the signal. We can get the precise data when the picture was being shown to the subject by utilising the tools in the top menu and selecting decreased signals, time, or sampling rate. The channels that were utilised in the experiment are chosen, and the signal is calculated by calculating the precise seconds of the start and finish times of showing the picture. For example, the first image of this subject had been exhibited from 358 to 362. Figure 7 illustrates how to get the signal for the first image.

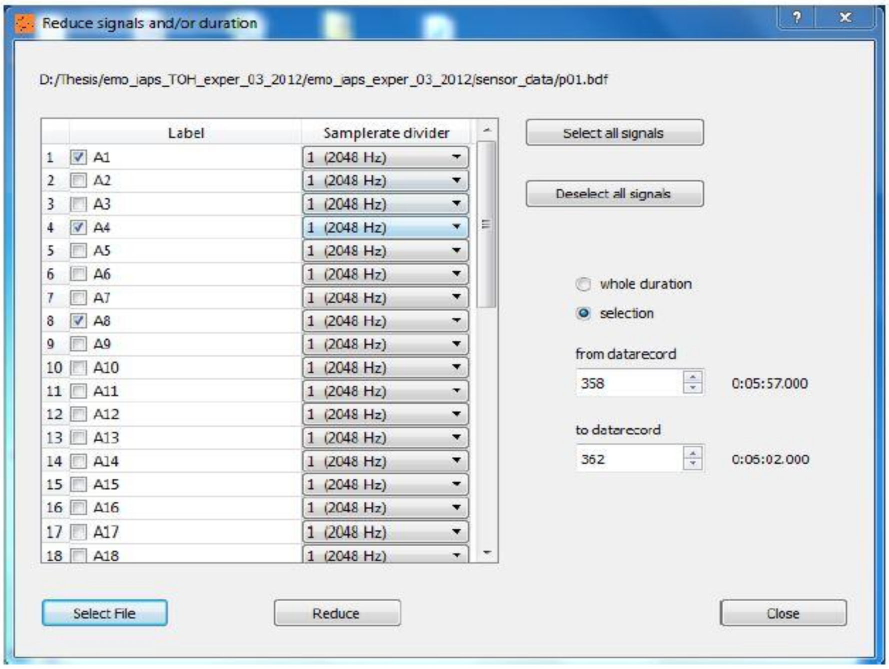

Figure. 7 – Reduce the number of signals or their length.

The data is saved for five seconds after hitting the "Reduce" button, and then reopened for further processing. Figure 8 depicts this signal as well. According to the data, the signal starts at 12:46:40 and lasts for 5 seconds.

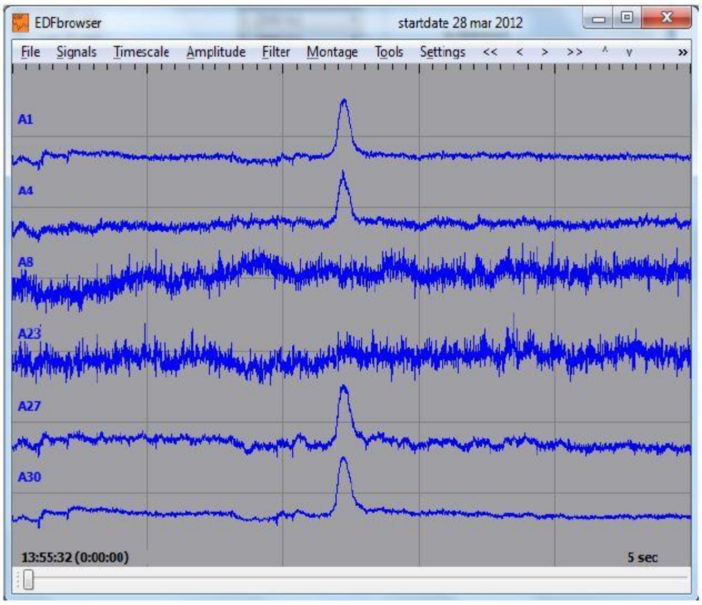

Figure. 8 –signal reduction

2.2.3 Feature Extraction of EEG

Now that the noise and artefacts have been removed from the data, we must pick which attributes to keep in order to construct data sets that will be fed into a deep learning tool to educate it what type of emotion these signals represent. In the first step, the 6 characteristics listed below were chosen to be extracted from each five-second section of the signal spectrum.

