**ORIGINAL ARTICLE**

# Human attention detection system using deep learning and brain–computer interface

S. Anju Latha Nair[1] · Rajesh Kannan Megalingam[1]

## Abstract
Brain–Computer Interface is tested as a successful method in improving human cognitive functions such as attention and memory. Attention plays a significant role in areas ranging from a person's day-to-day life to educational domain and professional activities. When attention is evaluated using camera-based techniques, the users may suffer privacy issues. Using Brain–Computer Interface (BCI) to obtain a measure of attention will be useful in this regard. The paper proposes a Human Attention Recognition System (HARS) in which EEG signal acquisition is used to obtain the attention of the individual, Renyi's entropy-based mutual information method is used for feature selection and a deep learning-based classifier is used to classify the signals. HADS is not camera-based: therefore, faces of the subjects are not revealed. EEG signals were collected using the Ultracortex Mark III dry electrodes and were visualised using OpenBCI GUI (Graphical User Interface). The experiment is validated using the publicly available Confused Student EEG dataset from Kaggle, giving an accuracy of 99.21%. The results indicate that the proposed method can identify attention levels accurately and can be effectively used in educational systems, biofeedback and medical research.

**Keywords** Brain–computer interface · EEG · OpenBCI · Student attention · Mutual information · Deep learning

## 1 Introduction

Attention can be described as focussing cognitive resources on information, while avoiding distraction [1]. To know whether a person is attentive or not, the facial expressions and/or eye movement may be observed manually, or by automated methods. Zaletelj proposed a novel approach in assessing students' attention in classrooms in which Kinect One sensor was used to capture body movements and facial features of students [2]. Zhang et al. [3] collected information on students' attention using wearable devices. These techniques have the limitation that they fail to respond if there is no obvious shift in attention, resulting in a change in eye/head movement or facial expressions. This made researchers to apply Brain–Computer Interface (BCI) in attention analysis. BCI is defined as a system which gives their users communication and control channels that do not depend on the brain's normal output channels of peripheral nerves. BCI-based communication has a unique feature that no traditional means of communication is required. BCI can be used to acquire the emotional state as well as the cognitive state of a person. The signals from the brain can be acquired with the help of specialised electrodes. The electrodes can be invasive, partially invasive or non-invasive. The first two types need surgical procedure by which electrodes can be placed inside the brain. Non-invasive electrodes are placed on top of the head, and are popular because they are easy to use and will not cause any long-term impacts. The raw EEG signals from the brain are applied to a feature extraction technique. This work used Fast Fourier Transform to extract the features. The extracted features are applied to a mutual information-based feature selection algorithm.

Apart from many possibilities of BCI like medical research and biofeedback, it can also be used to improve cognitive performance. EEG is a voltage signal which

✉ S. Anju Latha Nair
anjulatha@am.amrita.edu

Rajesh Kannan Megalingam
rajeshm@am.amrita.edu

1  Department of Electronics and Communication Engineering, Humanitarian Technology Labs (HuTlabs), Amrita Vishwa Vidyapeetham, Amritapuri, India

measures the neural activity of the brain and this activity fluctuates according to cognitive activity and mental state [4]. Sezer et al. [5] used Support Vector Machine (SVM) to classify EEG signals to track attention and obtained an accuracy of 70.62%. Hassan et al. [6] devised a novel attention recognition model with advanced machine learning algorithms obtaining data using a single channel EEG device with accuracy of 89%. Certain disorders such as attention deficit hyperactivity disorders (ADHD) can be detected effectively with the help of BCI Technology [7]. ADHD is a neurobehavioral disorder which can cause lack of attention and focus, along with other issues in controlling normal behaviour [8]. Cho et al. [9] proved in their work that EEG biofeedback can be used to enhance attention in children suffering from ADHD. Another set of stakeholders of EEG-based attention detection is visually impaired students, with the difference that auditory stimulus should be presented to them, instead of visual stimulus [10]. Application of deep learning models in classifying EEG data is emerging in fast pace in recent years [11, 12]. Deep neural networks are widely used with high success rates in various areas like bioinformatics, medical imaging and health monitoring [13], especially in the recognition of mental states like Alzheimer's disease [14] and depression [15, 16]. Toa et al. used deep learning to analyse brain signals, combined with eye-gaze for attention and obtained an accuracy of 92% [17].

In real-life, however, a change in attention is not necessarily associated with a change in eye gaze or head movement. Camera-based systems rely upon changes in facial expression for recognising a lack of attention. In many cases, the person's attention may be lost even when there is no change in facial expression. On the other hand, another person may maintain a neutral expression even when highly attentive. BCI measures the signals directly from the brain, not considering eye gaze or facial expression for detecting the attention state. This gives an objective measure of attention. BCI-based systems can also account for individual variability in the manifestation of attention. In addition to that, when attention analysis is required in domains other than education, for instance, in medical research or rehabilitation, using face detection may turn up as a privacy breach. Hence there is a lack of research in exploring the possibility of considering brain waves alone for the development of an attention recognition model, without compromising the accuracy. Keeping that in mind, this research is devoted to the development of a deep learning neural network model for analysing the brain features in an advanced, effective and accurate way.

BCI will collect only the brain signals, and will not expose the face, surroundings or audio information of its users, unlike camera-based systems. It typically requires the consent and engagement of the user for its working. So,

the users will have control over when and how the data are collected. This is a significant advantage over camera-based systems that may collect images or even videos, without the consent from the users. BCI-based systems are less susceptible to surveillance, because it primarily collects brain data and there are no cameras involved. Another advantage is that it is easy to anonymize brain data as it does not include facial or audio data of a person. This ensures the privacy of the users, especially in cases of medical and research settings. Another advantage of BCI is data minimization. BCI collects only specific brain signals; and therefore the overall data amount of data is less when compared to continuous video recordings.

The rest of the paper is organised as follows: Sect. 2 provides state-of-the-art studies in our research area. Section 3 provides the methods we have used to carry out the research. It provides the pre-processing techniques, feature extraction, dataset creation and the machine learning algorithms used in the study. Section 4 consists of the results.

# 2 Related work

In this section, a summary of previous studies in our research field is provided. This chapter is structured as follows: Initially, we present the various works in attention analysis. Then we discuss the use of machine learning in attention detection. The next sub-section presents some works which used BCI for the analysis of attention. In the final sub-section, we present the use of SSVEP BCI as a communication paradigm between humans and an external device, doing away with the conventional interactive methods such as keyboard or voice commands.

## 2.1 Attention analysis methods

In Zaletelj's model, head-motion, pen-motion and visual focus were integrated, thereby producing a multimodal system to analyse the attention of students [2]. Here, information is collected with the help of cameras, gyroscopes and accelerometers and students' behaviour is collected. Machine vision-based approach was used to obtain good estimations of manual ratings. Automated analysis was used to improve correlations between manual ratings and post-test variables. Farhan et al. [18] presented the use of Internet of Things (IoT) framework in attention analysis. They proposed an Attention Scoring Model (ASM) whose algorithm can be implemented using any programming language. A camera is used to monitor the students' activities while watching a video lecture. If a face is detected, a face recognition score is logged in and when eyes are seen opened, an eye detection score is logged in.

Goldberg et al. [1] propose a proof of concept in which a machine vision-based approach is used for analysing students' engagement or disengagement in class. The authors extracted direction of gaze, head pose and facial expressions and performed an automated analysis. A pilot study on camera-based attention systems was done by Renawi et al. [19] which used a webcam, a standard computer and computer vision algorithms to estimate the level of attention of students in a classroom. All these works used changes in facial expressions captured using camera for detecting inattentive states. In most of the real-life scenarios, a loss of attention, even when it is deliberate or not, may not reflect in a change in facial expression. So, it is evident that there is a requirement of techniques other than camera-based ones for analysing attention.

