

Article

IoMT Based Facial Emotion Recognition System Using Deep Convolution Neural Networks

Navjot Rathour ¹, Sultan S. Alshamrani ^{2,*}, Rajesh Singh ¹, Anita Gehlot ¹, Mamoon Rashid ^{3,*}, Shaik Vaseem Akram ¹ and Ahmed Saeed AlGhamdi ⁴

- ¹ School of Electronics and Electrical Engineering, Lovely Professional University, Jalandhar 144001, India; navjot.16885@lpu.co.in (N.R.); rajesh.23402@lpu.co.in (R.S.); anita.23401@lpu.co.in (A.G.); vaseem.11814442@lpu.in (S.V.A.)
- ² Department of Information Technology, College of Computer and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
- ³ Department of Computer Engineering, Faculty of Science and Technology, Vishwakarma University, Pune 411048, India
- ⁴ Department of Computer Engineering, College of Computer and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; asjannah@tu.edu.sa
- * Correspondence: susamash@tu.edu.sa (S.S.A.); mamoon.rashid@vupune.ac.in (M.R.); Tel.: +91-781-434-6505 (M.R.)



check for updates

Citation: Rathour, N.; Alshamrani, S.S.; Singh, R.; Gehlot, A.; Rashid, M.; Akram, S.V.; AlGhamdi, A.S. IoMT Based Facial Emotion Recognition System Using Deep Convolution Neural Networks. *Electronics* **2021**, *10*, 1289. <https://doi.org/10.3390/electronics10111289>

Academic Editors: Abdelmalik Taleb-Ahmed and Abdeldjalil Ouahabi

Received: 21 April 2021
Accepted: 24 May 2021
Published: 28 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Facial emotion recognition (FER) is the procedure of identifying human emotions from facial expressions. It is often difficult to identify the stress and anxiety levels of an individual through the visuals captured from computer vision. However, the technology enhancements on the Internet of Medical Things (IoMT) have yielded impressive results from gathering various forms of emotional and physical health-related data. The novel deep learning (DL) algorithms are allowing to perform application in a resource-constrained edge environment, encouraging data from IoMT devices to be processed locally at the edge. This article presents an IoMT based facial emotion detection and recognition system that has been implemented in real-time by utilizing a small, powerful, and resource-constrained device known as Raspberry-Pi with the assistance of deep convolution neural networks. For this purpose, we have conducted one empirical study on the facial emotions of human beings along with the emotional state of human beings using physiological sensors. It then proposes a model for the detection of emotions in real-time on a resource-constrained device, i.e., Raspberry-Pi, along with a co-processor, i.e., Intel Movidius NCS2. The facial emotion detection test accuracy ranged from 56% to 73% using various models, and the accuracy has become 73% performed very well with the FER 2013 dataset in comparison to the state of art results mentioned as 64% maximum. A *t*-test is performed for extracting the significant difference in systolic, diastolic blood pressure, and the heart rate of an individual watching three different subjects (angry, happy, and neutral).

Keywords: deep convolution neural networks; facial emotion recognition (FER); IoMT; raspberry-pi; *t*-test

1. Introduction

Internet of Medical Things (IoMT) is an emerging technology that is widely spread in health care management applications for assisting patients in real-time scenarios [1,2]. IoMT is the amalgamation of smart sensors, wirelessly connected devices, and medical devices. At present, it is also monitoring the emotional, physiological, and vital states with the assistance of wearable devices and non-invasive off-the-shelf hardware [3]. In the modern era, the amount of stress and anxiety is huge due to success and failure in their respective work, and it further leads to the suicide of the person. Stress and anxiety are sensed through the facial emotion of an individual. Facial expressions have maximum magnitude over the words during a personal conversation. Researchers have recently recommended building robust and dedicated devices for distinguishing moods

and emotions [4]. Distinct procedures are utilized for constructing the automated tools for facial emotion recognition (FER) that are implemented in surveillance systems, expression recognition, interviews, and aggression detection [5,6].

Generally, the facial emotion of an individual in few studies has been realized through the computer vision (CV) based device [7,8]. Yet developing a hardware-based CV device with maximum accuracy to detect facial emotions are challenging due to deficiency of adequate data and complexity in the computation unit. A conceptual framework for the recognition of facial expressions of suspicious activity with a Raspberry-Pi (Farnell, Leeds, UK) is implemented as a part of security enhancement in a smart city [9–11]. However, it is examined that there are major issues in facial recognition where the videos/images don't provide accurate and reliable support for identifying the head pose, illumination variation, and subject dependence for FER [7]. To enhance the reliability and accuracy in FER, the sensor data need to be correlated with pure image/video processing. Face sensors, non-visual sensors (depth, audio, and EEG sensors), and infrared thermal sensors assist FER. With the technological advancements in smart sensors and wireless wearable devices, IoMT can also be implemented for the FER of an individual based on certain parameters to save a life.

Furthermore, advancements in artificial intelligence (AI) algorithms empower interpreting the data obtained from health-related wearable and other wireless devices [12]. The modern DL algorithms are capable of recognizing the phenomena from real-time vision and sensor data with outstanding accuracy in IoMT related applications [13]. With this advantage, we have proposed a conceptual framework and device that utilizes the power of the deep learning (DL) algorithm for analyzing the non-invasive blood pressure sensor data and computer vision data for accurate FER of an individual [14]. The idea of implementing the DL algorithm in the edge device with a neural computing stick is to establish a resource-constrained portable device that detects the FER with a cost-effective real-time scenario [15]. A conceptual framework of the proposed study is shown in Figure 1. The major contribution of the proposed study are as follows:

- Design and implementation of IoMT based portable edge device for recognizing the facial emotions of an individual in real-time.
- A customized wrist band and vision mote are designed and developed for recognizing facial emotions by correlating the sensor data and visuals data.
- A pre-trained deep network has been imported on Raspberry-Pi itself using Co-Processor Intel Movidius neural compute stick (Mouser Electronics, Mansfield, TX, USA).
- Deep convolution neural network model for delivering outstanding accuracy in the edge device for FER.
- Real-time experiments are performed on a distinct individual by considering three facial emotion subjects, including happy, angry, and neutral.
- A *t*-test is performed for extracting the significant difference in systolic, diastolic blood pressure, and heart rate of an individual during watching three different subjects (angry, happy, and neutral).

The organization of the study is as follows: Section 2 provides the related works. Section 3 covers the proposed methodology that provides the methods implemented in this study. Section 4 covers the system development, where the customized vision mote and wrist band are addressed. Section 5 covers results and validation of real-time experimental setup, where the results are obtained from the distinct individuals in real-time, and *t*-test validation is also discussed. Section 6 concludes the paper.

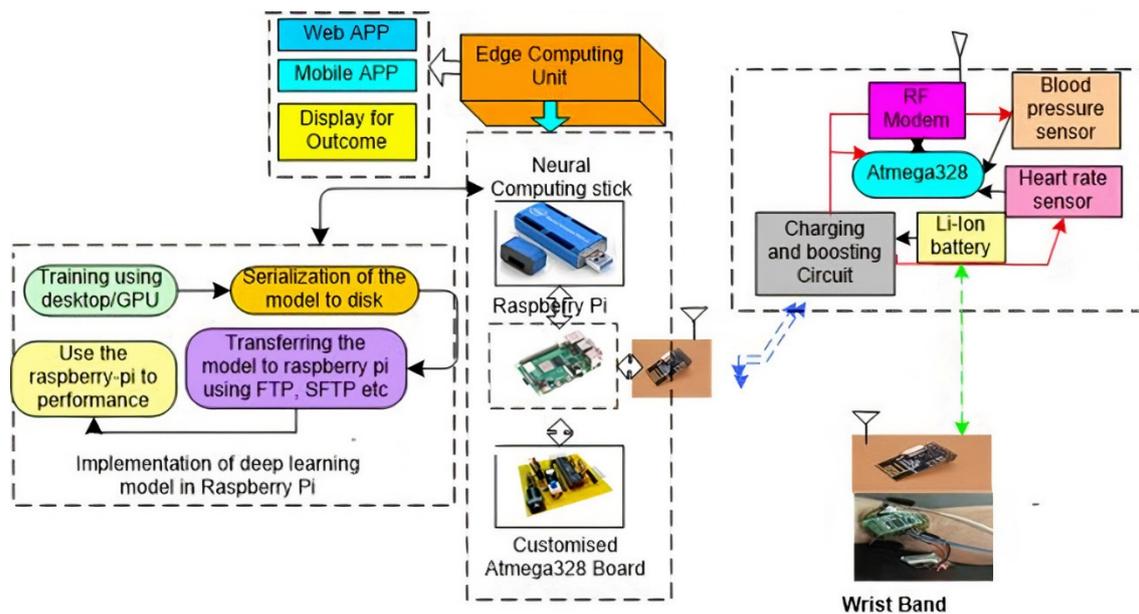


Figure 1. A conceptual framework of the proposed study.

2. Related Work

Deep learning techniques can also be used to improve efficiency. For example, a synthetic data generation unit was designed to secretly generate faces with varying expressive saturations using a 3D convolutional neural network (CNN) [16,17]. Many descriptive approaches to interaction forms of emotions are included in the classification of the input data, and the CNN network is an effective algorithm of deep learning [18,19]. The CNN modeling program was used to describe a model, which was taught using different FER datasets, and to show the capacity to classify networks that provide emotional training in both data sets and actual FER activities [20]. The first derives spatial existence from picture collections, while the second derives spatial structure from temporal facial landmarks. These two versions are fused using a modern convergence method to improve facial speech recognition efficiency. A system in which face expressions can be employed to obtain the relevant content of a video stream in which characteristics are retrieved utilizing a descriptor of histogram-oriented gradients (HOG) with an even local ternary pattern (U-LTP) [21].