1. The greatest possible worth
2. The very bare minimum.
3. The difference between the mean value of the baseline and the sample mean of the raw signal.

calculating the confidence interval

4. The average of the signal's initial difference's actual values.
5. The median of the signal's second difference's real values.

To extract the aforesaid features, we must translate or convert the filter output, which has been saved as a BDF file, to ASCII. We may achieve this by using the tool from the EDF browser's top button. By selecting export BDF to ASCII from the tool option, the result will be translated into a text file containing statistical data that may be utilised in MATLAB.

3. Results And Discussion

This section contains a description of the categorization methods as well as a discussion of them. When each database is complete, it is submitted to WEKA, a classification software that employs all of WEKA's default settings. 10 fold cross-validation is used as the significance level for cross-validation. The data was classified using the SVM method. For dataset classification, the SVM was used because research has shown that it is a better choice for sentiment categorization from EEG data.

The entire datasets are then categorised again, but this time only on the polarity axis, i.e. Positive, Neutralor Negative valence. For data A1, 37 % for databases A2, 38 percent for databases A3, and up to 60% for data A4, the class labels were % for data A1, 37 percent for databases A2, 38 percent for databases A3, and up to 60% for sets of data A4. All of the results are summarised in Figure 9.

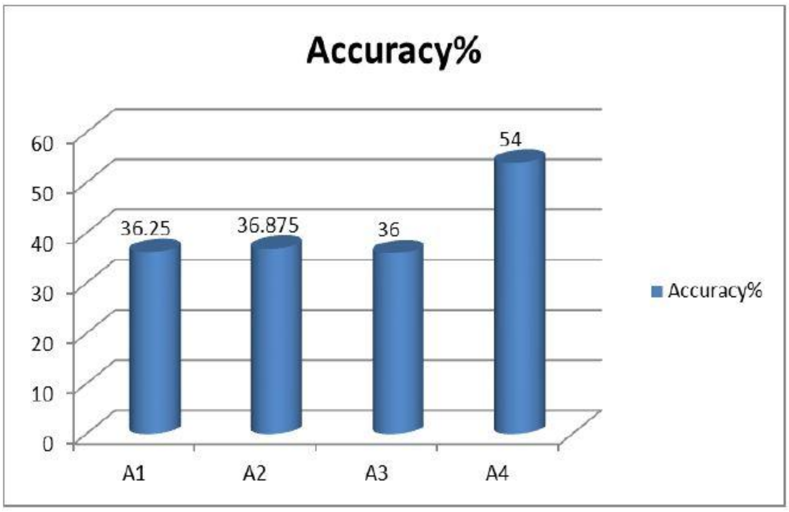

Figure. 9 –Results accuracy for three emotional responses utilising a variety of data sets

Our third categorization focused on the binary class score. When employing binary class ratings, feelings are defined as either positive feeling or not positive emotion, neutrality or not neutrality, and negative sensitivity or not negative sensibility. We categorised these 3 pairs of two binary base classifiers for each of the 4 data sets. When detecting feelings that are either good valence or not good valence, the dependability results for datasets A1, A2, A3, and A4 are 65%, 70 per cent, 58 %, and 71 %, respectively. The graphical representation of the result is shown in Figure 10.

