IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 March 2023, accepted 3 May 2023, date of publication 15 May 2023, date of current version 18 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3276244

RESEARCH ARTICLE

Real-Time On-Chip Machine-Learning-Based Wearable Behind-The-Ear Electroencephalogram Device for Emotion Recognition

NGOC-DAU MAI, HA-TRUNG NGUYEN, AND WAN-YOUNG CHUNG, (Senior Member, IEEE)
Department of Artificial Convergence, Pukyong National University, Busan 48513, South Korea

Corresponding author: Wan-Young Chung (wychung@pknu.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government through Ministry of Science and ICT (MSIT) under Grant NRF-2019R1A2C1089139.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Pukyong National University Institutional Review Board under Application No. 1041386-202003-HR-11-02.

ABSTRACT In this study, we propose an end-to-end emotion recognition system using an ear-electroencephalogram (EEG)-based on-chip device that is enabled using the machine-learning model. The system has an integrated device that gathers EEG signals from electrodes positioned behind the ear; it is more practical than the conventional scalp-EEG method. The relative power spectral density (PSD), which is the feature used in this study, is derived using the fast Fourier transform over five frequency bands. Directly on the embedded device, data preprocessing and feature extraction were carried out. Three standard machine learning models, namely, support vector machine (SVM), multilayer perceptron (MLP), and one-dimensional convolutional neural network (1D-CNN), were trained on these rich emotion classification features. The traditional approach, which integrates a model into the application software on a personal computer (PC), is cumbersome and lacks mobility, which makes it challenging to use in real-life applications. Besides, the PC-based system is not sufficiently real-time because of the connection latency from the EEG data acquisition device. To overcome these limitations, we propose a wearable device capable of performing on-chip machine learning and signal processing on the EEG data immediately after the acquisition task for the real-time result. In order to perform on-chip machine learning for the real-time prediction of emotions, 1D-CNN was chosen as a pre-trained model using the relative PSD characteristics as input based on the evaluation of the set results. Additionally, we developed a smartphone application that alerted the user whenever a negative emotional state was identified and displayed the information in real life. Our test results demonstrated the feasibility and practicability of our embedded system for real-time emotion recognition.

INDEX TERMS Electroencephalogram (EEG), emotion recognition, tiny machine learning, real-time EEG system, power spectral density (PSD), multilayer perceptron (MLP), support vector machine (SVM), one-dimensional convolutional neural network (1D-CNN).

I. INTRODUCTION

Emotion plays an important role in daily human life because they directly affect our ability to make decisions [1], [2]. Much attention has been paid to investigating and exploring the different methods of interaction and communication between humans and machines. Of particular interest has been the area of enabling intelligent machines to understand human emotions. Over the centuries, most methods for recognizing human emotions have been based on facial expressions, speech, and gestures [3]. Although these techniques have produced good results, they are still constrained by a number of practical issues and are subject to human control. Methods using bio-signals have recently emerged as an area

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kafiul Islam.

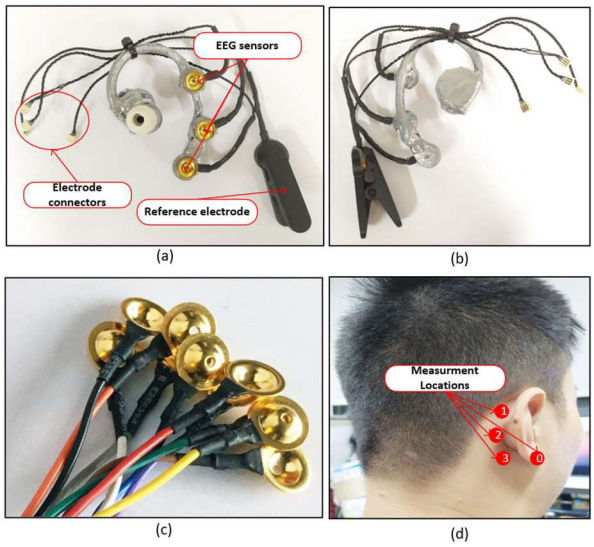

FIGURE 1. (a), (b) Back and the front side of the ear-hook with behind-the-ear EEG sensors; (c) The proposed device; (d) EEG measurement locations behind the ear.

TABLE 1. Summary of the technical parameters of the proposed device.

PCB Part	Features	Specification
Portenta H7	MCU	STM32H747 dual-core processor (480Mhz for Cortex-M7, 240Mhz for Cortex-M4) with 2Mbytes of Flash, 1Mbyte of SRAM
	Wireless Communication	BLE 5.2 with 2M PHY
	PMU	MC34PF1550A4EP
PH7-EEG Carrier	ADC	ADS1299 (eight-channel, low-noise, 24-bit, simultaneous-sampling delta-sigma (ΔΣ) analog-to-digital converters)
	Sampling Rate	250Hz
	Gain	24
	Number of Channels	8
	Electrode type	Dry gold cup

of immense interest in emotion recognition [4]. The commonly used bio-signals include body temperature, heart rate, electrocardiogram, and electroencephalogram (EEG) [5]. EEG signals have demonstrated their potential in emotion recognition [6]. EEG is an electrophysiological monitoring approach to record the cerebral electrical activity on the skin, typically by placing electrodes on the scalp [7]. Ear EEG is a technique that uses electrodes positioned in and around the ear to monitor brain activity [8]. Its superiority over the traditional EEG measurements that use electrodes placed on the scalp is its greater invisibility and wearer mobility; however, ear EEG has low signal amplitude [8]. Ear-EEGs are divided into two main groups based on the two measurement locations: (i) in-the-ear EEG that measures the signals from the areas within the concha and the ear canal [9] and (ii) behind-the-ear EEG that measures signals from different positions behind the ear lobe [10]. Currently, almost all emotion detection systems using EEG signals are computer-based and consist of two main parts: (i) an EEG-acquiring device having wearable wireless capability (Bluetooth) for data transmission and (ii) a computer for performing the emotion classification task [11]. This system is cumbersome and inconvenient because of the latency of the wireless technology; the system is also not real-time enough for the applications that require immediate results. Power spectral density (PSD) for EEG-based emotion recognition is an important feature that has proved effective in numerous studies [12], [13]. In this study, the PSD approach is used to decompose each EEG signal into five distinct frequency ranges, namely, delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50Hz). Then, the percentage of PSD that each EEG band occupies is calculated for feature extraction. By combining the functionality of an on-chip data processing system into a single device, the system may be utilized for real-time applications. Furthermore, using the behind-the-ear EEG instead of the scalp-EEG also makes the device more compact and comfortable for the daily user [14]. Machine learning has been increasingly integrated with EEG in various domains, including emotional detection, neural feedback training, epilepsy, rehabilitation, mental workload, and other fields [15]. In the realm of emotional detection, machine learning algorithms can be employed to analyze EEG signals for the identification and classification of diverse emotional states such as happiness, sadness, anger, and anxiety [16]. In stroke management, real-time health monitoring systems like HealthSOS [17] have incorporated machine learning techniques to predict the prognosis of stroke. Moreover, machine learning has been implemented in advanced driver-assistance systems to identify neurological biomarkers induced by driving [18]. In sleep studies, machine learning has been utilized to assess EEG-biomarkers to predict different sleep stages [19]. Tiny ML is a growing area of interest for implementing machine learning to embedded systems and investigating various models that can operate on low-powered devices such as mobile phones or microcontrollers [20]. Therefore, in this study, we developed an embedded device that deploys a tiny machine learning (Tiny ML) model using EEG signals measured by electrodes placed behind the ear for emotion detection applications. We investigated the performance of three common models for emotion classification: the support vector machine (SVM), multilayer perceptron (MLP), and one-dimensional convolutional neural network (1D-CNN). The model with the highest level of performance accuracy was chosen for use in our proposed embedded device.

