

Research Article

Controlling Embedded Systems Remotely via Internet-of-Things Based on Emotional Recognition

Mohammad J. M. Zedan ¹, Ali I. Abduljabbar ¹, Fahad Layth Malallah ¹,
and Mustafa Ghanem Saeed ²

¹Computer and Information Department, Electronics Engineering College, Ninevah University, Mosul 41002, Iraq

²Department of Computer Science, Cihan University Sulaimaniya, Sulaimaniya 46001, Iraq

Correspondence should be addressed to Mohammad J. M. Zedan; mohammad.jassim@uoninevah.edu.iq

Received 15 March 2020; Revised 4 August 2020; Accepted 18 November 2020; Published 7 December 2020

Academic Editor: Thomas Mandl

Copyright © 2020 Mohammad J. M. Zedan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, much research attention is focused on human-computer interaction (HCI), specifically in terms of biosignal, which has been recently used for the remote controlling to offer benefits especially for disabled people or protecting against contagions, such as coronavirus. In this paper, a biosignal type, namely, facial emotional signal, is proposed to control electronic devices remotely via emotional vision recognition. The objective is converting only two facial emotions: a smiling or nonsmiling vision signal captured by the camera into a remote control signal. The methodology is achieved by combining machine learning (for smiling recognition) and embedded systems (for remote control IoT) fields. In terms of the smiling recognition, GENKI-4K database is exploited to train a model, which is built in the following sequenced steps: real-time video, snapshot image, pre-processing, face detection, feature extraction using HOG, and then finally SVM for the classification. The achieved recognition rate is up to 89% for the training and testing with 10-fold validation of SVM. In terms of IoT, the Arduino and MCU (Tx and Rx) nodes are exploited for transferring the resulting biosignal remotely as a server and client via the HTTP protocol. Promising experimental results are achieved by conducting experiments on 40 individuals who participated in controlling their emotional biosignals on several devices such as closing and opening a door and also turning the alarm on or off through Wi-Fi. The system implementing this research is developed in Matlab. It connects a webcam to Arduino and a MCU node as an embedded system.

1. Introduction

To simplify lifestyles, scientific researches regarding controlling the electronic devices remotely by biosignals are of interest to researchers of the following fields: human-computer interaction (HCI), machine learning (ML), and embedded systems (ES) subjects. Device control operation can be divided into two types: touched (contacted) and untouched (contactless) controls [1]. Each one has its advantages and disadvantages. For the contactless control that can be performed such as using hand gesture [2] and pose, head, or face gesture, where video stream is recorded by the camera, then, several processing operations will be handled to extract the useful signal to be sent remotely for taking effects [2], which is also named vision-based control [1]. This

type of control is useful as no possibility of infection risk might be transferred to healthy people, especially for the general devices and tools having keypad or push button, where are existing in public areas such as bus or train stations. Nowadays, especially, protection is highly needed against some types of contagion, for instance, coronavirus [3], which is distributing uncontrollably. However, this method of remote control requires several complex devices and programming skills to be installed. One of the contactless controls is facial emotions; there are six essential classes of emotions [4], which are as follows: disgust, anger, happiness, fear, surprise, and sadness. However, in the current research, only two classes of emotions will be exploited to be considered as biosignals, which are happiness (smiling) and sadness (nonsmiling). These two classes of

emotions are deemed the scope of this research. It is worth mentioning that emotion is accompanied by the psychological and physiological states of cognition and consciousness that play an important role in human communication and have several applications in the Internet of Things, IoT [5, 6]. With the proliferation of IoT [7], the biosignal can also be transferred via wireless access points to a remote distance for applications, such as the healthcare sector, for instance, patient monitoring technologies, especially for patients that are treated in-home and also for disabled people [8, 9].

The main objective of this paper is how to transfer human facial emotional vision captured by the camera to the Wifi signal as an IoT application. In other words, we propose a new biosignal (smiling and nonsmiling) used to control electronic devices because machines might no longer be limited to explicit commands and can interact with people in a way more similar to how people interact with each other. At the same time, computers could be able to automatically detect symptoms of depression, anxiety, and sadness, allowing an early response to such conditions [10]. For example, when the biosignal is sad, an alarm device is turned on; otherwise, the alarm system is turned off. Another application is switching on and off motor devices used to open and close a door or any required tool. However, remote controlling devices at a satisfactory level is still a challenging task to be performed. To run the objective, a proposed system is constituting to combine machine learning trained from big data with IoT systems. The machine learning part will be trained to recognize the smiling faces as a sample of the facial emotional recognition system. Moreover, the output will be fed to the IoT system, whereas the proposed biosignal, which will be used to control the remote embedded device, is considered the contribution knowledge of current research.

This research paper has six sections organized as follows. Section 2 presents the literature review related to the recent similar works. In Section 3, the research methodology design is explained. Then, the experiment and testing are described in Section 4. In Section 5, the result and discussion are reported. Finally, in Section 6, the conclusion is presented followed by the list of references.

2. Literature Review

Recent works related to contactless or remote control are critically reviewed in this section. Electronic devices that might be controlled remotely can be classified in different sectors, such as healthcare, home appliances, devices in public transportation, and industrial devices. The procedure is happening when the camera captures the image of video streams and then the proposed system analyzes the images and then gives commands either to connected devices or remote devices. Here, the remote device might be using wireless technology through a Bluetooth module, infrared LED, and radiofrequency signal. For instance, controlling electronic devices remotely by voice and brain waves is explained in [11], and this work has also combined machine learning by using HMM with the embedded system using the

Arduino board. These signals are captured using a headset, so that, after various processing, biosignal will be converted to actions.