## 2.2 Use of machine learning in attention detection

Machine learning has been used for the classification of attention in many works. Numerous works which assessed students' attention with the help of machine learning techniques have been reviewed by Villa et al. [20]. Li et al. [21] proposed a machine learning-based approach, which is a novel multimodal assistant system to infer the attention of students during formative assessment. K-Nearest Neighbours was used by them for real-time recognition of attention by developing a Self-Assessment Manikin (SAM) model. Instead of using neural networks, simpler methods such as support vector machines (SVM) can lead to less training time and speedy convergence. They obtained an average accuracy of 57.03% [22]. So, when simpler methods are used for attention analysis, it results in lowering accuracy.

## 2.3 Use of BCI in human attention detection

Recently, brain waves are used for the recognition of emotional states of an individual. Electroencephalography (EEG) sensors are commonly used to capture the brain waves and advanced machine learning techniques are used to recognise attention level. One of the earliest attempts in this was made in a work by Hassan et al. [6]. The work used frequency decomposed EEG and devised a novel machine learning attention recognition model. A hybrid model combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) was used for sequence classification. A study of students' attention levels at real classroom settings was done by Sezer et al. [5] with the help of NeuroSky MindWave which is a commercial EEG device used to measure brain waves. The results of this study indicated that teaching methods using digital media helped in an increased attention compared to

lectures without video or PowerPoint presentations. SVM was used to classify EEG data collected from students in classroom with the help of mobile sensors [23]. The classification accuracy obtained was 70.62%. The use of artificial intelligence in multilevel attention recognition is explored by Parui et al. [24].

## 2.4 Use of SSVEP BCI as a communication paradigm

BCIs can be classified as endogenous and exogenous. Endogenous BCI allows intentional modulation of neuronal activity by the users whereas exogenous BCIs depend on stimulus applied externally to the user. Former includes paradigms like Motor Imagery. SSVEP and P300 paradigms fall under the latter. A study of both kinds is presented by Ravi et al. [25]. Our work is focussed on SSVEP-based BCI. Visual Evoked Potentials (VEP) are used widely in EEG-based BCIs. The design of a suitable visual stimulator has an important role in using VEPs for BCI. Wang et al. proposed the design of a visual stimulator for use in SSVEP BCI [26]. Their approach was to use computer monitor flickers to elicit steady state visual evoked potential at a flexible frequency. BCIs are used by people of all ages. The dependence of age in various subjects in the performance of BCI is reviewed by Volosyak et al. [27] and the work concluded that there will be a performance drop in SSVEP BCI in people with advanced age and suggested that GUIs should be modified for elderly users. SSVEP-based BCI can be used in neuro orthosis for cases like tetraplegia [28] and for controlling robotic arms [29] and robotic wheelchairs [30, 31]. It can also be used in combination with motor imagery [32], electromyography [33], eye gaze [34] or event related synchronisation [35] and event related desynchronisation [36]. When BCI is combined with another system or when more than one BCIs are involved, the system is called hybrid BCI [37]. Numerous attempts have been made in improving the performance of SSVEP BCI, such as Task Discriminant Component [38] and other optimization techniques [39]. From all the related work, it is obvious that the development of an accurate attention recognition model is a requisite, in diverse domains in which human cognition has a role.

## 3 Materials and methods

A BCI device establishes a communication between human brain and an external device. The architecture of the proposed BCI system is presented in Fig. 1. The objective of this system is to acquire EEG signals from the brain of the user, pre-process them, select the relevant features from

**Fig. 1** Block diagram of the human attention recognition system
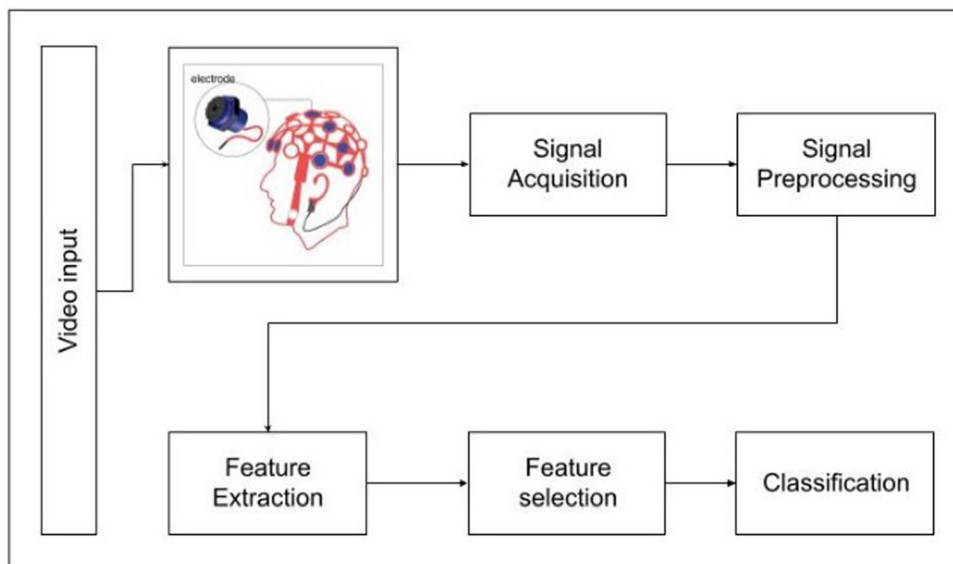


them according to the Human Attention Recognition algorithm and classify those signals using a deep learning classifier to recognise the attention levels. An EEG-based Brain Communication Interface is used to estimate the mental attention states of humans. A Human Attention Recognition Algorithm is developed and a 10-layered deep learning neural network is used for classification. The challenge is aimed at using machine learning to estimate the attention of students. The dataset is the publicly available Confused Student EEG brainwave data from Kaggle which consists of EEG data collected from 10 volunteers who were made to watch MOOC video clips [40]. Each student watched 10 videos, which means there are 100 data points. It also contains demographic data and video data, but for this work, we have used only EEG data as the benchmark. The creators of the dataset claim that the dataset is well suited for binary classification. It consists of Video ID, subject ID, EEG frequency bands and also User-defined and pre-defined labels.

The Electroencephalography (EEG) signal measures the electrical movement of the cerebrum [41]. EEG uses either dry electrodes or wet ones, depending on the application. The wet electrodes come with an abrasive paste and an electrolyte gel, used to reduce the skin impedance, to a range of 5–20 kΩ, which is an acceptable value when compared to MΩ range without the use of the gel. The paste and gel are minimally invasive and harmless, but are sticky and make the scalp dirty. Apart from that, when the gel begins to dry up, transductive properties start disappearing which makes the wet electrodes not suitable for long term measurements. Therefore, in recent decades, the use of dry electrodes has been increased, which resolves the limitations of the wet ones [42].

This work used an 8-channel headset in which dry electrodes are installed in Ultracortex nodes. Recording of EEG signals is done with Ultracortex Mark IV which is a 3D-printable headset. Among the electrodes, six of them are named spikey electrodes, designed to be used in areas of the head with hair and two of them, named non-spikey electrodes, are designed for the forehead. These electrodes are placed in a 3D printed headset. 5 more comfort units, without electrodes were also used to distribute the weight of the headset. An OpenBCI Cyton 8-channel board was used for EEG signal acquisition. Ultracortex Mark IV electrodes do not require conductive gel or adhesive paste as they are dry electrodes. Also, there is no need of skin preparation. There are two ear clips which act as the reference electrodes. Electrodes were placed according to the 10–20 International standard for EEG electrodes [43].

### 3.1 EEG recording

The 10–20 EEG system was introduced three decades back by Homan and since then had been the standard for locating EEG electrodes on the scalp [41]. It measures external cranial landmarks for locating electrodes. The measurement is based on the assumption that the scalp electrode locations and underlying cerebral structures maintain a consistent correlation. For 8-channel system, the

**Table 1** Electrode names and their positions on the brain

| Notation | Position |
| --- | --- |
| FP1, FP2 | Pre-frontal |
| C3, C4 | Central |
| P7, P8 | Parietal |
| O1, O2 | Occipital |

standard mentions electrode positions as described in the Table 1. The electrodes mounted on the helmet is shown in Fig. 2. This can be worn comfortably by a user for recording his/her EEG data. The acquired signals were visualised and recorded using OpenBCI Graphical User Interface.

