IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 July 2022, accepted 18 July 2022, date of publication 29 July 2022, date of current version 5 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3195028

RESEARCH ARTICLE

Discriminating Fake and Real Smiles Using Electroencephalogram Signals With Convolutional Neural Networks

MOSTAFA M. MOUSSA[1], (Student Member, IEEE), USMAN TARIQ[2], (Member, IEEE), FARES AL-SHARGIE[2], (Member, IEEE), AND HASAN AL-NASHASH[2], (Senior Member, IEEE)

[1]Biomedical Engineering Program, American University of Sharjah, Sharjah, United Arab Emirates
[2]Department of Electrical Engineering, American University of Sharjah, Sharjah, United Arab Emirates

Corresponding author: Hasan Al-Nashash (hnashash@aus.edu)

This work was supported in part by the Research Grant EFRG18-BBR-CEN-02, Grant EFRG-EN0244, and Grant FRG20; and in part by the Open Access Program Grant through the American University of Sharjah, United Arab Emirates, under Grant OAP22-CEN-102.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the American University of Sharjah. The IRB protocol was entitled Cognitive Assessment and Enhancement under Application No. IRB-19-513 and was approved on 31 March 2020.

ABSTRACT Genuineness of smiles is of particular interest in the field of human emotions and social interactions. In this work, we develop an experimental protocol to elicit genuine and fake smile expressions on 28 healthy subjects. Then, we assess the type of smile expressions using electroencephalogram (EEG) signals with convolutional neural networks (CNNs). Five different architectures (CNN1, CNN2, CNN3, CNN4, and CNN5) were examined to differentiate between fake and real smiles. We transform the temporal EEG signals into normalized gray-scale images and perform three-way classification to classify fake smiles, genuine smiles, and neutral expressions in the form of subject-dependent classification. We achieved the highest classification accuracy of 90.4% using CNN1 for the full EEG spectrum. Likewise, we achieved classification accuracies of 87.4%, 88.3%, 89.7%, and 90.0% using Beta, Alpha, Theta, and Delta EEG bands respectively. This paper suggests that CNNs models, widely used in image classification problems, can provide an alternative approach for smile detection from physiological signals such as the EEG.

INDEX TERMS Smile, emotion, electroencephalogram (EEG), convolutional neural networks (CNNs), machine learning.

I. INTRODUCTION

Smiling plays an important role in our daily lives and may be thought of as a result of our perception of our surroundings. However, people can fake smiles to convey happiness, e.g. when they smile due to social pressure or when they do not want to express their true feelings. Hence, it is important to differentiate between a real smile from a fake one.

Applications of smile recognition are many, such as mental health monitoring, human-computer interaction, human behavioral studies, and patient monitoring [1], [2]. But if smiles can be faked how do we differentiate between a fake smile from a real one expressing pleasure? There has been some work done in the past, for instance in [3], Hoque *et al.* study temporal patterns in facial movements related to smiles that express delight. In this work, we look for a modality that cannot be faked. We propose to look into brain signals via electroencephalography (EEG) as opposed to facial expression recognition systems, especially considering the recent spread of face masks. Although most EEG reading and analysis systems require dedicated stationary hardware and software, non-contact, wearable wireless EEG electrodes have been in development for some time [4]–[6]. This will make our proposed work translatable to real-life applications. Note that, video-based systems require full facial visibility, i.e. they require processing person identifiable information, which can raise serious privacy concerns.

This work can also be applied to psychology, psychiatry, and other commercial applications. For instance, the presented work, when deployed to practice, can be used for

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

mental health monitoring, which can alert a patient when it is the time to see a professional. This can also aid in reflecting the mental state after recovery from depression or traumatic disorder. Organizations and companies may befit from devices to ascertain the level of happiness among their employees, etc.

Smiling emotions in general are expressed in terms of valence and arousal [7]. Valence ranges from negative to positive (describes pleasantness and attractiveness) and from sad to happy. Arousal ranges from calm or bored to excited or alert [8], [9]. These ranges encompass the six basic emotions described by Ekman and Friesen, wherein each specific emotion would be a point in this affect space, where happiness for instance would have high valence and moderate arousal [8], [10]. In this work, however, we are concerned with one of those six emotions, happiness, and as hinted earlier, we plan to measure it by first identifying real smiles from fake ones. Smiles can be socially motivated, as Crivelli *et al.* study [11] showed or smiles can be spontaneous.

Note that, smiling spontaneously/genuinely can be considered a proxy for happiness (a discrete emotion) for the sake of the study. An example of a study that uses EEG to directly classify depression is Seal's *et al.* [12] wherein they use their framework, DeprNet to classify depression using the Patient Health Questionnaire 9 score and EEG signals, and though they acquire a high accuracy of 99.4% and area-under-curve (AUC) of 99.9%, this questionnaire is a rather subjective method of acquiring ground truth. Another study that is more pertinent to our work is done by Hoque *et al.*, who use an experimental procedure to elicit frustration and delight through a video, and expressing these affective states when recalling situations, i.e. their fake class [3]. Smiles were classified as a binary classification problem using Support Vector Machines (SVMs), Hidden Markov Models (HMM), and Hidden-state Conditional Random Fields (HCRF). They achieved an accuracy of 92 % with a dynamic SVM [3]. Distinguishing fake from real smiles is a challenging problem for computers, though people may be able to do it instinctively with reasonable accuracy.

As mentioned earlier, we can divide emotions based on their arousal and valence scores. We are using this emotion model in the study. We selected images from the Geneva Affective Picture Database (GAPED) image dataset [13], [14], as stimuli. These images are designed to elicit a certain kind of emotional response from the subjects. We selected images that can have a high valence and low arousal, and medium valence and low-to-medium arousal effect on the subjects. For further details about the emotion model, the readers are referred to the one explained by Liu *et al.* [15].

The main contribution of this work lies in the use of an automated feature extraction technique via multiple convolutional neural networks (CNN). In [13], the features were extracted manually using the discrete wavelet transform (DWT) and empirical mode decomposition (EMD). In addition, we have validated the proposed CNN networks on the raw data using multiple other techniques like long short-term memory (LSTM) networks, shallow artificial neural network (ANN), and support vector machines (SVM). Furthermore, the present study performed three-way classification (True vs Fake vs Neutral); which is an extension to the work published in [13] that performed binary classification. The simplicity of the proposed CNN architectures is also part of our contribution. The proposed method required less computation, which can enable us to deploy this algorithm in wearable devices.

Studies in emotion recognition, apart from visual channel, often involved the use of voice, eye-tracking mechanisms (for visual stimuli), functional near-infrared spectroscopy (fNIRS), and functional magnetic resonance imaging (fMRI), and data related to the autonomous nervous system (galvanic skin response, heart rate, etc.) [16]–[20]. For example, Zhang *et al.* [21] investigated the use of involuntary facial expressions to detect deceit in a system that represents expressions as facial Action Units (AUs). The study detected deceit with accuracies of 86.02%, 73.16%, 80.46%, and 90.15% for anger, enjoyment, fear, and sadness respectively. Meanwhile, the study in [22] used two physiological signals (pupillary response and Galvanic skin response) of subjects viewing videos in the MAHNOB [23] and AFEW [24] datasets to detect genuine smiles using machine learning and observers' self-judgement. The study classified smiles with an average accuracy of 97.2% using machine learning models and 58.9% to 68.4% based on observers' judgement [22]. Another study in [25] utilized the temporal features of four physiological signals namely: pupillary response (PR), electrocardiogram (ECG), galvanic skin response (GSR), and blood volume pulse (BVP) to distinguish between genuine and fake smiles and achieved 98.8% accuracy.