Bissoli Alexandre et al. [12] proposed an IoT smarthome controlling system based on eye-gaze tracking to help people with disabilities. The system depends on the Eye Tribe software application to realize where the user is gazing, and depending on that, a controlling signal is generated as messages and sent asynchronously to the clients (devices) using the transmission control protocol (TCP). Zedan Mohammad JM et al. [13] proposed a human-computer interface (HCI) system based on the hand gesture of a smart glove which is connected to the IoT platform. The movement of finger and palm is captured depending on the output of a flex and accelerometer sensors that are included in the smart glove. The control actions were translated from the hand movement in three axes to control up to three devices using wireless communication through a web server.

A similar approach has been implemented by Espinosa-Aranda et al. in [14]. A convolutional neural network (CNN), which is a deep learning technology, is used for a real-life facial emotions application based on of the Eyes of Things (EoT) device, which is a computer vision platform for analyzing images locally and take control of the surrounding environment accordingly. The achieved emotional recognition model is trained to categorize the six elementary emotions in addition to the neutral face; furthermore, the Cohn-Kanade extended database (CK+) is used for the purpose of the validations.

Another work for patient monitoring as health smart homes (HSH) is presented in [15, 16] by using camera and image processing on IoT, in which the patient images are used for the detection to assist patients and elderly people within an in-home healthcare context. Another controlling work, which is about automatic device control using emotional recognition, is implemented based on human expressions as follows: sad, neutral, smiling, and surprised. The intended system recognizes these expressions using the regions of the eye and the mouth and the analyzed information is used to recognize various facial expressions. The types of control are as follows: lighting, heating, ventilation, air-conditioning, and so on. Then, the output information obtained is fed to a microcontroller that helps in controlling various electronic devices as in [17]. Moreover, numerous electrical devices are controlled in smart homes using radiofrequency (RF) and Wi-Fi. The Arduino Ethernet Shield Wi-Fi module is used to transmit control signals from a central processor to the electrical devices in the smart house in [18]. In addition, remote control devices are done based on a wearable device that can be used to control the computer from a distance by hand gestures, which use the technique of sweep frequency capacitive sensing (SFCS) [19]. Another idea for remote control is using a kind of bio-potentials, namely, surface electromyography (sEMG) [20], which is the electrical potential generated by muscle cells and captured by surface electrodes. Accordingly, in [21], a multivariate Gaussian classifier is exploited to train some (sEMG) signals obtained from one person. These signals represent the facial expressions: neutral, smiling, frowning,

and wrinkled nose, in which wireless signals are used to control the remote devices by biopotentials. It is worth saying that this type of emotional recognition is expensive compared with the facial image processing method so we focused on it in this current research. However, the superiority of the facial sEMG lies in it being unconstrained by the lighting conditions and head pose and its potential for being used in wearable or portable devices [21].

3. Methodology

The proposed idea can be described by dividing it into three main phases as depicted in Figure 1. The first phase is dedicated to the training operation, the second phase is for the testing, or in other words, predicting in real time, and the third phase is dedicated to transmitting and receiving the resulting signal using WAN to be an application of the Internet of Thing (IoT).

3.1. Phase One. The methodology starts by training the model of smiling and nonsmiling using the GENKI-4 database, which is well-explained [22]. The general block diagram is shown in Figure 1, as it is clear that the first step is reading the images, which are used for the training. The used images are applied to some preprocessing operations, such as scaling to a suitable size $[50 \times 50]$ and then image enhancement in terms of image contrast to achieve better accuracy by the equalization operation. After that, an algorithm of the face detection named Viola-Jones [23] is exploited in this research to crop only the face. After that, a feature extraction named histogram orientation gradient (HOG) was applied, which was introduced by Dalal and Triggs at the CVPR conference in 2005 as explained in [24, 25].

After implementing the HOG, each sample will have a number of features (F) that are arranged in a vector. The basic implementation of the HOG descriptor is as follows: divide the image into small connected regions named cells, and for each cell, compute a histogram of gradient directions or edge orientations for the pixels within the cell. To compute the gradient and direction of each pixel, consider an image A and gradient estimation filters $h_x = [-1, 0, 1]$, and $h_y = [-1, 0, 1]^T$. Let g_x and g_y represent the gradient images generated by $g_x = A * h_x$ and $g_y = A * h_y$, where $*$ represents the convolution symbol. The magnitude of the gradient at each pixel can be calculated as in

$$G(i, j) = \text{SQRT}[g_x(i, j)^2 + g_y(i, j)^2], \quad (1)$$

and the dominant gradient angle at each pixel can be estimated by

$$\theta(i, j) = \tan^{-1} \left[\frac{g_y(i, j)}{g_x(i, j)} \right]. \quad (2)$$

Using the gradient orientation obtained, discretize each cell into angular bins. Each cell's pixel contributes a weighted gradient to its corresponding angular bin. Adjacent cells are grouped into blocks in the spatial region, which forms the basis for grouping and normalization of histograms. The normalized

group of histograms represents the block histogram and the set of these block histograms represents the descriptor.

In this work, HOG has been applied to the face image using $[2 \times 2]$ block size and 14-cell size with 9-bin histogram per cell resulting in the 1260-dimensional feature vector. The HOG features provide us with the edge information of the face images. The features for all training images are computed and stored in the database. Feature vector comes from the result of the multiplication of five cells per column multiplied by seven cells per row multiplied by four blocks with 9 bins of histogram, and the result is: $FV_HOG = 1, 2, 3, \dots, 1260$.