The voltage levels from each of the electrodes along with information like accelerometer data and timestamp are automatically saved as a text file by the OpenBCI application.

## 3.2 Experimental setup

The Kaggle dataset provided pre-processed data which could be applied directly to the deep learning model. To generalise the model, we created our own dataset in which eye-blinks and power supply noise are eliminated using filters. The acquired raw data were converted into frequency sub-bands, Alpha1, Alpha2, Beta1, Beta2, Gamma1, Gamma2, Delta and Theta. Before the experiment, the participants were given a brief introduction on the nature of the experiment. They were seated in a cabin, with a table, chair, a laptop and a mobile phone. Then they were allowed to wear the EEG equipment, with the help of one of the authors and were asked to relax so that they can get used to the equipment. Each of the participant were to watch a 2 min video in a day. There were 10 such videos and it took 10 days per participant to complete the experiment. While the participants watched the video, the EEG signal from their brain were acquired and recorded. The collected raw EEG was input to the deep learning network for binary classification. The detailed process of the dee



**Fig. 2** Experimental set up

learning model will be explained in a subsequent subsection.

Given the EEG data from 10 participants, our task was to determine their attention using deep learning methods. The participants were assigned to watch 10 videos containing pre-defined labels classifying them as 'easy to focus' (label 1) and 'difficult to focus' (label 2). The easy videos contained topics which are familiar to an engineering student, whereas the difficult topics were taken from the middle of video clips containing topics which are not familiar to them, removing the introduction part. The participants wore an 8-channel EEG headset which is connected to the OpenBCI software which was used to extract the focus data.10 healthy male volunteers, each with normal or corrected to normal vision participated in the experiment. The videos had an average length of 2 min. The volunteers wore the Ultracortex Mark IV headsets while watching the videos and the EEG waves were recorded with the help of OpenBCI software. They were all given awareness of the procedure and purpose of the study. Before the commencement of the experiment, consent was obtained from each of the participants and once the signals were acquired, their names were anonymised, so that their privacy is not affected. The data are not made public and is available with the authors only, ensuring its ethical use. Data are collected from 10 volunteers, watching 10 videos, leading to 100 data points in more than 12,000 rows. The sampling frequency was 256 Hz. Each volunteer watched one video per day. The experiments were monitored by one of the authors so as to ensure that no significant disruptions took place.

We used K-means to cluster the data due to the fact that it guarantees convergence. Although the number of target classes is known to be 2, we verified that using the elbow method, which finds the optimal value of the number of clusters, k by a graphical visualisation-based technique [45]. In this method, variation is plotted against the number of clusters and the optimal value is picked as the elbow of the curve. Figure 3 shows the variation Vs number of clusters and the elbow of the curve was obtained at $k = 2$. Hierarchical Model, Gaussian Mixture Model (GMM) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) were also applied, but K-means gave a better performance in demarcating the datapoints.

## 3.3 Data pre-processing

We used a mutual information-based feature selection algorithm to determine the mental attention states of subjects based on the EEG signals acquired from their brains using a BCI device. A band pass filter is applied so that frequencies of range 1.5–50 Hz is included. A notch filter is applied for 50 Hz frequency to eliminate noise due to
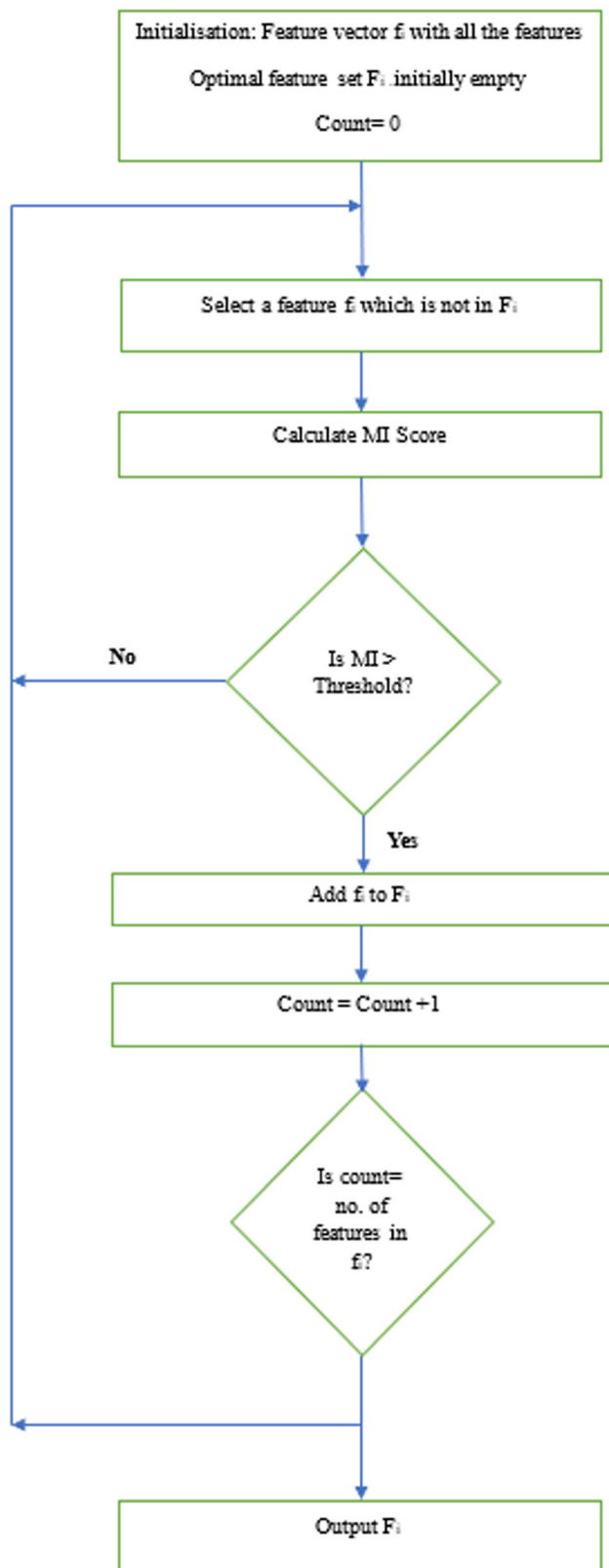
Fig. 3 Flowchart of the proposed algorithm

supply power. K-means clustering is a classical algorithm which is used to divide data into classes or clusters [44].

Our work combines feature–feature mutual information and feature-class mutual information so that an optimal subset of relevant features is obtained. MI-based feature selection was chosen for our work because of its effectiveness in evaluating multiple classifiers in multiple datasets. Hoque et al. evaluated MI-based feature selection method over 12 datasets having varying dimensionalities and the performance was compared with 8 other feature selection methods like Chi square and symmetric uncertainty [46]. EEG signals were classified using maximised mutual information by Aci et al. [11] for classification of emotions.

**Algorithm 1** Mutual Information-based Feature Selection

Input: $f_i$ feature vector, k no. of clusters
Output: Optimal Feature set $F_i$
Initialise count=0
Calculate MI according to equation 8.
MI previous= Initial MI
While count < i
      If MI > 1
            $f_i \in F_i$
      Else $f_i$ not an element of $F_i$
      End If
End While

For the dataset $D$, let $F$ be the feature set

$$F = \{f_1, f_2, f_3, \ldots, f_d\}$$

An optimal subset of the features $F'$ is to be selected such that.

$F' \subset F$.

For F', a classifier should give maximum classification efficiency.

The proposed method uses Mutual Information theory for finding out a feature subset of maximum relevance and minimum redundancy.

For any pair $(f_i, f_j), \in F'$

The formal definition of Mutual Information is given as follows:

Mutual Information

$$\mathrm{I}(f_i, f_j) = \frac{\frac{1}{F^2} \sum_{fi,fj \in F} p(f_i, f_j) \log p(f_i, f_j)}{p(f_i) p(f_j)} \tag{1}$$

where $p(f_i, f_j)$ is the joint probability density function of $f_i$ and $f_j$ and $p(f_i)$ and $p(f_j)$ are the marginal probability density functions for $f_i$ and $f_{jLiu}$.