Techniques that depend on audiovisual features can have plenty of limitations. However, EEG signal acquisition can be affected only by motion, eye blinking, or experimenter error, all of which can be simply controlled, resolved, or mitigated, as opposed to involuntary bodily responses or environmental factors.

We conjecture that EEG signals can provide a more reliable insight into the participants' emotions and consequently provide better means of smile detection. This can be used as a measure of a person's own emotional well-being which they can track with a measurable sensor. EEG data reflects the electrical activities of the brain which can reveal the true emotional state of a person. EEG signals are characterized by frequency bands named Delta (0.5-4 Hz), theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-35 Hz) and Gamma (>35 Hz). These frequency bands are associated with mental states [26].

EEG signals and machine learning have been extensively used in the field of affective computing to predict different emotions. Miranda and Patras [27] used convolutional and recurrent neural networks to predict positive and negative affective behavior and obtained average F1-scores of 0.59, and 0.61 respectively using EEG while the subjects watched videos (audiovisual stimuli). Xu and Plataniotis [28]

investigated semi-supervised learning using a stacked denoising autoencoder and deep belief networks and reported higher average F1-scores, at 86.60% for valence and 86.67% for arousal. Meanwhile, Williams' investigation on deep learning and transfer learning used multi-layer perceptrons, convolutional neural networks (CNN), Long short-term memory (LSTM), and LSTM-fully convolutional networks to classify positive/negative emotions and achieved maximum accuracy of 64.36% [29]. Gao *et al.* used a core-brain-network-based convolutional neural network with differential entropy and time-frequency representation of the EEG data to classify emotional states for 15 subjects of the SJTU emotion EEG dataset (SEED) [30] yielding an average accuracy of 91.45% [31].

Furthermore, studies in [32]–[35], that used the AMIGOS dataset [36] or DEAP dataset [37] employed convolutional neural networks or LSTM with different types of features and obtained classification accuracies in the range of 85.65% to 99.72%. The AMIGOS dataset consists of physiological signals based on electrocardiograms (ECG), galvanic skin response (GSR) and EEG signals that are recorded using emotional videos as the stimuli [26]. Meanwhile, DEAP dataset contains EEG, electromyogram (EMG), electrooculogram (EOG), GSR, respiration amplitude, skin temperature, and blood volume in response to video stimuli [37]. Despite manual feature extraction, the best among these methods for binary EEG emotion classification is [32] at 99.72%, which uses data from the DEAP dataset, 5 fully connected layers, and phase locking values as features.

A. NOVELTY OF THIS WORK

Our work, unlike the prior literature in this area, focuses on discriminating fake from genuine expressions; more specifically distinguishing fake from real smiles. To do so, we utilize a novel dataset collected in our lab [13] for three levels of smile expressions. In addition, we used minimal pre-processing of the EEG signals. Our work is one of the first of its kind to use EEG signals to discriminate genuine and fake smiles. It focuses on the use of pre-processed EEG data and multiple CNNs to classify the expressions into a genuine smile, a posed smile, or a neutral expression.

The next section, Section II presents the experimental setup and describes the methodology, including deep learning architectures and parameters selection. Section III describes in details the obtained results, and Section IV presents discussion of findings from the experiments. Finally, Section V concludes the paper.

II. MATERIALS AND METHODS

A. SUBJECTS, STIMULI, AND EXPERIMENTAL PROTOCOL

A pool of 28 subjects, 8 females and 20 males ranging between 18-26 in age, with no known mental illnessess have participated in the study. The subjects chose their preferred timing to perform the experiment between 8 am to 5 pm, to ensure they were in a more relaxed state. All subjects were briefed prior to the experiment, and were given detailed consent forms that included all the information as they showed up for the experiment. They were made aware that they can revoke their consent at any time. The study was approved by the Institutional Review Board (IRB) of the American University of Sharjah.

The stimuli, used for the sequence described in Figure 1, came from a set of 245 images, carefully selected from the Geneva Affective Picture Database (GAPED) image dataset [13], [14]. Three types of image sets were chosen to conduct this study (115 funny images, 70 neutral images, and 60 images of a plain book to prompt fake smile).

Images meant to elicit positive emotions include, but are not limited to babies and animals. The neutral set includes inanimate objects. The different image sets are intended to induce different emotions and by extension different smile expressions. For instance, positive images or images with a high valence and low arousal scores, are intended to induce a true smile, neutral images or images with a medium valence score and low-to-medium arousal score are intended to induce no response [38]. The book is an additional image used to indicate the participants when to fake a smile.

To begin with, the subjects are asked to read a brief description of the experiment and its outcomes. They then give their informed consent. Then, the experimenters equip the subjects with the EEG cap. The electrode impedance is then checked to ensure the recording process goes without error from the equipment. The EEG data is recorded using 64 Ag/AgCl scalp electrodes arranged according to the standard 10–20 system (ANT waveguard system and ASA Lab 4.9.2 acquisition software, ANT Neuro). The EEG data is sampled at 500 Hz. The impedances of all EEG electrodes are maintained below 10 kΩ using Ag/AgCl gel, and are referenced to the left and right mastoids. Figure 1(b) also shows an example of the hardware setup and data acquisition layout.

Afterwards, the subjects are given the instructions, as follows. They would firstly learn where three keys are on the keyboard, each corresponding to a response; ''P'' when they genuinely smiled, ''N'' for neutral responses (no smile), and ''Q'' for fake smiles. The subjects were asked to fake a smile and press the key whenever they would see the prompt image, which in this experiment was that of a book. There were a total of 245 trials in this experiment for 245 images. These were randomly presented and were meant to elicit one of the three aforementioned responses. Each image would remain onscreen for 2 seconds or until a key is pressed, and is followed by a cross for 1 second. The entire experiment lasted for about 13 minutes per subject.

B. DATA PRE-PROCESSING AND CLASSIFICATION

After we obtain the data from all subjects, it is pre-processed primarily with MATLAB and EEGLAB toolboxes [39] to filter out interference frequencies, remove eye blink artifact using Independent Component Analysis (ICA), and extract epochs to be used as features to the classifiers.

We use Finite Impulse Response (FIR) band-pass filtering, with a low cut-off frequency of 0.5 Hz, and the high

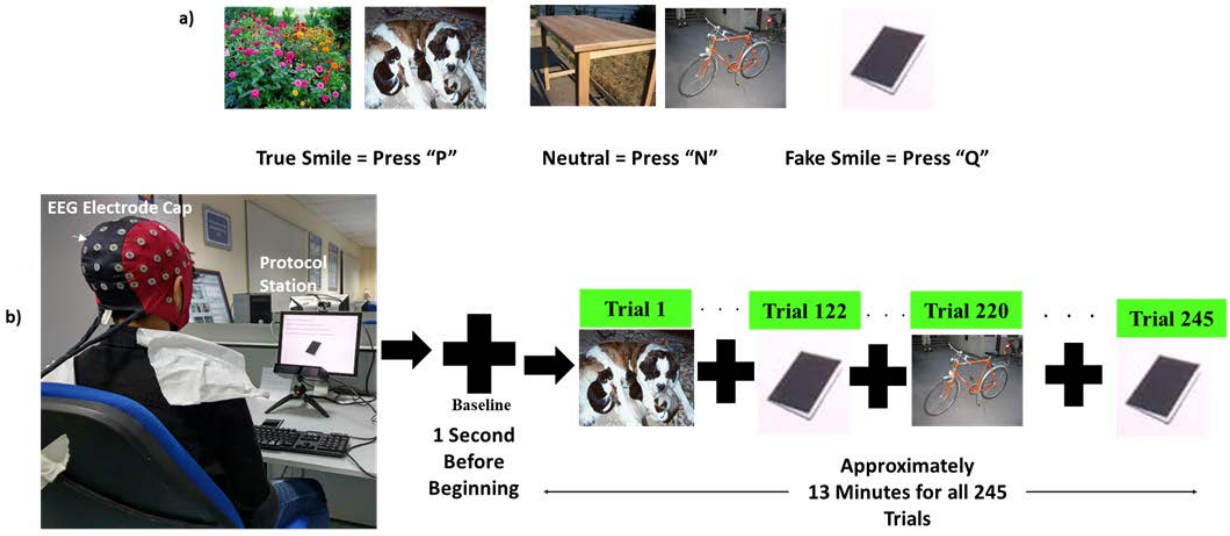

FIGURE 1. a) Type of stimuli and corresponding key. b) Timing window for the protocol.