In terms of classification, there are several algorithms used to perform classification such as CNN as the popular one. However, a support vector machine (SVM) is selected as it is a binary classification to train the model for implementing this small machine learning problem, Because it is just needed to predict whether the input tested image is smiling or nonsmiling; at the same time, it has acceptable results in terms of recognition rate. After SVM training, the reference model of the SVM will be stored in the server and used in the future for the prediction in real time. Let x represent the input space. By some nonlinear mapping, function φ is mapped into a feature space, b is biasing, and w is the weight vector as in

$$Y = f(x, w) = WT\varphi(x) + b. \quad (3)$$

As illustrated in Figure 2, SVM can several samples into two classes in case binary SVM mode, by the hyperplane vector, which is decided after several attempts to stop at the smallest targeted error. In Figure 2, there are two classes: the first class exists if the result of $w \cdot x + b \geq +1$ and the second class exists if the result of the calculation of $w \cdot x + b \leq -1$, where the separated plane is $w \cdot x + b = 0$.

3.2. Phase Two. In terms of future predicting which is based on the SVM model, it will be considered in phase two as depicted in Figure 1. The operation is kicked off by reading a real-time video and then capturing snapshot images, and the same operation that has been conducted in phase one will be applied to the snapshot image (tested image). Finally, predicting operation is applied by using SVM based on the trained SVM model.

Accordingly, a control signal is generated and sent through serial communication (USB) to the Arduino board using Matlab Support Package. The Arduino Uno device is demonstrated with little capacity and needs to add new devices for communication via Wi-Fi. Moreover, the NodeMCU platform has almost the same market value (functions) of the Arduino Uno; besides, it has the Wi-Fi feature internally.

While the Arduino board is not utilized as a final controller for the servo motor, the control signal is sent again serially from Arduino to the NodeMCU platform (transmitter part) that supports wireless communication through a web server. Such a device is used to make a wire-free control system to increase portability, and hence various technical applications for device control can be implemented.

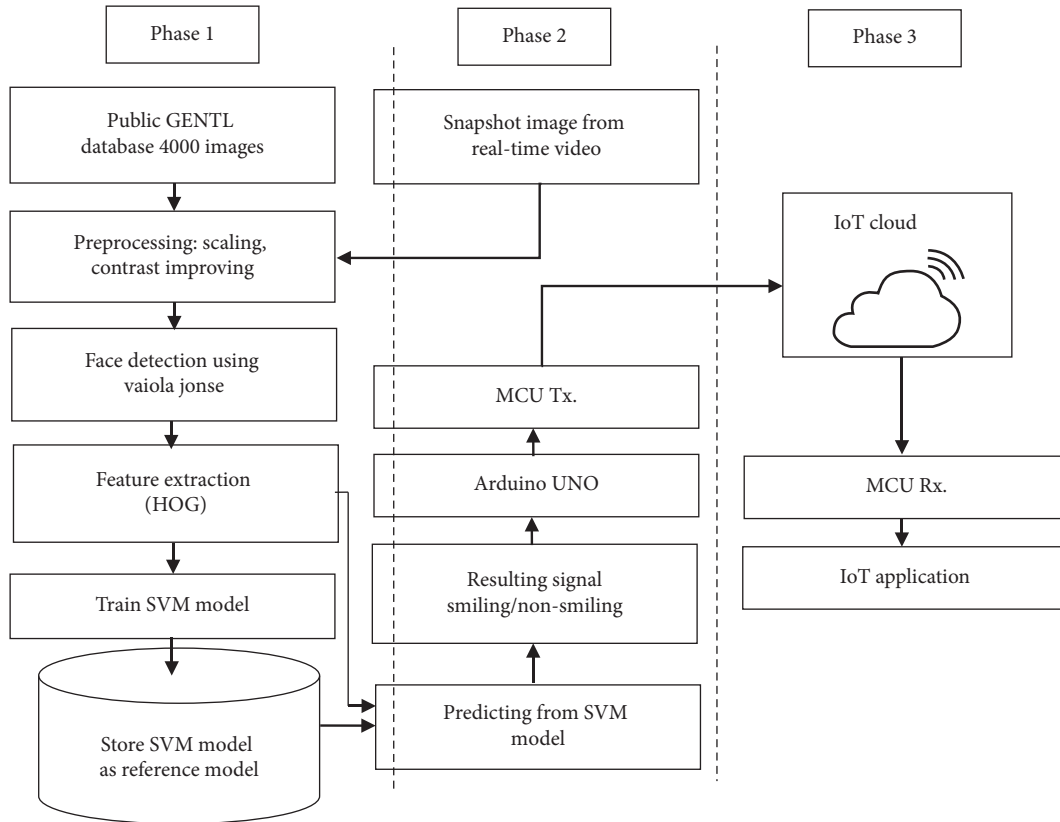


FIGURE 1: Block diagram illustrates the methodology of the proposed work.

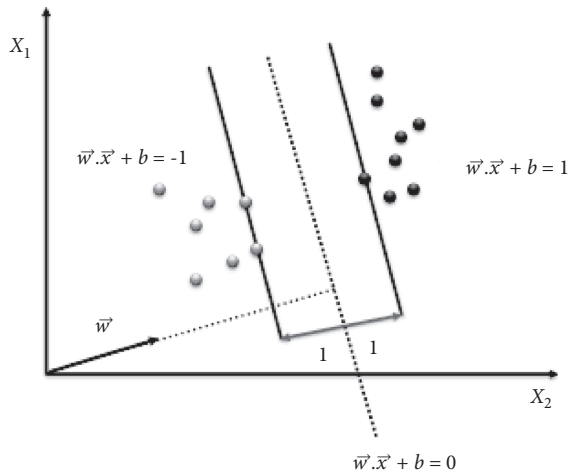


FIGURE 2: Illustrating the SVM classes with their question types.