The main contributions of this research can be summarized as follows: Firstly, a thorough hardware and firmware design of our wearable customed-designed behind the ear EEG device was developed for direct on-chip data

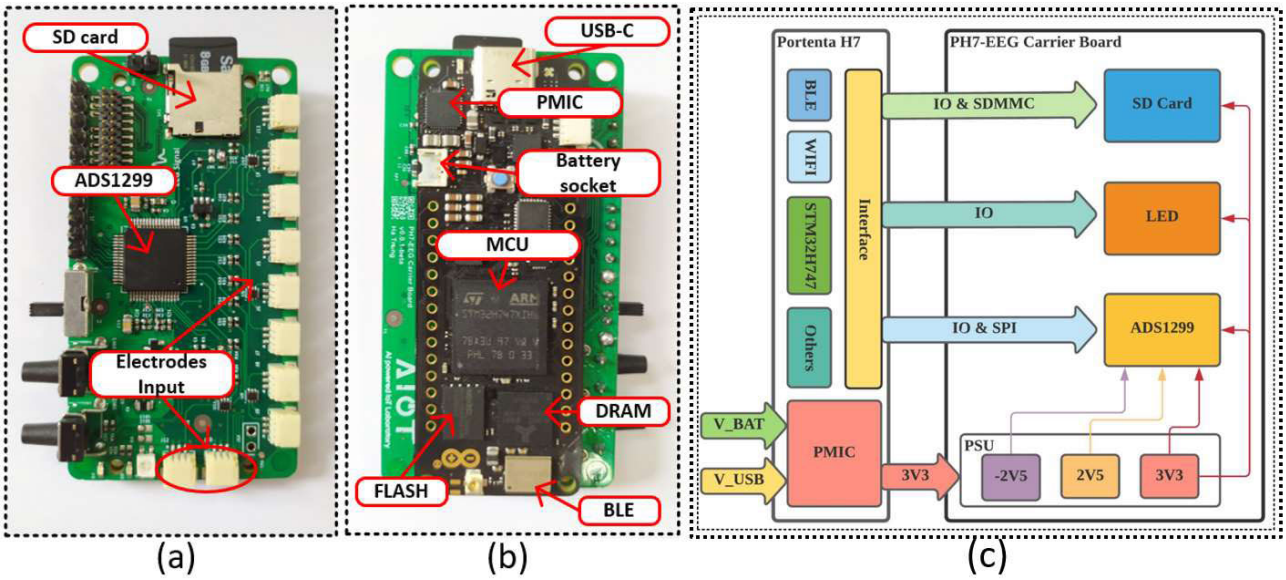

FIGURE 2. (a) First PCB part of the proposed device; (b) Second PCB part of the proposed device; (c) The block diagram of the proposed device.

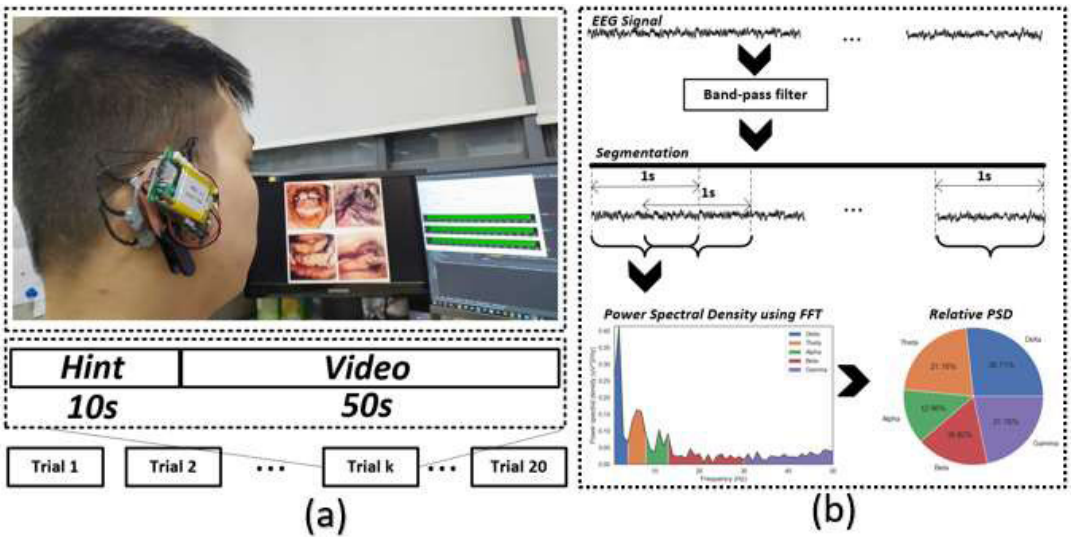

FIGURE 3. (a) Experiment Protocol, (b) Feature extraction from the relative PSD using FFT.

collection and processing. Secondly, the performance of a proposed 1D-CNN model with hyperparameter tuning was evaluated and compared with two other proposed models, MLP and SVM, for emotion recognition using ear-EEG signals collected from the device, on both subject-dependent and subject-independent cases. Finally, the practical application of a real-time behind-the-ear EEG-based emotion recognition system was demonstrated. The entire process of data collection, preprocessing, and deploying and running the model was performed directly on the real-time on-chip device. In addition, a smartphone application was developed for receiving information transmitted from the device through Bluetooth low-energy (BLE) wireless technology, and for displaying and performing a negative-state warning task for the user.

II. PROPOSED SYSTEM

A. BEHIND-THE-EAR EEG SENSORS

A scalp-EEG system is often required for a headset to provide the most comprehensive coverage of a user's head [21]. This drawback leads to a cumbersome system that is inconvenient to the user; the setup stage is time-consuming, and the system restrains the EEG potential only for clinical-grade applications. The ear EEG method is a novel approach to brain signal acquisition that overcomes the constraints of traditional EEG-based BCI systems. The EEG signals in this study are obtained using passive electrodes similar to those used for scalp-EEGs, except they are positioned around the ears. As a result, ear EEG setup is significantly simpler and less time-consuming than scalp EEG. Our proposed ear-EGG approach

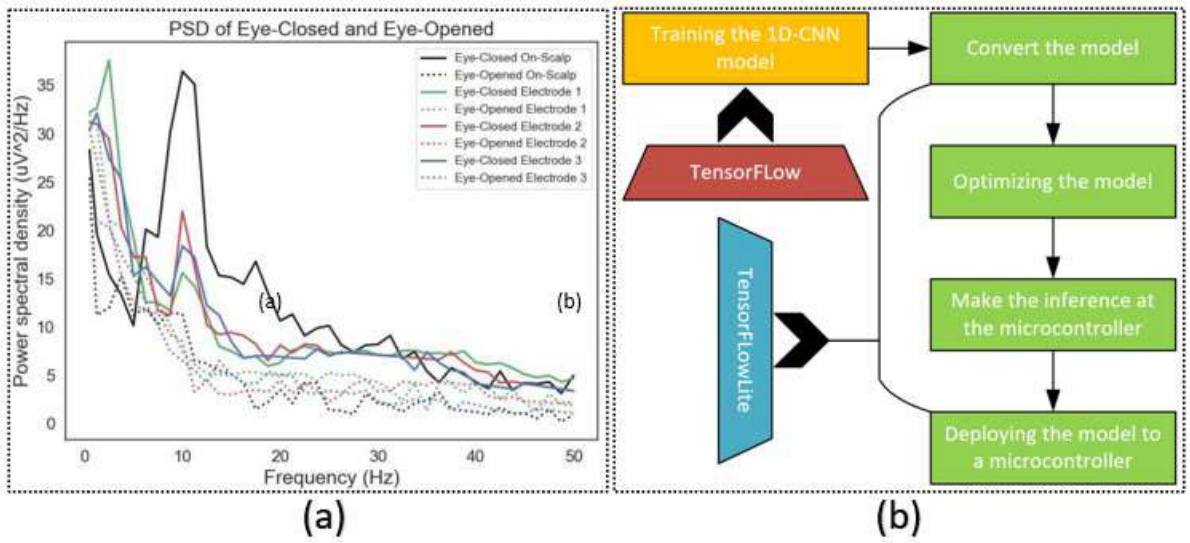

FIGURE 4. (a) Power spectral densities (PSDs) during eyes-closed and eyes-opened resting state from On-Scalp and the behind-ear locations; (b) Use of TensorFlow and TensorFlow Lite to deploy a model on our embedded device.