cutoff frequency of 40 Hz [13]. After filtering, the data is re-referenced to the computed average of the channels for each subject. Then, to remove eye-blink artifacts, ICA is used to extract independent components and the dominant ones corresponding to eye-blinks are identified and removed manually. The data are then filtered to obtain one of the four clinical frequency bands corresponding to brain waves Delta (0.5-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz) and Beta (12-35 Hz). To obtain the full spectrum of the EEG data, the signals are band-pass filtered between 0.5 to 40 Hz. In this context, gamma band was inherently presented within the full spectrum. Finally the data is segmented into epochs. Each epoch starts 0.5 second before each event and lasts till 1 second after the event to retain useful information. This results in a feature matrix with dimensions $5 \times 62 \times 750 \times N$ (5 is the number of frequency bands- alpha, beta, theta, delta and the full spectrum-, 62 is the number of electrodes, 750 is the number of data points, and N is the number of epochs for each subject). This data is then used as an input to the proposed structures of the CNNs. We have also tested other algorithms like Support Vector Machines (SVM), artificial neural networks (ANN) and long short-term memory (LSTM) networks to evaluate the effectiveness of the CNN structure proposed in this work.

It is worth pointing out here the difference between machine learning and deep learning. In traditional machine learning, features (a measurable quantity used to describe the data), are in general learned independent of the end-goal of the algorithm (e.g., classifiers in case of classification problems). Let us take the example of object recognition. In traditional approaches, we would first extract useful hand-crafted features, such histogram of gradients, orientation of edges, etc. Then we would pass these extracted features to a classifier learning algorithm, that would then learn classifiers to classify objects. Some of the traditional machine learning algorithms include support vector machines (in which you try to optimally separate the data in the feature space), decision trees (where you make a tree-like structure that makes decisions on every branch, based upon the learned thresholds in different feature dimensions), random forests (a collection of many trees), shallow feed-forward neural networks (which consist of three layers; an input layer, a hidden layer and the output layer; the hidden and output layers consist of artificial neurons, that compute a learned weighted sum of the inputs and then apply a non-linear function; the final layer may give probability of a certain label corresponding to certain input instance), etc. [40].

In deep learning algorithm we learn the relevant features from the data itself, jointly with the end-goal of the learning algorithm (e.g., learning classifiers). This yields data specific features that have shown to give superior results. Deep learning is essentially a sub-field of machine learning. It consists of algorithms that put an emphasis on learning successive layers of meaningful representations of the data. The higher the number of layers, the greater the *depth* of the learned model. In most cases, deep learning algorithms are variants of neural networks. Some examples of deep learning algorithms include, deep neural networks (consisting of multiple hidden layers of neurons), deep convolutional neural networks (that use convolution operations), deep LSTM networks (specialized for time sequenced data), deep auto-encoders (that generate the output using the learned representation of inputs with some bottlenecks), generative adversarial networks (that generate artificial data), etc. [40].

We used five different deep convolutional neural network models (CNN1-CNN5) with different layers, parameters and structures, selected in the same fashion as in [32]. CNN1 was the first architecture designed, and it was selected after finding that a single convolutional layer works best for the validation data. CNN2 and CNN3 were further

adaptations of CNN1. The remaining two architectures were selected to observe the effect of increasing architecture complexity and adding a convolutional layer on the results and on the computation time. Table 1 provides details of the architectures and model parameters.

In each model, the CNN has two main structures in the network; convolution and fully connected layers. The first three models (CNN1-CNN3) have a single convolutional layer and two fully connected layers. Meanwhile, CNN4 and CNN5 utilize two convolutional layers and two fully connected layers as shown in Figure 2.

The convolutional layers in the five models have different parameters, which can be seen in Table 1. The LSTM network structure is similar to the CNN1 architecture, but with the EEG data presented as a sequence. The LSTM layer has 128 hidden units and the format of output set to *last*. The training options used are similar to CNN1.

Here, we describe the details of the layers shown in Table 1 and Figure 2. These details are also known as *hyper-parameters* of the networks. These *hyper-parameters* are user defined settings, that describe how the networks are constructed. Three of our architectures have a single convolutional layer, while the others have two convolutional layers. Convolutional layers perform convolution of the preceding layers with learned filters. With all our networks, we use the same training options with a mini-batch size of 128, 40 epochs with 3 iterations per epoch, but without employing dropout layers, as L2 regularization and batch normalization was sufficient. For instance, let us take CNN1 (note that, the same explanation also applies to other CNN architectures). It has one convolution layer which has a *kernel size* of 11×11. The *kernel size* refers to the size of the filter. Hence, in this case the input layer is convolved with filters of size 11×11. This results in a two-dimensional output (both input matrix and filter are two-dimensional, hence the convolution result is also two-dimensional). This output is known as the *feature map* of the corresponding filter. Since, we have 128 such filters for CNN1. This implies that convolving the input with 128 filters outputs 128 feature maps. In signal and image processing [41], we usually perform convolution in strides of 1. However, we can also replicate the same process in strides of 2, or more. Here we used strides of 4 in both horizontal and vertical dimensions. This results in an output size that is 4 times smaller in both the dimensions. This reduces the computation in the proceeding layers. We then apply a non-linear activation function, *Rectified Linear Unit (ReLU)*. This adds non-linearity to the structure and increases the expressive power of the network. We also use Batch normalization layers. These standardize the output of the convolution layers across batches, i.e., make them mean-centered and normalize their standard deviation for further computation. This has shown to boost learning [42].

We then concatenate the output of the convolution layers into a vector which is then fed to fully connected layers of the network, as shown in Figure 2. We use *softmax* activation function in the final layer. This forces the output of the network to show the probability of the input belonging to each of the three classes (true smile, fake smile, or neutral expression). It is worth mentioning here, that the filter weights and the weights in fully connected layers, are all learned from the training EEG data with backpropagation algorithm using a *cross-entropy* cost function [40].

The softmax activation function is given by Equation 1. In this equation, n_i corresponds to the output of the i^{th} neuron in the final layer and y_{n_i} is the output of the softmax function for this neuron.

$$y_{n_i} = \frac{e^{n_i}}{\sum_{i=1}^{K} e^{n_i}} \tag{1}$$

Note that y_{n_i} is in the range [0, 1], and $\sum_{i=1}^{K} y_{n_i} = 1$ for the three classes ($K = 3$). This shows that the output of the softmax layer applied to the neurons in the final layer, gives rise to a probability distribution across the three classes. These show the probability of the input belonging to each of the three classes.