The programming process for the NodeMCU platform is done using Arduino IDE software, where different libraries and instructions are used to configure the transmitter node to work as a server as elaborated in Figure 3(a). Once the program is loaded, a power bank can be utilized to provide a continuous power supply.

3.3. Phase Three. The proposed system uses the Internet-of-Things (IoT) platform for communication, where the transmitter and receiver can communicate with each other

using the built-in Wi-Fi, which is part of the NodeMCU platform for a distance up to 100 meters. However, an access point is used to cover a larger distance such as hospital blocks, in case it required to broaden up the coverage area.

The receiver part consists of the NodeMCU device, which is connected to any terminal device, that is, servo motor.

Depending on the connected terminal device, the receiver should be programmed to control the function of that terminal. Many instructions are used to configure the operation of the second node to work as a client as depicted in Figure 3(b). The servo motor is connected to the NodeMCU through a PWM pin, which is a digital pin. Therefore, the output either 0 or 5 V. However, this pin can output intermediate voltage values between 0 and 5 V.

The communication between the two nodes is performed by using the HTTP protocol exploiting the ESP8266.h library and the integrated development environment Arduino_IDE.

The hypertext transfer protocol (HTTP) is an application level in TCP/IP protocol for distributed, collaborative, hypermedia information systems [26]. Moreover, it can transfer plain-text, hypertext, audio, images, and Internet-accessible information. The client-server environment means communication across a network between multiple programs. The client triggers the communication by sending a request to the server. The HTTP request is sent to the server consisting of a query line, query header fields, and query body as shown in Figure 4.

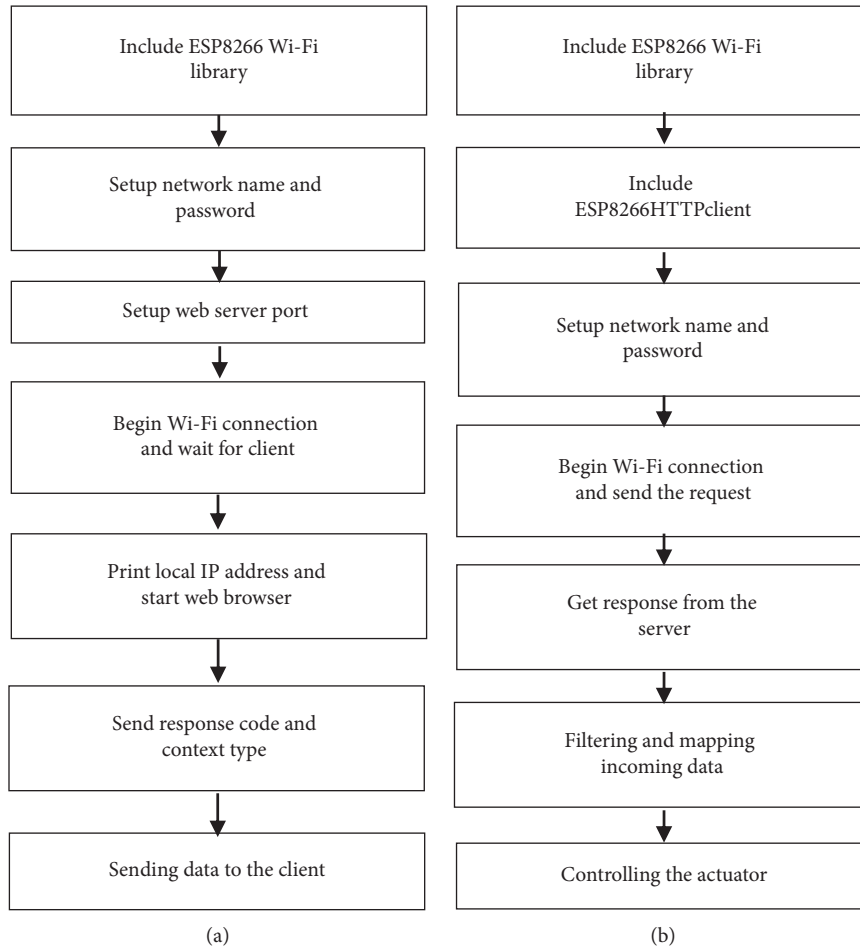


FIGURE 3: The operational flow. (a) Transmitter node (server). (b) Receiver node (client).

4. Experiment

The procedure of the experiment can be divided into two parts: the first part is the machine learning operation by preparing a suitable model to be dependent on it, in case of classification for the future predication whether the face is smiling or not. The second part is referring to the embedded system operation by connecting and applying commands for any device remotely based on the classifier decision.

In terms of the first part, machine leaning is performed by using training and testing to the proposed model based on a database, in which the GENTI1-4k database is used. Here, we have involved 1081 samples to train the model on smiling faces and 919 image samples for nonsmiling faces, as explained in Table 1, for training and testing information details.

The classifier used in this paper is named support vector machine (SVM), which has been used as a binary classification as the negative and positive for nonsmiling and smiling, respectively. In other words, the positive user is referred to the (+1) class, while the negative class is referred to as the (-1) class for both training and testing. The threshold used here for deciding between the two classes is (0) because it is an unbiased separation between -1 and +1 predicted scores.

The experiment has been conducted according to the following steps.