The proposed system is effective for face coding through interpreting the distinct emotions both individually and simultaneously through the contraction of the facial muscles and their relaxation [22]. It reveals that several muscle movements are treated as action units that can track expressions, and these units can be combined in different emotions to determine people's moods. The first and most important step in all FER systems is face detection, which is still a difficult task due to a variety of issues such as image compression artifacts, high lighting, or low resolutions, and so on [23]. Contextual Multi-Scale Region-based Convolution Neural Network (CMS-RCNN) method is applied in which it detects the face even when it is inverted or at a very bleak angle. The amalgamation of SIFT and CNN features to recognize facial expressions and authors conclude that the proposed system can be trained with fewer images for obtaining tremendous accuracy [24]. Proposed a Human-Robot Interaction for facial expression recognition, as well as a conditional generative adversarial network (cGAN). It was utilized to identify the expression through discriminative representations achieved in 3D [25].

A CNN-based mechanism for facial expression recognition is proposed, and a feature extraction module, the focus module, the rebuilt module, and the classification module, four forms of processing were carried out [26]. A neural hybrid deep learning method is proposed especially for socially beneficial robots to recognize the emotions from facial

expressions. To track humanoid robots in real-time, Deep CNN has been used [27,28]. The extraction of functions for both regulated and uncontrolled images inspired by human vision was introduced, and the Gabor filter was used for reducing computer expenses and vector lengths for large features [29]. Presenting multi-face characteristics and supporting vector machines to enhance the analytical efficiency of facial expression recognition. Three system forms were employed, namely the discrete transform cosine (DCT), angular transform radial (ART), and the Gabor filter (GF) [30].

Face Expression Recognition Function Fusion Network (FFN-FER) is proposed, where its focus on a common IC channel and inter-category feature distinctions (ID) channel within the category, and the IC was used to find out the standard features and ID to analyze features [31]. Attempts to define and recognize emotions in a few typical Hollywood film clips suggested a multi-layer cognitive framework, adopted BP algorithm to optimize and learn network weights, and used the spatial relationship projection to minimize model parameters to boost training efficiency [32]. A system is proposed for real-time monitoring of dustbins using image processing and Raspberry-Pi for overcoming the overflow of waste from the bins [33]. In this paper, we presented an approach for real-time facial landmark detection and feature extraction, which is the most critical prerequisite for emotion recognition systems through Raspberry-Pi [34]. An image and video capturing-based system is proposed for monitoring the sanitation in the premises of a hospital through Raspberry-Pi and Arduino (Unique India, Delhi, India) [35]. In order of recognizing the facial emotion through speech data, where memory and computer requirements are to be limited, three different state-of-the-art selection methods were examined: ILFS, ReliefF, and Fisher, and compared them with the proposed 'Active Feature Selection' (AFS) features selection process [36,37]. Based on robust features and machine learning from audio language, new emotion recognition is proposed. The audio data used as input in the device from which Mel Frequency Cepstrum Coefficients (MFCCs) were measured as features for a person-independent emotional recognition system [38].

The associated facial emotion factors and corresponding tests to validate these factors were investigated. Furthermore, some researchers have tried to capture body characteristics with ultrasonic radio frequency signals, and the cell phone and emotional recognition algorithms would differ based on variations with acquisition devices [32,39]. In [40], the method for the enhancement of the less complex processing essence of programmed human-machine interactions (HMI) in health monitoring is extended to a multi-modal visualization analysis (MMVA). The proposed method is designed particularly to define a patient's facial expressions using input visualization facial expressions and textures. This paper introduces a new WiFi-based method of facial expression recognition called WiFE. Our main insight is that in various expressions, facial muscular activity produces distinctive waveform patterns in channel state information (CSI) time series in Wireless Local Area Network (WLAN) signals [41]. A deep tree-based model is proposed in a cloud environment for automatic facial recognition, and the proposed deep model is less costly computationally without affecting its accuracy [42]. The traditional face recognition method fails to predict the exact facial characteristics that minimize the accuracy of face recognition. The method of fabricating facial points maximizes calculating sophistication. This work is then used to predict and fit the face of the database with an accurate and artificial intelligence Internet based on a face expression detection system [43,44].

From the above literature, it is identified that computer vision devices need to be correlated with the sensor data of an individual for implementing an effective facial emotion recognition system. In order to do that first, the human facial emotion is captured using Raspberry-Pi using a timestamp. At the same time, the physiological sensor values are recorded for the same subject. The expression captured is then validated via the physiological value recorded via the device itself. If both values are found to statistically significant, then the expression is considered a valid expression.

3. Proposed Methodology

In this method, facial emotion recognition has been conducted on Raspberry-Pi itself instead of the cloud. The pre-processing step is similar in this method also. The deep network has been trained and imported on Raspberry-Pi, and the process has been speedup via co-processor, i.e., Intel Movidius Neural Compute Stick 2. The detailed architecture using this method is shown in Figure 2.

Step 1. Selection of Deep Convolutional Neural Network: Working with a resource-constrained device, such as Raspberry-Pi, also needs architectures that do not require less power and occupy lesser space and perform fast processing. Therefore, working with VGG and ResNet will not be suitable as they require 200–500 MB that is huge for resource-constrained devices due to their sheer size and capability to perform computations. So, working architectures, such as Mobile Nets, will work here, which are different from conventional CNNs as they use depth-wise separable convolution. So, the Mini_Xception model [45] has been used to train the FER 2013 dataset (<https://www.kaggle.com/msambare/fer2013>, accessed on 5 July 2018) as this network splits the convolution into two stages. The first stage is a 3×3 depth-wise convolution, and the second stage performs 1×1 pointwise convolution. That is the major point that helps in reducing the number of parameters in the network. The only thing that needs to be compromised is accuracy because these networks are not as accurate as actual CNNs.

Step 2. Training of pre-trained network and importing: OpenCV 3.3 launched back in 2017 And was a highly improved deep neural network module. This module supports several frameworks, which include Caffe, Tensor Flow, and PyTorch/Torch as well. The Caffe module has been used in this to import on Raspberry-Pi. The network has been trained via the FER 2013 dataset using Google Co-Lab with K80 GPU (NVIDIA, Santa Clara, CA, USA). Once the network has been trained, Prototxt files that define the model itself having all layers in it and the Caffe model file that contain all the weights of actual layers are imported and parse command-line arguments. First, the model is loaded via Prototxt and model file paths and then stored the model as a net.

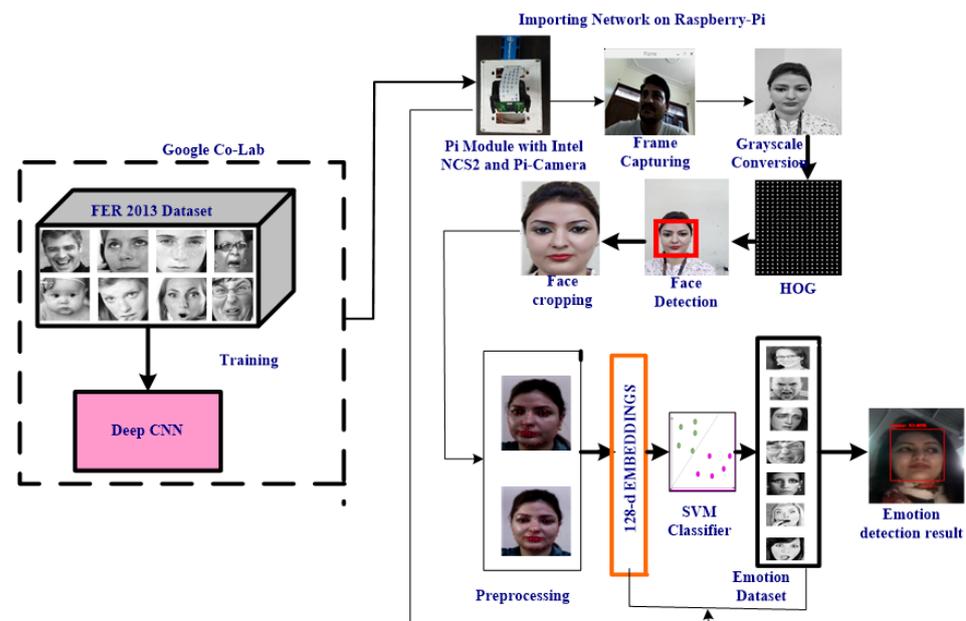


Figure 2. Architecture for face recognition and facial emotion recognition in real-time using Raspberry-Pi and Intel Movidius NCS2.

Step 3. Feeding the pre-processed image: The next step is to feed the pre-processed image into the network. The pre-processing stages are already explained in the first method. Pre-processing includes setting the blob dimensions and normalization.

4. System Development

This section explains the detailed steps of hardware development of the system, which is capable of detecting the real-time facial emotions of human beings. The complete hardware development is divided into two parts. The first part is the vision node, which has a camera, Raspberry-Pi, and Servo motors to provide pan and tilt and co-processor implements the deep network; the other part of the system is a wrist band, which has two sensors, i.e., the Heart Rate and BP sensors (Sunrom, Ahmedabad, India), to record the physiological values of the person and correlated with facial expression. The expressions recorded with this device were verified by the physiological values and gave a close relationship between the device and the values recorded by it.

The main part of the vision node is Raspberry-Pi and the Pi camera, as shown in Figure 3. In the vision node, servo motors were used to pan and turn the device to monitor a face in real-time, and moreover, this node contains both RF Modem and Wi-Fi to store the data in the cloud.