Since the input can only belong to one of the three classes, the true labels can be encoded as one-hot binary vectors, which have the value 1 for dimension representing the true label and 0 for the others. Hence, these label vectors can also be thought as probability distributions. Thus, in this context a good cost function to compare the true probability distribution with the one at the output of the network, can be a *cross-entropy loss function*. We used the weighted cross-entropy function in our experiments, as shown in Equation 2.

$$Loss = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{K} w_i t_{n_i} \, ln \, y_{n_i} \tag{2}$$

In this equation, N is the number of training samples, K is the number of classes, w_i is the weight of class i, t_{n_i} is the ground truth or the indicator that sample n is in class i, and y_{n_i} is the output of sample n for class i or the likelihood that the network classifies sample n in class i, given by Equation 1. In Equations 1 and 2, parameters such as the number of samples N, the number of classes K, and the input t_{n_i} are dependent on the EEG data and classification problem, while others are set to the MATLAB default, such as the class weight w_i.

The input signals to the networks consist of EEG data represented as floating point images with pixel values that ranges from 0.0 to 1.0. These normalized pixel values correspond to the amplitudes of EEG signals. This is achieved by converting the multi-channel EEG data, which is a function of time, into a 3-D array, whose size is $C \times T \times N$. In this context, C is the number of channels, T is the time points, and N is the number of images, which is equivalent to number of epochs/events. Each image contains the normalized floating pixel value that corresponds to the amplitude of EEG signals, as discussed earlier.

In the convolutional layers, features are learned from the input EEG signals for each of the five frequency bands separately when the input is only a specific band, or all

TABLE 1. Descriptions of the five CNN architectures.

Model		CNN 1	CNN 2	CNN 3	CNN 4	CNN 5
Architecture	Convolutional layers	1	1	1	2	2
	Fully Connected layers	2	2	2	2	2
	Total Number of layers	10	10	10	13	13
Kernel size(s)		[11x11]	[3x3]	[5x5]	[5x5] & [11x11]	[2x2] & [3x3]
Stride(s)		[4x4]	[4x4]	[4x4]	[4x4]	[3x3]
Number of filters in convolutional layer(s)		128	70	90	128 & 90	70 & 40
Number of hidden units in fully connected layers		150 & 3	150 & 3	150 & 3	150 & 3	150 & 3

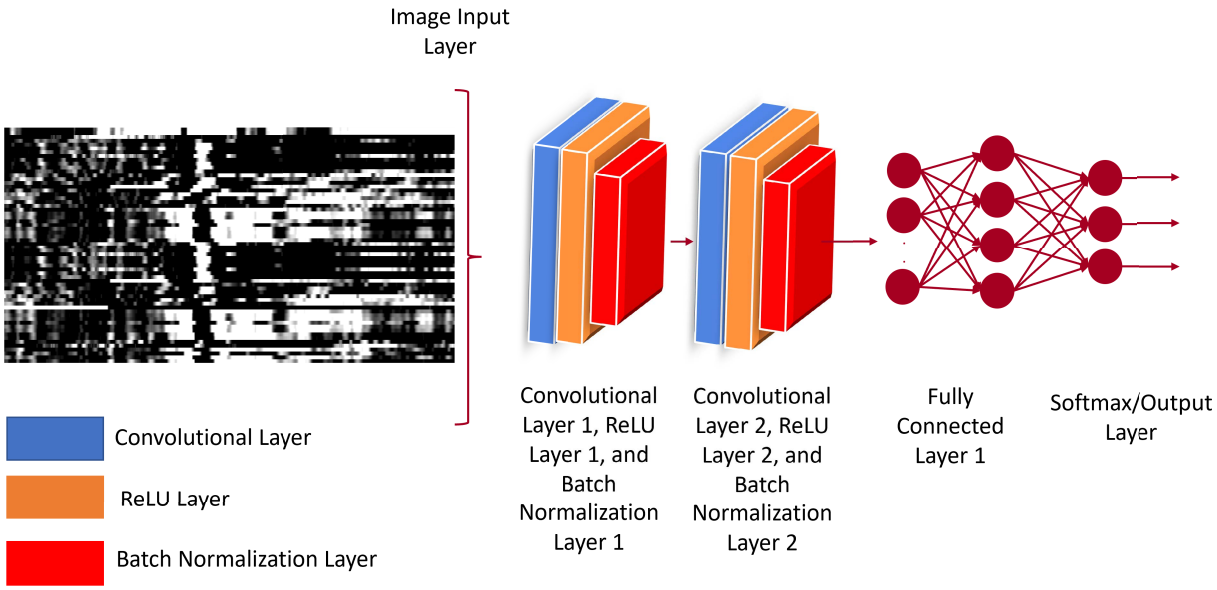

FIGURE 2. CNN5 Architecture. Processed EEG signals are input to the network through the input layer, the convolutional layer(s) reduce the size of the data, and the fully connected layer(s) are where learning basically occurs.

the bands jointly when the input in the full spectrum (see Subsection II-B for the description of the bands). The input signals are mean-centered and over-sampled to remedy class imbalance by augmenting with scaling and replication.

The output from the convolutional layers extract features from the input data. The fully-connected layers then categorize the input data into various classes based on the training data. In particular, the first fully connected layer has 150 neurons and the second fully connected layer has 3, corresponding to the number of classes. Hence, the neurons of the final layer forecast the outcome of the input signal as true smile, fake smile or neutral expression.

All convolutional network architectures have the following training options: Stochastic Gradient Descent with Momentum of 0.95 as the training function, L2-regularization with a factor of 0.0005, and piece-wise learning rate schedule with an initial learning rate of 0.01. The default learnable weights in MATLAB are initialized by the Glorot weight initializer, which assigns the weights from a zero-mean uniform distribution with a variance dependent on the size of input and output to the hidden layers [43]. The CNNs were trained and tested using 4-fold cross-validation. We chose $K = 4$ to balance between the amount of experimentation and the amount of training and testing data in each fold.

The proposed CNNs were compared with ANN, SVM, and LSTM in Section III. In addition, to evaluate the effectiveness of the proposed models, we used several evaluation metrics namely; accuracy, sensitivity, specificity, precision, and F1-score.

III. EXPERIMENTAL RESULTS

The overall classification accuracy, sensitivity, specificity, precision, and F1-score of the proposed models are summarized in Table 2. The classification performance results are given in the form of mean $\pm$ standard deviation across subjects. From Table 2, we obtain the following significant points:

- Regarding classification models, the best classification performance in the form of accuracy, sensitivity, specificity, precision, and F1-score is achieved using CNN1 compared to all other CNN architectures, LSTM as well as the SVM model.
  However, ANN produced comparable classification performance to CNNs when using the full spectrum. Aside from accuracy, we can see that CNN1 yields the highest F1-score in contrast with the ANN and the other CNN architectures, and consequently the highest harmonic mean of sensitivity and precision. This tells us that,

TABLE 2. Subject-dependent results of the full spectrum (0.5-40Hz), Beta (12-35Hz), Alpha (8-12Hz), Theta (4-8Hz), and Delta (0.5-4Hz) Bands for all algorithms.