- (1) The SVM training matrix is built consisting of images from 2162 individuals: 1081 image samples were smiling, while 919 image samples were nonsmiling. Each sample has several feature vectors which varied based on the HOG setting; however, the best accuracy is achieved when the feature vector length is 756 features. Accordingly, the training matrix size is $[2000 \times 756]$ (756 features for each sample with 2000 samples (1081 + 919) for the trained reference model).
- (2) The performance is calculated for both FAR and FRR errors.
- (3) The performance of the result is calculated by extracting the false accept rate (FAR) and the false reject rate (FRR) for the proposed model, while the testing matrix is built similar to the way and number $[2000 \times 756]$ of the training matrix.
- (4) The threshold that has been used is zero.
- (5) FRR is calculated when the smiling tested samples are recognized incorrectly by the system as

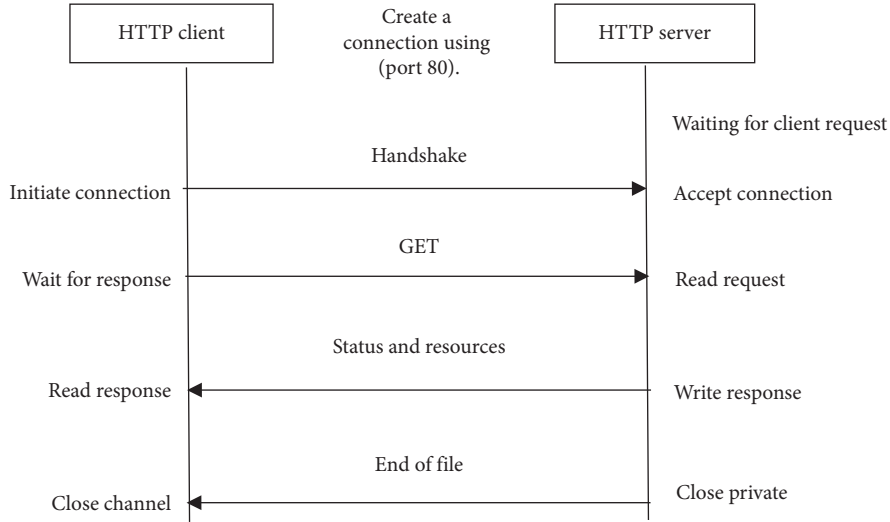


FIGURE 4: Communication between the HTTP client and HTTP server [26].

TABLE 1: A number of sample images used for the smiling and nonsmiling training reference models.

Image type	Training number	Testing number
Smiling	1081	919
Nonsmiling	1081	919
All	2162	1838

nonsmiling. Therefore, each sample, which is incorrectly recognized, will be increased by 1 ($FR = FR + 1$), and the error rate of this error is calculated as in

$$FRR = \frac{FR}{1081} \times 100\%. \quad (4)$$

- (6) Conversely, FAR is calculated when the nonsmiling tested samples are recognized incorrectly by the system as smiling, so that each sample is recognized incorrectly and the counter will be increased by 1 ($FR = FR + 1$), and then, the error rate of this error is calculated as in

$$FAR = \frac{FA}{919} \times 100\% \quad (5)$$

- (7) The accuracy of the system model is computed by taking into consideration all individuals in the GENKI-4K database, and an average of the 2000 individuals' accuracy is computed by using

$$User_{accuracy}\% = 100 - \frac{FRR + FAR}{2}. \quad (6)$$

The parameters used in the SVM model are as follows: the polynomial is the activation function and the ISDA optimization algorithm is used for the training of selecting

the most suitable weights and biasing for the best separation line between the two classes, smiling and nonsmiling.

It is worth mentioning that the percentage of the training to the testing is 50% for training and 50% for testing from this database. Also, another partitioning will be applied in the research to measure the accuracy of the proposed recognizing system, which is using 10-fold cross-validation. The smiling samples are from 1 to 2162 and the nonsmiling samples are from 2163 to 4000. Then, applying this labeled matrix to the 10-fold. Here, k -fold $k = 10$. This partitioning is performed randomly of the dataset, in which each fold will have 400 random samples of both classes. It means that in each round the trained model will be based on $k-1$ fold, and the remaining fold will be used for validation.

5. Results and Discussion

In general, the result can be divided into two parts: in the first part, the result is related to the training and testing of the machine learning that consists of the feature extraction HOG and classifier SVM. The second part of the result is regarding the control of the remote embedded device based on the trained model.

In terms of the first part, several experiments have been conducted to get the best recognition rate by modifying parameters of HOG and SVM as elaborated in the tables.

The size of images of the detected face is considered two sizes those used in the experiments as $[100 \times 100]$ pixels, and second-type $[80 \times 80]$ selection is based on the extensive experiments that have been performed to ultimately agree on these two sizes to give the best recognition rate. Moreover, the parameter block size (bs) of HOG is fixed at two pixels for all experiments. Now, by alternating the type of training methodology of the SVM between least squares (LS) and SMO, performance is selected based on the best recognition rate.

The learning algorithm (LS) is used, and the results are shown in Table 2. It is clear that the best recognition rate is

TABLE 2: HOG and SVM optimized parameters for training and testing with the LS learning algorithm.

Exp. No.	HOG-CS	HOG-bs	Image size	Feature length/image	Training time	Testing time	Accuracy	ConMat	
1	11	2	100	864	4.6	1.1	88.6000	809	110
								118	963
2	7	2	100	2808	7.3	4.6	87.9000	801	118
								124	957
3	9	2	100	1440	7.31	3.7	88.4500	808	111
								120	961
4	13	2	100	432	5.3	0.3	87.7	802	117
								129	952
5	15	2	100	360	6.40	0.4	88.20	807	112
								124	957

TABLE 3: HOG and SVM optimized parameters for training and testing with the SMO.