The minimum redundancy condition

For the feature set $F$,

$$\min(F_i, F_j) = \frac{1}{F^2} \sum_{fi,fj \in F} I(f_i, f_j) \tag{2}$$

where $I(f_i, f_j)$ is the mutual information (MI) between the features $f_i$ and $f_j$.

The Maximum relevance condition

$$\frac{1}{F} \sum_{fi,fj \in F} I(C, f_i) \tag{3}$$

where $C$ defines the classes and $I(C, f_i)$ is the relevant feature $f_i$ for a class $C$.

The final set should satisfy both the conditions simultaneously.

$$I(F_i, F_j) = p(f_i f_j) \log p(f_i f_j) p(f_i) p(f_j) \tag{4}$$

Feature–feature mutual information is calculated for all the 14 features. The selected features should have this value lower than the threshold, which is set to be 1.

Feature-class mutual information should be high for each class C.

MI score is calculated by including both the conditions.

Calculation of mutual information involves computational difficulties and when higher dimensions are involved, it will become more and more complex. A solution to this problem is to use non-parametric entropy estimators to calculate Mutual information. Such an estimator is the Renyi's entropy. Renyi's entropy can be used to formulate mutual information between features and target classes. A transformation function is applied to feature–feature mutual information and feature-class mutual information and a discrete value is assigned for each feature.

Entropy is the uncertainty in a variable. A high value for entropy implies almost the same probability of occurrence for every event, whereas a low entropy implies widely different probability of occurrences. The information one random variable possess about another random variable is termed as the mutual information of those variables [60]. This measure is used in feature selection to quantify the relevance of a feature subset with respect to the target class. Discarding irrelevant features and selecting the relevant ones will help to reduce computation time. Mutual information-based feature selection used in this work corresponds to finding an optimal feature subset with minimum redundancy and maximum relevance (mRMR). It finds applications in various fields like machine learning, information theory and image processing, in feature extraction and clustering analysis.

Renyi's entropy is a generalised formulation for order α. Selecting $\alpha = 2$ makes the measure positive for all the values of $i$ and $j$. So, we use Renyi's quadratic formulation for measuring entropy.

Let μ be the median and σ be the standard deviation of a particular feature

$$f'(i) = 1 \; if \; i \geq \mu z + \sigma/2$$

$$0 \; if \; i < \mu - \sigma/2$$

The entropy or uncertainty of a class label is given as follows:

$$H(C) = -\Sigma P(c) \log P(c) \tag{5}$$

For a feature vector y, the uncertainty of the class identity is known as the conditional entropy [17].

$$H(C, Y) = yp(y)(cp(c/y) \log(p(c/y) \mathrm{d}y \tag{6}$$

The amount by which the class uncertainty is reduced, over the feature vector is called mutual information, which is given by:

$$I(C, Y) = H(C) - H(C/Y) \tag{7}$$

$$= \Sigma P(c) \log P(c) - yp(y)(cp(c/y) \log(p(c/y) \mathrm{d}y \, p(c,y)$$

$$= p(c/y) p(y) \, \mathrm{and} \, P(c) = yp(c,y) \mathrm{d}y \tag{8}$$

**Lemma 1** *The joint density of C and Y can be factored as product of marginal densities P(c) and p(y).*

**Proof of Lemma 1**

$$I(C, Y) = cyp(c,y) \log p(c,y)/P(c)p(y) \mathrm{d}y \tag{9}$$

When $C$ and $Y$ are independent of each other,

$$I(C, Y) = 0$$

$$w = \operatorname{argmax} w(I\{c_i, y_i\}) : y_i = w^T x_i \tag{10}$$

Here, $w^T w = I$

Columns of D X d matrix W span Rd.

For visualisation of the EEG data, we used OpenBCI GUI. OpenBCI has an in-built feature of saving the electrode readings in a text file with timestamp added for each entry. During data acquisition, if the recording option is turned on, a text file is formed in a folder named 'Recordings' in the OpenBCI folder of the Documents folder of a Windows PC, by default. To convert the data from time domain to frequency domain, fast Fourier transform (FFT) was used. The raw data obtained in this manner was converted into different frequency bands with the help of a Butterworth filter. FFT was applied to every signal.

i.e., $cyp(c,y) \log p(c,y)/P(c)p(y) \mathrm{d}y = 0$

which implies

$$\log[p(c,y)/P(c)p(y)] = 0$$

$$p(c,y) = P(c)p(y) \tag{11}$$

This means that joint density of $C$ and $Y$ can be factored as product of marginal densities $P(c)$ and $p(y)$.

To maximise the mutual information between the feature $y_i$ and class $C$, the partial derivative with respect to y needs to be equated to zero [30].

Let $J_p$ be the number of samples in a class $C_p$

$$ITyi = yi[(i) + (ii) + (iii)]$$

To find a linear transform with reduced dimensions, there should be a subspace Rd where $d < D$

## 3.4 Classification

Our work used Python Keras; an open-source deep learning framework to create a Sequential model [47]. Dense layer, which is a regular deeply connected neural network was used for classification of data. In deep learning, input is analysed layer-by-layer. 10 such layers were used. Each layer used ReLU activation and is followed by a dropout layer to handle overfitting and batch normalisation layer to normalise the batch with its mean and standard deviation. Sci-kit Learn's StandardScaler was used to scale the data. The model was optimised using Adamax optimiser and Keras binary cross entropy loss class was used to compute the cross entropy loss between true labels and predicted labels. The results are tabulated in Table 2.

## 3.5 Feature extraction

The section explains the different methods by which features were extracted. The methods used are statistical methods, Welch periodogram [48] and fast Fourier Transform (FFT). FFT was used to implement the separation of frequencies into different frequency bands. This method of feature extraction was applied on every electrode signal separately. With 10 signals multiplied by eight frequency bands, 80 features were obtained. With the feature extraction methods, a time-independent dataset was created. The dataset contains a file for each participant which consists of values at each electrode with timestamps. For benchmarking classification of data, machine learning algorithms [49, 50] logistic regression (LR), linear discriminant analysis (LDA), K-neighbours classifier (KNN), decision tree classifier (DT), Naïve Bayes (NB), and Support Vector Machine (SVM) are used.

## 3.6 Visualisation using OpenBCI

We used OpenBCI GUI, a powerful tool from OpenBCI for recording and visualising the EEG data from the OpenBCI Cyton Board. The GUI consists of mini tools called 'widgets' that fit into the interface panes. There are numerous widgets available in the tool, but the work used only four of them. The most important widget which displays the EEG data is the 'Time series' widget. It displays eight graphs, representing the voltage detected by the eight electrodes of the EEG acquisition device. The wires of the Ultracortex matches with the colour-codes of the GUI, so that it is easy to keep track of the electrode-channel mapping. If an electrode is providing poor signals, the GUI gives us a 'railed' warning, so that the electrodes can be checked for proper positioning and contact. Another visualisation feature is the 'FFT plot' widget. It displays frequencies on the x-axis and the corresponding amplitudes on the y-axis. This widget is also colour-coded to match the channels of the time series widget. The head plot widget shows the regions of brain with more activity in deep red colour, with intensity lowering with decreasing activity. It also contains the band power widget which displays the relative voltages of the different frequency sub-bands (Tables 3 and 4).

## 3.7 Novel contributions

The novel contributions of this paper are presented as under:

1. A Spherical Gaussian kernel-based quadratic entropy model for the binary classification of EEG.
2. A mutual information-based deep learning sequential network with 10 dense layers for the classification of EEG
3. Creation of an SSVEP dataset with educational and non-educational videos as visual stimulus
4. Validation of the entropy model by benchmarking the proposed model against traditional machine learning models.

