Biomedical Signal Processing and Control 62 (2020) 102094

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Human activity recognition using deep electroencephalography learning

Amirsaleh Salehzadeh*, Andre P. Calitz, Jean Greyling

Department of Computing Sciences, Nelson Mandela University, Port Elizabeth, 6019, South Africa

ARTICLE INFO

Article history:
Received 18 October 2019
Received in revised form 18 June 2020
Accepted 19 July 2020
Available online 26 June 2020

Keywords:
EEG artifact classification
Deep learning
Human activity recognition
EEG wearable device

ABSTRACT

Electroencephalography (EEG) signals can be contaminated by the noise raised from a non-cerebral artifact and vary in magnitude due to physiological differences between individuals. Hence, EEG has not been extensively applied to the design of Human Activity Recognition (HAR) systems. HAR involves classifying the activities of an individual that are captured by sensory technologies. To address this issue, this paper introduces a deep learning-based framework for classifying EEG artifacts (FCEA), based on a person's physiological activity. The FCEA proposes an approach for designing a processing pipeline involving a Convolutional Neural Network and a Long Short-Term Memory Recurrent Neural Network, in order to classify raw EEG signals based on the EEG artifacts. To demonstrate the performance of the FCEA, a 3-class dataset of EEG activities was created. A consumer-grade EEG wearable device was used to collect the data from two prefrontal EEG channels, from 8 participants whilst speaking loudly, reading printed text and watching a TV program. These activities include jaw-clenching and head and eye movements, that are part of a wide range of daily human activities. Moreover, these activities cannot easily be detected by using the sensory technologies that have commonly been used in HAR. The comparative performance analysis results demonstrate that the 1- and 2-channel models trained by the FCEA outperform competitive state-of-the-art deep-learning models for HAR and raw EEG data. The deep-learning approach proposed by the FCEA improves raw data processing to obtain a better generalization performance.

1. Introduction

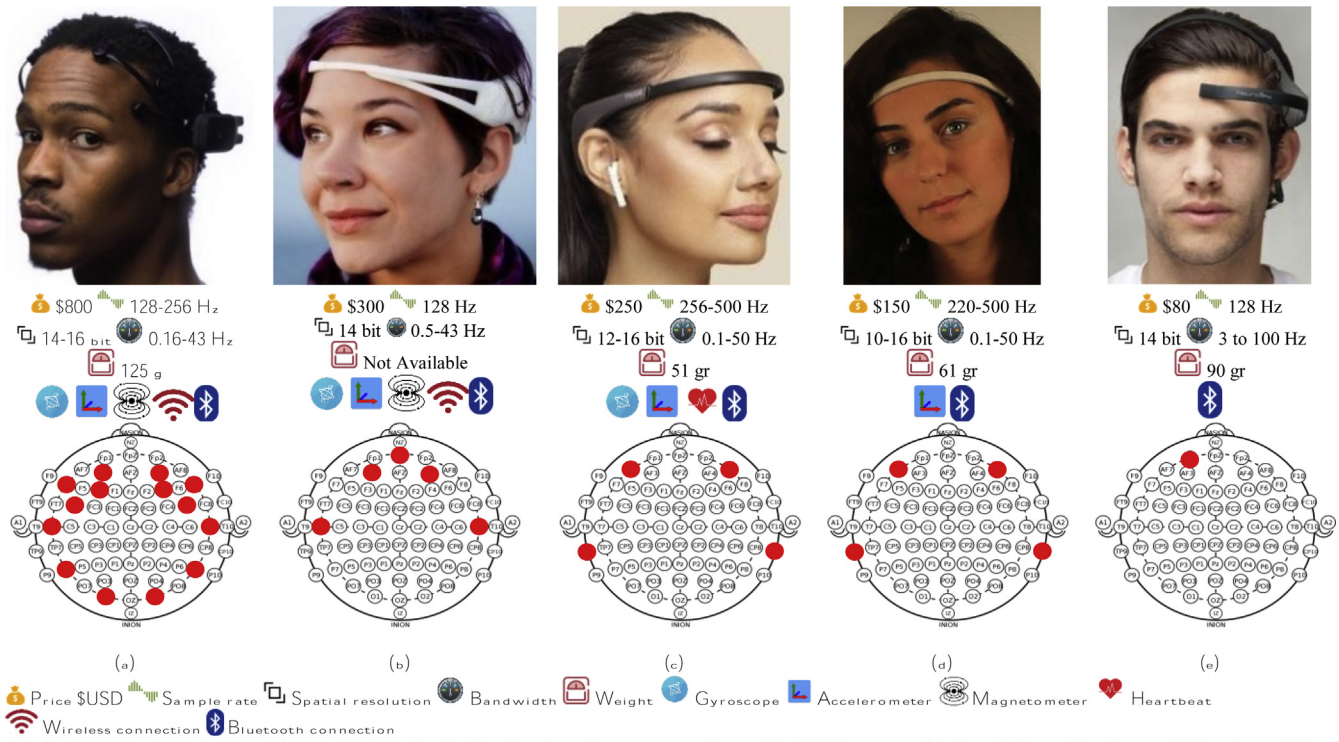
Cameras, accelerometers, gyroscopes and sound sensors have been used to detect the movement of users in automatic Human Activity Recognition (HAR) systems [1–4]. In recent years, various biosensors, such as Electromyography (EMG) [5–8], Electrooculography (EOG) [9] and Electroencephalography (EEG), have been used to detect user activities in HAR systems. However, EEG-based HAR systems function differently and can detect motor imagery activities in the brain [10–13]. The use of wireless wearable biosensors in HAR has led Human-Computer Interaction (HCI) researchers to develop more natural user interfaces with a greater degree of accessibility [7].

Non-invasive EEG electrodes placed on the scalp of a user are used to capture the electric fields generated by the activity of a group of neurons in the neighborhood of the electrodes. Fig. 1 provides details of the consumer-grade EEG wireless wearable devices, which are available on the market today, indicating their user-friendly designs. They are economical brain-sensing devices, which have been used in healthcare, HCI and for medical research purposes, however, they have not been commonly applied to HAR system design. The price, sampling rate, spatial resolution, bandwidth, weight, number of channels, connectivity method and other sensors that are used in the devices are also shown in Fig. 1. The devices support prefrontal EEG, which has the potential to be affected by the activities of the human face muscles. The EEG of the prefrontal lobe demonstrates the intellectual activities that are associated with this area, such as cognitive [14] and emotional control [15].

In HAR system design, Machine Learning (ML) algorithms have made significant progress in the extraction of human activity features from sensor data. The conventional ML algorithms that have been utilized in modeling sensor data, rely on feature extraction and data preprocessing [1,2]. In addition, while preprocessing biosignals, it is often necessary to cancel or reduce the noise artifacts [16–19]. In recent years, however, the emergence of deep-learning (DL) techniques addressed the limitations of the conventional ML algorithms. DL algorithms enable the inferring of complex nonlinear and hidden relationships in raw data from sensors by aligning signal features to a more separable space. DL algorithms have

* Corresponding author.
E-mail address: Amirsaleh@Mandela.ac.za (A. Salehzadeh).

https://doi.org/10.1016/j.bspc.2020.102094
1746-8094/



2. An overview of the sensing and ML technologies that have been commonly used in HAR are provided, and the applications of DL to knowledge

Fig. 1. Popular consumer-grade EEG headsets and the EEG channels of the devices based on the location of intermediate 10 % EEG electrodes standard [24]; (a) Emotiv Epoc+ [25]; (b) Emotiv Insight [25]; (c) Muse 2 (2018) InteraXon [26]; (d) Muse InteraXon (2014) [27]; (e) Neurosky Mindwave [28].

been incorporated into DL models. The training of DL models still requires a considerable amount of computing power [1–3,20,21].

Convolutional Neural Networks (CNNs) are DL algorithms, which are based on the Convolution Theorem, states that the convolution of two signals is the pointwise product of their transformations, i.e. the convolution in the time domain equals point-wise multiplication of the signals in the frequency domains, and vice versa [21]. In the convolutional layers of a CNN, several trainable kernel windows slides over the input data and each of the kernels populates one output feature map by operating the convolution process. A Long Short-Term Memory (LSTM) network is a type of Recurrent Neural Network (RNN) that enables the modeling of the long temporal dependencies in data and does not face the vanishing gradient problem that occurs when using traditional RNNs [22]. The CNN and LSTM networks have been used in the state-of-the-art models of human activities by using raw accelerometer and gyroscope data [1,2]. However, CNNs discover deep dependencies in both the spatial and temporal dimensions of data, and therefore, attain a generalization performance of HAR by using biosignals [10–12] and video data [23].