Classification Model	Band	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
CNN1	Full Spectrum	**90.4 ± 3.1**	89.4 ± 5.5	92.9 ± 4.2	**91.0 ± 4.4**	**90.0 ± 3.8**
	Beta	87.4 ± 2.7	86.5 ± 5.8	91.0 ± 4.1	87.5 ± 4.0	86.9 ± 4.4
	Alpha	88.3 ± 3.0	88.2 ± 6.1	90.7 ± 4.4	87.5 ± 4.0	87.8 ± 4.3
	Theta	89.7 ± 2.5	89.1 ± 6.2	91.4 ± 4.1	88.8 ± 3.2	88.8 ± 3.8
	Delta	90.0 ± 2.8	89.3 ± 4.8	92.5 ± 4.0	90.0 ± 3.5	89.6 ± 3.6
CNN2	Full Spectrum	89.4 ± 2.8	87.6 ± 6.3	92.0 ± 4.3	88.8 ± 4.4	88.0 ± 4.1
	Beta	86.1 ± 2.4	86.1 ± 5.8	89.9 ± 4.1	86.1 ± 4.1	86.0 ± 4.3
	Alpha	86.7 ± 3.1	86.3 ± 6.6	90.1 ± 4.5	86.5 ± 3.8	86.3 ± 4.7
	Theta	88.4 ± 2.4	87.7 ± 4.3	91.9 ± 3.9	89.5 ± 3.0	88.5 ± 2.8
	Delta	88.7 ± 2.6	87.5 ± 5.2	91.8 ± 3.9	88.9 ± 3.3	88.1 ± 3.7
CNN3	Full Spectrum	89.5 ± 2.3	88.0 ± 6.5	91.6 ± 3.8	88.3 ± 4.5	88.0 ± 4.4
	Beta	85.7 ± 2.3	86.4 ± 5.6	89.4 ± 4.6	85.3 ± 3.7	85.8 ± 4.3
	Alpha	86.4 ± 2.6	86.3 ± 6.1	89.8 ± 4.7	85.8 ± 4.0	86.0 ± 4.6
	Theta	87.9 ± 2.1	87.5 ± 5.0	90.8 ± 4.3	87.5 ± 2.7	87.4 ± 3.4
	Delta	88.5 ± 2.8	87.0 ± 5.3	91.9 ± 3.9	88.5 ± 3.8	87.7 ± 3.7
CNN4	Full Spectrum	89.3 ± 3.2	89.8 ± 7.1	89.5 ± 6.0	86.7 ± 6.3	87.9 ± 4.6
	Beta	87.4 ± 2.6	87.4 ± 9.0	88.3 ± 7.3	87.5 ± 5.1	87.0 ± 3.9
	Alpha	87.2 ± 3.2	86.3 ± 7.8	89.5 ± 6.2	87.6 ± 4.5	86.7 ± 4.6
	Theta	88.3 ± 3.4	87.6 ± 7.5	90.4 ± 5.4	88.5 ± 3.9	87.9 ± 4.5
	Delta	89.2 ± 3.3	87.7 ± 7.3	92.0 ± 4.3	89.8 ± 4.2	88.5 ± 4.7
CNN5	Full Spectrum	89.6 ± 3.1	**90.1 ± 7.3**	90.0 ± 5.5	87.9 ± 5.4	88.7 ± 4.6
	Beta	87.6 ± 2.7	89.2 ± 6.7	88.1 ± 6.6	85.6 ± 4.5	87.2 ± 3.9
	Alpha	88.1 ± 2.6	89.3 ± 5.4	89.6 ± 5.4	86.8 ± 4.1	88.0 ± 3.7
	Theta	89.4 ± 2.3	90.0 ± 5.0	90.9 ± 4.8	88.4 ± 3.5	89.1 ± 3.2
	Delta	89.5 ± 2.8	89.3 ± 5.4	91.9 ± 3.9	88.9 ± 4.5	89.0 ± 4.3
LSTM	Full Spectrum	79.2 ± 4.9	77.1 ± 10.8	85.1 ± 10.8	76.1 ± 7.1	76.0 ± 6.1
	Beta	87.6 ± 4.4	83.0 ± 11.1	93.6 ± 3.2	88.0 ± 4.6	85.1 ± 6.9
	Alpha	81.0 ± 6.4	77.2 ± 12.4	88.9 ± 5.9	79.9 ± 7.8	78.1 ± 8.8
	Theta	83.3 ± 5.6	79.3 ± 10.4	90.5 ± 4.6	82.4 ± 6.9	80.5 ± 7.5
	Delta	77.4 ± 4.9	72.4 ± 12.6	85.7 ± 9.3	75.0 ± 6.7	73.1 ± 7.8
ANN	Full Spectrum	88.9 ± 2.8	88.1 ± 2.3	**93.8 ± 1.2**	88.3 ± 2.6	88.2 ± 2.4
	Beta	86.2 ± 2.6	84.3 ± 1.6	92.0 ± 0.9	84.8 ± 2.7	84.5 ± 2.1
	Alpha	83.6 ± 3.4	82.4 ± 2.9	91.0 ± 1.5	82.8 ± 3.0	82.6 ± 2.9
	Theta	86.9 ± 3.4	86.0 ± 3.0	92.8 ± 1.5	86.1 ± 3.2	86.0 ± 3.1
	Delta	86.7 ± 3.0	85.9 ± 2.6	92.8 ± 1.4	86.1 ± 2.5	86.0 ± 2.5
SVM	Full Spectrum	47.8 ± 8.2	63.4 ± 42.9	63.4 ± 42.9	45.8 ± 10.9	56.4 ± 18.6
	Beta	48.4 ± 7.8	65.2 ± 45.9	65.2 ± 45.9	44.1 ± 14.2	58.3 ± 19.0
	Alpha	48.2 ± 7.9	62.5 ± 41.9	62.5 ± 41.9	40.6 ± 18.2	57.0 ± 18.0
	Theta	47.7 ± 8.4	62.3 ± 42.3	62.3 ± 42.3	46.6 ± 11.2	54.0 ± 21.3
	Delta	47.4 ± 8.3	60.8 ± 39.5	60.8 ± 39.5	47.9 ± 14.7	52.1 ± 21.6

overall, CNN1 is the most suitable model for smiles classification.

- Considering the type of frequency band, the full spectrum produced the highest classification performance followed by delta and theta frequency bands. In particular, we found that the full spectrum in CNN1 performs the best with an average classification accuracy of 90.4%, outperforming CNN2, CNN3, CNN4, and CNN5, which yielded classification accuracies of 89.4%, 89.5%, 89.3%, and 89.6%, respectively.

To further investigate what CNN1 classifies erroneously, we looked at its confusion matrix. Figure 3 shows the confusion matrices of CNN1 for the full spectrum and the four EEG channels. It shows a higher rate of confusion between the neutral class and true smiles than other class pairs. This is also true vice-versa. That is, for true smiles that are misclassified, more of them are confused with neutral expression than with fake smiles. The misclassification for fake smiles, on the contrary, is distributed similarly across the true smile and neutral expression classes. The reported confusion matrix clearly shows the superior performance of CNN1 for all the cases. Figure 4 shows the training progress in terms of accuracy and loss of one fold with CNN1.

IV. DISCUSSION

In this study, we measured brain activity using EEG data, due to its simplicity compared to other brain imaging modalities, and utilized the EEG signals to distinguish between genuine smiles, fake smiles and neutral expression. We classified these smiles using CNN with different architectures. To the best of our knowledge, this is the first study to use the EEG signals to classify genuine and fake smiles with deep learning. We achieved the highest classification accuracy of 90.4% using CNN1. The obtained result outperformed ANN, LSTM and SVM classifiers.