**Table 2** Evaluation metrics

| Metric | Purpose | Description | Equations |
| --- | --- | --- | --- |
| Accuracy | Classification | The number of properly classified instances divided by the number of all instances | $\frac{TP+TN}{TP+FP+TN+FN}$ |
| Mean squared error | Prediction | The sum of squared differences between the predicted and actual values of the label | $\frac{\sum_{n=1}^{i}(yi-pi)^2}{n}$ |

**Table 3** Loss and accuracy

| Number of epochs | Loss | Accuracy | Validation loss | Validation accuracy |
|---|---|---|---|---|
| 1 | 85.96 | 56.74 | 73.33 | 53.92 |
| 10 | 56.55 | 71.61 | 45.56 | 77.67 |
| 25 | 31.67 | 85.85 | 18.90 | 93.35 |
| 50 | 17.76 | 93.07 | 06.66 | 98.26 |
| 75 | 9.19 | 96.63 | 3.30 | 99.29 |
| 100 | 6.43 | 97.83 | 2.80 | 99.52 |
| 125 | 4.02 | 98.56 | 3.03 | 99.52 |
| 150 | 2.97 | 98.90 | 3.04 | 99.64 |

**Table 4** Comparison of machine learning algorithms

| Sl. No | Algorithm | Accuracy % |
|---|---|---|
| 1 | Logistic regression | 97.32 |
| 2 | Linear discriminant analysis | 94.82 |
| 3 | KNN | 97.32 |
| 4 | Decision tree | 93.04 |
| 5 | Naïve Bayes | 93.04 |
| 6 | SVM | 96.07 |
| 7 | Proposed algorithm | 99.21 |



**Fig. 4** Elbow method to find the number of clusters



**Fig. 5** Sample welch periodogram

# 4 Results

In this section, we present the experimental results of the participants' EEG data. Since 10 volunteers participated in the experiment for 10 videos in 10 days, we obtained 100 EEG data samples. The performance of the deep learning-based classifier is evaluated using tenfold cross validation [51]. Cross validation is used as a standard method in machine learning for evaluating accuracy of classification and regression [52, 53]. In cross validation, a specified number of available data (1 for every 10, in this case) is kept away while training the machine leaning model so that the model will be unaware of that subset of data. After finishing the training, this unseen set of data is applied to the model and accuracy of the model is evaluated.
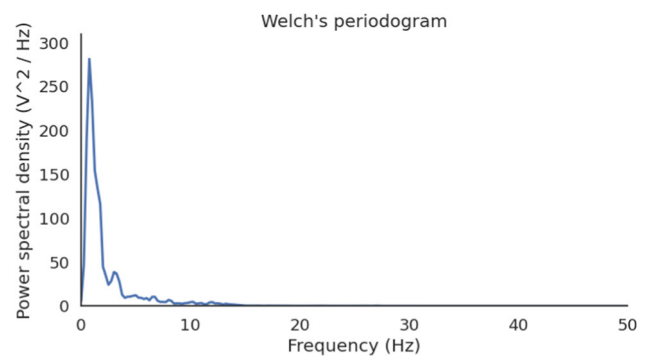
## 4.1 Evaluation metrics

The metrics used for classification and prediction are listed and described in the Table 2.

All the experiments conducted for this study were performed by healthy individuals. They were all informed about the objectives and procedures of the experiments conducted. Figure 9 shows the sample EEG data acquired using Ultracortex Mark IV equipment visualised with the help of OpenBCI GUI. The data so obtained were applied to a K-means clustering algorithm and it was revealed that there were only 2 target classes which could be identified.

Elbow method was used to obtain the number of classes, as shown in Fig. 3. To obtain the correlation between the features, Canonical Correlation Analysis was done. K-means clustering technique was used to cluster the data into two, attentive and inattentive. A mutual information-based feature selection is used after preparing a heatmap with Pearson correlation technique. The Fig. 4 Elbow method to find the number of clusters.

The clusters can be seen as $K = 2$ classifier, Naive Bayes Classifier, SVM, Decision designed system is
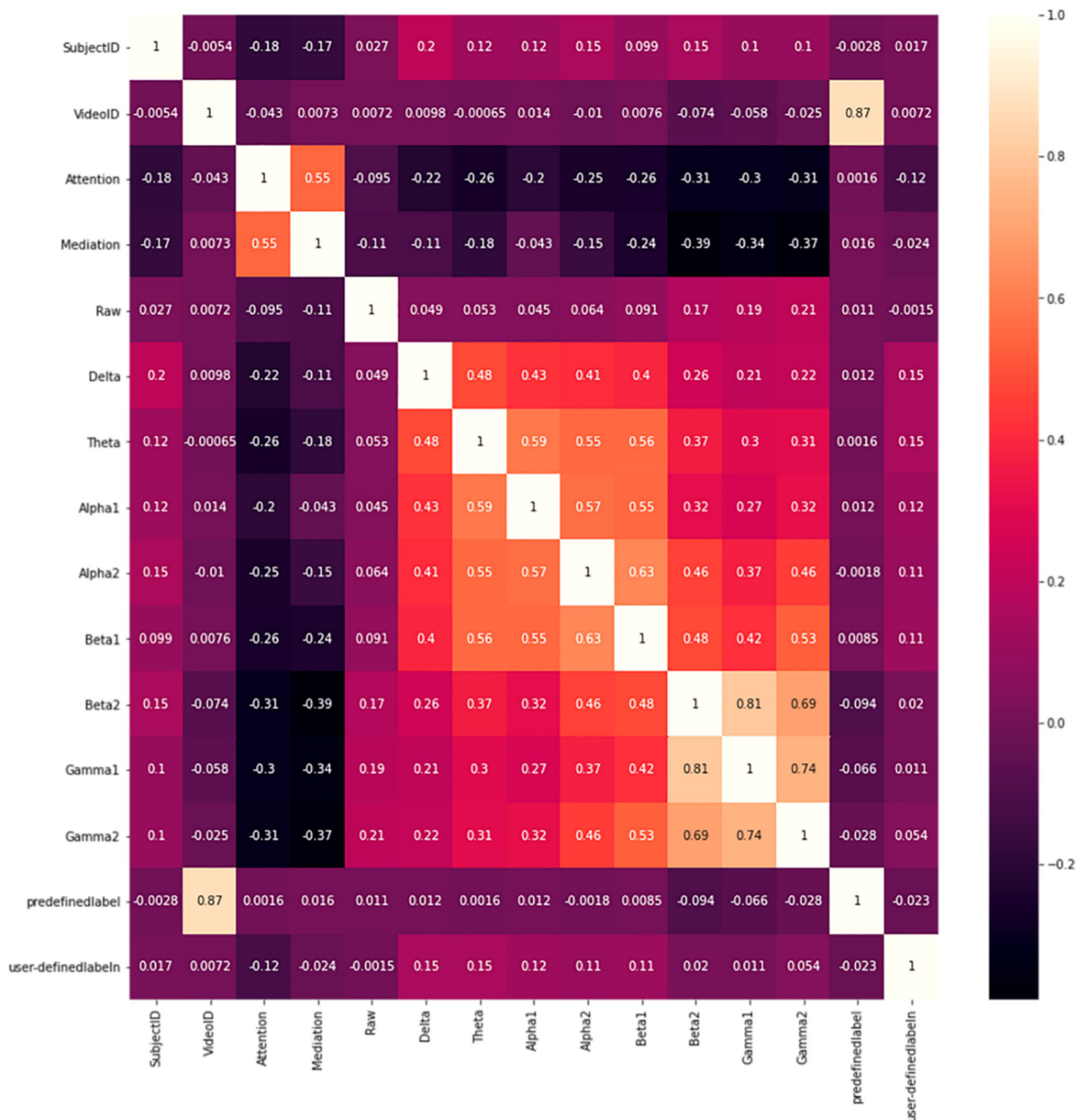
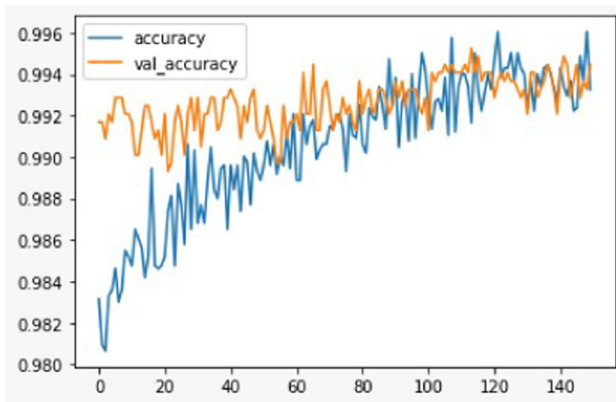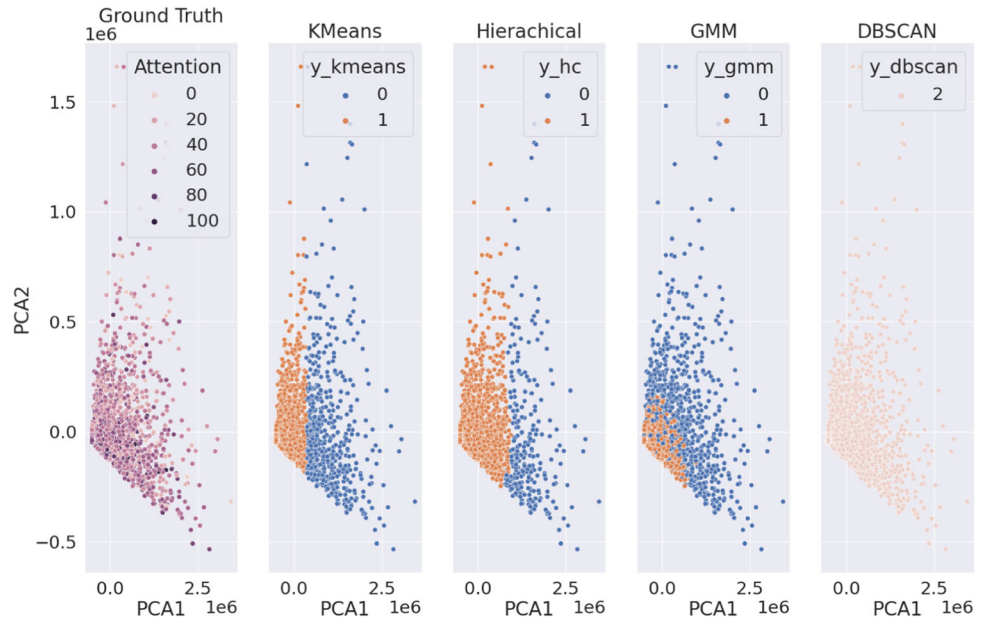**Fig. 6** Heatmap for calculating the correlation between features