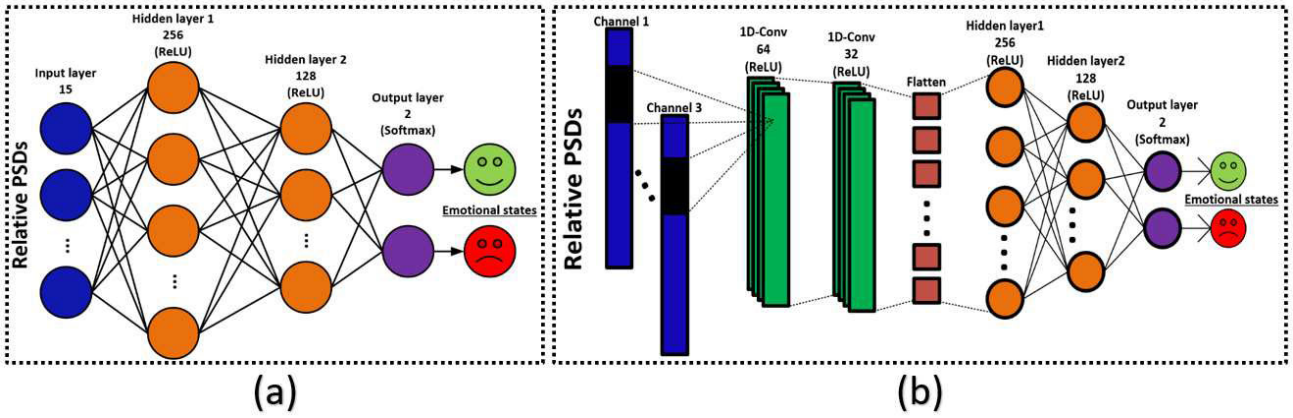

FIGURE 5. (a) Proposed MLP model for on-chip classification; (b) Proposed 1D-CNN model for on-chip classification.

used three electrodes placed behind the right ear, where the potential amplitude for the visual stimuli was excellent. EEG signals were acquired from three distinct locations located in the mastoid region positioned posterior to the right ear. These locations were selected due to the ease of signal acquisition compared to on-scalp locations, as well as the reduced susceptibility to interference from extraneous sources such as ocular movements or muscular artifacts. The precise locations of the measurement points are shown in Fig. 1.

B. DEVICE SPECIFICATION

The wearable device (shown in Fig. 1 (c)) combines a high-performance microprocessor for edge machine learning with an accurate analog front-end (AFE) for EEG data acquisition. The device is appropriate for the wearable scenario since it is battery-powered and has an archival size of 6.5 cm 3.5 cm 0.5 cm. Two printed circuit board (PCB) plates were used in the proposed device. The first plate of the device (see Fig. 2 (a)) was the PCB used to acquire the EEG signal using the AFE for analog-to-digital conversion (TI ADS1299); the device featured eight signal pick-up channels (only three of which worked in the study) [22]. The ADS1299 was powered using bipolar supply settings in this configuration. As a result, the power supply unit of this PCB is responsible for converting the primary voltage supply of 3V3 from the second part to three voltage levels: 3V3, 2V5, and 2V5. The passive dry contact gold cup electrode was utilized for data collection; this electrode is extremely effective and simple to use. In addition, an SD card was included in the design for data logging, and three light-emitting LEDs, namely, power-on, wireless connection, and battery, were employed to designate device information. The PortentaH7 module (shown in Fig. 2 (b)) is the device's second plate and has three major functions: power supply, control, and wireless communication. A dual-core 32-bit microcontroller STM32H747XI with a maximum frequency of 480 Mhz

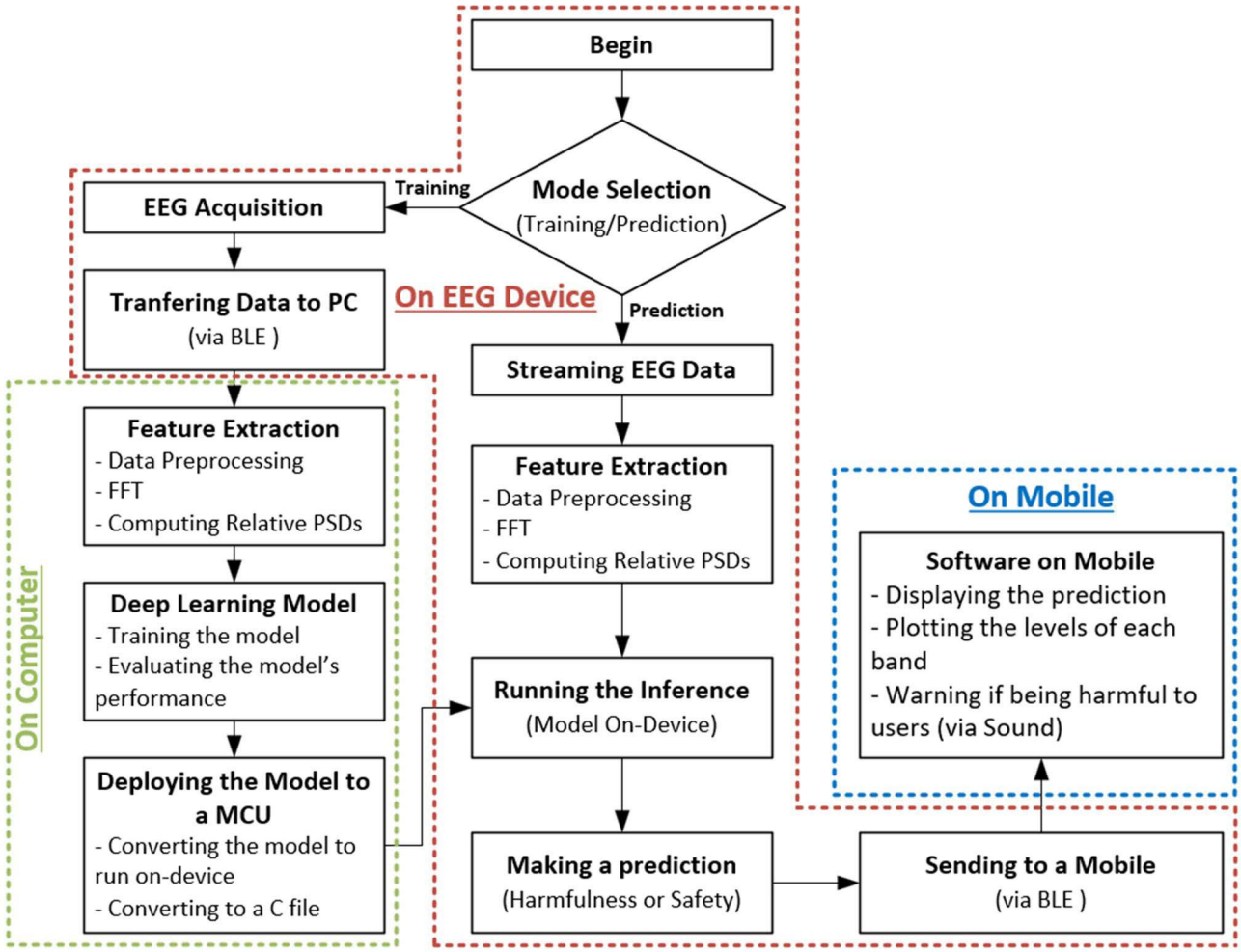

FIGURE 6. Functional block diagram of the proposed intelligent EEG embedded system.

and 2 Mb SDRAM performs the control function. In real-time, PortentaH7 can effectively run signal processing tasks created with TensorFlow Lite. Furthermore, the radio module Murata 1DX allows for 5.2 BLE communication and WiFi 802.11b/g/n 65 Mbps on the PortentaH7. To increase the energy efficiency of the device, a power supply unit was designed using the power management (PMU) integrated chip MC34PF1550A4EP, which functioned as a buck converter, a low-dropout regulator, and a battery charger. This solution allowed the device to operate with power from both the USB cable and the battery. A micro-USB connector is used to charge the LiPo battery. Fig. 2(c) shows the device's block diagram. Table 1 shows the hardware technical specifications of our EEG equipment in further detail.

We conducted a closed-eye and open-eye evaluation to analyze the obtained signal quality of our device. Signals were obtained from three measurement locations 1, 2, and 3 behind the ears, as well as two measurement locations O1-O2 on the scalp. The Welch technique is used to compute the power spectral density (PSD) using a 0.5s window on both the EEG signals behind the ears and on the scalp.

TABLE 2. t-Test to determine whether there is a significant difference between the two emotional states.