Roy, et al. (2019) examined literature relevant to the recent applications of DL for the processing of raw EEG data. They identified trends, approaches and remaining challenges in using these technologies. DL has been generally utilized in processing, augmenting and classifying EEG data. According to the Roy, et al. (2019), more than half of the studies used publicly available datasets and trained the models for raw or preprocessed EEG data. Furthermore, 7% of the studies used only a combination of CNNs and RNNs, whereas, 40 % used only CNNs and 14 % used only RNNs. They highlighted that “most studies often suffer from poor reproducibility due to the unavailability of the data and code” [13].

In this paper, a framework for training a classification model for EEG artifacts (FCEA), that are generated by facial activities, is presented. The FCEA shows the procedures of implementing a stem-to-stern EEG data-processing pipeline. A 3-class EEG artifact dataset was created for the purpose of conducting and evaluating an experiment using 8 participants. Data in the dataset were recorded from two prefrontal EEG channels (AF7-AF8) of a Muse headband (Fig. 1.d). The dataset classes were generated by recording three daily activities of participants whilst reading printed text, speaking out loudly and watching a TV program. A comparative analysis was performed to evaluate the performance of 1- and 2-channel EEG models, which were trained by following FCEA and some state-of-the-art DL approaches to classifying EEG and HAR. Thorough research on the relevant literature yielded no related articles that reported a semi-similar approach or framework to the FCEA.

The theoretical and practical contributions of the research study presented in this paper are as follows:

1 The results show that the EEG artifact classification, by using a combination of CNN and LSTM, provides a considerable improvement of generalization performance over CNN-based and LSTM-based models. It also reduces the number of trainable parameters in the network and training time.
2 An overview of the sensing and ML technologies that have been commonly used in HAR are provided, and the applications of DL to knowledge extraction from EEG is presented in this paper. The paper aims to express the significance and the novelty of the proposed approach by illustrating how interpreting data from a consumer-grade EEG in HAR would benefit the scientific community.
3 The activities investigated in this study are representative of a wide range of activities in the daily lives of humans. These activities create EEG artifacts of eye and head movement whilst reading text and jaw clenching whilst speaking. The HAR systems reported in the literature most often use ambient, smartphone or wrist-worn sensory devices to detect user movements. These technologies restrict the user movement in the environment,

they are expensive and do not provide data about the facial activities.

The paper is organized as follows. Section 2 presents related work and Section 3 shows components of the proposed framework. Section 4 discusses the methodology that was followed to evaluate the performance of the framework. The results are presented in Sections 5 and 6 discusses the way this research study gives insight into new discoveries and provides recommendations for future work. The conclusion of this article is presented in Section 7.

2. Related work

This section presents research which applies ML to interpret sensor data. Section 2.1 discusses common approaches followed in HAR and Section 2.2, specifically shows the application of DL to analyze and interpret EEG data.

2.1. Human activity recognition approaches

Human activities entail performing a sequence of actions that include short muscular movements [20]. The technologies used for detecting human actions are commonly based on heterogeneous sensors placed throughout the environment, such as cameras and wearable sensors equipped with accelerometers, gyroscopes and biosensors [1–4].

Cameras provide video data in form of a 3D matrix. Two vectors of the matrix represent the spatial dimensions for the pixel matrix and the third dimension represents the time domain. Each pixel of an image often includes extra dimensions indicating the channel of the image (i.e. RGB, depth or thermal density information). In contrast, the accelerometer and gyroscope data have lower dimensionality. Spatial values of accelerometer data are often given by the acceleration of gravity on Earth (9.81 m/s^2) and gyroscope data are commonly given by angular velocity per second.

The related state-of-the-art work across HAR which used cameras, accelerometers and gyroscopes in measuring metrics indicating human actions is presented in Table 1. In addition, the approach followed for training a data-classification model in each of the studies is shown in Table 1. The CNNs are more appropriate for learning and interpreting camera and motion-detection sensor data. The LSTM networks are not capable of classifying video data alone. They, however, attain a desirable classification performance when they are used for modeling accelerometer and gyroscope raw data, as well as the feature maps populated by a CNN [2].

Using cameras restricts user movements in an environment to an area within the camera's angle of view. In addition, it requires the light in the environment to be within a certain range of density. Wireless wearable devices give more freedom of movement to users [7]. Moreover, the smartphone and smartwatches which have been mostly used in HAR, are not suitable for recognizing facial activities. To address this issue, Farooq (2018) proposed a smart-goggles prototype equipped with accelerometer sensors. Farooq classified food intake based on the chewing pattern of users from accelerometer data [31].

Biosensor-based technologies enable a greater level of accessibility in HCI, in comparison with vision- and sensor-based HAR [7]. The research studies in the field of HAR that reported the utilization of EMG, Electrocardiography (ECG) and EOG for measuring human actions are presented in Table 2. The research reported in Table 2, indicates the lack of the generalization of the models. The biosignals are non-stationary and vary in the spatial dimension due to the location of sensors on the skin, size of the muscles, amount of body fat under the skin and other factors. However, using DL has addressed these issues and has enabled Côté-Allard, et al. (2018)

Table 1
Sensory and ML technologies used for HAR.

Metrics	Year	Activity	Sensor type	Modeling algorithm	Accuracy	Ref.
Hand-finger movements	2018	3 different hand gesture datasets	Depth-aware camera	CNN + Fully connected dense layers	Above 90 % for all of the datasets	[23]
Hand-finger RGB images	2017	9 wrist gestures	RGB camera	CNN + LSTM	Mean = 80 %	[29]
3D head acceleration	2016	Head movement directions of surgeon	Accelerometer of a Muse headband (Fig. 1.d)	A static model for head movement speed	---------	[30]
Wrist acceleration and orientation	2018	Walking, climbing stairs, descending stairs, sitting, standing, and lying down	Accelerometer and gyroscope of smartphone	LSTM	Mean = 90.97 %	[1]
Wrist acceleration and orientation	2018		Accelerometer and gyroscope of smartphone	Processed data + CNN	Mean = 97.63 %	[2]
The vibrations of the head generated by chewing	2018	Food intake	Accelerometer attached to a smart glasses prototype	k-nearest neighbor	Mean = 87.9 %	[31]

Table 2
Biosensor and ML technologies used for HAR.

Activity	Year	Sensor type	Metrics	Modeling algorithm	Accuracy	Ref.
Sitting, lying, go boating, bicycle riding	2014	ECG and SPO_2 electrodes placed on the body	Heart rate and oxygen level in the blood	Coupled hidden Markov model	Mean = 95 %	[32]
Reading, writing, and resting	2018	Dried biosensors placed around the eyes	EOG activities generated by eye movements	Support Vector Machine	Mean = 88 %	[9]
Smiling, frowning, and neutral face expressions	2014	EMG sensors attached to a wearable headset prototype	EMG activities generated by the facial expressions	Independent component analysis + Fully connected dense layers	Above 90 % for all of the emotional states	[33]
6 hand-finger gestures	2018	EMG sensors of a commercial armband	EMG activities generated by the forearm muscles when moving the hand fingers	Fully connected dense layers	Mean = 95 %	[7]
Taking 5 foods	2016	EMG sensors attached to a smart glasses prototype	EMG activities generated by clenching from the face muscles	Random Forest and Linear Discriminant Analysis	Mean = [63 %–84 %] for 8 participants	[6]

to attain a classification accuracy of 98 % when classifying 7 hand-wrist gestures, captured by EMG sensors placed on the forearms of participants. Côté-Allard, et al. (2018) used a deep CNN model and evaluated its performance by comparing subjects [34].

2.2. Deep EEG learning

According to Roy, et al. (2019), when dealing with the processing of EEG signals, researchers utilized DL for three general purposes, namely classifying, augmenting and/or preprocessing EEG data [13]. Applying DL to extract features from raw EEG data before classification reduces the amount of effort required for the preprocessing of the data.

Zhang, et al. (2018) used an Emotiv Epoch headset (Fig. 1.a) to capture EEG signals from 7 participants while they were imagining the upward arrow, downward arrow, leftward arrow, rightward arrow and cycle actions. They also noted that recording the time when the subjects stayed relaxed with their eyes closed. They attained an average classification accuracy of 93 % by using a combination of a Duelling-Q-Network [35] and an LSTM network. They trained and used a model for designing a Brain-Computer Interface (BCI) for controlling devices in a smart environment and in a direct brain text-typing system [12]. Lawhern, et al. (2018) evaluated the performance in the classification and generalizability of the EEGNet [10], Shallow ConvNet and Deep ConvNet [11], the state-of-the-art CNN-based architectures for raw EEG data. When classifying 4 publicly available BCI datasets, they removed the noise from the raw EEG data before training the models, and therefore, the performances they observed were not due to artifact or noise sources in the data [10].