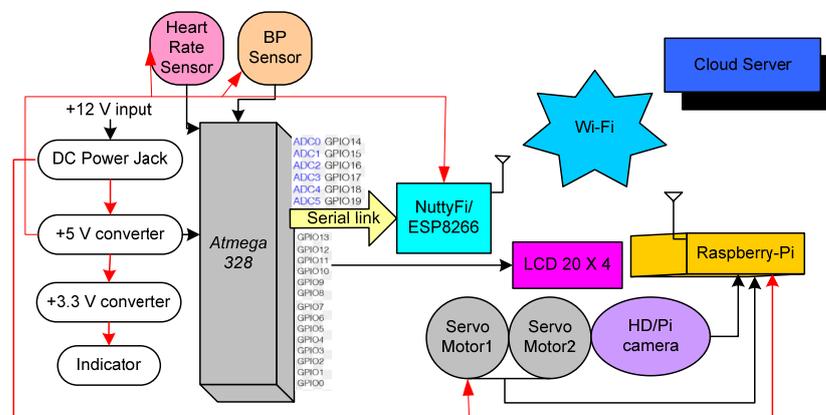
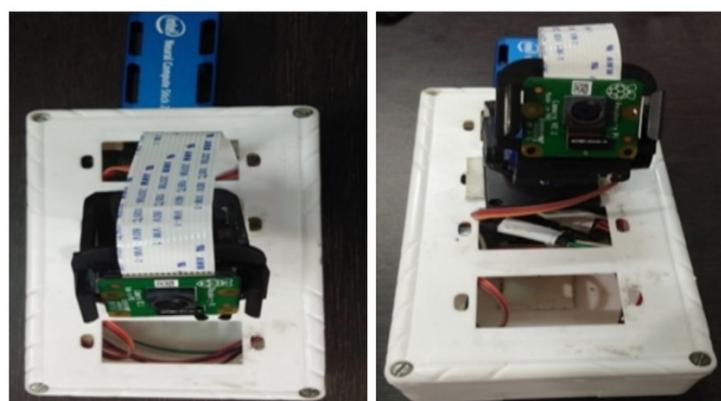


Figure 3. Block diagram of vision mote.

The RF modem is used to collect the wearable band data and to transfer the information to the server. The LCD is used to display the captured values from the wristband. The wrist band has physiological sensors on it, i.e., BP sensor and Heart Rate sensor. The sensor values are received by Raspberry-Pi using Pyfirmata and combined with the facial emotion images that are captured via the Pi camera in real-time. The data gathered from both the camera and sensors are then correlated, and conclusive expression and confidence have been extracted to understand the facial emotion of the person in real-time. Figure 4 presents the customized vision mote with the complete package.



(a) Top view of vision mote

(b) Front view of vision mote

Figure 4. Customized vision mote.

Figure 5a illustrates the block diagram of the wrist band that can collect the physiological values of the person via two different sensors, i.e., the BP sensor and the Heart Rate sensor. The BP sensor can detect the systolic and diastolic values of heart rate and display the same on the LCD as well as on the cloud. The Heart Rate sensor will sense the values of heart rate and display the same on the LCD and send all the values to the cloud also. The physiological sensor values collected along with the real-time emotion detected images and conclusion are taken by looking into the threshold and recording values on Raspberry-Pi itself.

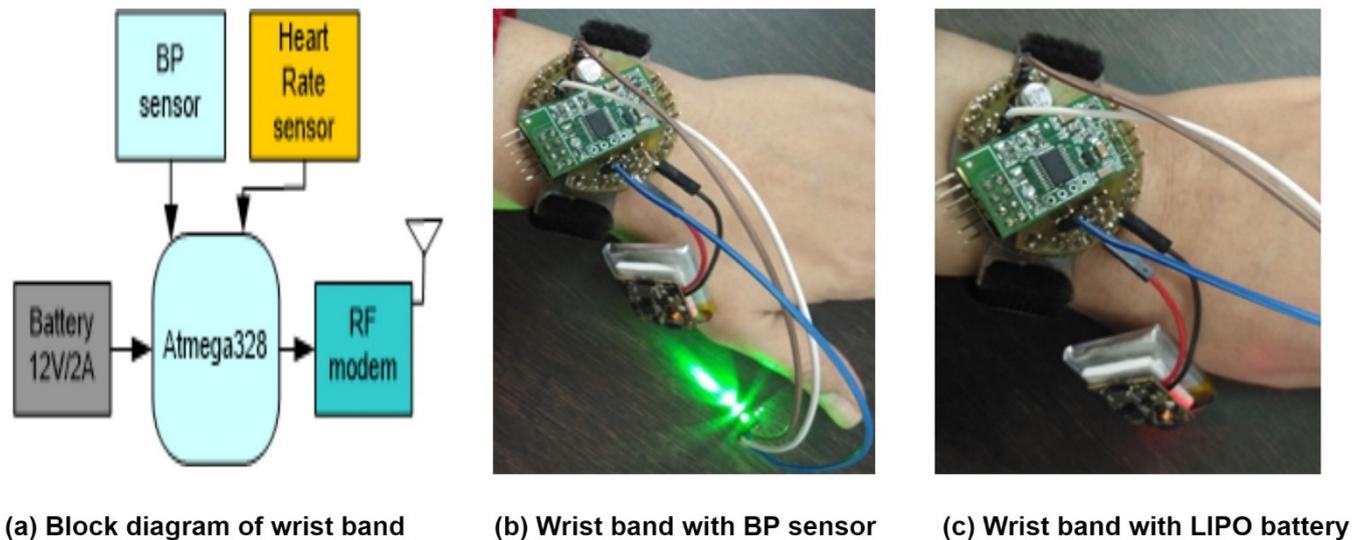


Figure 5. Block diagram of the wrist band.

Figure 5b,c illustrates the images of the wrist band having both physiological sensors, i.e., Heart Rate and BP, attached with it. This provides an RF Modem that has the capability to send to an RF modem in another part of the room. Moreover, the other node has a Wi-Fi module also that is uploading the recorded values on the cloud. The band is working with a LIPO battery, so it has quite a long backup and can work for long. The battery is rechargeable, and very low power is required to perform that. The vision mote is shown in Figure 6, which is displaying the real-time values of BP and heart rate and uploading the same on the cloud also via a Wi-Fi module.

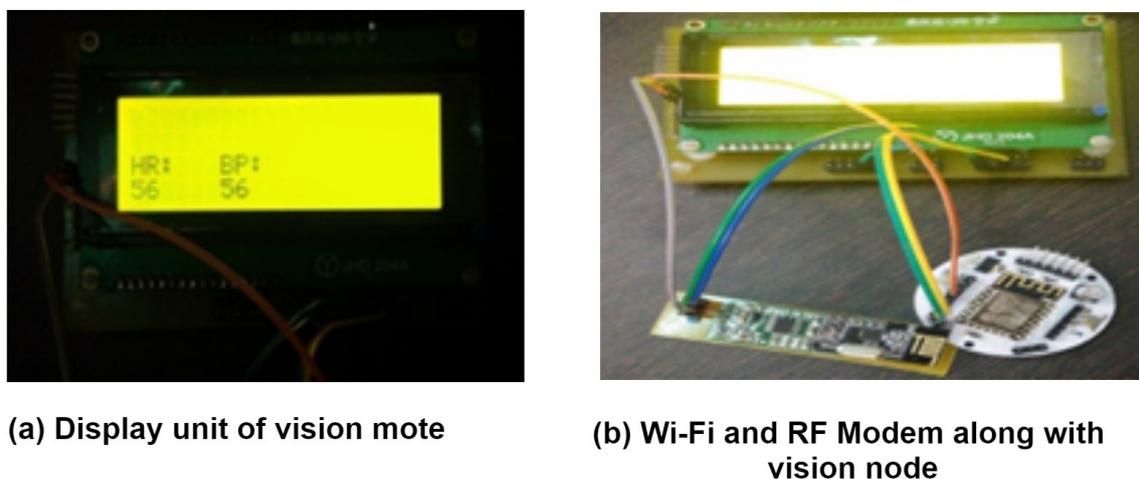


Figure 6. Real-time value in vision mote.

The complete system has been designed by customization of the boards, and the bit map of the RF modem, i.e., part of the vision mote, is shown in Figure 7a. The bit map of the wristband is also shown in Figure 7b, which has an RF modem in itself also. The threshold value that is used to detect the criticality of the situation is shown in Table 1.

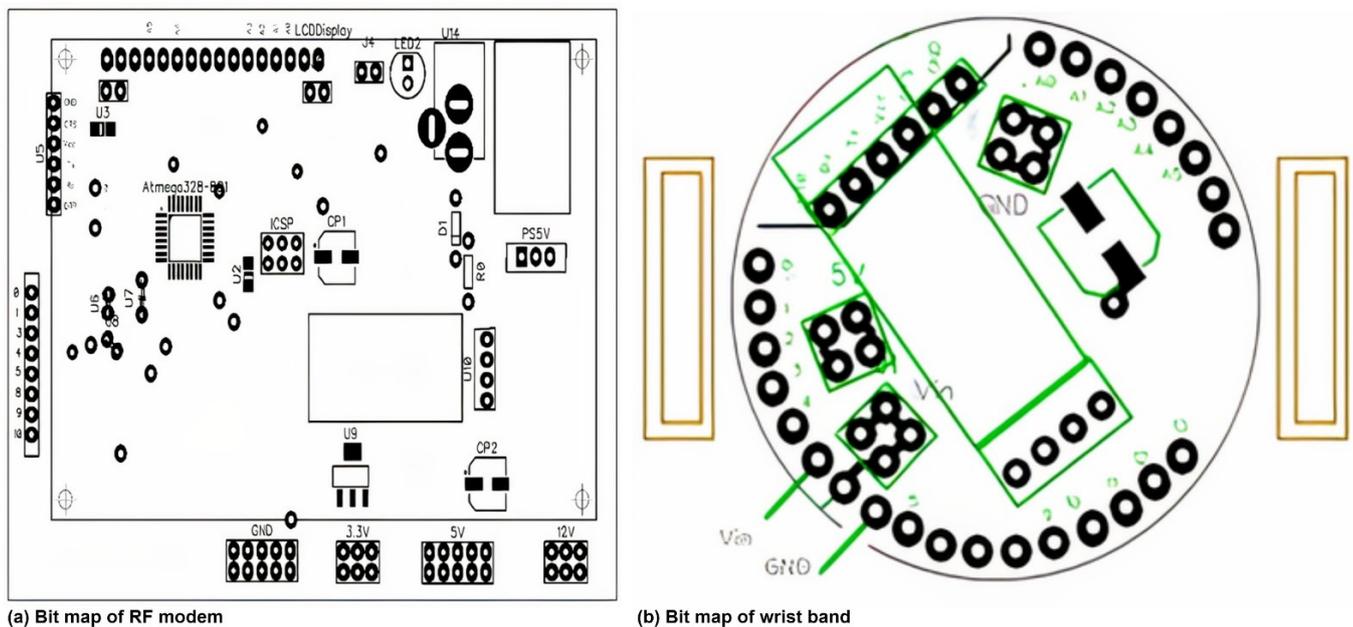


Figure 7. (a) Bit map of the RF Modem and (b) Bit map of the wrist band.