compared with K-neighbours Tree classifier and logistic regression. K-fold validation with $K = 10$ was used to validate the results. The deep learning network is trained with the EEG characteristic of a few subjects. In the final stage, the EEG signals are classified as either 'attentive' or 'non-attentive'. Figure 4 shows the clustered data after applying K-means. The mutual information-based network was initially applied on the public dataset and the accuracy obtained was 99.21%. The dataset being labelled,

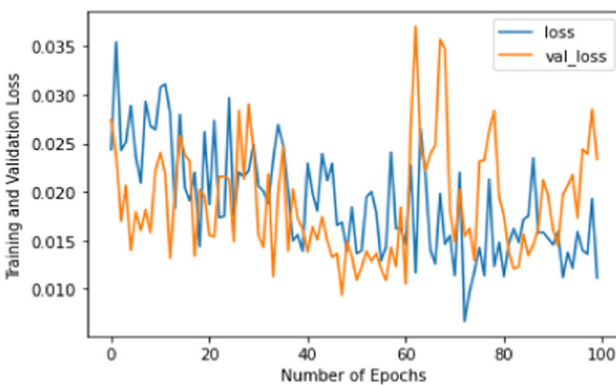traditional machine learning algorithms were also applied for classification.

Fig shows the ground truth against the clustering models K-means, Hierarchical, GMM and DBSCAN methods. It is obvious from the figure that the performance is better with K-means, compared to other models (Figs. 5 and 6).

At the first epoch, accuracy was obtained as 55.70% which kept on increasing up to 99.41% at epoch 150. K-fold validation was used with number of folds = 10.

**Fig. 7** Comparison of the different clustering models



Figure 8 shows the loss with training data and validation data plotted separately. A few representative values for different epochs are as shown in Table 2. It can be inferred that Accuracy is increasing and loss is decreasing with increasing number of epochs.



**Fig. 8** Accuracy and validation accuracy plots

## 5 Conclusion and future work

In the proposed work, we developed a Spherical Gaussian kernel-based quadratic entropy model for the binary classification of EEG. We collected a new dataset related to individuals who listened to a set of videos. The dataset consists of EEG data from 10 individuals watching 10 videos, which gives 100 datapoints for analysis. We demonstrated detection of attention levels with high accuracy which reached 99.81% (best) and 99.21% (average). The entropy model was validated using a public dataset. The proposed model was benchmarked against 6 other machine learning algorithms and our method outperformed them with respect to accuracy. The mutual information-based deep learning EEG model can be of use to detecting attention levels in students as well as detecting attention related disorders such as ADHD [55]. The mutual information based deep learning EEG model used in this work can be generalised for detection of attention levels in different circumstances. A few of such examples in which generalisation of this study can be used are detection of Alzheimer's disease and ADHD. In earlier works, human actions were detected, analysed and controlled using numerous methods [56] like exemplar-based methods [57], bag of visual words [58] and also BCI-based methods [59, 60].



**Fig. 9** Training loss and validation loss

After compiling the model, the accuracy and validation accuracy are plotted as shown in Figs. 7 and 8. Figure 7 shows the accuracy for training set and validation set.
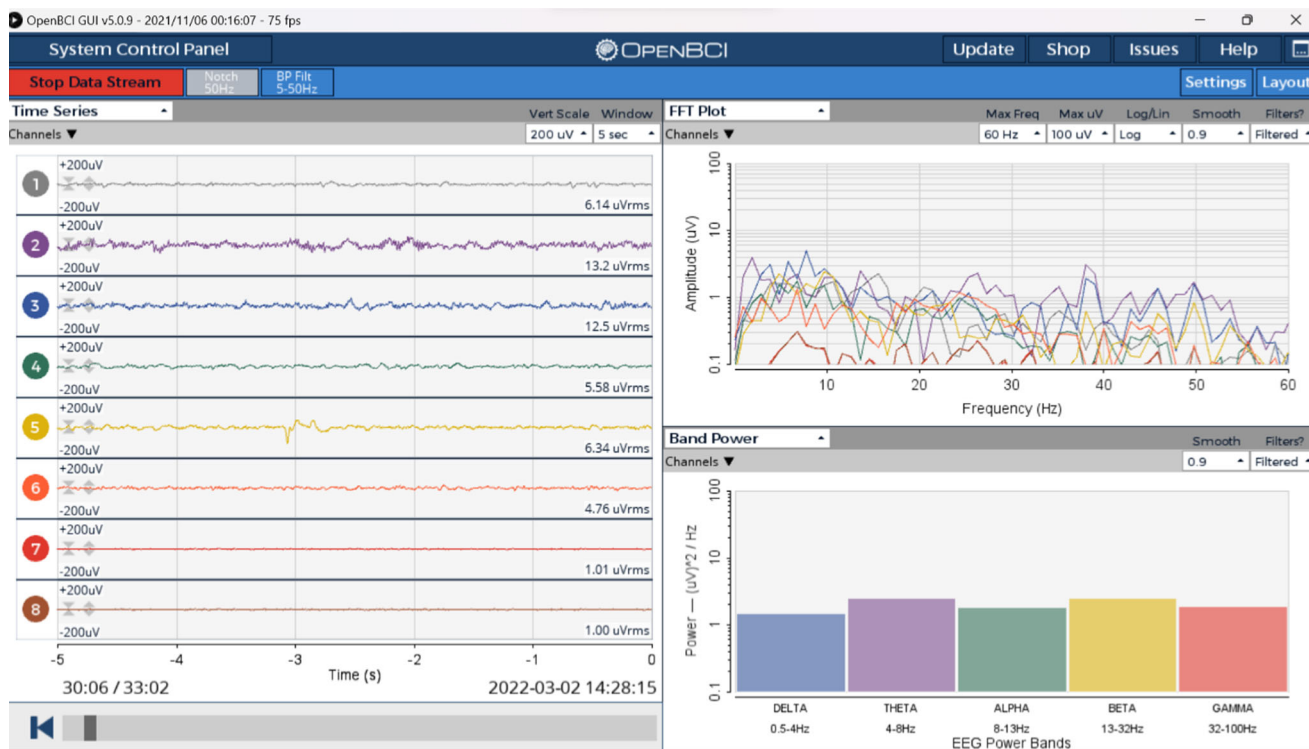
**Fig. 10** Sample screenshot of OpenBCI GUI. It shows the time series, FFT plot and band power from an EEG reading