Sundhara Kumar, et al. (2019) designed a BCI for mapping EEG response to emotional arousal in every scene of a horror and comic movie by classifying the EEG signal from a Neurosky Mindwave headset (Fig. 1.e). They used a fully connected deep neural network with 4 layers to measure high and low power bands from the EEG signals. They collected the data from 50 participants and obtained an inter-subject prediction accuracy of 85 % [36]. Alhagry, et al. (2017) proposed an LSTM network to classify the EEG power bands extracted from data in the Dataset for Emotion Analysis by using EEG, Physiological and video signals datasets. They attained an average classification accuracy of 85.65 %, 85.45 %, and 87.99 % with arousal, valence and liking emotional states, respectively [15].

DL has also been applied to augmenting EEG signals and thus improving classification performance when a small training dataset or limited data are available. Zhang and Liu (2018) proposed a processing pipeline involving a conditional Deep Convolutional Generative Adversarial Network, for generating artificial EEG signals, automatically. To verify the quality of the artificial data, they evaluated the performance of two CNN models. One model was trained with motor-imagery EEG data in an open dataset and the other model with artificial data. The results demonstrated an improvement of accuracy from 83 % to 86 % when using the artificial data [37]. Schwabedal, et al. (2019) targeted the problem of class imbalance of under-represented sleep stages, by extracting Fourier Transform features from raw EEG data in an open source dataset. They illustrated up to 24 % improvement in classification accuracy for some classes [38].

DL and other ML techniques have also been utilized in the handling of EEG artifacts. Yang, et al. (2016) used EOG signals and trained an Autoencoder Network model for removing the noise generated by eye movement from EEG signals. Autoencoder Networks are highly effective in learning particularly statistical moment-information, automatically. Therefore, they have been successfully used in training an EEG model for noise reduction and cancellation. The authors also used the model to automatically remove the noise from EEG signals without the interference of

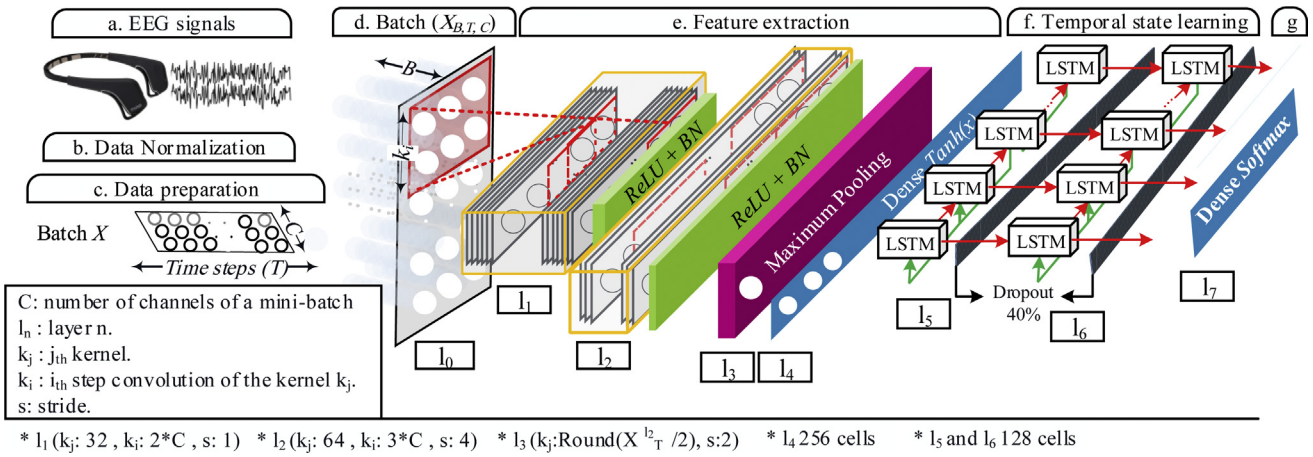


Fig. 2. The proposed framework and processing pipeline architecture.

EOG [39]. In 2018, they improved the precision of their approach, by utilizing a Deep Autoencoder Network [40]. Other researchers have also attained generalizability performance by utilizing ML in removing ECG [41] and EMG [42] artifacts from EEG signals.

3. The proposed framework

This section identifies the components of the framework for training a classification model for EEG artifacts (FCEA), as proposed for developing a stem-to-stern EEG data processing pipeline, as is shown in Fig. 2. Section 3.1 explains the data preparation process. The proposed processing pipeline involves a CNN to populate feature maps from raw EEG data, which is explained in Section 3.2. An LSTM network is then used for modeling the temporal state of the EEG feature maps, as described in Section 3.3. The training of a model and parameters in the pipeline are described in Section 3.4.

3.1. Data preparation

Vector $X \in \mathbb{R}^{T,C}$ denotes a portion of data, called a mini-batch (Fig. 2.c). X is a matrix of $T \times C$, showing EEG amplification in μV from n-channels X_c, for time steps T. The Min-Max normalization method transforms EEG voltage values of a given sequential raw data X (Fig. 2.a) into the range of 0 and 1 (Fig. 2.b), by calculating:

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

3.2. Feature extraction

1D convolutional layer l slides j number of kernels $K \in \mathbb{R}^{i,j}$ over the input matrix X^{l-1} received from the layer $l-1$, for i steps. It is for populating the feature map vector X_j^l by convolving the kernel K_j with the filter regions X_j^{l-1}. This process is mathematically expressed by [22]:

$$X_j^l = ReLU\left(\sum_i X_i^{l-1} * K_{i,j}\right) \tag{2}$$

Activation function Rectified Linear Unit $(eLU\left(X_j\right) = max(0, X_j))$ determines threshold rate for the feature map value X_j. Dissimilar to the convolutional layer, a maximum pooling layer l, slides pooling windows K over the input matrix X^{l-1}, and takes the largest number at each step i to populate feature maps X^l. The pooling process holds the dimensionality of vector X_j^{l-1} in the size of vector X_j^l, however, it reduces the feature map resolution. The size of feature maps vector X_T^l is calculated by [43]:

$$X_T^l = \left(\frac{X_T^{l-1} - X_j^l}{s}\right) + 1 \tag{3}$$

Stride s controls movement of the kernel K on a convolutional or pooling layer. Batch normalization $BN(X^l)$ shifts the internal covariate of the feature map X^l and maintains the mean and standard deviation of activation function, respectively close to 0 and 1, by measuring [44]:

$$BN_{\gamma,\beta}(X^l) = \gamma\left(\frac{X^l - \bar{X}^l}{\sqrt{Var(X^l)}}\right) + \beta \tag{4}$$

$\bar{X}^l$ is the mean value of the matrix of feature maps X^l, and $Var(X^l)$ it is the variance of X^l. γ and β are trainable parameters. ReLU and BN improve generalization performance and accelerate the training process [10–13].

Fig. 2.e shows the CNN proposed for extracting a non-linear decisive feature map matrix X^{l4} from the input mini-batch X^{l0} by processing two 1D-convolutional layers (l_1 and l_2), one pooling layer (l_3) and one dense layer (l_4). The layer l_1 convolves 32 kernels (X_j^{l1}) on the temporal dimension X_T^{l0} using kernels, double sized of the number of channels in the input data ($K_j^{l1} = 2 \times X_C^{l0}$). Selecting stride $s^{l1} = 1$ enables sliding the kernel with the most possible overlap and considers all values in X^{l0} during the convolution process. In the layer l_2, a greater stride ($s^{l2} = 4$), kernel length ($K_j^{l2} = 3 \times X_C^{l0}$) and 64 kernels ($X_j^{l2}$) are used. It allows for creating discrete feature maps by using a 50 % larger kernel length that slides with less overlap. Layers l_1 and l_2, each normalizes the activation value by using a BN. The maximum pooling layer l_3 reduces the dimensionality of vector X_T^{l3} and aggregates the data by using the kernel length $K_j^{l3} = Round(X_T^{l2}/2)$ sliding with strides $s^{l3} = 2$. Finally, dense layer l_4 with 512 neurons classifies and transforms the matrix of feature maps $X_{j,T}^{l3}$ into $X_{512,T}^{l4}$, using a hyperbolic tangent $tanh(x)$ activation function, between the range of -1 and 1, where:

$$\tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{5}$$

3.3. Long short-term memory recurrent neural network

Fig. 3 shows the components of an LSTM block that consists of the memory cell c and three gates for forgetting the current cell state c_{t-1} (forget gate f), reading input value (input gate i) to