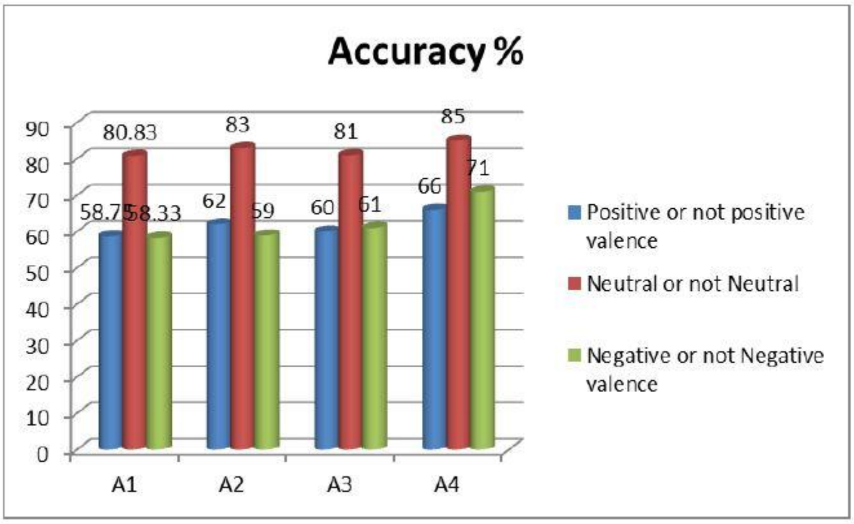

Figure. 10 – For binary emotive reactions, the classification result was

As a consequence, the graph above shows that different databases yield varied levels of accuracy. The same variation can be noticed, in this case dramatically, depending on the number of feelings considered for categorisation. The first argument for the differential inaccuracy is that it is due to a different set of data, which indicates that the sentiment analysis characteristics are different. As a result of the bulk nature of the data, which makes it difficult to examine in this manner, picture retrieval is crucial in emotion recognition. As a result, the issue of which characteristics should be taken into account and which will be discarded arises.

The reliability of the KNN classifier was demonstrated to be great when compared to other approaches. The KNN classifier performs well for a limited number of samples. As a result, the KNN classifier's efficiency is thoroughly investigated. Table 1 shows the efficiency of the Classification algorithm with emotional overlapping, which is visually illustrated in Figure 11.

.

Table 1. The emotion was detected using a KNN classifier with 24 features.

Emotion	KNN Accuracy
Sad	76.99
Anger	67.99
Happy	67.08
Calm	89.04

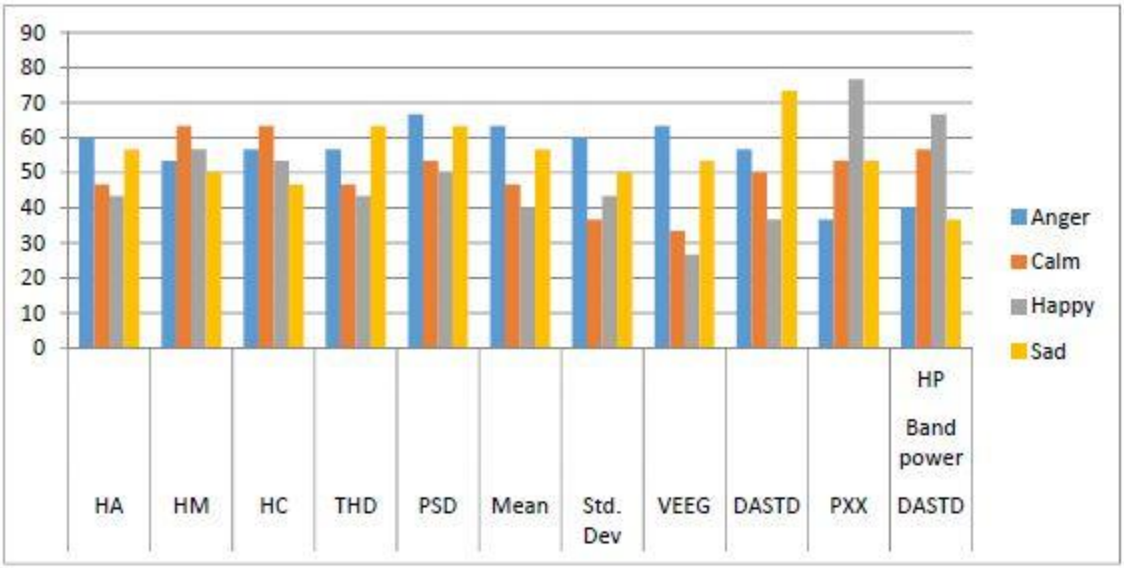

Figure 11. The graphical representation of the emotion accuracy rate utilising the KNN Classifier's essential features.

Accuracy improves greatly when mood overlapping was taken into account. Table 2 shows the reliability of KNN when highly rated attributes are used, and Figure 12 shows it graphically.