Electrode	t-test (p-value)				
	Delta	Theta	Alpha	Beta	Gamma
1	0.0263	0.0712	0.0029	0.0135	0.4309
2	0.0036	0.0039	0.0012	0.0063	0.0041
3	0.0037	0.0213	0.1575	0.0994	0.0092

The EEG alpha attenuation responses from these measurement sites with eyes open and closed are shown in Fig. 4 (a). For all four signals, the neuronal alpha wave peaks at around 10 Hz. However, PSD varies markedly between signals recorded at the scalp and behind the ear. The PSD of the signal from O1-O2 (on the scalp) is more prominent than the PSDs of the other three signals. Previous studies [10] have also shown that the signal recorded at the ear is less influenced by artifacts caused by blinking or opening and closing of

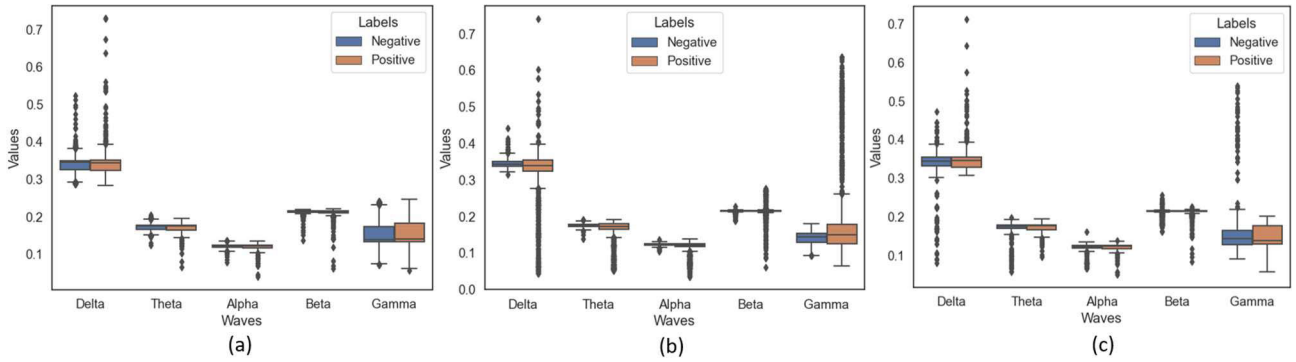

FIGURE 7. Features extracted from three electrodes for two emotionals states: (a) electrode 1, (b) electrode 2, (c) electrode 3.

the eyes. The PSD intensity with the eye open is relatively comparable for all signals.

C. EXPERIMENT PROTOCOL

Fourteen healthy people (7 males and 7 females aged 24 to 45 years) took part in the experiment; none of them had any visual impairment. Each participant was comfortably seated in front of a computer. A photo, a highlighted title, or a brief descriptive sentence provided as the emotional stimulus material. To begin a trial, the individual was required to comprehend the instructions displayed on the screen. For the first 10 seconds, the trigger of a new trial was displayed. Following that, the stimuli were exhibited on the screen at random for 50 seconds. The 20 videos given to the subjects were produced and chosen from a variety of sources, including YouTube, Twitter, Instagram, Facebook, Pinterest, BBC News, The New York Times, Breaking News, and others. All participants were presented with an equal number of photographs depicting each of the two emotional states, and the resulting emotional responses were recorded while they viewed the emotional stimuli. The experiment was given permission by the Pukyong National University Institutional Review Board with the number 1041386-202003-HR-11-02.

D. PROCESSING ON THE EEG DEVICE

The EEG-based embedded device allowed two modes: training and prediction. In the training mode, the EEG data collected from the points behind the ear were sent to the computer for processing, building models, and converting into a format that the embedded device could interpret. For the prediction mode, the data was collected in real-time. These data were initially preprocessed and then utilized to extract important features from the relative PSDs using the fast Fourier Transform (FFT) after a specified duration. These features are fed into the previously trained model, which is then deployed in the embedded device. Following that, an emotional state prediction was produced using TensorFlow Lite to execute the inference. Finally, the output was sent through the BLE protocol to a pre-built and installed smartphone application. Fig. 6 depicts the system's functional block diagram.

E. PROCESSING ON THE COMPUTER

The data from the embedded device was stored and processed on the computer. To begin, preprocessing and feature extraction were carried out in the same manner as on the EEG equipment. Following that, specific models for training on the rich features obtained were proposed. The models were then developed using TensorFlow. Finally, the models were evaluated and assessed in order to determine the most suitable model for deploying the embedded device. TensorFlow Lite was used to aid with model deployment.

F. PROCESSING ON THE MOBILE

A smartphone application built with Android Studio and then pre-installed on the phone received and displayed the data sent to the embedded device via the BLE protocol. The information. shown and executed included the predicted emotional state and its accuracy and the ratio of the relative PSDs from five frequency bands. Furthermore, it alerted the user immediately if any malicious content was detected.

III. METHODOLOGY

A. FEATURE EXTRACTION

First, EEG signals are band-pass filtered from 1 Hz to 50 Hz to remove unwanted noise. Then, all the EEG data obtained is divided into segments having 1-s length and 50% overlap. Feature extraction is an essential part of pattern recognition. Various methods have been proposed to extract features from physiological signals. In this study, FFT is used to derive the spectral features from the five EEG frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50 Hz). Then, we take the squared magnitude of the FFT to compute the PSD for each frequency band. Finally, the relative PSD of each EEG band is calculated as follows:

$$rPSD_i = \frac{PSD_i}{\sum_1^5 PSD_i}, \tag{1}$$

where PSD_i and $rPSD_i$ are the PSD and relative PSD values of the ith band ($i = 1, 2, 3, 4$, and 5 corresponding to the delta, beta, alpha, theta, and gamma bands, respectively).

TABLE 3. Model hyperparameter tuning for 1d-cnn and mlp models.

Model name	Layer no	Layer and parameter type	Hyper-parameters for tuning	The best Hyper-parameters after tuning	Activation
1D-CNN	0	Input	(None, 3, 5)	(None, 3, 5)	
	1	Conv1D	[64, 32, 16] * 3	64* 3	Relu
	2	MaxPooling1D	2	2	
		Dropout	[0.25, 0.5, 0.7]	0.5	
	3	Conv1D	[64, 32, 16] * 3	32* 3	Relu
	4	MaxPooling1D	2	2	
		Dropout	[0.25, 0.5, 0.7]	0.25	
	5	Flatten			
	6	Dense	[256, 128, 64]	256	Relu
		Dropout	0.5	0.5	
	7	Dense	[256, 128, 64]	128	Relu
		Dropout	[0.25, 0.5, 0.7]	0.25	
	8	Output	2	2	Softmax
		Optimizer	[Adam, SGD]	Adam	
		Learning Rate	[0.01, 0.001, 0.0001]	0.001	
		Loss	Categorical cross entropy	Categorical cross entropy	
		Batch size	[32, 64, 128]	64	
		Epoch	[500, 1000, 1500]	1000	
MLP	0	Input	(None, 15)	(None, 15)	
	1	Dense	[256, 128, 64]	256	Relu
		Dropout	0.5	0.5	
	2	Dense	[256, 128, 64]	128	Relu
		Dropout	[0.25, 0.5, 0.7]	0.5	
	3	Output	2	2	Softmax
		Optimizer	[Adam, SGD]	Adam	
		Learning Rate	[0.01, 0.001, 0.0001]	0.001	
		Loss	Categorical cross entropy	Categorical cross entropy	
		Batch size	[32, 64, 128]	64	
		Epoch	[500, 1000, 1500]	1000	

B. CLASSIFICATION

1) SUPPORT VECTOR MACHINE

The SVM is a powerful machine learning model for classification; this model separates a given set of binary labeled training data and finds a hyperplane such that the margin between the two classes is the maximum possible [23]. In this study, Linear, the most generalized form of kernelization, is used as the kernel of the SVM classifier because it scales well to a large number of training samples.

2) NEURAL NETWORK-BASED MODELS

In recent years, neural network-based models have shown great promise in extracting features and learning patterns from complex data. These models have improved our understanding of EEG signals because of their powerful ability to learn valuable information from the brain. Two standard models based on neural networks, including MLP and 1D-CNN, are used in this study. MLP is an artificial feedforward neural network class [24] with at least three layers: an input layer, a hidden layer, and an output layer. MLP uses backpropagation as a supervised learning approach for training models. Meanwhile, 1D-CNN is a variant of CNN [25], a regularized version of MLP, which is commonly used for biological signals, such as EEG signals. A 1D-CNN model can automatically learn important information from EEG signals and then use them for classification via a classifier to return results, unlike the traditional hand-engineered approaches. One of the crucial components of a 1D-CNN network is the Conv layer, which does not appear in the MLP network discussed above. The Conv layer consists of learnable filters that reduce the number of free parameters, making the network more profound.