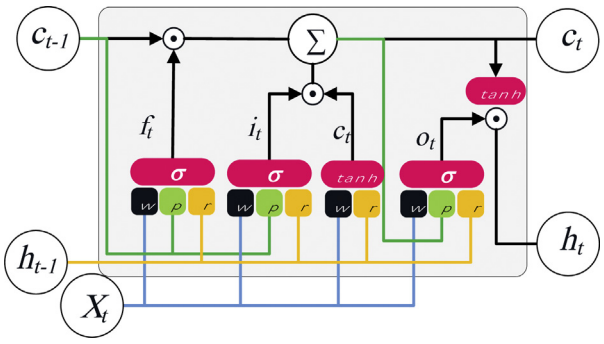


Fig. 3. Architecture of a Long Short-Term Memory cell.

determine value of c_t, and populating output value (output gate o). The gates regulate the flow of data across an LSTM layer [22]. The LSTM block receives the input matrix of X_t at timestamp t from the weighted connections of w and weighted recurrent connections of r. Peephole-weighted connection (p) enables the gates to determine the cell state c_{t-1}. A block predicts the output value of h_t for the given input of X_t. Mathematically, the three gate values (f_t, i_t, o_t), cell state (c_t) and output value h_t at the timestamp t are given by [45]:

$$h_t = o_t \odot tanh(c_t) \tag{6}$$

Sigmoid (σ) and *tanh* (eq. 5) are activation functions for calculating threshold rates, where:

$$\sigma(x) = \frac{2}{1 + e^{-x}} \tag{7}$$

The Sigmoid function is used as the gating function that thresholds a value between 0 and 1 at the 3 gates (, i and o) in the cell to either *let no flow*, or *complete flow* of data through the gates. The *tanh* thresholds a value between -1 and 1 that sustains a longer range before hitting the zero point. This property of the *tanh* function enables overcoming the vanishing gradient problem. The *tanh* function is used to determine the cell state c and output value h.

Fig. 2.f shows an LSTM network with 2 layers (l_5 and l_6), proposed for learning the sequential trend of incoming feature map matrix X^{l_4} on the time domain. Both layers include 128 LSTM units. A dropout layer with a fraction rate of 40 % was also placed after each of the layers to prevent overfitting and complex co-adaptations while training a model. Layers l_5 and l_6 respectively populate the matrix of full feature map sequence $X^{l_5}_{128,T}$ and only the last output vector $X^{l_6}_{128}$ in the output sequence vector $X^{l_6}_{T}$.

3.4. Training a classification model

A model trained by the FCEA assigns label y_n to an unseen mini-batch X. y_n corresponds to the class label of X and designates a value from a set of class labels Y indicating human activities forming EEG artifacts. For instance, in this study, $\forall y_n : y_n \in Y = \{y_1 =$ "Reading", $y_2 =$ "Speaking", $y_3 =$ "Watching"$\}$.

The batch $B = \{(X_1, y_1), \ldots, (X_b, y_b)\}$, includes b labeled mini-batches and is required for training and evaluating the performance of a model. For a training batch B, shuffling the dimension of X_C of the mini-batch X_b, converts the data into a random sample, rather than using the data that was highly correlated.

The dense layer l_7 (Fig. 2.g) is the final layer of the network, where the vector $X^{l_6}_{128}$ is classified categorically, by using a Cross-Entropy Loss with SoftMax activation function, which determines the probability distribution over mutually exclusive output class labels Y. For a network with the neurons Y in the output layer, the value of each neuron ranges between [0, 1] and all values sum to 1. The SoftMax function for the input vector X_j is calculated by [22]:

$$S(X_j)_n = \frac{e^{X_{j_n}}}{\sum_j e^{X_j}}. \tag{8}$$

The loss function determines a penalty for an incorrect classification of a given input vector. The cross-entropy loss function measures uncertainty of two probability distributions S_n and y_n over the same underlying set of events, by measuring the average number of bits needed to identify an event drawn from the set. For discrete distributions of S_n and Y, the cross-entropy of H is measured by [46]:

$$H(Y, S) = -\sum_n y_n log(S_n) \tag{9}$$

The batch-learning approach determines value of network parameters after presenting the complete training set to the network. In this method, at least one training epoch is required to adjust the parameters [47]. Optimization algorithms minimize the loss function by updating the value of parameters during the learning process. The Adam optimization algorithm can be used for gradient-based optimization of stochastic functions, based on adaptive estimates of lower-order moments. It is computationally an efficient algorithm, that requires limited memory and is invariant to diagonal rescaling of the gradients. The Adam algorithm is also well-known for non-stationary objectives and problems with noisy and/or sparse gradients [48].

4. Evaluation of the framework

Section 4.1 presents the Muse EEG wearable device used for acquiring data during 3 experiments that were conducted to create a dataset of human activities among 8 participants (Section 4.2). Section 4.3 describes the dataset and Section 4.4 shows the procedures followed during the evaluation experiment.

4.1. Data acquisition

The Muse headband (Fig. 1.d) from the InteraXon Inc., Toronto, Canada, is a low cost, lightweight and wireless EEG wearable device which has a user-friendly design. The device captures the brainwaves using four EEG channels from seven EEG electrodes. There is a reference electrode on the forehead, two grounds, and two reading channels on the forehead. Two rubber conductive sensors are also placed behind the ears of a subject for reading the differential between the prefrontal and temple areas. The EEG signal from the Muse is down-sampled to a sampling rate of 220 *Hz*, that is measured in microvolts (μV), between the range of [0, 1682.815] μV [26].

4.2. Participants

Four male and four female participants, all university students (age between 18 and 34 years), participated in the study. Research ethics approval was obtained from the Nelson Mandela University Ethics Committee (Reference Number: H18-SCI-CSS-001) and the participants were required to sign written informed consent forms. A headband was used to collect information from participants whilst performing reading and speaking tasks and watching TV. The headband was used to record brain signals and assisted with the collection of data for the human activity EEG dataset. The data collected was anonymized and were used only for research purposes. The data is shared available to the public [49].

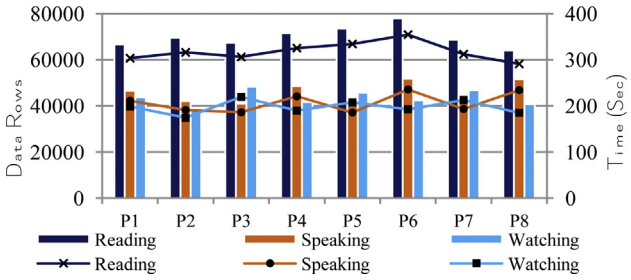


Fig. 4. The data rows recorded from each of the participants.

4.3. Describing the dataset

The participants were asked to wear the headband, sit in a comfortable position and perform the activities of reading printed text, speaking out loudly and watching a TV program. In order to perform the reading activity, they were asked to read three full A4 pages of text. They were then requested to talk about a random event loudly, for about four to five minutes and watch a three minutes video on a large TV screen, located two meters away from their seat.

The data from the AF7 and AF8 channels were recorded at the sample rate of 220 *Hz*. In total, 1,272,459 samples were recorded in the CSV format for the reading, speaking and watching classes. The bar chart in Fig. 4 shows the number of data rows recorded for each task, where the line chart shows the task completion time, in seconds. Due to the differences in task-completion times, as shown in Fig. 4, the dataset is imbalanced. It consisted of 44 % the reading data, 29 % speaking data and 27 % watching data.

4.4. Experimental procedures

Two experiments were conducted to evaluate the classification performance of two models trained by each of the FCEA and the DL models that are described in Section 4.4.1. Section 4.4.2 describes two different sets of data which were prepared for the evaluation experiments. An nVidia GTX 1080 Ti GPU with 11GB of memory and 3584 CUDA cores was used to train the models by using the Keras API [50] in the Python programming language. To prevent overfitting, the auto early-stopping technique of the API was used to stop the training process. Five hundred mini-batches were presented to the models during the training process. The Adam algorithm with the setting of hyper-parameter values (Learning rate = 0.0005, $\beta 1$ = 0.9, $\beta 2$ = 0.9999, epsilon = 1e-08) was used to minimize the cross-entropy loss function [48]. Section 4.4.3 demonstrates the metrics used for evaluating performance and accuracy.