Although BCI is a promising technology and this work has used it successfully in recognising attention with a good accuracy, it has a few demerits as well. One major demerit of the BCI is the inconsistency of brain output from person to person and from time to time. When the subjects were stressed or tired, more time was taken to get a steady output, before the video stimulus could be applied (Fig. 9). The recorded EEG data needed numerous steps for preprocessing to obtain good accuracy, so a real-time analysis could not be performed. But if the preprocessing steps can be done in a fast and efficient way, real-time attention monitoring can be done. The subjects put on the EEG cap for a few minutes only, including the time for obtaining stable signals and the time for which video stimuli were applied. Wearing the head cap for long time may cause inconvenience to the users (Fig. 10).

BCI is an interdisciplinary domain which involves research in biology, engineering, applied mathematics and computer science. HARS was devoted to detecting the attention of a human being accurately. The attention data can be used in BCI-based control of electronic devices ranging from a simple gaming vehicle to a complex electronic wheelchair. Incorporating BCI with Internet of Things (IoT) can help humans control devices in home or office with their brain signals. Currently, evoked potentials from local electrodes are used in BCI. In future, development of a comfortable thin layer of EEG harnessing equipment can be developed, which can decode more brain waves than those acquired from the electrodes. Further application of BCI in attention and action recognition gives hope to lot of physically compromised, but mentally active people to lead a better life.

**Data availability** Data generated during the work is available from the authors on reasonable request.

## Declarations

**Conflict of interest** The authors declare no conflict of interest that are relevant to the content of this article. The authors have no relevant financial or non-financial interests to disclose. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Goldberg P, Sümer Ö, Stürmer K, Wagner W, Göllner R, Gerjets P, Trautwein U (2021) Attentive or not? toward a machine

learning approach to assessing students' visible engagement in classroom instruction. Educ Psychol Rev 33:27–49. https://doi.org/10.1007/s10648-019-09514-z

2. Zaletelj J, Košir A (2017) Predicting students' attention in the classroom from Kinect facial and body features. EURASIP J Image Video Process 2017:1–12. https://doi.org/10.1186/s13640-017-0228-8

3. Zhang X, Wu CW, Fournier-Viger P, Van LD, Tseng YC (2017) Analysing students' attention in class using wearable devices. In: 2017 IEEE 18th international symposium on a world of wireless, mobile and multimedia networks (WoWMoM), pp 1–9. https://doi.org/10.1109/WoWMoM.2017.7974306

4. Carini RM, Kuh GD, Klein SP (2006) Student engagement and student learning: testing the linkages. Res Higher Educ 47:1–32. https://doi.org/10.1007/s11162-005-8150-9

5. Sezer A, İnel Y, Seçkin AÇ, Uluçınar U (2015) An investigation of university students' attention levels in real classroom settings with NeuroSky's MindWave mobile (EEG) device. In: International Educational Technology Conference–IETC 2015

6. Hassan R, Hasan S, Hasan MJ, Jamader MR, Eisenberg D, Pias T (2020) Human attention recognition with machine learning from brain-EEG signals. In: 2020 IEEE 2nd Eurasia conference on biomedical engineering, healthcare and sustainability (ECBIOS), 16–19

7. Chen He, Song Y, Li X (2019) A deep learning framework for identifying children with ADHD using an EEG-based brain network. Neurocomputing 356:83–96

8. Moghaddari M, Lighvan MZ, Danishvar S (2020) Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG. Comput Methods Progr Biomed 197:105738

9. Cho BH et al (2002) Attention enhancement system using virtual reality and EEG biofeedback. In: Proceedings IEEE virtual reality 2002, pp 156–163. https://doi.org/10.1109/VR.2002.996518

10. Ghasemy H, Momtazpour M, Sardouie SH (2019) Detection of sustained auditory attention in students with visual impairment. In: 2019 27th Iranian conference on electrical engineering (ICEE), pp 1798–1801

11. Acı Çİ, Kaya M, Mishchenko Y (2019) Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods. Exp Syst Appl 134:153–166. https://doi.org/10.1016/j.eswa.2019.05.057

12. Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A (2021) The state of the art of deep learning models in medical science and their challenges. Multimed Syst 27(4):599–613. https://doi.org/10.1007/s00530-020-00694-1

13. Xiao C, Choi E, Sun J (2018) Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inf Assoc 25(10):1419–1428. https://doi.org/10.1093/jamia/ocy068

14. Mirzaei G, Adeli H (2022) Machine learning techniques for diagnosis of Alzheimer disease, mild cognitive disorder, and other types of dementia. Biomed Signal Process Control 72:103293

15. Sarkar A, Singh A, Chakraborty R (2022) A deep learning-based comparative study to track mental depression from EEG data. Neurosci Inf 2(4):100039. https://doi.org/10.1016/j.neuri.2022.100039

16. Safayari A, Bolhasani H (2021) Depression diagnosis by deep learning using EEG signals: a systematic review. Med Novel Technol Dev 12:100102. https://doi.org/10.1016/j.medntd.2021.100102

17. Toa CK, Sim KS, Tan SC (2021) Electroencephalogram-based attention level classification using convolution attention memory neural network. IEEE Access 9:58870–58881

18. Farhan M, Jabbar S, Aslam M, Hammoudeh M, Ahmad M, Khalid S, Han K (2018) IoT-based students interaction framework using attention-scoring assessment in eLearning. Future Gener Comput Syst 79:909–919. https://doi.org/10.1016/j.future.2017.09.037

19. Renawi A, Alnajjar F, Parambil M, Trabelsi Z, Gochoo M, Khalid S, Mubin O (2022) A simplified real-time camera-based attention assessment system for classrooms: pilot study. Educ Inf Technol 1–18

20. Villa M, Gofman M, Mitra S, Almadan A, Krishnan A, Rattani A (2020) A survey of biometric and machine learning methods for tracking students' attention and engagement. In: 19th IEEE international conference on machine learning and applications (ICMLA), 2020, pp 948–955. https://doi.org/10.1109/ICMLA51294.2020.00154

21. Li Y et al (2011) A real-time EEG-based BCI system for attention recognition in ubiquitous environment. In: Proceedings of 2011 international workshop on Ubiquitous affective awareness and intelligent interaction

22. Djamal EC, Dewi PP, Dea AD (2016) EEG-based recognition of attention state using wavelet and support vector machine. In: 2016 International seminar on intelligent technology and its applications (ISITIA). IEEE

23. Liu NH, Chiang CY, Chu HC (2013) Recognizing the degree of human attention using EEG signals from mobile sensors. Sensors 13(8):10273–10286

24. Parui S, Basu D, Mansoor W, Ghosh U (2021) Artificial intelligence induced multi-level attention states recognition from brain using EEG signal. In: 2021 4th international conference on signal processing and information security (ICSPIS), pp 1–4. https://doi.org/10.1109/ICSPIS53734.2021.9652419