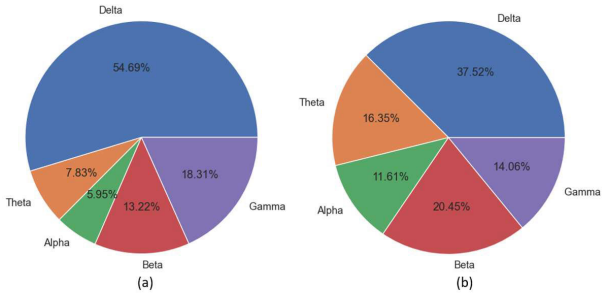

FIGURE 8. The average distribution of relative PSD from all electrodes over five frequency bands between 2 emotional states: (a) negative and (b) positive.

Optimizing the values of hyperparameters, such as learning rate, batch size, and number of layers, is a crucial step in creating a deep learning model that performs well [26]. Hyperparameter tuning can be a challenging and time-consuming process as it requires exploring a vast search space to find the

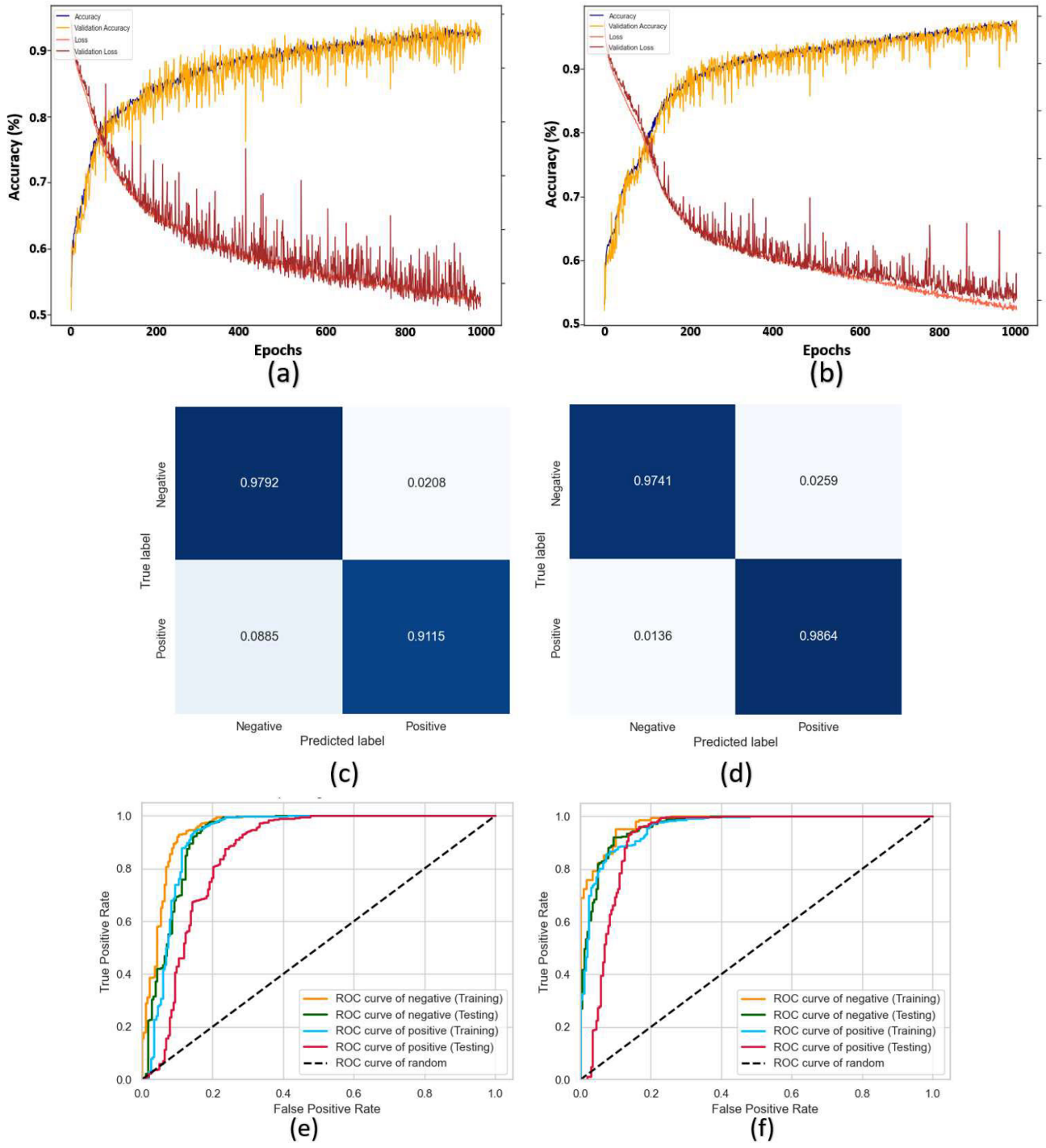

FIGURE 9. Classfication results for MLP (a, c, e), and 1D-CNN (b, d, f) models on Subject 2.

optimal values. To achieve the best performance and minimize error rates, various techniques such as grid search, random search, and Bayesian optimization can be used. Proper hyperparameter tuning is essential to building a deep learning model that can generalize well and efficiently handle new data. In this study, Neural Network Intelligence (NNI) [27] with Bayesian optimization was used to tune hyperparameters for two proposed deep learning models. By defining a search space for the hyperparameters and specifying an objective function, NNI iteratively selects the next set of hyperparameters based on the probabilistic model and previous evaluation results. The hyperparameters that yielded the best performance for each model, as well as other model parameters, are presented in Table 3.

TABLE 4. Comparison of classification accuracy (%) between SVM, MLP, and 1D-CNN across ten users for user-dependent and user-independent cases.

User	Gender	Classification Accuracy (%)					
		User-Dependent Cases			User-Independent Cases		
		SVM	MLP	1D-CNN	SVM	MLP	1D-CNN
1	Female	88.93	91.59	95.19	77.54	82.89	85.57
2	Female	89.67	94.54	98.03	79.33	84.70	87.63
3	Female	79.65	89.27	93.24	70.58	82.38	84.32
4	Female	88.85	91.39	96.03	78.89	81.89	81.92
5	Female	86.45	90.92	95.03	76.07	80.25	85.46
6	Female	84.91	90.13	94.34	74.89	80.13	85.02
7	Female	89.84	93.78	97.78	80.15	83.54	86.46
8	Male	87.56	91.01	94.76	78.87	82.01	81.51
9	Male	82.56	87.56	93.42	73.43	78.94	80.67
10	Male	84.93	90.23	93.91	75.02	80.54	83.94
11	Male	81.34	91.56	93.23	76.52	82.13	85.43
12	Male	83.29	90.43	95.73	77.45	80.19	85.67
13	Male	78.95	91.24	94.17	74.97	81.95	86.12
14	Male	85.76	92.13	93.29	76.92	79.97	85.90
Mean		85.19	91.13	94.87	76.47	81.54	84.68
Std		3.50	1.67	1.52	2.48	1.54	1.94

The proposed MLP model includes four layers shown in Fig. 5 (a) with two hidden layers. The size of the input layer varies with the number of channels used during feature extraction. The first and second hidden layers consist of 256 neurons and 128 neurons, respectively. The proposed 1D-CNN model includes two layers of 1D-Conv with 256 neurons, followed by two hidden layers with 256 and 128 neurons (see Fig. 5 (b)). Rectified linear unit activation is used after each hidden layer's neuron, and SoftMax is accepted as the activation function of the output neurons. After each convolutional layer, a max-pooling layer is applied to reduce the computational time and prevent the model from overfitting. The Adam optimizer with an initial learning rate of 0.001 and categorical cross-entropy loss was chosen for this study.