Exp. No.	HOG-CS	HOG-bs	Image size	Feature length/image	Training time	Testing time	Accuracy	ConMat	
1	7	2	100	2808	97.74	4.0	87.80	800	119
								125	956
2	9	2	100	1440	92.01	2.5	88.45	808	111
								120	961
3	11	2	100	864	3.32	0.5	88.60	809	110
								118	963
4	13	2	100	432	3.2	0.5	87.70	801	118
								128	953
5	15	2	100	360	2.95	0.3	88.10	806	113
								125	956

up to 88.6% among five conducted experiments, when the feature vector length is 864, to get this feature length, HOG (bs) is set at 2, and HOG cell size (cs) is set at 11 with an image size of $[100 \times 100]$, and here, the confusion matrix is also shown in Table 2, as the output 809 nonsmiling samples are correctly recognized, while the nonsmiling samples that are not recognized correctly are 110, out of the total 919 samples. Notwithstanding, the smiling samples used in the testing, which is correctly recognized, is 963, while the wrongly recognized smiling samples are 118 out of the total 1081 samples.

Moreover, experiments have been performed with another learning algorithm, which is sequential minimal optimization (SMO), and the results are shown in Table 3. It is clear that the best recognition rate was up to 88.6% (in Experiment 3) among the five conducted experiments, when the feature vector length is 864; to get this feature length, HOG (bs) is set at 2, and HOG cell size (cs) is set at 11 with an image size of $[100 \times 100]$. The confusion matrix is also shown in Table 3 for each experiment

That means changing the learning algorithm type from LS to SMO does not affect the recognition rate either positively or negatively. However, there are some modifications in the extracted information such as the training and testing time; the SMO learning algorithm is faster than LS using the same feature vector length as shown in both Tables 2 and 3.

Now, to decrease the error rate as much as possible, one of the discovered solutions is changing the detected face image size to $[80 \times 80]$ pixels, in which this size is used with several experiments by exploiting the learning algorithm

SMO because it consumes less training and testing time as shown in Table 4.

It is clear that the best recognition rate is achieved up to 88.9% among five conducted experiments, when the feature vector length is 756, to get this feature length, HOG (bs) is set at 2, and HOG-cell size (cs) is set at 9 with an image size of $[100 \times 100]$, and here the confusion matrix is also shown in Table 4, which has the output of 817 nonsmiling samples that are correctly recognized, while the nonsmiling samples that are not recognized correctly are 102, out of the total 919 samples. Notwithstanding, the smiling samples used in the testing, which are correctly recognized, are 961, while the wrongly recognized smiling samples are 120 out of the total 1081 samples. It is worth mentioning that decreasing the image size leads to decreasing the trained time because the feature represented vector for the image will be decreased as well. Therefore, the trained time in this experiments is 3.2 sec as listed in Table 4 for more details.

For the calculations of the experimental results of experiment number two in Table 4, the FRR and FAR and the overall accuracy are computed as (4)–(6), respectively. According to the highest achievement of the testing, the error for FRR = 11.1% detailed as in $120/1081 * 100$, and FAR = 11.1% detailed as in $102/919 * 100$. The overall accuracy of the 2000 sample tested images is as follows: $100 - (FRR + FAR)/2$, which is 88.9% as in Table 4.

As a result of the 10-fold of smiling versus nonsmiling,, the accuracy is up to 89.2%. The training matrix is $[4000 \times 756]$, and the confusion matrix is as in Table 5.

In this experiment of Table 5, the training dataset is not the same as the testing set in terms of the number of samples

TABLE 4: HOG and SVM optimized parameters for training and testing with SMO.

Exp. No.	HOG-CS	HOG-bs	Image size	Feature length/image	Training time	Testing time	Accuracy	ConMat	
1	7	2	80	1440	93.8	2	88.70	803	116
2	9	2	80	756	3.26	0.8	88.90	110	971
3	11	2	80	432	2.83	0.5	88.25	817	102
4	13	2	80	360	2.76	0.3	88.35	120	961
5	15	2	80	144	3.28	0.1	88.30	801	118
								117	964
								799	120
								113	968
								800	119
								115	966

but according to the k-fold cross-validation. It selects ten-fold, which means that each fold contains 400 samples randomly taken as smiling and nonsmiling samples. The algorithm of this cross-validation is to take $k-1$ fold to training, and the remaining k-fold is used for the test; in other words, here, 400 random samples have been used for testing and 3600 samples were used for training. As a percentage, 10% is used for testing and 90% for training.

The second part of the results is about the control of the remote embedded device based on the trained model. As shown in Figure 5, there are two NodeMCUs left and right; the right one represented the server node (transmitter part) and was connected to the computer where the smiling/nonsmiling detection occurs. The server node is responsible for sending the control signal remotely.

The left NodeMCU represents the client (receiver node), where it receives the signal via Wi-Fi. As a result, when the face is smiling, the motor will have a specific turn as a vertical position. Here, the speed response of the motor depends on the type of used NodeMCU and computers that are responsible for emotional detection to move the motor based on the detected biosignal.

Similarly, as it is as illustrated in Figure 6 that the face is nonsmiling, the motor has turned around its blades to the horizontal position. The motor position is easily programmed to a required position based on the biosignal detected from the server-side computer.