4.4.1. The-state-of-art deep learning models

Table 3 shows the DL architectures which attained the state-of-the-art performance, as described in Section 2. There are 3 general approaches that researchers have used in the DL architecture (CNN, LSTM or CNN + LSTM). CNNs are most frequently used in processing raw EEG data (EEGNet, Deep ConvNet and Shallow ConvNet), as well as accelerometer and gyroscope signals (CNNHAR). The LSTM network has also been used for this purpose (EEGLSTM and LSTMHAR). In a similar manner to the FCEA, researchers have utilized CNNs to reduce the dimensionality of raw EEG and used an LSTM network to generalize a model for EEG data (IoB). The train-validation-test set division size that was originally used by the authors is also shown in Table 3.

4.4.2. Data preparation

Two batches $X^{1D} \in \mathbb{R}^{B,T,C}$ and $X^{2D} \in \mathbb{R}^{B,1,T,C}$ were created (220 Hz) for each of the experiments, when a sliding window of 1500 milliseconds (ms) was used ($T = 330$). X^{2D} included an extra vector which indicated the batch number, that was set equal to one. X^{1D} was presented to networks with the LSTM and 1D-CNNs (CNNHAR, EEGLSTM, LSTMHAR and FCEA). X^{2D} was presented to the 2D-CNNs (EEGNet, Shallow ConvNet, Deep ConvNet and IoB).

Data from the AF7 and AF8 channels were used ($C = 2$) in the first experiment. In the second experiment, AF7 channel data were used ($C = 1$). A split of 60:20:20 was chosen to create the train-validation-test set. The validation data were used to provide an unbiased evaluation of a model fit on the training dataset. The test dataset was used to provide an unbiased evaluation of a final model fit [51].

		Predicted Values: *Positive*	Predicted Values: *Negative*	
Actual Values	*Positive*	a	b	Precision = $a/(a+b)$
	Negative	c	d	Negative Predictive Value = $d/(c+d)$
		Sensitivity $a/(a+c)$	Specificity $d/(b+d)$	Accuracy $(a+d)/(a+b+c+d)$

Fig. 5. The values of a confusion matrix.

4.4.3. Evaluation metrics

The confusion matrix, F1-score, the Area under Receiver Operating Characteristics Curve Score, training time and number of parameters were the metrics used in evaluating the performance of the models. A *confusion matrix* shows the classification performance of a model. Each row of the matrix represents the instances in a predicted class, while each column represents the instances in an actual class and vice versa [52]. The matrix gives insight into the errors being made by a classifier and the types of errors that have been made. It presents the way a model would confuse class prediction [53]. Below are the definitions used for a confusion matrix:

Accuracy: The ratio of the number of correct predictions divided into the total number of samples.

Precision: The number of positive cases that were correctly predicted.

Negative Predictive Value: The number of negative cases that were correctly predicted.

Sensitivity: The number of actual positive cases which are correctly predicted.

Specificity: The number of actual negative cases which are correctly predicted.

Fig. 5 indicates the values in a confusion matrix.

F1-score is the Harmonic Mean between precision and sensitivity that has a value in the range of [0, 1]. The F1-score shows how precisely a classifier performs (how many instances it classifies correctly), and how robust it is (it does not miss a significant number of instances). The greater the F1-score, the better is the performance of a model. The F1-score is measured by:

$$F1\ Score = 2^{*}\frac{Precision^{*}Sensitivity}{Precision + Sensitivity}$$

Receiver Operating Characteristic Curve (*ROC*) presents the independent changes in the proportion of responders. The ROC curve is a plot between sensitivity and specificity for a number of given threshold rates. The threshold rates determine the coordinates of points in the curve. Each point on the ROC curve represents a sensitivity-specificity pair corresponding to a specific decision threshold. The ROC curve with perfect discrimination, passes through the upper left corner (100 % sensitivity, 100 % specificity) [54]. To adjust the ROC curve down to a single point, the area under this curve (AUC) can be plotted. The AUC of a classifier is equal to the probability that the classifier rated positive predictions higher than a random chosen value and negative predictions.

Table 3
The DL architecture designs and train-validation-test dataset size proposed for modeling raw EEG and accelerometer signals. The hyper-parameter values of convolutional (number of kernels, kernel length, strides) and pooling (kernel length, strides) layers are shown. For the LSTM and dense layers, the number of neurons in a layer and activation functions are shown.

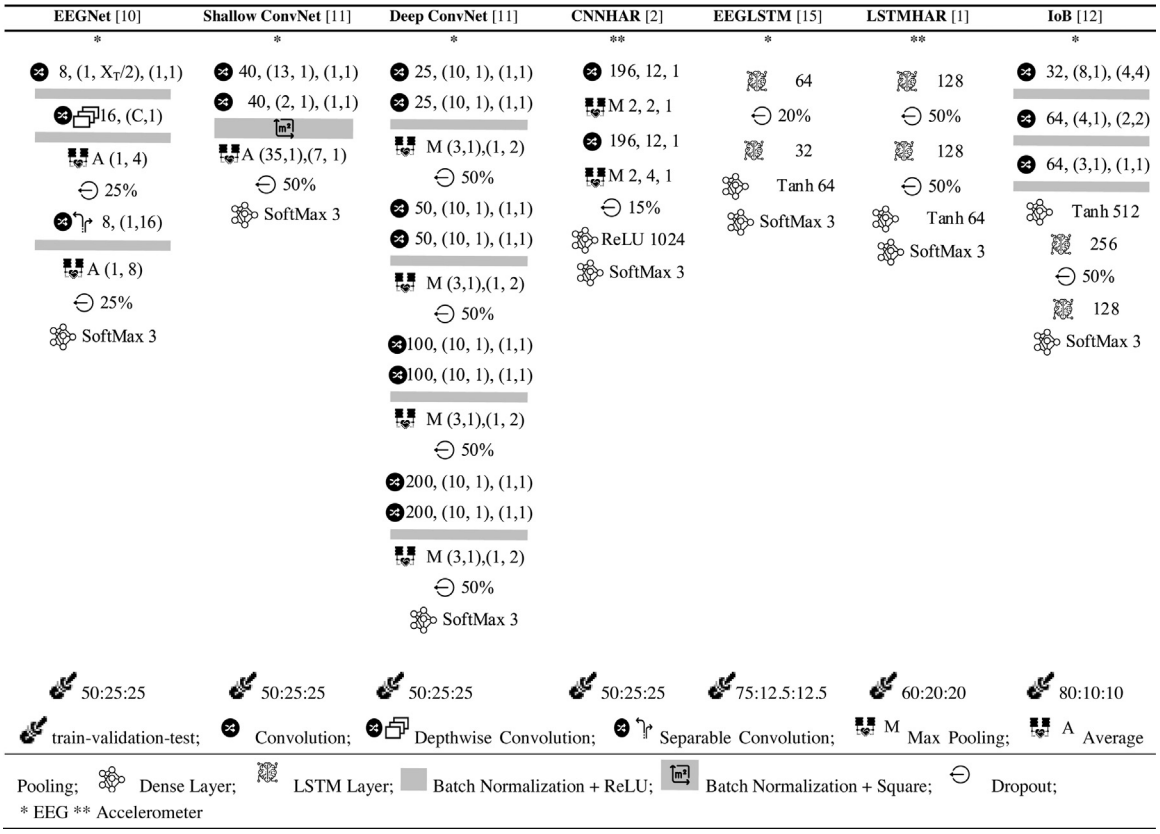


Table 4
Overall comparison of the models.