C. MODEL DEPLOYMENT ON THE EMBEDDED DEVICE

For applying machine learning to embedded systems, Tiny ML is being increasingly used because it explores the types of models that can be operated on inadequate and low-powered devices, such as mobiles or microcontrollers [20]. Tiny ML offers many benefits, such as low latency, low power consumption, low bandwidth, and a high level of privacy [20], [25]. Tiny ML is an emerging field that has multiple applications in diverse areas, such as healthcare, entertainment, education, and security. The process of building and training models in this study uses a complete, flexible ecosystem of tools and libraries called TensorFlow [28]. To use a TensorFlow-based trained model for obtaining input data to generate an output, we need to apply TensorFlow's interpreter. However, TensorFlow's interpreter is designed to run models on advanced computers and servers, which

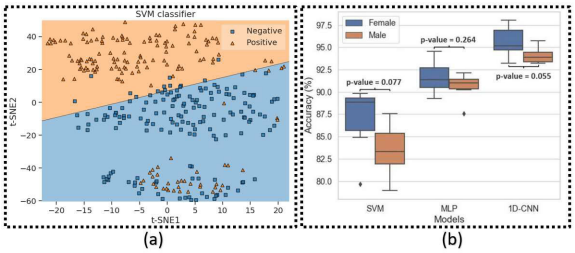

FIGURE 10. (a) Data visualization using t-SNE and SVM; (b) Comparison of classification accuracy between male and female subjects.

are more potent than embedded systems such as microcontrollers. TensorFlow Lite [29] has emerged to address this demand. TensorFlow Lite is used as an influential interpreter for running models on tiny and low-powered devices. The two main components of TensorFlow Lite when used for microcontrollers are TensorFlow Lite Converter and TensorFlow Lite Interpreter. TensorFlow Lite Converter is used to convert a TensorFlow model into a format proper for memory-constrained devices with useful optimizations, such as quantization to reduce the model size and latency with minimal or with no accuracy loss on tiny devices. The second principal component is TensorFlow Lite Interpreter, which executes an on-device TensorFlow Lite model to make predictions based on the input data.

IV. RESULTS

A. EEG FEATURES

Using the same method, the feature extraction procedure was carried out on a computer and an embedded device. After collecting data from the sensors mounted behind the ear, the

TABLE 5. Comparison of the relevant information between our proposed system and previous studies.

Author	Experimental Data Set	Methodology	Experimental Results	Operating on PC or on chip
Bhosale et al. [33]	DEAP	An adaptation method based on meta learning without requiring any fine-tuning of the pre-trained model	2 classes, subject-independent modelling (Valence: 76.46%, Arousal: 75.81%)	On PC
Liu et al. [36]	DEAP	A domain adaption by subject clustering	2 classes, subject independent modelling (Valence: 73.9%, Arousal: 68.8%	On PC
Wang et al. [37]	SEED, DEAP	A few-label adversarial domain adaption (FLADA)	subject-independent modelling (DEAP-2 classes:68.0%, SEED-3 classes: 89.32%)	On PC
Zhang et al. [38]	SEED	A LSTM-Capsule based knowledge distillation framework	3 classes, subject-independent modeling: 91.07%	On PC
Mohammadi et al. [39]	DEAP	Discrete wavelets transform-based features and KNN/SVM classification	2 classes, subject-dependent modeling (Valence: 86.75%, Arousal: 84.05%)	On PC
Li et al. [40]	SEED	2D map of differential entropy feature and hierarchical CNN model	2 classes (subject dependent: 0.882, subject independent: 0.802)	On PC
T. -H. Nguyen et al [34]	Self-made scalp EEG device	Power spectral density (PSD)-based frequency point features and MLP model	2 classes, subject-dependent modeling: 86.8%, subject-independent modeling: 61.2%	On PC
Athavipach C et al [35]	Self-made in-ear EEG device	Six statistical parameters and SVM classification	4 classes: 71.07% accuracy (valence), 72.89% accuracy (arousal), and 53.72% (all four emotions)	On PC
Aya_Hassouneh et all [41]	Emotiv EPOC device	Raw signals and LSTM model	6 classes: 87.25%	Realtime on PC
Our proposed system	Self-made behind-ear EEG device	Relative power spectral density (PSD) and 1D-CNN model with hyperparameter tuning	2 classes: subject-dependent modeling: 94.87%, subject-independent modeling: 84.68%	Realtime on chip

band-pass filter eliminated the noise to obtain the desired signals ranging from 1 Hz to 50 Hz. To ensure temporal correlation between the data points, the filtered data was divided into segments using a sliding window of 256 data points (1 s) and a 50% overlap during data preprocessing on computers. The segments were then split into five frequency bands using FFT to obtain an estimate of the relative PSD for each band: delta (1-4 Hz), theta (4-8 Hz), alpha (8-16 Hz), beta (16-32 Hz), and gamma (32-50 Hz). Fig. 3(b) depicts the entire feature extraction procedure. The study involved participants watching 20 videos designed to evoke emotions, with ten videos for each emotional state - Negative and Positive. Each video was 50 seconds long, and data were segmented using a 1-second window with a 50% overlap. Therefore, we have 20 (videos) $*$ 3(electrodes) $*$ [50 s (each video) $*$ 1 (1 s window) $*$ 2 (50% overlap) $-$ 1] $=$ 5,940 segments per participant, which was used for feature extraction. With 14 participants, the total data segments for feature extraction are 14 (subjects) $*$ 5,940 (segments for each subject) $=$ 83,160. After feature extraction, 83,160 (data segments) $*$ 5 (5 frequency band-based relative PSD features) $=$ 415,800 samples are used as the dataset for building and evaluating models.

Fig. 7 depicts the relative PSD distribution data for two emotional states acquired from the five frequency bands for each measurement site. The t-test [30], an inferential statistic, was employed to analyze the difference between the data of the two groups of positive and negative emotions. Table 2 displays the results of comparing the two emotional groups on the five frequency bands of the three electrodes. On all three electrodes, the relative PSD values of the alpha and theta waves demonstrate a significant difference between negative and positive emotions ($p < 0.05$). For electrode 2, all five bands differ significantly between the two states. In contrast, there was no difference ($p > 0.05$) between the theta and gamma waves of electrode 1 and the alpha and beta waves of electrode 3. These t-test results indicate that classification between the two emotional states is feasible.

In Fig. 8, the average distribution of relative Power Spectral Densities (PSDs) across five frequency bands and two emotional states (positive and negative) is compared. The results show that delta waves consistently had the highest relative PSDs in both emotional states, at approximately 54.69% and 37.53%, respectively. In contrast, alpha waves had the lowest relative PSDs, at 5.95% and 11.61%, respectively. These

findings suggest that delta waves may contain more useful information for emotion recognition when presented with emotional stimuli.

In this study, we have used t-distributed stochastic neighbor embedding (t-SNE) [31] on the relative PSDs of three measuring locations; t-SNE is a method in dimensionality reduction and visualization of high-dimensional datasets. Dimensionality reduction typically causes a reduction in the accuracy of the models in classification; therefore, t-SNE is used for data visualization in this study with the SVM classifier (see Fig. 10 (a)). The data point distribution between the two emotional states is considerably pronounced, which proves the feasibility of classifying the collected data after the experiment.

B. CLASSIFICATION RESULTS

We evaluated our proposed emotion recognition models utilizing the user-dependent and user-independent approaches for the relative PSD features obtained following feature extraction using FFT. To evaluate the performance of the proposed model and avoid overfitting, a 4-fold cross-validation method is utilized. The dataset is randomly split into four parts, where three parts are combined to create a training dataset, and the remaining fold is kept as a testing dataset. The model is trained using the training dataset, and the testing dataset is used to evaluate the model's average error and accuracy. In the case of user-dependent data, training and testing data are obtained from the same subject but for different trials. 20 trials are conducted, with 15 trials used for training and the remaining five for testing. In the user-independent scenario, a leave-one-user-out cross-validation approach is used for all participants to make the model more practical for real-life applications. This method may result in lower accuracy but is more suitable for practical scenarios. This approach uses one user's EEG signals as a testing set and the remaining users' signals as the training set. The training and testing data for the user-independent case are derived from different participants. In this work, we employed three models to detect EEG-based emotional states: SVM, MLP, and 1D-CNN. Table 4 summarizes the results obtained by utilizing the models.

Table 4 shows that the proposed 1D-CNN model outperforms SVM and MLP in the user-dependent case, with average accuracy of 85.19 %, 91.13 %, and 94.87 %, respectively. Table 4 results demonstrated that accuracy varied based on the model, user, and case (whether user-dependent or user-independent). In the user-dependent case, the highest accuracy achieved by 1D-CNN was 98.03 % for data from User 2. The results obtained after training the 1D-CNN and MLP models on User 2 for the user-dependent case are shown in Fig. 9. The confusion matrix was used to calculate the sensitivity, specificity, and accuracy using the following formulas:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{2}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{TP}}, \tag{4}$$

To assess the classification performance of the models, we also utilized the Receiver Operating Characteristic (ROC) curve [32]. This graphical representation demonstrates the performance of a classification model across all possible classification thresholds, as shown in Fig. 9 (e), (f). The ROC curve illustrates two crucial parameters: True Positive Rate (TPR) and False Positive Rate (FPR). The Area Under the ROC Curve (AUC) quantifies the overall performance of the classification model, and it represents the two-dimensional Area beneath the entire ROC curve, ranging from (0,0) to (1,1). An increase in the AUC indicates an enhancement in the model's ability to distinguish between different classes.