Table 1. Blood pressure categories for ages (18 years and older) [46].

Physical State	Systolic (mm Hg)	Diastolic (mm Hg)
Hypotension	<90	<60
Desired	90–119	60–79
Prehypertension	120–139	80–89
Stage 1 Hypertension	140–159	90–99
Stage 2 Hypertension	160–179	100–109
Hypertensive crisis	≥180	≥110

The two-dimensional model based on valance and arousal is shown in Figure 8. The model explains the 4 basic emotional states and corresponding primary and tertiary emotions. Figure 9 shows the experimental setup that is established to capture the real-time facial emotion of the subjects along with the physiological values that include heart rate and blood pressure. This experimental setup includes a wrist band with physiological sensors, such as heart rate and BP, and a vision node as well. The vision node is capturing the facial emotions in real-time as the subjects are made to watch the videos that can take them to various emotional states, and proper time has been given to all the subjects to carry out this work efficiently. It generally takes time to switch from one emotional state to another state. Therefore, proper care has been taking in that direction while conducting this experiment.

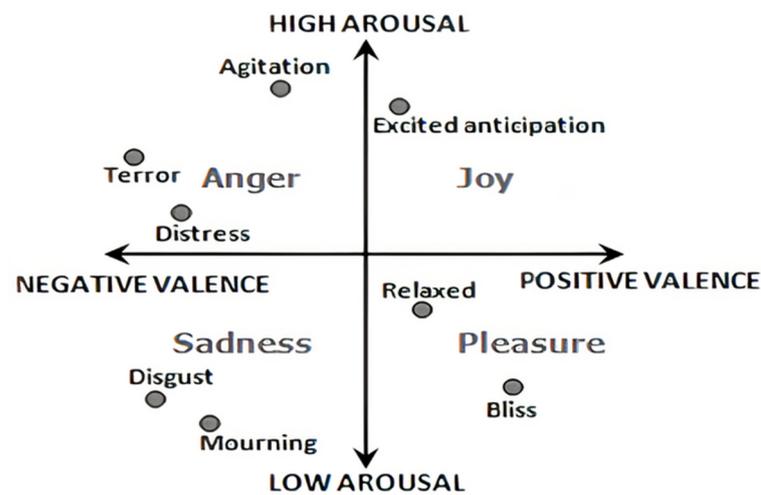


Figure 8. Illustration of 2D model based on valence and arousal (Russell, 1976).



Figure 9. Experimental setup done to capture the facial emotions and physiological sensor values in real-time on different subjects.

5. Results and Validation

5.1. Results

In this section, a detailed description of all the experiments is given. The performance of various models is explained in this section. The experiments that have been displayed in this chapter are done on Google-CoLab with a 12GB NVIDIA Tesla K80 GPU (NVIDIA, Leeds, UK). Another model that has been used for training is the Mini_Xception model. This is a modified depth-wise separable convolutional neural network. As compared to the conventional convolution neural network, this particular model does not require performing convolution across all the channels. This particular thing makes this model lighter and also reduces the connections that are very few in comparison to conventional models. The model architecture of the Mini_Xception model is shown in Figure 10.

The major benefit of this architecture is that it does not contain any fully connected layers, and the inclusion of depth-wise separable convolutions helps to reduce the number of parameters. The introduction of residual models also enables the gradients to perform better in backpropagation to lower layers. The network is trained via Google Co-Lab in batch mode, using SGD and Adam optimizer separately, and achieves an accuracy of 69% with 35 epochs with Adam optimizer. The efficiency came maximum with Adam as SGD is more locally unstable. Instead of the Mini_Xception and Mobilenet_V2 models, the dataset has been trained on Densenet161 and the Resnet Model also. The results for these models are quite less as compared to the previous two models. Therefore, these two models are not considered for further consideration as the accuracy is quite less in comparison to the accuracy with the other two models. A brief description of the model is shown in Table 2, which explains the name of the model, accuracy, learning rate, test accuracy, and the optimizer used for the model.

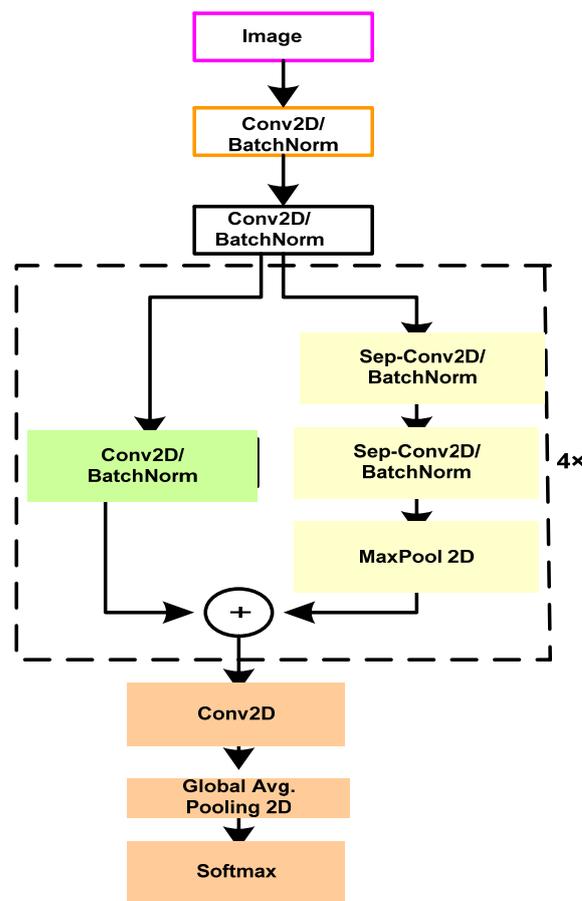
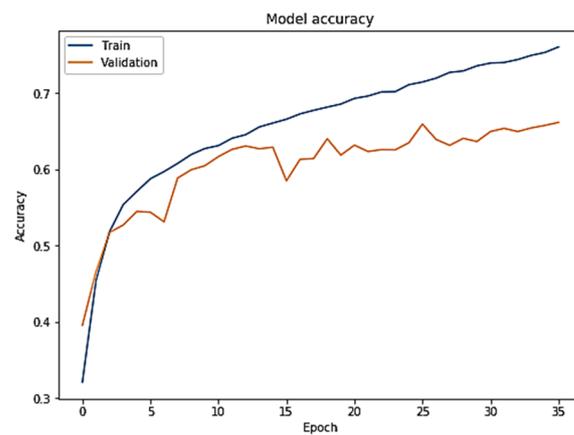


Figure 10. Mini_Xception model architecture.

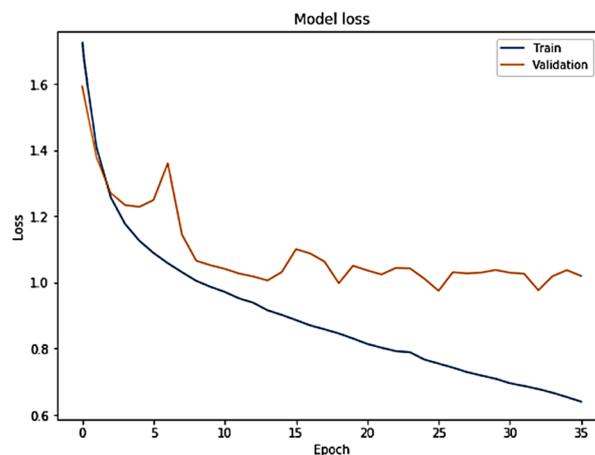
Table 2. Various Models tested on FER 2013 Database.

Model	Accuracy	Learning Rate	Test Accuracy	Optimizer
Mini_Xception	73%	0.005	69%	Adam
Densenet161	59%	0.001, 0.001, 0.005	43%	SGD
Resnet38	68%	0.0001	60%	SGD
Mobilenet_V2	72.5%	0.0001, 0.001	64%	Adam

Figure 11 shows the training loss, and from the graph, it is visible that loss is decreasing exponentially, and till the 35th epoch, the loss has reduced to a minimum. From the confusion matrix, it has been seen that the disgusted faces are misclassified as angry faces, and the reason behind that is the count of disgusted faces in a dataset is the least. The major reason behind the misclassification is a non-uniform dataset; that is how the FER 2013 dataset is distributed. The model accuracy shown in Figure 11 has reached the training accuracy of 73%. The accuracy that has been achieved with a model using 35 epochs is quite high and can be considered for deployment on the system for real-time facial emotion detection.



(a) Training accuracy



(b) Training loss

Figure 11. Training accuracy and loss of Mini_Xception model Architecture.

The setup has been used to detect facial emotions in real-time and validate the same via physiological sensors. The wrist band is designed to record the physiological values such as heart rate and blood pressure of the subject under various situations. The source of empirical data in this experimentation is the facial emotions of the subjects with a timestamp. At the same time, the data gathered via physiological sensors at the same time.

The subjects are set to see the videos that help them to enter various emotional states, and then the heart rate and blood pressure of those subjects, along with facial expression, are captured at the same time. The recorded values with time stamps and facial images with a timestamp are then used further to validate the expression recorded via the system. The system has been designed particularly to detect emotions with two tiers of validation. Firstly, via facial images only and then for further validation of the extracted emotions, physiological sensors have been used. Table 3 shows the description of videos that have been used to carry out the experiment using different subjects. As the proposed system is being tested and designed for edge devices, and it has been found in the prior literature that real implementation on a resource-constrained device such as Raspberry-Pi is challenging [9]. So, to begin with, initially, four expressions have been recorded and validated. In the future, more powerful embedded boards can be used to increase the efficiency of the system. The different videos that can bring any person to a happy, sad, and angry state are mentioned in this Table 3. Radar plots for variation in blood pressure under happy, angry, and neutral states are illustrated in Figure 12. The experiment has been recorded for 20 different people, but for validation purposes, only three subjects with

five different observations under four different emotional states have been used, as shown in Table 4.