Model			Accuracy (%)	Precision	Sensitivity	F1	Parameters	Training Time
AF7-AF8	**CNN**	*Shallow ConvNet*	83.1	0.84	0.83	0.84	29,463	01:39:46
		Deep ConvNet	76.06	0.82	0.76	0.75	425,916	05:26:06
		EEGNet	63.24	0.64	0.63	0.63	15,003	01:15:21
		CNNHAR	76.21	0.76	0.76	0.75	13,714,299	03:37:47
	LSTM	*EEGLSTM*	57.57	0.33	0.58	0.42	31,875	00:27:33
		LSTMHAR	63.6	0.70	0.64	0.65	199,043	02:45:00
		IoB	91.21	0.93	0.91	0.91	1,302,051	02:55:16
		FCEA	***94.66***	***0.95***	***0.95***	***0.95***	***510,563***	***00:53:07***
AF7	**CNN**	*Shallow ConvNet*	75.59	0.76	0.76	0.75	14,749	01:00:06
		Deep ConvNet	57.57	0.33	0.58	0.42	27,987	04:11:12
		EEGNet	69.05	0.73	0.69	0.70	307,383	00:54:13
		CNNHAR	71.8	0.74	0.72	0.72	7,091,067	02:08:42
	LSTM	*EEGLSTM*	57.57	0.33	0.58	0.42	31,619	00:19:55
		LSTMHAR	63.96	0.64	0.64	0.63	198,531	01:52:00
		IoB	76.97	0.74	0.72	0.72	1,039,139	01:11:31
		FCEA	***84.53***	***0.86***	***0.85***	***0.83***	***502,179***	***00:41:19***

5. Result analysis

Table 4 reports that the models trained by the FCEA attained the highest accuracy, precision, sensitivity and F1-score values, where both AF7 and AF7-AF8 models had an acceptable number of parameters and training time. This highlights the performance of the FCEA models in labeling test data by the means of accuracy, exactness and completeness. The accuracy is indicated by the percentage. Moreover, the FCEA models have an acceptable number of trainable parameters which reduces the training time. The test data were temporally and spatially varied, which had never previously been presented to the models. This indicates the generalizability performance of the models.

The FCEA (AF7-AF8: 95.44 %, AF7: 84.53 %) and IoB (AF7-AF8: 91.21 %, AF7: 76.97 %) models that use a combination of CNN and LSTM networks attained the highest accuracy rates. The CNN-based models achieved an accuracy higher than the LSTM-based models. However, the low accuracy attained by the Deep ConvNet (57.57 % for the AF7 and AF7-AF8 models – the Keras API stopped the training when the training did not have any improvement) shows that EEG artifact classification does not necessarily require a deep network for high-level feature extraction. Utilizing an LSTM network alone, the authors could not achieve a desirable result. Increasing the number of LSTM units in layers improves the classification accuracy, as it can be interpreted by comparing the performance of the EEGLSTM and LSTMHAR models.

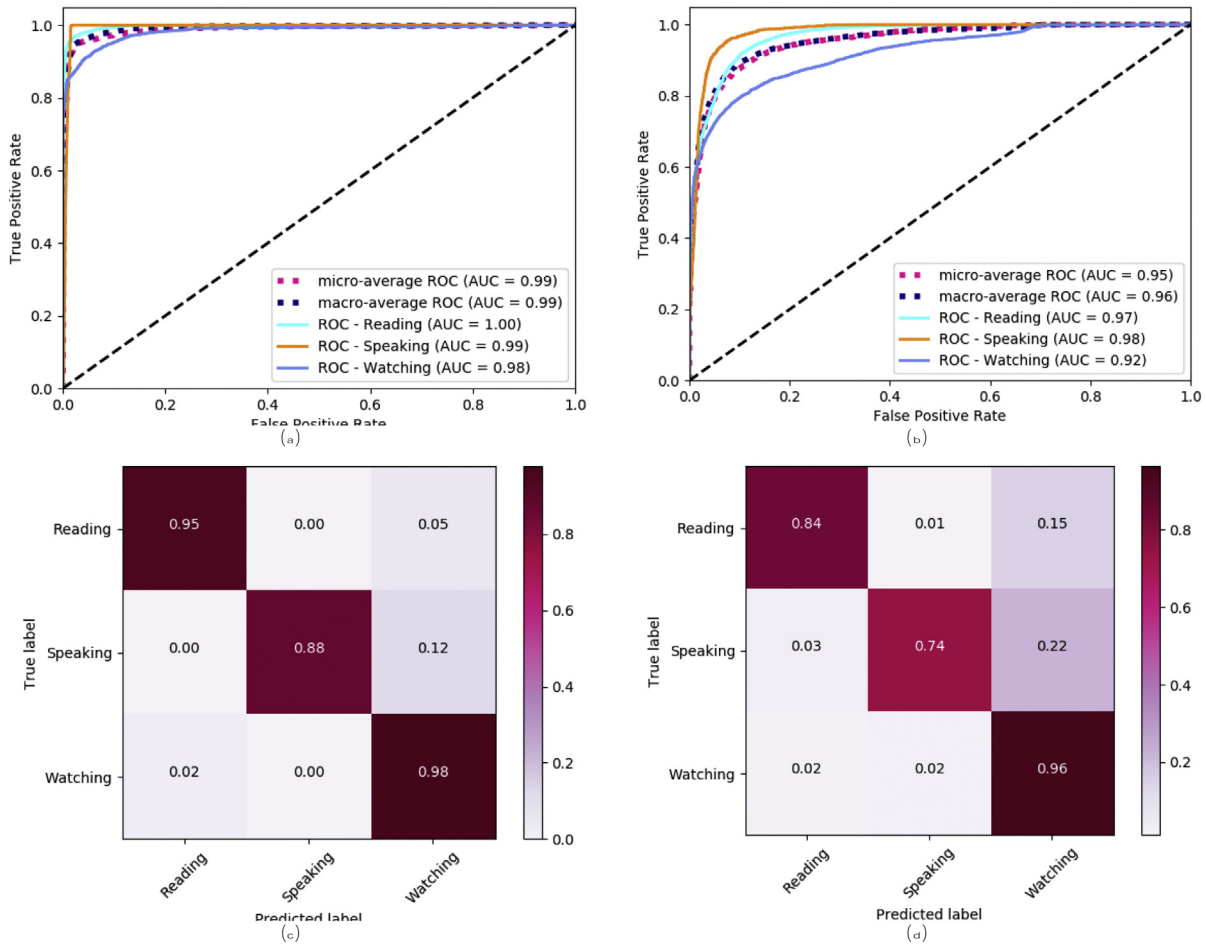


Fig. 6. ROC curves with AUCs for the AF7-AF8 (a) and AF7 (b) models. Confusion matrices for the AF7-AF8 (c) and AF7 (d) models.

Fig. 6 shows the ROC curves, AUCs and confusion matrices for the FCEA models. The results demonstrate that classification of the watching activity obtained a better performance, compared with the other two classes, although, the training dataset included the least data for the watching activity. This demonstrates that a physiological activity that thresholds fewer artifacts is the simplest to be classified, which is rational whereas the other activities involve the artifacts generated by moving the jaw, head and eyes.

6. Discussion

This study investigated common sensory technologies and ML algorithms used in HAR. Section 6.1 discusses the DL architectures evaluated in this study empirically. Section 6.2 presents and compares the sensory technologies for HAR. Section 6.3 discusses how using commercial EEG benefits HAR.

6.1. Deep-learning model architectures

The empirical evaluation presented in Section 4 focused on evaluating different DL architectures used in HAR and EEG data processing. The results approach gives an insight into proposing appropriate DL architecture for modeling human activity on raw EEG data. Either the CNNs or RNNs have been more commonly used in the state-of-the-art DL models for HAR and EEG data classification. The evaluation results in this study, indicate that stacking an LSTM network against a CNN improves generalizability, accuracy, exactness and completeness of models significantly. It also reduces the number of trainable model parameters. The AF7-model (accuracy = 84.5 %; AUC = 0.96; F1 = 0.83) and AF7-AF8 model (accuracy = 94.6 %; AUC = 0.99; F1 = 0.95) trained by following the FCEA, outperformed the competitive models during the performance evaluation experiment.* The experiments conducted in this study only used 8 subjects, performing 3 activities in a practical environment, although, a larger dataset can perhaps improve the single-channel model's accuracy in future research.

Theoretically, an LSTM network enables learning from long-term temporal dependence data, in which the spatial feature extraction capability of CNNs is supplemented. Extracting EEG artifact features requires detecting abnormal voltage generated by the non-cerebral activities in the signal, however, extracting cerebral features from the raw EEG data possibly requires deeper temporal and spatial feature extraction. In addition, a 60:20:20 train-validation-test splitting was used for developing and testing the models. For future research, the depth and width of a CNN can be modified to reduce spatial dimensionality of raw EEG data, whereas, modifying the LSTM network would handle the complexity of data across the time domain. Implementing various train-validation-test splitting procedures and other evaluation methods, such as cross-validation, may improve the validity of this approach.

* The anonymous data and source codes for this research project are published in [49].