Fig. 9 (e),(f). displays the ROC curves for the 1D-CNN and MLP models on both the training and testing datasets for each class (Negative and Positive). For the training set, the AUC for 1D-CNN was 0.9766 and 0.9586 for the Negative and Positive classes, respectively. In contrast, MLP obtained an AUC of 0.9522 and 0.9216 for the Negative and Positive classes, respectively. On the other hand, for the testing set, 1D-CNN achieved an AUC of 0.9635 and 0.9187 for the Negative and Positive classes, respectively, while MLP obtained 0.9253 and 0.8573 for the Negative and Positive classes, respectively.

Fig. 10 (b) compares classification accuracy between the two groups of male and female groups with three models. Overall, emotion recognition accuracy is higher in the female group than in the male group. However, statistically, there is no notable difference between them for each model $p = 0.077$, $p = 0.2$, and $p = 0.055$ for SVM, MLP, and 1D-CNN, respectively (all greater than 0.05).

In addition, under the same classification models, the performance of the user-dependent case is higher than that of the user-independent case. This is most evident when evaluated with the 1D-CNN model with average accuracies of 94.87% and 84.68% for the user-dependent and user-independent cases, respectively. This clarifies that the difference in data obtained from different users significantly affects the emotion recognition process. Therefore, in the process of implementing models for practical applications, a user-independent approach needs to be adopted.

C. MODEL DEPLOYMENT

Because it performed better than other models, 1D-CNN was chosen for deployment on the device for real-time emotion. Furthermore, due to its feasibility and effectiveness in real-world applications, the user-independent method was used. The following procedures are used to deploy a model on the microcontroller (see Fig. 4 (b)). To begin, TensorFlow is used to train the 1D-CNN model. The trained model is then transformed to a TensorFlow Lite model with the help of a TensorFlow Lite Converter. We employed helpful improvements such as quantization to make the model appropriate

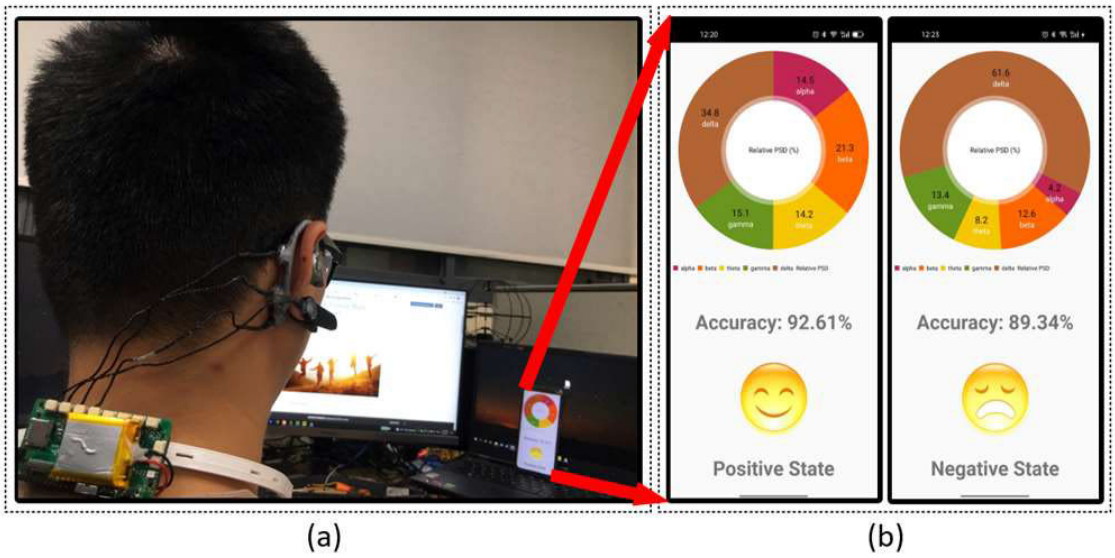

FIGURE 11. (a) Process of using our real-time EEG-based embedded system; (b) Information about the user's emotional state is displayed in the smartphone app.

for memory-constrained devices; this helped minimize model size and latency with low or no loss in accuracy on tiny devices. Because the multiple microcontroller platforms lack local file system capability, we converted the model into a C/C++ byte pattern and stored it in a read-only program memory on a device using standard tools. Finally, we used the embedded C/C++ library to build and run the inference on the microcontroller. To improve comfort and simplicity of use, a smartphone application with a user-friendly and simple interface was developed and installed on smartphones, which was connected to the embedded device through the BLE protocol. The application was divided into three major functional groups. The first functional group is concerned with displaying the user's emotional state. The proportion of the relative PSD per frequency band is shown in the second functional group. The third functional group employs audio to notify people immediately when a negative state is detected. Fig. 11 depicts the entire deployment procedure of our real-time EEG-based embedded system for emotion detection.

V. CONCLUSION

In this study, we developed a real-time on-chip embedded system with a 1D-CNN model using behind-the-ear EEG. EEG data acquired from locations behind the ear are of reasonably high quality and are easier to get than data obtained from the scalp. Using FFT, the preprocessed signals are extracted into valuable features from the relative PSDs across five frequency bands. These rich retrieved features were used to implement the three proposed models, namely SVM, MLP, and 1D-CNN, for classifying emotional states. The collected results showed that the 1D-CNN model had the highest performance accuracy in both user dependent and user independent cases. As a result, we selected the 1D-CNN model with recognition. user-independent method to deploy in our embedded system for real-time emotion recognition. TensorFlow Lite was used to deploy the chosen 1D-CNN model to the device. A smartphone application was also developed to make it easier and more comfortable for users. The primary function of this application is to show the output of the EEG-based embedded device with the installed 1D-CNN model through the BLE protocol and then to notify users when a negative state is identified. This application can be used to prevent emotional disorders.

Several studies have investigated the use of electroencephalography (EEG) signals and machine learning models for emotion classification, including our own research. One such study by Bhosale et al. [33] introduced an adaptation method based on meta-learning for emotion recognition using EEG signals with two classes - Valence and Arousal - on the DEAP dataset. Many other studies have also utilized the DEAP dataset for their research in emotion classification. In addition to employing existing datasets such as DEAP or SEED, some studies have utilized commercial devices or self-made devices to collect and process EEG data for emotional classification. For example, Nguyen and Chung [34] developed a self-made device to collect EEG data from the scalp. Other studies have used EEG signals collected from the ear due to the advantages they offer over scalp EEG. Athavipach et al. [35] developed an in-ear EEG device to classify four emotional classes. However, these studies commonly processed signals and ran machine learning and deep learning models on a PC, without real-time execution. This limitation restricts the applicability of the studies for real-time and portable applications over an extended period. To address these challenges, our study proposes a lightweight, wearable EEG device that is comfortable for prolonged use, and performs data collection, processing, and deep learning model execution directly on the device in real-time on chip.

Table 5 presents a detailed comparison of the relevant information between our proposed system and previous studies.

Despite these benefits, our system still has some limitations that need to be considered in future research. Firstly, although this study was conducted on a group of 14 participants (7 males and 7 females) with varying ages, and each participant underwent 20 trials evenly divided between 2 emotional stimuli (negative and positive states) to ensure balance in the dataset, in order to effectively apply our findings to real-world applications, we will conduct experiments on more subjects with more trials, and expand our research to include various other emotional states. Another issue in this study is related to signal processing. Here, we applied a simple bandpass filter to eliminate unwanted signals outside the frequency range of 1 to 50 Hz, in addition to instructing the participants to sit comfortably and avoid unnecessary movements that could cause noise interference in the experiment results. The primary purpose of using such a simple filter in signal preprocessing was to reduce the computation time for this process, in order to allocate more time for other important tasks, such as feature extraction or running machine learning models to balance the accuracy achieved and the real-time nature of the system. However, this simple filter has limitations in processing other artifacts caused by the participants during the experiment, such as electromyography (EMG) caused by muscle movements, as it is located in close proximity to the cheek. Therefore, in future research, we will focus on implementing more advanced and efficient signal processing techniques, such as blind source separation (BSS) or independent component analysis (ICA), to reduce noise from body movements, blinks, muscles, heart, and power lines.

REFERENCES

[1] A. Etkin, C. Büchel, and J. J. Gross, "The neural bases of emotion regulation," *Nature Rev. Neurosci.*, vol. 16, no. 11, pp. 693–700, Nov. 2015.