This technique is applied to around 40 individuals at the university by asking them to turn on or turn off the embedded system (a motor in our case), by using their smiling and nonsmiling faces in front of a video camera; all of them could transfer their emotion to a signal that is controlling the motor (see the video in the Supplementary Materials). However, this experiment is required to perfect the illumination of the room to control the lightness captured by the camera, that is, white balancing. Experiments are performed indoors as the scope of this technique, whereas it could be performed outdoors, but it needs some image processing tools to add white balancing.

Typically, deep neural networks perform better for most tasks compared with SVM. However, CNN needs a high hardware capability for processing such as multiprocessor CPU; otherwise, it uses much more time and power consumption, while SVM requires less time processing compared with CNN, and this is very important especially that

TABLE 5: Confusion matrix of the whole dataset recognition rate of the two classes.

	Predicted values		
	Total samples	Non-smile	Smile
Non-smile	1838	1612	226
Smile	2162	206	1956

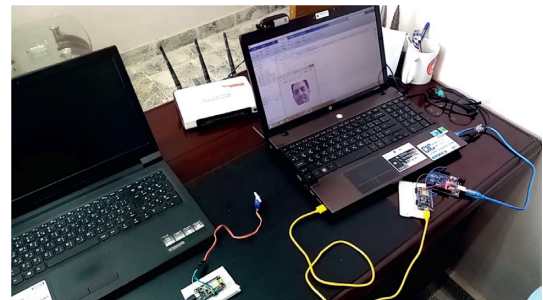


FIGURE 5: Communication between the HTTP server (right PC) and HTTP client (left PC); smiling state with the vertical motor position.

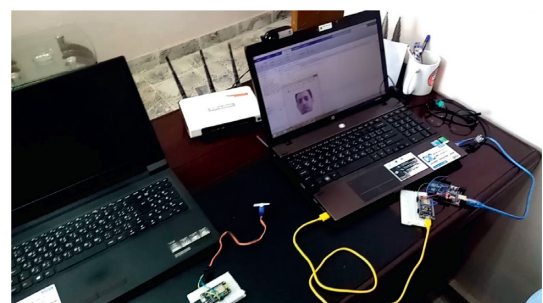


FIGURE 6: Communication between the HTTP client (right PC) and HTTP server (left PC); the nonsmiling state with the horizontal motor position.

the hypothesis of this research is applying the notion to embedded system devices which are close to the light-weighted devices. Thus, the lower processing time is highly needed to work well in the real time. Another reason is that SVM can be employed easier than a CNN for small pattern recognition problems. The problem discussed here comprises of two classes (smiling and nonsmiling), and a SVM works well with binary pattern recognition problems.

However, CNN does not outperform other machine learning classifiers in all cases, and this is determined according to the feasibility of the case. For instance, a comparison between SVM and CNN has been performed in this paper [27] for the goal of waste separation. Then, SVM achieved a higher classification accuracy of 94.8% than CNN, which achieved only 83%.

A similar approach has been implemented by Kamath et al. as in [17], which is about automatic device control using four biosignals of emotional recognition: sad, neutral, smiling, and surprised, but here it is without any clear testing and percentages, and no public database was used. However, our proposed system is two biosignals implemented on a public dataset with a clear explanation of the experiment and the accuracy results.

Another similar work as in [21], which used a multivariate Gaussian classifier to train offline from one person's surface EMG (sEMG) signals comprising of four facial expressions: neutral, smiling, frowning, and wrinkled nose to control remote devices. However, this type of emotional recognition is expensive compared with the notion of this paper, which depended on the facial image processing method as well as normal traditional camera for inputting the data stream.

6. Conclusion

Facial emotional recognition can be used to control any embedded system devices of different life sectors. The recognition must be performed in real time to send remotely to control the device which is connected by the Internet as IoT. In other words, smiling and nonsmiling, which are in a turn called biosignals, could be exploited to control the remote device. The benefit here constitutes several applications, such as protection from any contagion disease by removing the touching control from life, especially some kind of disease that causes pandemic risk, for instance, coronaviruses. Technically, the operation is performed by dividing the work into two parts, which are machine learning by using pre-processing, HOG and SVM, and then handling the extracted signal to the embedded device to control any desired devices. These devices might be remotely located or nearby the person according to the application, but in any case, the control is performed in an untouched manner. The experiment is applied to 40 persons. Remarkably, all persons could use their face to turn the motor vertically or horizontally based on their face by smiling or not. In future work, emotional recognition might be boarded to more than two signals for having better controllability and flexibility in the control.

Data Availability

A 4,000-image subset of the Public GENKI-4K database is available at <http://mplab.ucsd.edu>.

Disclosure

The research and publication costs were funded by the authors of this paper.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Supplementary Materials

The supplementary materials contain codes of the implemented research and real-time video results. (*Supplementary Materials*)