Table 2. The KNN classifier uses 24 characteristics to detect emotions.

Emotion	Band power	PXX	Mean	DASTD
Sad	56.09	78.09	897.56	78.98
Happy	68.98	87.98	56.76	56.87
Calm	63.98	45.98	45.65	67.23
Anger	67.3	76.45	89.98	54.87

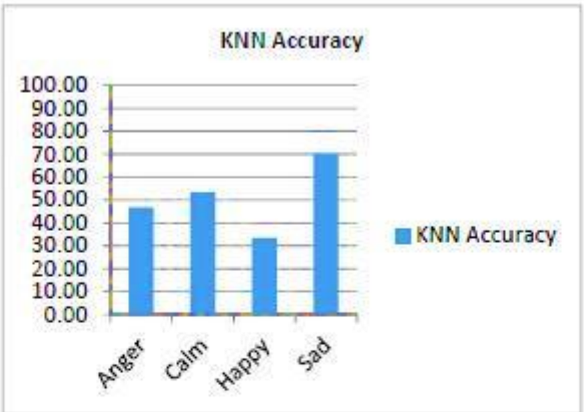

Figure 12. The schematic representation of emotion identification accuracy using the KNN classifier using 24 features.

4.Conclusion

1. The growth of signal processing and feature separation processes, and also the development of detectors and signal recording devices, has enhanced the possibilities for using messages derived from human body parts, like heart or brain signals, to recognize a person's situation and thus identify pathological or psychological conditions in living beings.
2. As a result, the task of categorizing signals became necessary for improving productivity in the classification of situations based on indications.

3. Effective data extraction from EEG sensors, in light of the increased interest in encephalography to improve HCI and build BCIs for real-time applications.

4. Due to its ability to create appropriate feature models from hybrid data, deep learning has shown significant promise in helping to make meaning of EEG signals in recent methodologies.

Reference :

[1] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," Emotion, vol. 8, no. 10, pp. 355–358, 2017.

[2] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," J. Neural Eng., vol. 16, no. 3, p. 031001, 2019.

[3] H. Dose, et al., "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," Expert Syst. Appl., vol. 114, pp. 532–542, 2018.

[4] L. Farsi, et al., "Classification of alcoholic EEG signals using a deep learning method," IEEE Sens. J., vol. 21, no. 3, pp. 3552–3560, 2020.

[5] O. Faust, et al., "Deep learning for healthcare applications based on physiological signals: A review," Comput. Methods Programs Biomed., vol. 161, pp. 1–13, 2018.

[6] R. Hussein, et al., "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals," Clin. Neurophysiol., vol. 130, no. 1, pp. 25–37, 2019.

[7] F. Li, et al., "A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning," Appl. Sci., vol. 10, no. 5, p. 1605, 2020.

[8] S. L. Oh, et al., "A deep learning approach for Parkinson's disease diagnosis from EEG signals," Neural Comput. Appl., vol. 32, no. 15, pp. 10927–10933, 2020.

[9] P. Sandheep, et al., "Performance analysis of deep learning CNN in classification of depression EEG signals," in Proc. TENCON 2019 – IEEE Reg. 10 Conf., pp. –, 2019.

[10] T. Shi, L. Ren, and W. Cui, "Feature recognition of motor imaging EEG signals based on deep learning," Pers. Ubiquitous Comput., vol. 23, no. 3, pp. 499–510, 2019.

[11] S. Toraman, S. A. Tuncer, and F. Balgetir, "Is it possible to detect cerebral dominance via EEG signals by using deep learning?," Med. Hypotheses, vol. 131, p. 109315, 2019.

[12] K. M. Tsiouris, et al., "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," Comput. Biol. Med., vol. 99, pp. 24–37, 2018.

[13] J. M. Williams, “Deep learning and transfer learning in the classification of EEG signals,” M.S. thesis, 2017.
[14] Y. Yuan, et al., “A multi-view deep learning framework for EEG seizure detection,” IEEE J. Biomed. Health Inform., vol. 23, no. 1, pp. 83–94, 2018.
[15] X. Zheng, et al., “Ensemble deep learning for automated visual classification using EEG signals,” Pattern Recognit., vol. 102, p. 107147, 2020.