25. Ravi A, Beni NH, Manuel J, Jiang N (2020) Comparing user-dependent and user-independent training of CNN for SSVEP BCI. J Neural Eng 17(2):026028

26. Wang Y, Wang YT, Jung TP (2010) Visual stimulus design for high-rate SSVEP BCI. Electron Lett 46(15):1

27. Volosyak I, Gembler F, Stawicki P (2017) Age-related differences in SSVEP-based BCI performance. Neurocomputing 250:57–64

28. Ortner R, Allison BZ, Korisek G, Gaggl H, Pfurtscheller G (2010) An SSVEP BCI to control a hand orthosis for persons with tetraplegia. IEEE Transact Neural Syst Rehabil Eng 19(1):1–5. https://doi.org/10.1109/TNSRE.2010.2076364

29. Chen X, Zhao B, Wang Y, Gao X (2019) Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm. J Neural Eng 16(2):026012

30. Müller SMT, Bastos-Filho TF, Sarcinelli-Filho M (2011) Using a SSVEP-BCI to command a robotic wheelchair. In: 2011 IEEE international symposium on industrial electronics, Gdansk, Poland, 2011, pp 957–962. https://doi.org/10.1109/ISIE.2011.5984288

31. Na R, Hu C, Sun Y, Wang S, Zhang S, Han M, Zheng D (2021) An embedded lightweight SSVEP-BCI electric wheelchair with hybrid stimulator. Digit Signal Process 116:103101. https://doi.org/10.1016/j.dsp.2021.103101

32. Horki P, Solis-Escalante T, Neuper C, Müller-Putz G (2011) Combined motor imagery and SSVEP based BCI control of a 2 DoF artificial upper limb. Med Biol Eng Comput 49:567–577. https://doi.org/10.1007/s11517-011-0750-2

33. Lin K, Cinetto A, Wang Y, Chen X, Gao S, Gao X (2016) An online hybrid BCI system based on SSVEP and EMG. J Neural Eng 13(2):026020

34. Putze F, Weiß D, Vortmann LM, Schultz T (2019) Augmented reality interface for smart home control using SSVEP-BCI and eye gaze. In: 2019 IEEE international conference on systems,

man and cybernetics (smc) (pp 2812–2817). IEEE. https://doi.org/10.1109/SMC.2019.8914390

35. Pfurtscheller G, Solis-Escalante T, Ortner R, Linortner P, Muller-Putz GR (2010) Self-paced operation of an SSVEP-based orthosis with and without an imagery-based brain switch: a feasibility study towards a hybrid BCI. IEEE Transact Neural Syst Rehabil Eng 18(4):409–414. https://doi.org/10.1109/TNSRE.2010.2040837

36. Allison BZ, Brunner C, Altstätter C, Wagner IC, Grissmann S, Neuper C (2012) A hybrid ERD/SSVEP BCI for continuous simultaneous two dimensional cursor control. J Neurosci Methods 209(2):299–307. https://doi.org/10.1016/j.jneumeth.2012.06.022

37. Pfurtscheller G, Allison BZ, Bauernfeind G, Brunner C, Solis Escalante T, Scherer R, Zander TO, Mueller-Putz G, Neuper C, Birbaumer N (2010) The hybrid BCI. Frontiers in neuroscience, p 3

38. Yin E, Zhou Z, Jiang J, Yu Y, Hu D (2014) A dynamically optimized SSVEP brain–computer interface (BCI) speller. IEEE Transact Biomed Eng 62(6):1447–1456. https://doi.org/10.1109/TBME.2014.2320948

39. Liu B, Chen X, Shi N, Wang Y, Gao S, Gao X (2021) Improving the performance of individually calibrated SSVEP-BCI by task-discriminant component analysis. IEEE Transact Neural Syst Rehabil Eng 29:1998–2007. https://doi.org/10.1109/TNSRE.2021.3114340

40. Wang H, Li Y, Hu X, Yang Y, Meng Z, Chang KM (2013) Using EEG to improve massive open online courses feedback interaction. In: AIED workshops

41. Homan W (1988) The 10–20 electrode system and cerebral location. Am J EEG Technol 28(4):269–279. https://doi.org/10.1080/00029238.1988.11080272

42. Herwig U, Satrapi P, Schönfeldt-Lecuona C (2003) Using the International 10–20 EEG system for positioning of transcranial magnetic stimulation. Brain Topogr 16:95–99. https://doi.org/10.1023/B:BRAT.0000006333.93597.9d

43. Lopez-Gordo MA, Sanchez-Morillo D, Valle FP (2014) Dry EEG electrodes. Sensors 14(7):12847–12870

44. Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Vaughan TM (2000) Brain-computer interface technology: a review of the first international meeting. IEEE Transact Rehabil Eng 8(2):164–173

45. Chong B (2021) K-means clustering algorithm: a brief review. Acad J Comput Inf Sci 4:37–40

46. Hoque N, Bhattacharyya DK, Kalita JK (2014) MIFS-ND: a mutual information-based feature selection method. Exp Syst Appl 41(14):6371–6385. https://doi.org/10.1016/j.eswa.2014.04.019

47. Dürr O, Sick B, Murina E (2020) Probabilistic deep learning: with python, keras and tensorflow probability. Manning Publications

48. Sun Z et al (2019) Mutual information based multi-label feature selection via constrained convex optimization. Neurocomputing 329:447–456

49. Gupta A, Agrawal RK, Kirar JS, Andreu-Perez J, Ding WP, Lin CT, Prasad M (2019) On the utility of power spectral techniques with feature selection techniques for effective mental task classification in noninvasive BCI. IEEE Transact Syst Man Cybern Syst 51(5):3080–3092. https://doi.org/10.1109/TSMC.2019.2917599

50. Lashgari E, Liang D, Maoz U (2020) Data augmentation for deep-learning-based electroencephalography. J Neurosci Methods 346:108885. https://doi.org/10.1016/j.jneumeth.2020.108885

51. Kumari P, Deb S (2020) EEG cross validation of effective mobile technology by analyzing attention level in classroom. In: 2020 International conference on electronics and sustainable communication systems (ICESC), 2020, pp 961–965. https://doi.org/10.1109/ICESC48915.2020.9155588

52. Ghosh SM, Bandyopadhyay S, Mitra D (2021) Nonlinear classification of emotion from EEG signal based on maximized mutual information. Exp Syst Appl 185:115605

53. Hosseini MP, Hosseini A, Ahi K (2020) A review on machine learning for EEG signal processing in bioengineering. IEEE Rev Biomed Eng 14:204–218. https://doi.org/10.1109/RBME.2020.2969915

54 Koprinska I (2009) Feature selection for brain-computer interfaces Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, Heidelberg

55. Gao Z, Sun X, Liu M, Dang W, Ma C, Chen G (2021) Attention-based parallel multiscale convolutional neural network for visual evoked potentials EEG classification. IEEE J Biomed Health Inf 25(8):2887–2894. https://doi.org/10.1109/JBHI.2021.3059686)

56. Megalingam RK (2021) Human action recognition: a review. In: 2021 10th international conference on system modelling & advancement in research trends (SMART), MORADABAD, India, 2021, pp 249–252. https://doi.org/10.1109/SMART52563.2021.9676211

57. Nair SAL, Megalingam RK (2022) Fusion of bag of visual words with neural network for human action recognition. In: 2022 12th International conference on cloud computing, data science & engineering (Confluence). IEEE

58. Lin J-S, Jiang Z-Y (2017) An EEG-based BCI system to facial action recognition. Wireless Pers Commun 94:1579–1593

59. Megalingam RK, Thulasi AA, Krishna RR, Venkata MK, BV AG, Dutt TU (2013) Thought controlled wheelchair using EEG acquisition device. In: 3rd International conference on advancements in electronics and power engineering, pp 207–212

60. Vergara JR, Estévez PA (2014) A review of feature selection methods based on mutual information. Neural Comput Appl 24:175–186. https://doi.org/10.1007/s00521-013-1368-0