[2] P. Slovic, M. L. Finucane, E. Peters, and D. G. MacGregor, "Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality," *Risk Anal. Int. J.*, vol. 24, no. 2, pp. 311–322, Apr. 2004.

[3] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in *Handbook of Face Recognition*. New York, NY, USA: Springer, 2005, pp. 247–275.

[4] A. Haag, "Emotion recognition using bio-sensors: First steps towards an automatic system," in *Proc. Tutorial Res. Workshop Affect. Dialogue Syst.* Berlin, Germany: Springer, 2004, pp. 36–48.

[5] P. Kumari, L. Mathew, and P. Syal, "Increasing trend of wearables and multimodal interface for human activity monitoring: A review," *Biosensors Bioelectron.*, vol. 90, pp. 298–307, Apr. 2017.

[6] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.

[7] R. T. Pivik, R. J. Broughton, R. Coppola, R. J. Davidson, N. Fox, and M. R. Nuwer, "Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts," *Psychophysiology*, vol. 30, no. 6, pp. 547–558, Nov. 1993.

[8] K. B. Mikkelsen, S. L. Kappel, D. P. Mandic, and P. Kidmose, "EEG recorded from the ear: Characterizing the ear-EEG method," *Frontiers Neurosci.*, vol. 9, p. 438, Nov. 2015.

[9] J. J. S. Norton, D. S. Lee, J. W. Lee, W. Lee, O. Kwon, P. Won, S.-Y. Jung, H. Cheng, J.-W. Jeong, A. Akce, S. Umunna, I. Na, Y. H. Kwon, X.-Q. Wang, Z. Liu, U. Paik, Y. Huang, T. Bretl, W.-H. Yeo, and J. A. Rogers, "Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 13, pp. 3920–3925, Mar. 2015.

[10] Y. Gu, E. Cleeren, J. Dan, K. Claes, W. Van Paesschen, S. Van Huffel, and B. Hunyadi, "Comparison between scalp EEG and behind-the-ear EEG for development of a wearable seizure detection system for patients with focal epilepsy," *Sensors*, vol. 18, no. 2, p. 29, Dec. 2017.

[11] S. Kirill, E. Jablonskis, and C. Prahm, "Evaluation of consumer EEG device Emotiv EPOC," in *Proc. MEi, CogSci Conf.*, 2011, pp. 1–99.

[12] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, "A review on the computational methods for emotional state estimation from the human EEG," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–13, Jan. 2013.

[13] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: Principles, state-of-the-art, and challenges," *Brain-Comput. Interfaces*, vol. 1, no. 2, pp. 66–84, Apr. 2014.

[14] M. G. Bleichner, B. Mirkovic, and S. Debener, "Identifying auditory attention with ear-EEG: CEEGrid versus high-density cap-EEG comparison," *J. Neural Eng.*, vol. 13, no. 6, Dec. 2016, Art. no. 066004.

[15] 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.

[16] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.

[17] I. Hussain and S. J. Park, "HealthSOS: Real-time health monitoring system for stroke prognostics," *IEEE Access*, vol. 8, pp. 213574–213586, 2020.

[18] I. Hussain, S. Young, and S.-J. Park, "Driving-induced neurological biomarkers in an advanced driver-assistance system," *Sensors*, vol. 21, no. 21, p. 6985, Oct. 2021.

[19] I. Hussain, M. A. Hossain, R. Jany, M. A. Bari, M. Uddin, A. R. M. Kamal, Y. Ku, and J.-S. Kim, "Quantitative evaluation of EEG-biomarkers for prediction of sleep stages," *Sensors*, vol. 22, no. 8, p. 3079, Apr. 2022.

[20] P. Warden and D. Situnayake, *TinyML: Machine Learning With Tensorflow Lite on Arduino and Ultra-Low-Power Microcontrollers*. Sebastopol, CA, USA: O'Reilly Media, 2019.

[21] D. Yao, "A method to standardize a reference of scalp EEG recordings to a point at infinity," *Physiological Meas.*, vol. 22, no. 4, pp. 693–711, Nov. 2001.

[22] Instruments, Texas. (2017). *Ads1299-x Low-Noise, 4-, 6-, 8-Channel, 24-Bit, Analog-to-Digital Converter for EEG and Biopotential Measurements*. Accessed: May 12, 2017. [Online]. Available: http://www.ti.com/lit/ds/symlink/ads1299.pdf

[23] A. Géron, *Hands-on Machine Learning With Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA, USA: O'Reilly Media, 2019.

[24] M. Zare, H. R. Pourghasemi, M. Vafakhah, and B. Pradhan, "Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: A comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms," *Arabian J. Geosci.*, vol. 6, no. 8, pp. 2873–2888, Aug. 2013.

[25] L. Dutta and S. Bharali, "TinyML meets IoT: A comprehensive survey," *Internet Things* vol. 16, Dec. 2021, Art. no. 100461.

[26] M. Feurer and F. Hutter, "Hyperparameter optimization," in *Automated Machine Learning: Methods, Systems, Challenges*. 2019, pp. 3–33.

[27] Microsoft. *Neural Network Intelligence (Version v2.10)*. Accessed: Jan. 10, 2023. [Online]. Available: https://github.com/microsoft/nni

[28] J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, and R. A. Saurous, "TensorFlow distributions," 2017, *arXiv:1711.10604*.

[29] (2017). *TensorFlow Lite|TensorFlow*. [Online]. Available: https://www.tensorflow.org/lite

[30] T. K. Kim, "T test as a parametric statistic," *Korean J. Anesthesiol.*, vol. 68, no. 6, p. 540, 2015.

[31] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

[32] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Dec. 2006.

[33] S. Bhosale, R. Chakraborty, and S. K. Kopparapu, "Calibration free meta learning based approach for subject independent EEG emotion recognition," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103289.

[34] T. Nguyen and W. Chung, "Negative news recognition during social media news consumption using EEG," *IEEE Access*, vol. 7, pp. 133227–133236, 2019, doi: 10.1109/ACCESS.2019.2941251.

[35] C. Athavipach, S. Pan-ngum, and P. Israsena, "A wearable in-ear EEG device for emotion monitoring," *Sensors*, vol. 19, no. 18, p. 4014, Sep. 2019, doi: 10.3390/s19184014.
[36] J. Liu, X. Shen, S. Song, and D. Zhang, "Domain adaptation for cross-subject emotion recognition by subject clustering," in *Proc. 10th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2021, pp. 904–908.
[37] Y. Wang, J. Liu, Q. Ruan, S. Wang, and C. Wang, "Cross-subject EEG emotion classification based on few-label adversarial domain adaption," *Exp. Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115581.
[38] G. Zhang and A. Etemad, "Distilling EEG representations via capsules for affective computing," 2021, *arXiv:2105.00104*.
[39] Z. Mohammadi, J. Frounchi, and M. Amiri, "Wavelet-based emotion recognition system using EEG signal," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 1985–1990, Aug. 2017.
[40] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for eeg-based emotion recognition," *Cognit. Comput.*, vol. 10, pp. 1–13, Apr. 2018.
[41] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," *Informat. Med. Unlocked*, vol. 20, Jan. 2020, Art. no. 100372.

NGOC-DAU MAI received the B.E. degree in electronics and telecommunications from the Hanoi University of Science and Technology, Vietnam, in 2018, and the master's degree from the Department of Artificial Intelligence Convergence, Pukyong National University, South Korea, in 2021. His research interests include the development of wearable healthcare devices, signal processing, EEG-based affective computing systems, machine learning, deep learning, and computer vision.

HA-TRUNG NGUYEN received the B.E. degree in electronics and telecommunications from the Hanoi University of Science and Technology, Vietnam, in 2020. He is currently pursuing the master's degree with the Department of Artificial Intelligence Convergence, Pukyong National University, South Korea. His research interests include wearable healthcare device, EEG-based brain–computer interface systems, computer vision, and machine learning.

WAN-YOUNG CHUNG (Senior Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Kyungpook National University, Daegu, South Korea, in 1987 and 1989, respectively, and the Ph.D. degree in sensor engineering from Kyushu University, Fukuoka, Japan, in 1998. He was an Assistant Professor with Semyung University, from 1993 to 1999. He was an Associate Professor with Dongseo University, from 1999 to 2008. He has been a Full Professor with the Department of Electronic Engineering, Pukyong National University, South Korea, since 2008. His research interests include ubiquitous healthcare, wireless sensor network applications, and gas sensors.