References

- [1] M. F. Layth, N. Y. M. G. S. Khaled, S. D. Abdulameer, and A. W. Altuhafi, "Vision-based control by hand-directional gestures converting to voice," *International Journal Of Scientific & Technology Research*, vol. 7, no. 7, pp. 185–190, 2018.
- [2] F. L. Malallah, S. Baraa, M. D. Aso et al., "Hybrid hand-directional gestures for biometric based on area feature extraction and expert system," *Journal of Theoretical & Applied Information Technology*, vol. 95, no. 23, 2017.
- [3] J. A. Backer, K. Don, and W. Jacco, "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020," *Eurosurveillance*, vol. 25, Article ID 2000062, 2020.
- [4] A. Nandi, P. Dutta, and M. Nasir, "Recognizing human emotions from facial images by landmark triangulation: a combined circumcenter-incenter-centroid trio feature-based method," in *Algorithms in Machine Learning Paradigms*, pp. 147–164, Springer, Berlin, Germany, 2020.
- [5] W. Hua, F. Dai, L. Huang, J. Xiong, and G. Gui, "HERO: human emotions recognition for realizing intelligent Internet of Things," *IEEE Access*, vol. 7, pp. 24321–24332, 2019.
- [6] Z. Zhou, Y. Guo, Y. He, X. Zhao, and W. M. Bazzi, "Access control and resource allocation for M2M communications in industrial automation," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3093–3103, 2019.
- [7] R. M. Varaprasad, K. Srujan Raju, G. Vishnu Murthy, and B. Kavitha Rani, "Configure and management of internet of things," in *Data Engineering and Communication Technology*, pp. 163–172, Springer, Berlin, Germany, 2020.
- [8] F. Bahar, F. Firouzi, and K. Chakrabarty, "Healthcare IoT," in *Intelligent Internet of Things*, pp. 515–545, Springer, Berlin, Germany, 2020.
- [9] P. Maria, K. Fysarakis, V. Sakkalis et al., "Biosensors and Internet of Things in smart healthcare applications: challenges and opportunities," in *Wearable and Implantable Medical Devices*, pp. 25–53, Elsevier, Amsterdam, Netherlands, 2020.
- [10] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, New York, NY, USA, October 2016.
- [11] B. Djemaa, A. Ayoub, and D. Chogueur, "Controlling electronic devices remotely by voice and brain waves," in *Proceedings of the 2017 International Conference on Mathematics and Information Technology (ICMIT)*, pp. 38–42, Adrar, Algeria, December 2017.
- [12] B. Alexandre, D. Lavino-Junior, M. Sime, L. Encarnação, and T. Bastos-Filho, "A human-machine interface based on eye tracking for controlling and monitoring a smart home using the internet of things," *Sensors*, vol. 19, p. 859, 2019.
- [13] J. M. Zedan Mohammad, A. A.-J. Mamooun, and A. M. A. Mohammed, "Real-time environmental control system for disability access under node-MCU platform," *IOP*

- Conference Series: Materials Science and Engineering*, vol. 518, no. 4, 2019.
- [14] J. Espinosa-Aranda, N. Vallez, R.-S. Jose Maria et al., "Smart doll: emotion recognition using embedded deep learning," *Symmetry*, vol. 10, no. 9, p. 387, 2018.
 - [15] L. Y. Mano, B. S. Faiçal, L. H. V. Nakamura et al., "Exploiting IoT technologies for enhancing Health Smart Homes through patient identification and emotion recognition," *Computer Communications*, vol. 89-90, pp. 178–190, 2016.
 - [16] I. Mirja, Y. Gizatdinova, V. Surakka, and E. Vankka, "Head movement and facial expressions as game input," *Entertainment Computing*, vol. 5, pp. 147–156, 2014.
 - [17] K. D. Laxmana, R. Jackson, R. Jacob, C. C. Chacko, C. M. Soumya, and P. Remya Krishnan, "A novel method for automatic device control using emotion recognition," in *Proceedings of the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICI-CICT)*, pp. 1173–1176, Kannur, Kerala, India, July 2017.
 - [18] T. N. Dang and L. T. H. Le, "An improvement of remote control panel for numerous electrical devices in smart homes using RF and wifi," *International Journal of Online and Biomedical Engineering (ijOE)*, vol. 15, no. 15, pp. 127–133, 2019.
 - [19] S. Pallavi and R. Berlia, "Project touchport: a system consisting of a portable touchscreen to convert any computer screen to a touch sensitive screen and a wearable wireless device to remotely control your computer," in *Proceedings of the 2014 IEEE Global Humanitarian Technology Conference-South Asia Satellite (GHTC-SAS)*, pp. 72–76, Trivandrum, India, September 2014.
 - [20] G. Acar, O. Ozturk, A. J. Golparvar et al., "Wearable and flexible textile electrodes for biopotential signal monitoring: a review," *Electronics*, vol. 8, no. 5, p. 479, 2019.
 - [21] M. Jiang, T. N. Gia, A. Anzanpour et al., "IoT-based remote facial expression monitoring system with sEMG signal," in *Proceedings of the 2016 IEEE Sensors Applications Symposium (SAS)*, pp. 1–6, Catania, Italy, April 2016.
 - [22] K. S. Ebrahimi, P. Froumenty, and C. Pal, "Facial expression analysis based on high dimensional binary features," in *Proceedings of the European Conference on Computer Vision*, pp. 135–147, Cham, Switzerland, September 2014.
 - [23] V. Paul and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, Kauai, HI, USA, February 2001.
 - [24] N. Abbas, K. N. Yasen, K. H. A. Faraj, L. F. A. Razak, and F. L. Malallah, "Offline handwritten signature recognition using histogram orientation gradient and support vector machine," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 8, pp. 2075–2084, 2018.
 - [25] D. Navneet and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, pp. 886–893, San Diego, CA, USA, June 2005.
 - [26] F. Roy, J. Gettys, J. Mogul et al., *Rfc2616: Hypertext Transfer Protocol*, RFC Editor, Marina del Rey, CA, USA, 1999.
 - [27] E. Sakr George, M. Mokbel, D. Ahmad, M. N. Khneisser, and H. Ali, "Comparing deep learning and support vector machines for autonomous waste sorting," in *Proceedings of the 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*, Beirut, Lebanon, November 2016.