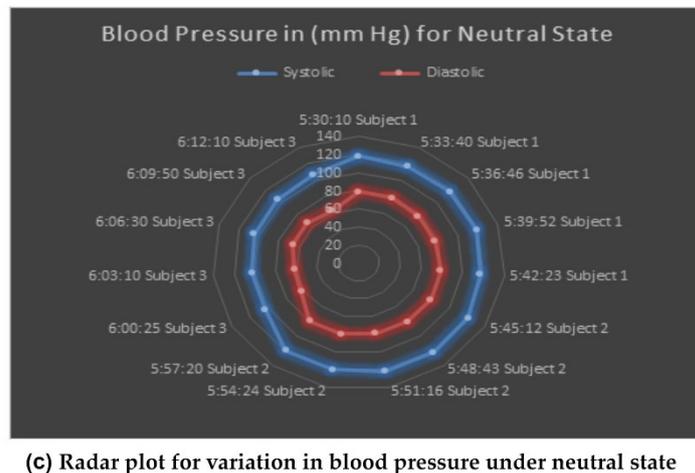
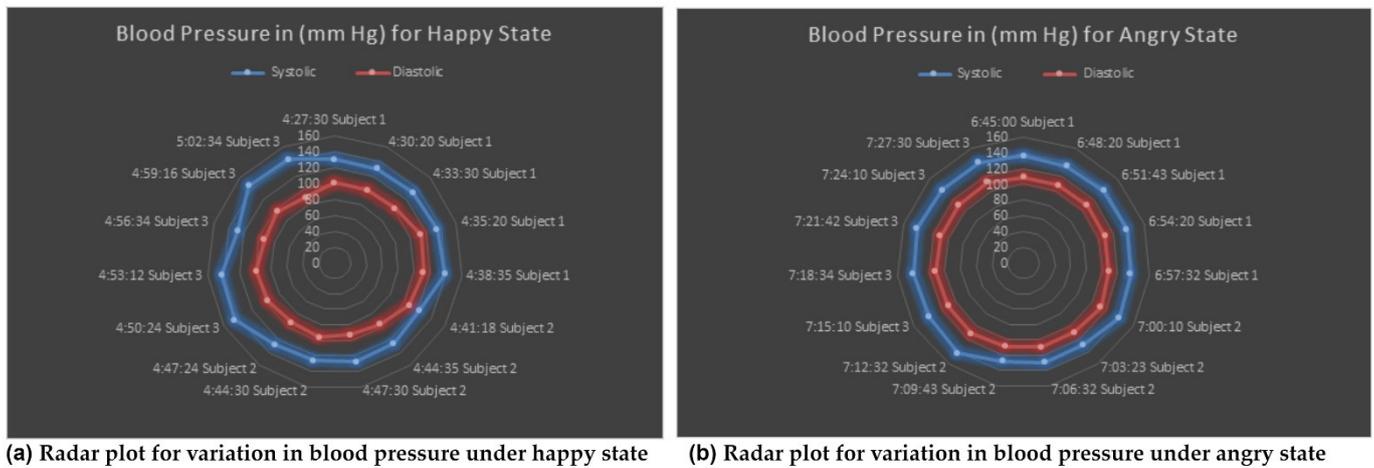


Figure 12. Radar plot for variation in blood pressure under happy, angry, and neutral state.

Table 3. Details of happy, sad, and angry videos for experimental analysis.

S.No.	Name of Video	Duration of Video (mm:ss)	Video Type
1	Tom and Jerry	25:59	Happy
2	Mr.Bean at the Dentist	27:20	Happy
3	Contagious Laughter Compilation	16:00	Happy
4	Sridevi Funeral video	14:54	Sad
5	Mumbai Terror Attack videos	16:00	Sad
6	She tells her story on why she fled away from North Korea	9:38	Sad
7	Best of Angry people	14:29	Angry
8	Nirbhaya’s Mothers Interview	17:25	Angry
9	Pit Bull Terrier Dog Attacks	9:38	Angry

Table 4. Experimental results of the three subjects under happy, neutral, and angry conditions.

S. No.	Subject Number	Age	Blood Pressure (mm Hg)		Heart Rate (BPM)	Expression	Timestamp (hh:mm:ss)
			Systolic	Diastolic			
1	Subject 1	20	130	101	91	Happy	4:27:30
2	Subject 1	20	130	100	91	Happy	4:30:20
3	Subject 1	20	132	102	94	Happy	4:33:30
4	Subject 1	20	135	114	91	Happy	4:35:20
5	Subject 1	20	139	112	102	Happy	4:38:35
6	Subject 2	22	122	108	98	Happy	4:41:18
7	Subject 2	22	126	96	80	Happy	4:44:35
8	Subject 2	22	128	93	83	Happy	4:47:30
9	Subject 2	22	126	96	80	Happy	4:44:30
10	Subject 2	22	128	93	90	Happy	4:47:24
11	Subject 3	21	145	97	79	Happy	4:50:24
12	Subject 3	21	143	99	83	Happy	4:53:12
13	Subject 3	21	128	93	70	Happy	4:56:34
14	Subject 3	21	145	97	75	Happy	4:59:16
15	Subject 3	21	143	90	79	Happy	5:02:34
16	Subject 1	20	118	79	75	Neutral	5:30:10
17	Subject 1	20	117	79	76	Neutral	5:33:40
18	Subject 1	20	118	77	76	Neutral	5:36:46
19	Subject 1	20	119	77	75	Neutral	5:39:52
20	Subject 1	20	117	79	76	Neutral	5:42:23
21	Subject 2	22	121	80	78	Neutral	5:45:12
22	Subject 2	22	122	80	77	Neutral	5:48:43
23	Subject 2	22	122	79	77	Neutral	5:51:16
24	Subject 2	22	120	80	76	Neutral	5:54:24
25	Subject 2	22	119	79	76	Neutral	5:57:20
26	Subject 3	21	103	63	61	Neutral	6:00:25
27	Subject 3	21	103	62	61	Neutral	6:03:10
28	Subject 3	21	105	66	70	Neutral	6:06:30
29	Subject 3	21	105	66	71	Neutral	6:09:50
30	Subject 3	21	107	63	74	Neutral	6:12:10
31	Subject 1	20	136	109	85	Angry	6:45:00
32	Subject 1	20	135	108	84	Angry	6:48:20
33	Subject 1	20	137	109	86	Angry	6:51:43
34	Subject 1	20	137	109	87	Angry	6:54:20
35	Subject 1	20	136	109	87	Angry	6:57:32
36	Subject 2	22	140	113	102	Angry	7:00:10
37	Subject 2	22	129	110	98	Angry	7:03:23
38	Subject 2	22	130	110	98	Angry	7:06:32
39	Subject 2	22	128	109	100	Angry	7:09:43
40	Subject 2	22	142	112	105	Angry	7:12:32
41	Subject 3	21	138	110	108	Angry	7:15:10
42	Subject 3	21	140	112	109	Angry	7:18:34
43	Subject 3	21	142	111	107	Angry	7:21:42
44	Subject 3	21	138	110	107	Angry	7:24:10
45	Subject 3	21	140	112	108	Angry	7:27:30

The experimental values of 20 different subjects under different emotional states are recorded, and Table 4 shows the recorded values of three different subjects for basic emotions such as anger, neutral, happy, and sad. The values shown in Table 4 are recorded under an experimental environment where the three different subjects sat and wore a wrist band that consists of a heart rate sensor and blood pressure sensor on it. Once the subject was wearing this sensor, the setup of Raspberry-Pi with a Pi camera has also started to capture the expression of the person in real-time along with the physiological values. Figure 13 illustrates the captured facial expressions with a time stamp.



Figure 13. Expressions captured via the experimental setup with a timestamp.

A time-synchronized algorithm has been used to capture the facial expression and physiological values of the participants, i.e., heart rate and blood pressure. To validate the results, various analyses have been conducted. It has been found in the literature that emotional arousal increases systolic and diastolic blood pressure. Moreover, it has also been found in the literature that happiness, anger, and anxiety increase blood pressure, and the level of variation is dependent upon the individuals. In order to visualize the physiological values, box plots have been plotted. Figure 14a shows the box plot of the systolic blood pressure of the participants for all four expressions. From the box plot, it is clear that sadness tends to decrease the systolic blood pressure of the participants to the lowest, and on the other hand, anger and happiness tend to increase the systolic blood pressure of the participants. The first quartile and third quartile for each expression are also shown, which shows 25% and 75% of the values are lying under these quartiles. The medians for all the recorded values are also labeled on the box plot of each expression, which depicts the distribution of the systolic values for that particular expression.

Figure 14b shows the box plot of the diastolic blood pressure of the participants for all four expressions. From the box plot, it is clear that sadness tends to decrease the diastolic blood pressure of the participants to the lowest, and on the other hand, anger and happiness tend to increase the diastolic blood pressure of the participants. The first quartile and third quartile for each expression are also shown, which shows 25% and 75% of the values are lying under these quartiles. The medians for all the recorded values are also labeled on the box plot of each expression, which depicts the distribution of the systolic values for that particular expression. One outlier, i.e., the fourth recorded value of subject 1 from Table 4, where the value is comparatively high as compared to other recorded values, i.e., 114, has been depicted as an outlier.