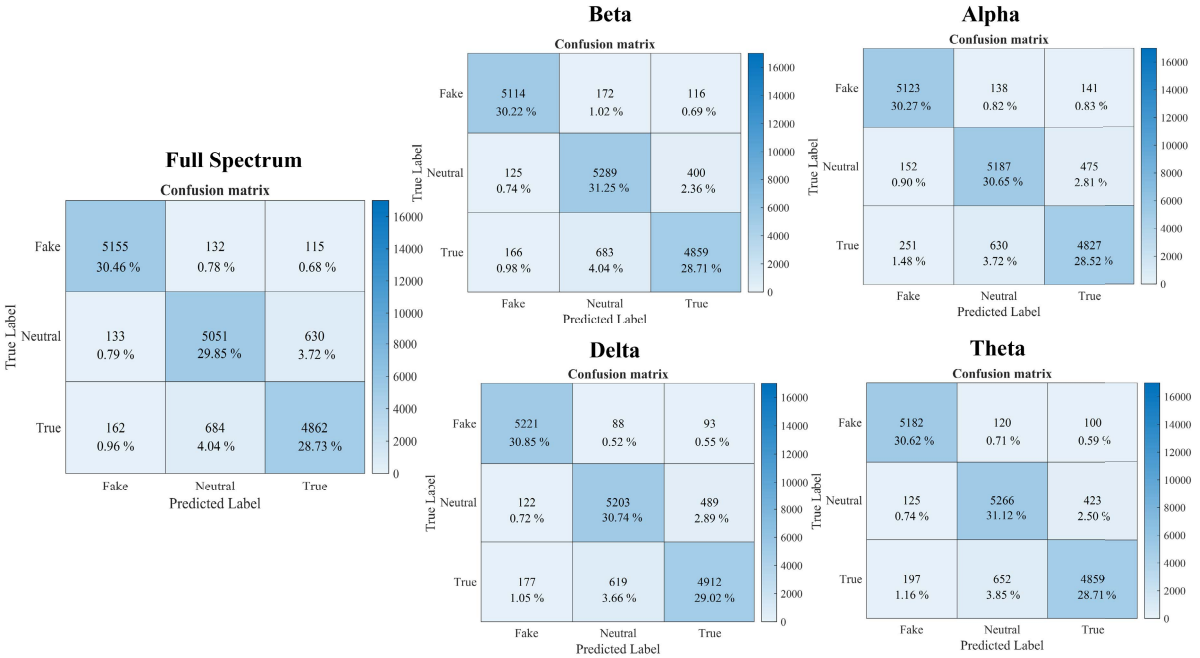

FIGURE 3. Confusion Matrices of CNN1 for all the frequency bands.

We compare our work with a subset of the aforementioned works, that are similar in scope or methodology, for various aspects, in Table 3. It is important to note that there is only one work on the discrimination of genuine vs acted smiles with EEG data [13]. The other works are attempting to solve other emotion related problems, such as recognition of arousal, valence and liking or arousal and valence. Apart from this, the authors of [13] perform only two-way classification and extract a multitude of features (power spectral density, phase locking value, etc). Since the study in [13] follows a different method of analysis, the results cannot be compared. The presented study performs three-way classification and learns the features algorithmically, which is a much more challenging problem.

We found that convolutional neural networks outperform other models in differentiating between the three classes of smiles as summarized in Table 2. It is also clear that CNN4 and CNN5 with the full spectrum yield a higher sensitivity than CNN1, but CNN1 performed better than the rest in terms of accuracy, specificity, precision, and F1-score. We found that, one convolutional layer proved better than two, as CNN1 outperformed CNN4 and CNN5. This could be due to the reduced number of parameters, in addition to the dataset size in our study. Furthermore, the standard deviations were more acceptable compared to the other architectures. In addition, reduced number of parameters imply shorter computation time and fewer required resources.

It is worth noting that our CNN testing shows that filter size has a small effect on classification performance than the number of filters. For instance, reducing the number of filters from 128 to 70 while keeping the kernel size at 11×11 reduced the classification accuracy from 90.45% to 90.21%, whereas reducing the kernel size from 11×11 to 3×3 reduced the accuracy from 90.45% to 89.15%.

Table 2 shows that CNNs perform marginally better than LSTM and ANNs. However, both are vastly superior to SVM, given the same pre-processing steps for this three-class problem. This is due to the lack of feature extraction for all algorithms, which is consistent with some aforementioned works in the literature ([7] and [22]).

When we examined the confusion matrix between the three expressions (genuine smile, fake smile and the neutral expression), we found the highest confusion between genuine smiles and neutral expression. The reason for the higher confusion rate between genuine smiles and neutral expression, seen in Figure 3, might be because the two may share similar features. A plain book image can be thought of as a neutral stimulus after all. Another reason could be due to the short time between stimuli; a fake smile may require a longer time interval than simply no response. These reasons may make it slightly more difficult for the network to distinguish those two classes (genuine smiles and neutral expression) than fake smiles.

For the four clinical frequency bands shown in Table 2, we can see that the average performance is, in fact, slightly sub-par when compared with the full spectrum at 0.5-40 Hz for all CNN architectures. This entails the loss of some features or patterns as some of those frequencies are filtered out. It is noted that despite all four bands showing slightly lower accuracy than the entire spectrum for all algorithms, theta and delta perform almost equivalently for the artificial neural network, and delta performs better than the other three bands for the all CNN architectures. We note the pattern of delta performing well with ANN and CNN1,

TABLE 3. Comparison with other similar works.

Aspects \ Works	Alhagry et al. [33]	Salama et al. [35]	Alex et al. [13]	This work
Dataset	DEAP [37]	DEAP [37]	Novel Dataset	Novel Dataset
Problem	Recognition of arousal, valence, and liking	Recognition of arousal and valence	Two-way classification between true smiles, fake smiles, and neutral expression	Multi-class classification between true smiles, fake smiles, and neutral expression
Feature Extraction and Selection	None	None	Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD), DWT-EMD	None
Pre-Processing	Data downsampled to 128 Hz, averaged to common reference, eye artifacts removed, high-pass filtering	High-pass filtering, band-stop filtering, channel normalization	Eye-blink artifact rejection with ICA, filtering (notch and band-pass to get brain waves), re-referencing to computed average, baseline DC offset removal, epoch extraction	None
Cross-Validation	4-Fold	5-Fold	5-Fold	4-Fold
Classification Methods	LSTM	3D-CNN	One-vs-One DWT, EMD, and DWT-EMD with SVM, KNN, ANN	SVM, ANN, CNN, LSTM
Maximum Subject-Dependent Accuracy	Arousal: 85.7%, Valence: 85.5%, Liking: 88.0%	Arousal: 87.4%, Valence: 88.5%	(using DWT-ANN) True vs Fake: 94.3 $\pm$ 2.2	(using CNN1) Fake vs Neutral vs True: 90.4% $\pm$ 3.1%
Maximum Subject-Dependent F-Score	Not computed	Arousal: 86.0%, Valence: 86.0%	Not computed	(using CNN1) Average F-Score: 90.0% $\pm$ 3.8%