6.2. Sensory technologies for human activity recognition

The human activities that were targeted in this study involved facial activities, including jaw-clenching and head and eye movements, that are part of a wide range of daily human activities. The three general sensory technologies that are commonly used for HAR were reviewed in Section 2.1, included, the smart wearable devices equipped with motion and orientation sensors, cameras and biosensors. Among the technologies discussed, only the smart-goggles and cameras (Table 1), as well as using EMG and EOG electrodes placed on the face (Table 2) have the potential of capturing facial activities. The use of cameras was introduced as a technology that lacks user-experience. For the rest of the studies, they either lacked the generalization or had not attained a desirable classification accuracy.

Utilizing a standard consumer-grade EEG wearable device (Fig. 1) in HAR is an original and economical technique which can be effectively applied to detect facial activities and addresses the limitations of the sensory technologies discussed above. The devices are economical, light-weight and user-friendly. The sampling rate and spatial resolution of these devices, built-in motion and orientation, and other biosensors used in these devices, enable the extraction of valuable knowledge related to user activities. However, further research is required to investigate the capabilities and limitations of this approach. The modeling of deeper EEG artifact features, such as, articulation, facial expressions and chewing would allow research that in future can extend the applications of this technology to designing fully-natural HAR, speech-to-text, eye-tracking, emotion recognition and diet-monitoring systems.

6.3. Human activity recognition using deep Electroencephalography learning

Before the availability of the EEG wearable devices on the market and rise of DL, EEG was known as a cumbersome, complex and costly procedure and interpreting its signal required the assistance of a therapist. As demonstrated in Section 2.2, utilizing DL simplifies classification, augmentation and preprocessing EEG signals, where, interpreting EEG signals allows for extracting the knowledge about mental state of a user. However, EEG artifacts impact the quality of a signal negatively and complicates interpreting it. Hence, a secondary technique is most often used to remove the artifacts. In addition to using the DL, the use of EMG, ECG and EOG sensors was shown as the techniques that researchers apply to remove the artifact from the signal. Most consumer-grade EEG wearable devices support prefrontal EEG which is correlated to the intellectual, emotional and cognitive states of a subject. However, in this study the potential of extracting knowledge of facial activities, only from a two-channel and single-channel EEG on the forehead of a user was shown.

Analyzing the human brain is perhaps the most fascinating scientific challenge in the 21st century. The feasibly of modern communication, sensory and computing technologies, such as, Big Data analytics, Internet of Things and cloud computing, improve the mapping of EEG signals with events taking place in an environment. As most of the consumer-grade EEG wearable devices support prefrontal EEG which is correlated to the intellectual, emotional and cognitive states of a subject, using DL to interpret the raw signals from these devices, could assist researchers with new possibilities. Future research should investigate the training of general and personal models for the human brain, by utilizing a large EEG dataset of various human physical, emotional and cognitive states in daily living conditions. Researchers could also apply the proposed approach to various domains, such as, Smart-City environments, neuroleadership, neuromarketing, health-care, educational, human-robot interaction and public and e-government applications.

7. Conclusions

In this paper, a DL-based framework to design a data processing pipeline for training and evaluating a classification model for EEG artifacts (called FCEA) was introduced. As was shown, collecting data from a consumer-grade EEG wearable device, and interpreting this data with the particular DL architecture proposed by the FCEA developed new and evolutionary perspectives on different aspects of HAR technologies and raw EEG signal classification. The evaluation results showed that applying a CNN to extract features from raw EEG data and then, feeding these features to an LSTM network results in a considerable improvement in classification performance (2–9 % for 1-channel and 8–12 % for 2-channel data) and results in generalizing a model. It was further discussed that how using a commercial EEG wearable device evolves the cutting-edge disciplines and science, as a wide range of future work was recommended.

Authorship contributions

Conception and design of study: Amirsaleh Salehzadeh, Andre P. Calitz, Jean Greyling.

Acquisition of data: Amirsaleh Salehzadeh, Andre P. Calitz, Jean Greyling.

Analysis and/or interpretation of data: Amirsaleh Salehzadeh, Andre P. Calitz, Jean Greyling.

Drafting the manuscript: Amirsaleh Salehzadeh, Andre P. Calitz, Jean Greyling.

Revising the manuscript critically for important intellectual content: Amirsaleh Salehzadeh, Andre P. Calitz, Jean Greyling.

Approval of the version of the manuscript to be published: Amirsaleh Salehzadeh, Andre P. Calitz, Jean Greyling.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. McDaniel, S. Quinn, Developing a start-to-finish pipeline for accelerometer-based activity recognition using Long short-term memory recurrent neural networks, in: Proc. 17th Python Sci. Conf., Austin, Texas, 2018, pp. 31–40, http://dx.doi.org/10.25080/Majora-4af1f417-005.

[2] A. Ignatov, Real-time human activity recognition from accelerometer data using Convolutional Neural Networks, Appl. Soft Comput. 62 (2018) 915–922, http://dx.doi.org/10.1016/J.ASOC.2017.09.027.

[3] J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu, Deep learning for sensor-based activity recognition: a Survey, Pattern Recognit. Lett. 119 (2018) 3–11, http://dx.doi.org/10.1016/J.PATREC.2018.02.010.

[4] O.D. Lara, M.A. Labrador, A survey on human activity recognition using wearable sensors, IEEE Commun. Surv. Tutorials 15 (2013) 1192–1209, http://dx.doi.org/10.1109/SURV.2012.110112.00192.

[5] A. Gruebler, K. Suzuki, Design of a wearable device for reading positive expressions from facial EMG signals, IEEE Trans. Affect. Comput. 5 (2014) 227–237, http://dx.doi.org/10.1109/TAFFC.2014.2313557.

[6] R. Zhang, S. Bernhart, O. Amft, Diet eyeglasses: recognising food chewing using EMG and smart eyeglasses, in: 2016 IEEE 13th Int. Conf. Wearable Implant. Body Sens. Networks, IEEE, 2016, pp. 7–12, http://dx.doi.org/10.1109/BSN.2016.7516224.

[7] A. Saleh Zadeh, A.P. Calitz, J.H. Greyling, Evaluating a biosensor-based interface to recognize hand-finger gestures using a Myo armband, in: Proc. Annu. Conf. South African Inst. Comput. Sci. Inf. Technol. - SAICSIT' 18, ACM Press, New York, USA, 2018, pp. 229–238, http://dx.doi.org/10.1145/3278681.3278709.

[8] M. Wand, A. Himmelsbach, T. Heistermann, M. Janke, T. Schultz, Artifact removal algorithm for an EMG-based silent speech interface, in: 2013 35th

Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE, 2013, pp. 5750–5753, http://dx.doi.org/10.1109/EMBC.2013.6610857.
[9] Y. Lu, C. Zhang, B.-Y. Zhou, X.-P. Gao, Z. Lv, A dual model approach to EOG-based human activity recognition, Biomed. Signal Process. Control 45 (2018) 50–57, http://dx.doi.org/10.1016/J.BSPC.2018.05.011.
[10] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: A Compact Convolutional Network for EEG-Based Brain-Computer Interfaces, 2018, http://dx.doi.org/10.1088/1741-2552/aace8c.
[11] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (2017) 5391–5420, http://dx.doi.org/10.1002/hbm.23730.
[12] X. Zhang, L. Yao, S. Zhang, S.S. Kanhere, Q.Z. Sheng, Y. Liu, Internet of Things Meets Brain-Computer Interface: A Unified Deep Learning Framework for Enabling Human-Thing Cognitive Interactivity, 2018, CoRR. abs/1805.0 http://arxiv.org/abs/1805.00789 (Accessed 1 August 2018).
[13] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T.H. Falk, J. Faubert, Deep Learning-based Electroencephalography Analysis: A Systematic Review, 2019 (Accessed 30 March 2019) http://arxiv.org/abs/1901.05498.
[14] K.G. Li, M.I. Shapiai, A. Adam, Z. Ibrahim, Feature scaling for EEG human concentration using particle swarm optimization, in: 2016 8th Int. Conf. Inf. Technol. Electr. Eng., IEEE, 2016, pp. 1–6, http://dx.doi.org/10.1109/ICITEED.2016.7863292.
[15] S. Alhagry, A.A. Fahmy, R.A. El-Khoribi, Emotion recognition based on EEG using LSTM recurrent neural network, Int. J. Adv. Comput. Sci. Appl. 8 (2017), http://dx.doi.org/10.14569/IJACSA.2017.081046.
[16] A. Gupta, P. Singh, M. Karlekar, A novel signal modeling approach for classification of seizure and seizure-free EEG signals, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2018) 925–935, http://dx.doi.org/10.1109/TNSRE.2018.2818123.
[17] S. Agrawal, A. Gupta, Fractal and EMD based removal of baseline wander and powerline interference from ECG signals, Comput. Biol. Med. 43 (2013) 1889–1899, http://dx.doi.org/10.1016/j.compbiomed.2013.07.030.
[18] R.R. Sharma, R.B. Pachori, Baseline wander and power line interference removal from ECG signals using eigenvalue decomposition, Biomed. Signal Process. Control 45 (2018) 33–49, http://dx.doi.org/10.1016/j.bspc.2018.05.002.
[19] A. Singhal, P. Singh, B. Fatimah, R.B. Pachori, An efficient removal of power-line interference and baseline wander from ECG signals by employing Fourier decomposition technique, Biomed. Signal Process. Control 57 (2020) 101741, http://dx.doi.org/10.1016/j.bspc.2019.101741.
[20] A. Almeida, G. Azkune, Predicting Human Behaviour with Recurrent Neural Networks, Appl. Sci. Basel (Basel) 8 (2018) 305, http://dx.doi.org/10.3390/app8020305.
[21] J.O. Smith, Mathematics of the Discrete Fourier Transform (DFT): With Audio Applications, 2nd ed., Julius Smith, 2007.
[22] A. Gibson, J. Patterson, Deep Learning: A Practitioner's Approach, O'Reilly Media, Sebastopol, California, United States of America, 2017 (Accessed 18 November 2018) http://shop.oreilly.com/product/0636920035343.do.
[23] S.F. Chevtchenko, R.F. Vale, V. Macario, F.R. Cordeiro, A convolutional neural network with feature fusion for real-time hand posture recognition, Appl. Soft Comput. 73 (2018) 748–766, http://dx.doi.org/10.1016/J.ASOC.2018.09.010.
[24] F. Sharbrough, E.G. Chatrian, P.R. Lesser, H. Luders, M. Nuwer, W.T. Picton, American Electroencephalographic Society Guidelines for Standard Electrode Position Nomenclature, J. Clin. Neurophysiol. 8 (1991) 200–202.
[25] EMOTIV, EMOTIV - Brainwear® Wireless EEG Technology, EMOTIV, 2016 (Accessed 6 June 2016) http://emotiv.com/.
[26] InteraXon Inc., Muse™ - Meditation Made Easy, Interaxon Off. Website, 2018 (Accessed 12 November 2018) https://choosemuse.com/muse-2/.
[27] Interaxon, MUSE™, Interaxon Co., 2015 (Accessed 11 July 2018) http://www.choosemuse.com/.
[28] NeuroSky, Biosensors, NeuroSky Co., 2016 (Accessed 6 June 2016) http://neurosky.Co.om/.
[29] E. Tsironi, P. Barros, C. Weber, S. Wermter, An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition, Neurocomputing 268 (2017) 76–86, http://dx.doi.org/10.1016/J.NEUCOM.2016.12.088.
[30] S. Viriyasiripong, A. Lopez, S.H. Mandava, W.R. Lai, G.C. Mitchell, A. Boonjindasup, M.K. Powers, J.L. Silberstein, B.R. Lee, Accelerometer measurement of head movement during laparoscopic surgery as a tool to evaluate skill development of surgeons, J. Surg. Educ. 73 (2016) 589–594, http://dx.doi.org/10.1016/J.JSURG.2016.01.008.
[31] M. Farooq, E. Sazonov, Accelerometer-based detection of food intake in free-living individuals, IEEE Sens. J. 18 (2018) 3752–3758, http://dx.doi.org/10.1109/JSEN.2018.2813996.
[32] Z. Wang, C. Zhao, S. Qiu, A system of human vital signs monitoring and activity recognition based on body sensor network, Sens. Rev. 34 (2014) 42–50, http://dx.doi.org/10.1108/SR-12-2012-735.
[33] J. Künecke, A. Hildebrandt, G. Recio, W. Sommer, O. Wilhelm, Facial EMG responses to emotional expressions are related to emotion perception ability, PLoS One 9 (2014) e84053, http://dx.doi.org/10.1371/journal.pone.0084053.
[34] U. Côté-Allard, C.L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, B. Gosselin, Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning, 2018 (Accessed 28 March 2019) http://arxiv.org/abs/1801.07756.
[35] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, N. de Freitas, Dueling Network Architectures for Deep Reinforcement Learning, 2015 (Accessed 10 December 2018) https://arxiv.org/abs/1511.06581.
[36] K.B. Sundhara Kumar, G. Krishna, N. Bhalaji, S. Chithra, BCI cinematics – a pre-release analyser for movies using H2O deep learning platform, Comput. Electr. Eng. 74 (2019) 547–556, http://dx.doi.org/10.1016/J.COMPELECENG.2018.03.015.
[37] Q. Zhang, Y. Liu, Improving Brain Computer Interface Performance by Data Augmentation With Conditional Deep Convolutional Generative Adversarial Networks, Arxiv, 2018 (Accessed 31 March 2019) http://arxiv.org/abs/1806.07108.
[38] J.T.C. Schwabedal, J.C. Snyder, A. Cakmak, S. Nemati, G.D. Clifford, Addressing Class Imbalance in Classification Problems of Noisy Signals by Using Fourier Transform Surrogates, 2019 (Accessed 31 March 2019) http://arxiv.org/abs/1806.08675.
[39] B. Yang, K. Duan, T. Zhang, Removal of EOG artifacts from EEG using a cascade of sparse autoencoder and recursive least squares adaptive filter, Neurocomputing 2014 (2016) 1053–1060.
[40] B. Yang, K. Duan, C. Fan, C. Hu, J. Wang, Automatic ocular artifacts removal in EEG using deep learning, Biomed. Signal Process. Control 43 (2018) 148–158, http://dx.doi.org/10.1016/J.BSPC.2018.02.021.
[41] V. Abolghasemi, S. Ferdowsi, EEG–fMRI: Dictionary learning for removal of ballistocardiogram artifact from EEG, Biomed. Signal Process. Control 18 (2015) 186–194, http://dx.doi.org/10.1016/J.BSPC.2015.01.001.
[42] J. Hu, C. Wang, M. Wu, Y. Du, Y. He, J. She, Removal of EOG and EMG artifacts from EEG using combination of functional link neural network and adaptive neural fuzzy inference system, Neurocomputing. 151 (2015) 278–287, http://dx.doi.org/10.1016/J.NEUCOM.2014.09.040.
[43] V. Dumoulin, F. Visin, A Guide to Convolution Arithmetic for Deep Learning, 2016 (Accessed 5 November 2018) http://arxiv.org/abs/1603.07285.
[44] S. Ioffe, C. Szegedy, Batch Normalization, Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015 (Accessed 22 April 2019) http://arxiv.org/abs/1502.03167.
[45] K. Greff, R.K. Srivastava, J. Koutnik, B.R. Steunebrink, J. Schmidhuber, LSTM: a search space odyssey, IEEE Trans. Neural Networks Learn. Syst. 28 (2017) 2222–2232, http://dx.doi.org/10.1109/TNNLS.2016.2582924.
[46] R.Y. Rubinstein, D.P. Kroese, The Cross-Entropy Method : a Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning, Springer Science and Business Media, New York, 2013.
[47] I.N. Da Silva, D.H. Spatti, R.A. Flauzino, L.H. Bartocci Liboni, S.F. Dos Reis Alves, Artificial Neural Networks: A Practical Course, Springer International Publishing, Switzerland, 2017, http://dx.doi.org/10.1007/978-3-319-43162-8_2.
[48] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, 2014 (Accessed 1 October 2018) http://arxiv.org/abs/1412.6980.
[49] A. Salehzadeh, amirsaleh-salehzadeh/BehaveNet: The BehaveNet, 2019, http://dx.doi.org/10.5281/ZENODO.2552600.
[50] F. Chollet, Keras, 2015 (Accessed 3 February 2018) https://github.com/fchollet/keras.
[51] J. Brownlee, What Is the Difference Between Test and Validation Datasets?, Mach. Learn. Mastery, 2017 https://machinelearningmastery.com/difference-test-validation-datasets/.
[52] D.M.W. Powers, Evaluation: From precision, recall and f-measure to roc., informedness, markedness and correlation, J. Mach. Learn. Technol. 2 (2011) 37–63.
[53] J. Brownlee, What Is a Confusion Matrix in Machine Learning, Mach. Learn. Mastery, 2016 (Accessed 16 November 2018) https://machinelearningmastery.com/confusion-matrix-machine-learning/.
[54] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clin. Chem. 39 (1993) 561–577 (Accessed 8 January 2019) http://www.ncbi.nlm.nih.gov/pubmed/8472349.