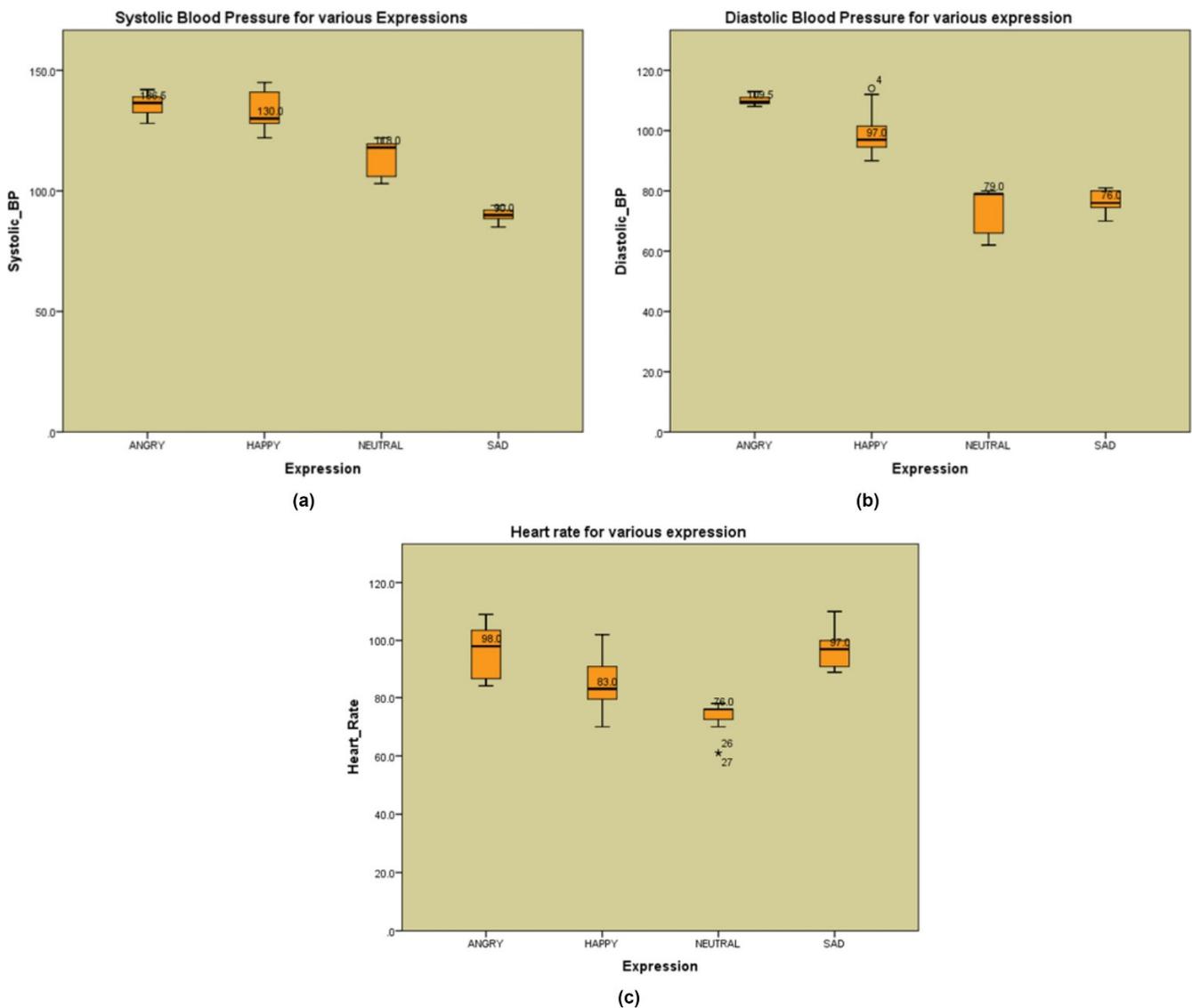


Figure 14. Box plots: (a) systolic blood pressure, (b) diastolic blood pressure, and (c) heart rate.

Figure 14c shows the box plot of the heart rate variation of the participants for all four expressions. From the box plot, it is clear that anger raises the blood pressure of participants to the maximum level while the neutral state has shown the minimum. The first quartile and third quartile for each expression are also shown, which shows 25% and 75% of the values are lying under these quartiles. The medians for all the recorded values are also labeled on the box plot of each expression, which depicts the distribution of the Heart rate values for that particular state. Two outliers for the neutral state, i.e., 26th and 27th value recorded and located at the 26th and 27th value in Table 4. Both the values are the same and comparatively low when compared to other recorded values, i.e., 61, hence depicted as an outlier. In order to validate the variation of physiological recorded values for various mental states, a paired-sample *t*-test has been applied. A paired-sample *t*-test is the statistical solution and is mainly used when we want to see if the mean difference between the two sets of observations has been found or not. Therefore, to validate our variation on the experimentally recorded values, this test has been utilized.

5.2. Validation

a. Paired-Sample *t*-test Analysis between Happy and Neutral States

H_1 = There is a significant decrease in the systolic blood pressure of participants when their emotional state is changing from happy to neutral.

$$H_1 : \mu_1 - \mu_2 < \partial_0 \quad (1)$$

where $\mu_1 - \mu_2$ is the difference between the hypotheses means, and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the systolic blood pressure of participants while watching different videos using an experimental setup for 20 participants for happy and neutral videos. There was a significant difference in systolic blood pressure while watching happy videos (M = 133.333, SD = 7.7429) and systolic blood pressure while watching neutral videos (M = 114.400, SD = 7.3853) conditions; $t(14) = 5.157, p = 0.000$ are shown in Table 5. There was a significant decrease in the systolic blood pressure when participants watched neutral videos after watching happy videos.

Table 5. Paired samples statistics for systolic BP (happy to neutral state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Systolic_BP_Happy	133.333	15	7.7429	1.9992
	Systolic_BP_Neutral	114.400	15	7.3853	1.9069

Hence enough evidence has been found that shows the mean difference between the systolic blood pressure of participants is statistically significant when their emotional state is changing from happy to neutral. Hence, the hypothesis is accepted, which says that there is a significant decrease in the systolic blood pressure of participants when their emotional state is changing from happy to neutral.

H_1 (Alternate Hypothesis) = There is a significant decrease in the diastolic blood pressure of participants when their emotional state is changing from happy to neutral

$$H_1 : \mu_1 - \mu_2 < \partial_0 \quad (2)$$

where $\mu_1 - \mu_2$ is the difference between the hypotheses means, and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the diastolic blood pressure of participants while watching different videos using an experimental setup for 20 participants for happy and neutral videos. There was a significant difference in diastolic blood pressure while watching happy videos (M = 99.400, SD = 7.0791) and diastolic blood pressure while watching neutral videos (M = 73.933, SD = 7.3918) conditions; $t(14) = 12.222, p = 0.000$ are shown in Table 6. There was a significant decrease in the diastolic blood pressure when participants watched neutral videos after watching happy videos.

Hence enough evidence has been found that shows the mean difference between the diastolic blood pressure of participants is statistically significant when their emotional state is changing from happy to neutral. Therefore, the hypothesis is accepted, which says that there is a significant decrease in the diastolic blood pressure of participants when their emotional state is changing from happy to neutral.

Table 6. Paired samples statistics for diastolic BP (happy to neutral state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Diastolic_BP_Happy	99.400	15	7.0791	1.8278
	Diastolic_BP_Neutral	73.933	15	7.3918	1.9085

H_1 (Alternate Hypothesis) = There is a significant decrease in the heart rate of participants when their emotional state is changing from happy to neutral.

$$H_1 : \mu_1 - \mu_2 < \partial_0 \quad (3)$$

where $\mu_1 - \mu_2$ is the difference between the hypotheses means, and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the heart rate of participants while watching different videos using an experimental setup for 20 participants for happy and neutral videos. There was a significant difference in Heart rate while watching happy videos ($M = 85.733$, $SD = 8.9400$) and diastolic blood pressure while watching neutral videos ($M = 73.267$, $SD = 5.4178$) conditions; $t(14) = 5.983$, $p = 0.000$ are shown in Table 7. There was a significant decrease in the diastolic blood pressure when participants watched neutral videos after watching happy videos.

Hence enough evidence has been found that shows the mean difference between the heart rate of participants is statistically significant when their emotional state is changing from happy to neutral. Therefore, the hypothesis is accepted, which says that there is a significant decrease in the heart rate of participants when their emotional state is changing from happy to neutral.

Table 7. Paired samples statistics for heart rate (happy to neutral state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Systolic_BP_Happy	133.333	15	7.7429	1.9992
	Systolic_BP_Neutral	114.400	15	7.3853	1.9069

b. Paired-Sample *t*-test Analysis on Neutral and Angry States:

H_1 (Alternate Hypothesis) = There is a significant increase in the systolic blood pressure of participants when their emotional state is changing from neutral to angry.

$$H_1 : \mu_1 - \mu_2 > \partial_0 \quad (4)$$

where $\mu_1 - \mu_2$ is the difference between the hypotheses means, and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the Systolic blood pressure of participants while watching different videos using an experimental setup for 20 participants for neutral and angry videos. There was a significant difference in systolic blood pressure while watching neutral videos ($M = 114.400$, $SD = 7.3853$) and diastolic blood pressure while watching angry videos ($M = 136.533$, $SD = 4.4379$) conditions; $t(14) = -8.137$, $p = 0.000$ are shown in Table 8. There was a significant increase in the Systolic blood pressure when participants watched neutral videos after watching happy videos.

Hence enough evidence has been found that shows the mean difference between systolic blood pressure of participants is statistically significant when their emotional state is changing from neutral to angry. Therefore, the hypothesis is accepted, which says that there is a significant increase in the systolic blood pressure of participants when their emotional state is changing from neutral to angry.

Table 8. Paired samples statistics for systolic BP (neutral to angry state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Systolic_BP_Neutral	114.400	15	7.3853	1.9069
	Systolic_BP_Angry	136.533	15	4.4379	1.1459

H_1 (Alternate Hypothesis) = There is a significant increase in the diastolic blood pressure of participants when their emotional state is changing from neutral to angry.

$$H_1 : \mu_1 - \mu_2 > \partial_0 \quad (5)$$

where $\mu_1 - \mu_2$ is the difference between the hypothesis means, and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the diastolic blood pressure of participants while watching different videos using an experimental setup for 20 participants for neutral and angry videos. There was a significant difference in Systolic blood pressure while watching neutral videos ($M = 73.933$, $SD = 7.3918$) and diastolic blood pressure while watching angry videos ($M = 110.200$, $SD = 1.4736$) conditions; $t(14) = -17.413$, $p = 0.000$ are shown in Table 9. There was a significant increase in the diastolic blood pressure when participants watched neutral videos after watching happy videos. Hence enough evidence has been found that shows the mean difference between diastolic blood pressure of participants is statistically significant when their emotional state is changing from neutral to angry. Hence the hypothesis is accepted, which says that there is a significant increase in the diastolic blood pressure of participants when their emotional state is changing from neutral to angry.