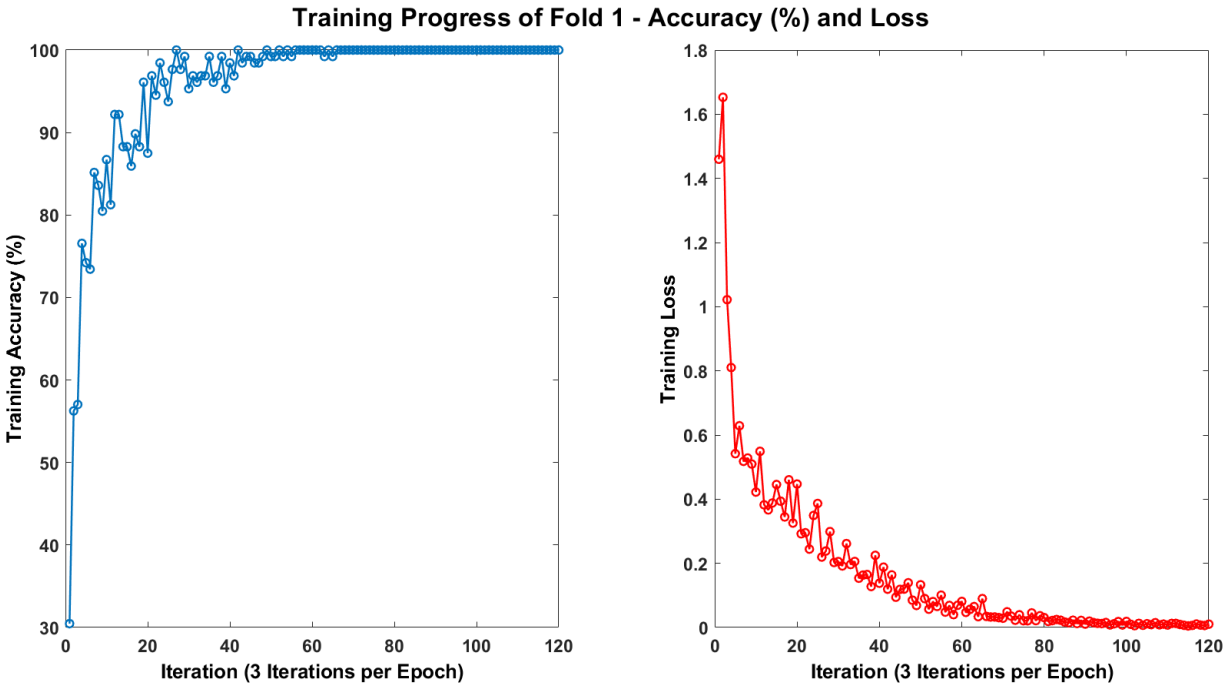

FIGURE 4. Progress of one training fold with CNN1.

despite theta coming close with the artificial network. Delta waves are usually associated with deep sleep, yet show the best, or close to the best, performance, excluding the full spectrum. We speculate this is due to the low EEG frequency of delta waves, meaning they are less affected by noise compared to the higher frequency bands. This translates to a lower error or higher accuracy. However, the superior performance obtained with full spectrum could mean smiles are related to gamma waves more closely than beta, alpha, theta, and delta waves. Gamma waves are associated

with brain hyperactivity, which the experimental protocol could fall under in terms of brain stimulation. This could also mean that smile expression is not frequency-specific, meaning that smiling would not necessarily be more prominently visible in any single band over another, and it may just follow different patterns than we would think from our knowledge about these brainwaves. We could just as well conclude that some features could be lost in the filtering process itself or could even leak from other bands, as perfect brick-wall filters are not realistic. Consequently, some information that could be included in the full spectrum might still be available in the theoretical range of any of the other bands.

The overall pattern obtained in our study indicates that convolution neural networks can discern EEG readings, at least in regards to smiling, with reasonable accuracy using minimal pre-processing prior to training. Our findings motivate further research into the application of Brain Computer Interfaces within the medical interventions in psychology, cognitive neuroscience or otherwise in applications pertaining to rehabilitation.

LIMITATIONS AND FUTURE WORK

Although, we have achieved high classification accuracy using CNNs on the temporal variations of EEG, the classification performance should be further improved for real-world applications. One possible way to improve the classification accuracy can be by combining the proposed CNNs models with Recurrent Neural Networks (RNN) or LSTMs which have good capability of learning from time series data. Another way to improve the classification accuracy of our work is to obtain the functional connectivity network and graph theory analysis and use them as an input to the proposed classification models as suggested in [44]. In addition, the proposed classification models should be generalized and tested with the new study samples, which may be captured under different conditions.

V. CONCLUDING REMARKS

We demonstrated one of the first studies to discriminate genuine smiles, fake smiles and neutral expressions with the usage of EEG signals and CNNs. Our model achived classification accuracy of 90.45% using EEG full spectrum data with subject-dependant experimentation. For the other bands, we achieved classification accuracies of 87.4%, 88.3%, 89.7%, and 90.0% using Beta, Alpha, Theta, and Delta bands respectively. We also compare the performance of the proposed models with existing classifiers such as ANN, LSTM and SVM and found that our CNN models outperform traditional classifiers.

ACKNOWLEDGMENT

The authors would like to thank the Biosciences and Bioengineering Research Institute for its support of this work and also would like to thank the Department of Electrical Engineering, American University of Sharjah, Sharjah, United Arab Emirates. This paper represents the opinions of the authors and does not mean to represent the position or opinions of the American University of Sharjah.

REFERENCES

[1] O. Déniz, M. Castrillon, J. Lorenzo, L. Anton, and G. Bueno, "Smile detection for user interfaces," in *Proc. Int. Symp. Vis. Comput.* Berlin, Germany: Springer, 2008, pp. 602–611.

[2] H. Wu, Y. Liu, Y. Liu, and S. Liu, "Fast facial smile detection using convolutional neural network in an intelligent working environment," *Infr. Phys. Technol.*, vol. 104, Jan. 2020, Art. no. 103061.

[3] M. E. Hoque, D. J. McDuff, and R. W. Picard, "Exploring temporal patterns in classifying frustrated and delighted smiles," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 323–334, Jul. 2012.

[4] T. J. Sullivan, S. R. Deiss, and G. Cauwenberghs, "A low-noise, non-contact EEG/ECG sensor," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, Nov. 2007, pp. 154–157.

[5] Y. M. Chi, S. R. Deiss, and G. Cauwenberghs, "Non-contact low power EEG/ECG electrode for high density wearable biopotential sensor networks," in *Proc. 6th Int. Workshop Wearable Implant. Body Sensor Netw.*, Jun. 2009, pp. 246–250.

[6] Y. M. Chi and G. Cauwenberghs, "Wireless non-contact EEG/ECG electrodes for body sensor networks," in *Proc. Int. Conf. Body Sensor Netw.*, Jun. 2010, pp. 297–301.

[7] M. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space," *IEEE Trans. Affect. Comput.*, vol. 2, no. 2, pp. 92–105, Apr. 2011.

[8] A. Frydenlund and F. Rudzicz, "Emotional affect estimation using video and EEG data in deep neural networks," in *Proc. Can. Conf. Artif. Intell.* Cham, Switzerland: Springer, 2015, pp. 273–280.

[9] L. F. Barrett, "Discrete emotions or dimensions? The role of valence focus and arousal focus," *Cognition Emotion*, vol. 12, no. 4, pp. 579–599, 1998.

[10] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Personality Social Psychol.*, vol. 17, no. 2, p. 124, 1971.

[11] C. Crivelli, P. Carrera, and J.-M. Fernández-Dols, "Are smiles a sign of happiness? Spontaneous expressions of judo winners," *Evol. Human Behav.*, vol. 36, no. 1, pp. 52–58, Jan. 2015.

[12] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma, and O. Krejcar, "DeprNet: A deep convolution neural network framework for detecting depression using EEG," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[13] M. Alex, U. Tariq, F. Al-Shargie, H. S. Mir, and H. A. Nashash, "Discrimination of genuine and acted emotional expressions using EEG signal and machine learning," *IEEE Access*, vol. 8, pp. 191080–191089, 2020.

[14] F. Al-Shargie, U. Tariq, M. Alex, H. Mir, and H. Al-Nashash, "Emotion recognition based on fusion of local cortical activations and dynamic functional networks connectivity: An EEG study," *IEEE Access*, vol. 7, pp. 143550–143562, 2019.

[15] Z. Liu, A. Xu, Y. Guo, J. U. Mahmud, H. Liu, and R. Akkiraju, "Seemo: A computational approach to see emotions," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–12.

[16] E. G. Krumhuber, A. Kappas, and A. S. R. Manstead, "Effects of dynamic aspects of facial expressions: A review," *Emotion Rev.*, vol. 5, no. 1, pp. 41–46, Jan. 2013.

[17] K. Pasupa, P. Chatkamjuncharoen, C. Wuttilertdeshar, and M. Sugimoto, "Using image features and eye tracking device to predict human emotions towards abstract images," in *Image and Video Technology* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2015, pp. 419–430.

[18] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and classification of emotion recognition based on EEG brain-computer interface system research: A systematic review," *Appl. Sci.*, vol. 7, no. 12, p. 1239, Dec. 2017.

[19] X. Hu, C. Zhuang, F. Wang, Y.-J. Liu, C.-H. Im, and D. Zhang, "FNIRS evidence for recognizably different positive emotions," *Frontiers Human Neurosci.*, vol. 13, p. 120, Apr. 2019.

[20] J. C. Britton, K. L. Phan, S. F. Taylor, R. C. Welsh, K. C. Berridge, and I. Liberzon, "Neural correlates of social and nonsocial emotions: An fMRI study," *NeuroImage*, vol. 31, no. 1, pp. 397–409, May 2006.

[21] Z. Zhang, V. Singh, T. E. Slowe, S. Tulyakov, and V. Govindaraju, "Real-time automatic deceit detection from involuntary facial expressions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–6.

[22] M. Z. Hossain and T. Gedeon, "Observers' physiological measures in response to videos can be used to detect genuine smiles," *Int. J. Hum.-Comput. Stud.*, vol. 122, pp. 232–241, Feb. 2019.
[23] S. Petridis, B. Martinez, and M. Pantic, "The MAHNOB laughter database," *Image Vis. Comput.*, vol. 31, no. 2, pp. 186–202, Feb. 2013.
[24] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Acted facial expressions in the wild database," Austral. Nat. Univ., Canberra, Australia, Tech. Rep. TR-CS-11, vol. 2, p. 1, 2011.
[25] M. Z. Hossain, T. Gedeon, and R. Sankaranarayana, "Using temporal features of Observers physiological measures to distinguish between genuine and fake smiles," *IEEE Trans. Affect. Comput.*, vol. 11, no. 1, pp. 163–173, Jan. 2020.
[26] D. O. Bos, "EEG-based emotion recognition," *Influence Vis. Auditory Stimuli*, vol. 56, no. 3, pp. 1–17, 2006.
[27] J. A. Miranda-Correa and I. Patras, "A multi-task cascaded network for prediction of affect, personality, mood and social context using EEG signals," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 373–380.
[28] H. Xu and K. N. Plataniotis, "Affective states classification using EEG and semi-supervised deep learning approaches," in *Proc. IEEE 18th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2016, pp. 1–6.
[29] J. M. Williams, "Deep learning and transfer learning in the classification of EEG signals," Ph.D. dissertation, Dept. Comput. Sci. Eng., Univ. Nebraska, Nebraska, 2017.
[30] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
[31] Z. Gao, R. Li, C. Ma, L. Rui, and X. Sun, "Core-brain-network-based multilayer convolutional neural network for emotion recognition," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
[32] S.-E. Moon, S. Jang, and J.-S. Lee, "Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2556–2560.
[33] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.
[34] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 57–67, 2018.
[35] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Shalaby, "EEG-based emotion recognition using 3D convolutional neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 329–337, Jan. 2018.
[36] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality and mood research on individuals and groups," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 479–493, Apr. 2021.
[37] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jun. 2012.
[38] E. S. Dan-Glauser and K. R. Scherer, "The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance," *Behav. Res. Methods*, vol. 43, no. 2, pp. 468–477, Jun. 2011.
[39] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
[40] E. Alpaydin, *Introduction to Machine Learning* (Adaptive computation and machine learning), 4th ed. Cambridge, MA, USA: MIT Press, 2020. [Online]. Available: https://mitpress.mit.edu/books/introduction-machine-learning-fourth-edition
[41] R. Gonzalez and R. Woods, *Digital Image Processing*, 4th ed. New York, NY, USA: Pearson, 2018.
[42] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[43] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[44] F. M. Al-Shargie, O. Hassanin, U. Tariq, and H. Al-Nashash, "EEG-based semantic vigilance level classification using directed connectivity patterns and graph theory analysis," *IEEE Access*, vol. 8, pp. 115941–115956, 2020.

MOSTAFA M. MOUSSA (Student Member, IEEE) received the B.S. degree in electrical engineering, in 2018, and the M.S. degree in biomedical engineering from the American University of Sharjah, United Arab Emirates, in 2020. He has worked as a Research Assistant for the duration of his master's and the following two years with AUS. His current research interests include human augmentation and rehabilitation, biomimicry, human–computer interaction, cognition, neuroscience, medical devices and robotics, computer-aided diagnosis (CAD), and deep learning in medical applications.

USMAN TARIQ (Member, IEEE) received the M.S. and Ph.D. degrees from the Electrical and Computer Engineering Department, University of Illinois at Urbana–Champaign (UIUC), in 2009 and 2013, respectively. He is currently a Faculty Member of the Department of Electrical Engineering, American University of Sharjah (AUS), United Arab Emirates. Before AUS, he worked as a Research Scientist with the Xerox Research Center Europe, Computer Vision Group, France. His research interests include computer vision, image processing, and machine learning, in general while facial expression recognition and face biometrics, in particular.

FARES AL-SHARGIE (Member, IEEE) received the B.S. and M.S. degrees in biomedical engineering from Multimedia University, Malaysia, and the Ph.D. degree in biomedical engineering from Universiti Teknologi PETRONAS, Malaysia. He worked closely with several biomedical engineering departments and companies, including Hitachi Ltd., Research and Development Group, Japan. He is the first author in more than 30 journals and conference papers, one book, and one book chapter. His current research interests include the assessment of mental stress, vigilance, and emotions via, EEG, fNIRS neuroimaging modalities, and eye tracking. He is a member of the Society of Functional Near-Infrared Spectroscopy.

HASAN AL-NASHASH (Senior Member, IEEE) is currently a Professor (a Former) Director of the Biosciences and Bioengineering Research Institute and the Former Chair of the Department of Electrical Engineering, American University of Sharjah. The main themes of his research work are in the general areas of neuroengineering, signal processing, and microelectronics. He designed and developed several electronic instruments to measure various biodynamic parameters. He is leading the effort to establish the Biosciences and Bioengineering Research Institute, AUS, and has previously led the effort to establish the M.S. graduate program in biomedical engineering. He played an active role in organizing several biomedical and electrical engineering conferences. He worked closely with several biomedical engineering departments and hospitals, including the National University of Singapore, Johns Hopkins University, and Rashid Hospital, Dubai. He is the author of more than 100 journals and conference papers, five book chapters, and two issued U.S. patents. He is the former Middle East and Africa representative on the IEEE-EMBS Administrative Committee.