Table 9. Paired samples statistics for diastolic BP (neutral to angry state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Diastolic_BP_Neutral	73.933	15	7.3918	1.9085
	Diastolic_BP_Angry	110.200	15	1.4736	0.3805

H_1 (Hypothesis) = There is a significant increase in the heart rate of participants when their emotional state is changing from neutral to angry.

$$(H_1 : \mu_1 - \mu_2 > \partial_0) \quad (6)$$

where $\mu_1 - \mu_2$ is the difference between the hypothesis means, and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the systolic blood pressure of participants while watching different videos using an experimental setup for 20 participants for neutral and angry videos. There was a significant difference in heart rate while watching neutral videos ($M = 73.267$, $SD = 5.4178$) and heart rate while watching angry videos ($M = 98.067$, $SD = 9.6471$) conditions; $t(14) = -7.170$, $p = 0.000$ are shown in Table 10. There was a significant increase in the heart rate when participants watched angry videos after watching neutral videos. Hence enough evidence has been found that shows the mean difference between the heart rate of participants is statistically significant when their emotional state is changing from neutral to angry. Hence the hypothesis is accepted, which says that there is a significant increase in the heart rate of participants when their emotional state is changing from neutral to angry.

Table 10. Paired samples statistics for heart rate (neutral to angry state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Heart_Rate_Neutral	73.267	15	5.4178	1.3989
	Heart_Rate_Angry	98.067	15	9.6471	2.4909

c. Paired-Sample *t*-test Analysis on the Angry and Sad States

H_1 (Alternate Hypothesis) = There is a significant decrease in the systolic blood pressure of participants when their emotional state is changing from angry to sad.

$$H_1 : \mu_1 - \mu_2 < \partial_0 \quad (7)$$

where $\mu_1 - \mu_2$ the difference between the hypotheses means and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the systolic blood pressure of participants while watching different videos using an experimental setup for 20 participants for angry and sad videos. There was a significant difference in systolic blood pressure while watching angry videos ($M = 136.533$, $SD = 4.4379$) and systolic blood pressure while watching sad videos ($M = 90.200$, $SD = 2.7568$) conditions; $t(14) = 31.535$, $p = 0.000$ are shown in Table 11. There was a significant decrease in the diastolic blood pressure when participants watched neutral videos after watching happy videos.

Hence enough evidence has been found that shows the mean difference between the diastolic blood pressure of participants is statistically significant when their emotional state is changing from happy to neutral. Hence the hypothesis is accepted, which says that there is a significant decrease in the systolic blood pressure of participants when their emotional state is changing from angry to sad.

Table 11. Paired samples statistics for systolic BP (angry to sad state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Systolic_BP_Angry	136.533	15	4.4379	1.1459
	Systolic_BP_Sad	90.200	15	2.7568	0.7118

H_1 (Alternate Hypothesis) = There is a significant decrease in the diastolic blood pressure of participants when their emotional state is changing from angry to sad.

$$H_1 : \mu_1 - \mu_2 < \partial_0 \quad (8)$$

where $\mu_1 - \mu_2$ is the difference between the hypotheses means, and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the diastolic blood pressure of participants while watching different videos using an experimental setup for 20 participants for angry and sad videos. There was a significant difference in diastolic blood pressure while watching angry videos ($M = 110.200$, $SD = 1.4736$) and in diastolic blood pressure while watching sad videos ($M = 76.53$, $SD = 3.563$) conditions; $t(14) = 33.722$, $p = 0.000$ are shown in Table 12. There was a significant decrease in the diastolic blood pressure when participants watched neutral videos after watching happy videos.

Hence enough evidence has been found that shows the mean difference between the diastolic blood pressure of participants is statistically significant when their emotional state is changing from angry to sad. Hence the hypothesis is accepted, which says that there is a significant decrease in the diastolic blood pressure of participants when their emotional state is changing from angry to sad.

Table 12. Paired samples statistics for diastolic bp (angry to sad state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Diastolic_BP_Angry	110.200	15	1.4736	0.3805
	Diastolic_BP_Sad	76.53	15	3.563	0.920

H_0 (Null Hypothesis) = There is no significant difference in the heart rate of the participants when their emotional state is changing from angry to sad.

H_1 (Alternate Hypothesis) = There is a significant difference in the heart rate of the participants when their emotional state is changing from angry to sad.

$$H_1 : \mu_1 - \mu_2 < \partial_0 \quad (9)$$

where $\mu_1 - \mu_2$ is the difference between the hypotheses means, and ∂_0 is the hypothesized difference.

A paired-sample *t*-test was conducted to compare the heart rate of participants while watching different videos using an experimental setup for 20 participants for angry to sad videos. There was no significant difference in the heart rate while watching angry videos ($M = 98.067$, $SD = 9.6471$) and while watching sad videos ($M = 96.067$, $SD = 5.8367$) conditions; $t(14) = 0.587$, $p = 0.566$ are shown in Table 13.

Hence enough evidence has been found that shows the mean difference between heart rate of not the participants is statistically significant when their emotional state is changing from angry to sad. Hence the null hypothesis is accepted, which says that there is no significant difference in the heart rate of the participants when their emotional state is changing from angry to sad.

Table 13. Paired samples statistics for heart rate (angry to sad state).

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Heart_Rate_Angry	98.067	15	9.6471	2.4909
	Heart_Rate_Sad	96.067	15	5.8367	1.5070

Table 14 illustrates the complete values of three kinds of pair states, including happy to neutral, neutral to angry, and angry to sad, concerning three parameters, namely systolic BP, diastolic BP, and heart rate.

Table 14. Paired-sample *t*-test parameters for validation.

S. No.	Pair State	Parameter	Paired Difference					
			Mean	St. Dev	St. Error	t	df	Sig 2-Tailed
1.	Happy to Neutral	Systolic BP	18.933	14.2200	3.6716	5.157	14	0.000
		Diastolic BP	25.467	8.0699	2.0836	12.22	14	0.000
		Heart Rate	18.933	14.220	3.6716	5.157	14	0.000
2.	Neutral to Angry	Systolic BP	−22.133	10.5347	2.7201	−8.137	14	0.000
		Diastolic BP	36.266	8.0664	2.0827	5.157	14	0.000
		Heart Rate	−24.800	13.3962	3.4589	−7.170	14	0.000
3.	Angry to Sad	Systolic BP	46.333	5.6904	1.4693	31.535	14	0.000
		Diastolic BP	33.666	3.8668	−0.9984	33.720	14	0.000
		Heart Rate	2.000	13.1909	3.4059	0.587	14	0.566

Table 14 also shows the close correlation between the data that has been captured via the physiological sensor and the expression that has been captured via the device. Enough evidence has been found in the experiment that shows that the paired state variation with parameter variation is statistically significant.

6. Conclusions

In this article, we have designed and implement an IoMT based portable FER edge device to recognize the facial emotion of an individual. FER is achieved by interfering with the systole, diastole, and heart rate sensor data of an individual with visuals capture through an edge device. The edge device is integrated with Intel Movidius neural com-

puting stick2 (NCS2), and a deep convolutional neural network implemented on NCS2 enables the edge device to recognize facial emotion accurately. The facial emotion detection test accuracy ranged from 56% to 73% using various models, and the accuracy became 73% and performed very well with the FER 2103 dataset in comparison to the state of art results with a 64% maximum. Finally, a *t*-test validation is conducted for identifying the significant differences in systolic, diastolic, and heart rate of an individual during watching the different subjects' visuals clips.

The primary goal of this work is to develop a system that can replace the existing bulky, wired and system-dependent system that almost makes the work of face and facial emotion detection impossible while walking on roads, airports, hospitals, and public places. One must spend a lot to receive the benefit of such a system. Moreover, it has also been observed from the literature review that studies have been carried out on various techniques that are required to achieve facial emotion recognition, but no literature has been found in the direction of designing and implementing a device (portable, cheap, and efficient) in real-time. This paper suggests the development of intelligence that can detect human faces and their emotions in real-time. A smart system and an IOT-based vision Mote device, which are designed for detecting the real-time behavior of a person. It is a small contribution to the social cause as the device is designed for detecting the real-time behavior of people under different situations. This device will automatically detect a human presence and capture the human face along with its facial emotions. Hence, it will collect real-time data and upload the captured emotions on the cloud, which can be accessed remotely.

The main aim behind this work is to develop a system that can understand human emotions at any point in time, irrespective of age, gender, and race. Moreover, successful efforts have been performed to make the system compact and cost-effective over existing heavy, costly, hefty, and complex facial emotion detection systems. In the future, the system can be implemented with the help of more powerful embedded boards that are available in the market, such as NVIDIA's Jetson Nano (NVIDIA, Leeds, UK) and Google Coral's Dev Board (Coral, Tuscaloosa, AL, USA). These boards may increase the cost a little but can make the existing system more efficient and capable of handling more complex Deep Neural Networks. However, to make the system maintenance, free solar batteries are also suggested. In this work, only those deep networks are optimized that can be easily deployable on Raspberry-Pi being a resource-constrained device, and the efficiency of 73% has been achieved, but in the future, with the help of embedded boards, various deep learning models can be used with better efficiency. The accuracy achieved via the propped system is sufficiently good as the system can measure physiological parameters of a human being via wrist band and, at the same time, capable of detecting facial emotions in real-time.

Author Contributions: N.R., S.V.A. and R.S. made contributions to conception and manuscript writing; S.S.A. and A.G. examined and supervised this research and the outcomes; M.R. and A.S.A. revised and polished the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Taif University Researchers Supporting Project number (TURSP-2020/215), Taif University, Taif, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Irfan, M.; Ahmad, N. Internet of medical things: Architectural model, motivational factors and impediments. In Proceedings of the 2018 15th Learning and Technology Conference (L&T), Jeddah, Saudi Arabia, 25–26 February 2018; pp. 6–13.
2. Nayyar, A.; Puri, V.; Nguyen, N.G. BioSenHealth 1.0: A Novel Internet of Medical Things (IoMT)-Based Patient Health Monitoring System. In *Lecture Notes in Networks and Systems*; Springer Nature: Cham, Switzerland, 2019; Volume 55, pp. 155–164.
3. Rahman, M.A.; Hossain, M.S. An Internet of medical things-enabled edge computing framework for tackling COVID-19. *IEEE Internet Things J.* **2021**. [[CrossRef](#)]
4. Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vis. Comput.* **2009**, *27*, 803–816. [[CrossRef](#)]

5. Kwong, J.C.T.; Garcia, F.C.C.; Abu, P.A.R.; Reyes, R.S.J. Emotion recognition via facial expression: Utilization of numerous feature descriptors in different machine learning algorithms. In Proceedings of the TENCON 2018-2018 IEEE Region 10 Conference, Jeju, Korea, 28–31 October 2018; pp. 2045–2049.
6. Rodríguez-Pulecio, C.G.; Benítez-Restrepo, H.D.; Bovik, A.C. Making long-wave infrared face recognition robust against image quality degradations. *Quant. Infrared Thermogr. J.* **2019**, *16*, 218–242. [[CrossRef](#)]
7. Canedo, D.; Neves, A.J.R. Facial expression recognition using computer vision: A systematic review. *Appl. Sci.* **2019**, *9*, 4678. [[CrossRef](#)]
8. Ruiz-Garcia, A.; Elshaw, M.; Altahhan, A.; Palade, V. A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots. *Neural Comput. Appl.* **2018**, *29*, 359–373. [[CrossRef](#)]
9. Sajjad, M.; Nasir, M.; Ullah, F.U.M.; Muhammad, K.; Sangaiah, A.K.; Baik, S.W. Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services. *Inf. Sci. N. Y.* **2019**, *479*, 416–431. [[CrossRef](#)]
10. Srihari, K.; Ramesh, R.; Udayakumar, E.; Dhiman, G. An Innovative Approach for Face Recognition Using Raspberry Pi. *Artif. Intell. Evol.* **2020**, 103–108. [[CrossRef](#)]
11. Gaikwad, P.S.; Kulkarni, V.B. *Face Recognition Using Golden Ratio for Door Access Control System*; Springer: Singapore, 2021; pp. 209–231.
12. Lin, H.; Garg, S.; Hu, J.; Wang, X.; Piran, M.J.; Hossain, M.S. Privacy-enhanced data fusion for COVID-19 applications in intelligent Internet of medical Things. *IEEE Internet Things J.* **2020**. [[CrossRef](#)]
13. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.A.S.; Asari, V.K. A state-of-the-art survey on deep learning theory and architectures. *Electronics* **2019**, *8*, 292. [[CrossRef](#)]
14. Jain, Y.; Gandhi, H.; Burte, A.; Vora, A. Mental and Physical Health Management System Using ML, Computer Vision and IoT Sensor Network. In Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020, Coimbatore, India, 5–7 November 2020; pp. 786–791. [[CrossRef](#)]
15. Zedan, M.J.M.; Abduljabbar, A.I.; Malallah, F.L.; Saeed, M.G. Controlling Embedded Systems Remotely via Internet-of-Things Based on Emotional Recognition. *Adv. Hum. Comput. Interact.* **2020**, 2020. [[CrossRef](#)]
16. Abbasnejad, I.; Sridharan, S.; Nguyen, D.; Denman, S.; Fookes, C.; Lucey, S. Using synthetic data to improve facial expression analysis with 3d convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 1609–1618.
17. Tümen, V.; Söylemez, Ö.F.; Ergen, B. Facial emotion recognition on a dataset using Convolutional Neural Network. In Proceedings of the IDAP 2017—International Artificial Intelligence and Data Processing Symposium, Malatya, Turkey, 16–17 September 2017. [[CrossRef](#)]
18. Saran, R.; Haricharan, S.; Praveen, N. Facial emotion recognition using deep convolutional neural networks. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 2020–2025. [[CrossRef](#)]
19. Cheng, H.; Su, Z.; Xiong, N.; Xiao, Y. Energy-efficient node scheduling algorithms for wireless sensor networks using Markov Random Field model. *Inf. Sci. N. Y.* **2016**, *329*, 461–477. [[CrossRef](#)]
20. Breuer, R.; Kimmel, R. A deep learning perspective on the origin of facial expressions. *arXiv* **2017**, arXiv:1705.01842.
21. Sajjad, M.; Shah, A.; Jan, Z.; Shah, S.I.; Baik, S.W.; Mehmood, I. Facial appearance and texture feature-based robust facial expression recognition framework for sentiment knowledge discovery. *Clust. Comput.* **2018**, *21*, 549–567. [[CrossRef](#)]
22. Zhang, L.; Verma, B.; Tjondronegoro, D.; Chandran, V. Facial expression analysis under partial occlusion: A survey. *arXiv* **2018**, arXiv:1802.08784. [[CrossRef](#)]
23. Zhu, C.; Zheng, Y.; Luu, K.; Savvides, M. CMS-RCNN: Contextual multi-scale region-based cnn for unconstrained face detection. In *Deep Learning for Biometrics*; Springer Nature: Cham, Switzerland, 2017; pp. 57–79.
24. Al-Shabi, M.; Cheah, W.P.; Connie, T. Facial Expression Recognition Using a Hybrid CNN-SIFT Aggregator. CoRR abs/1608.02833 (2016). *arXiv* **2016**, arXiv:1608.02833.
25. Deng, J.; Pang, G.; Zhang, Z.; Pang, Z.; Yang, H.; Yang, G. cGAN based facial expression recognition for human-robot interaction. *IEEE Access* **2019**, *7*, 9848–9859. [[CrossRef](#)]
26. Li, J.; Jin, K.; Zhou, D.; Kubota, N.; Ju, Z. Attention mechanism-based CNN for facial expression recognition. *Neurocomputing* **2020**, *411*, 340–350. [[CrossRef](#)]
27. Li, Q.; Liu, Y.Q.; Peng, Y.Q.; Liu, C.; Shi, J.; Yan, F.; Zhang, Q. Real-time facial emotion recognition using lightweight convolution neural network. *J. Phys. Conf. Ser.* **2021**, *1827*, 12130. [[CrossRef](#)]
28. Mellouk, W.; Handouzi, W. Facial emotion recognition using deep learning: Review and insights. *Procedia Comput. Sci.* **2020**, *175*, 689–694. [[CrossRef](#)]
29. Sadeghi, H.; Raie, A.-A. Human vision inspired feature extraction for facial expression recognition. *Multimed. Tools Appl.* **2019**, *78*, 30335–30353. [[CrossRef](#)]
30. Tsai, H.-H.; Chang, Y.-C. Facial expression recognition using a combination of multiple facial features and support vector machine. *Soft Comput.* **2018**, *22*, 4389–4405. [[CrossRef](#)]
31. Ji, Y.; Hu, Y.; Yang, Y.; Shen, F.; Shen, H.T. Cross-domain facial expression recognition via an intra-category common feature and inter-category distinction feature fusion network. *Neurocomputing* **2019**, *333*, 231–239. [[CrossRef](#)]
32. Zhang, T.; Liu, M.; Yuan, T.; Al-Nabhan, N. Emotion-Aware and Intelligent Internet of Medical Things towards Emotion Recognition during COVID-19 Pandemic. *IEEE Internet Things J.* **2020**. [[CrossRef](#)]

33. Rathour, N.; Gehlot, A.; Singh, R. Spruce-A intelligent surveillance device for monitoring of dustbins using image processing and raspberry Pi. *Int. J. Recent Technol. Eng.* **2019**, *8*, 1570–1574. [[CrossRef](#)]
34. Rathour, N.; Gehlot, A.; Singh, R. A standalone vision device to recognize facial landmarks and smile in real time using Raspberry Pi and sensor. *Int. J. Eng. Adv. Technol.* **2019**, *8*, 4383–4388. [[CrossRef](#)]
35. Rathour, N.; Singh, R.; Gehlot, A. Image and Video Capturing for Proper Hand Sanitation Surveillance in Hospitals Using Euphony—A Raspberry Pi and Arduino-Based Device. In *International Conference on Intelligent Computing and Smart Communication 2019. Algorithms for Intelligent Systems*; Springer: Singapore, 2020; pp. 1475–1486.
36. Haider, F.; Pollak, S.; Albert, P.; Luz, S. Emotion recognition in low-resource settings: An evaluation of automatic feature selection methods. *Comput. Speech Lang.* **2021**, *65*, 101119. [[CrossRef](#)]
37. Su, Y.-S.; Suen, H.-Y.; Hung, K.-E. Predicting behavioral competencies automatically from facial expressions in real-time video-recorded interviews. *J. Real-Time Image Process.* **2021**, 1–11. [[CrossRef](#)]
38. Uddin, M.Z.; Nilsson, E.G. Emotion recognition using speech and neural structured learning to facilitate edge intelligence. *Eng. Appl. Artif. Intell.* **2020**, *94*, 103775. [[CrossRef](#)]
39. Wang, S.; Guo, W. Robust co-clustering via dual local learning and high-order matrix factorization. *Knowl. Based Syst.* **2017**, *138*, 176–187. [[CrossRef](#)]
40. Altameem, T.; Altameem, A. Facial expression recognition using human machine interaction and multi-modal visualization analysis for healthcare applications. *Image Vis. Comput.* **2020**, *103*, 104044. [[CrossRef](#)]
41. Chen, Y.; Ou, R.; Li, Z.; Wu, K. WiFace: Facial Expression Recognition Using Wi-Fi Signals. *IEEE Trans. Mob. Comput.* **2020**. [[CrossRef](#)]
42. Masud, M.; Muhammad, G.; Alhumyani, H.; Alshamrani, S.S.; Cheikhrouhou, O.; Ibrahim, S.; Hossain, M.S. Deep learning-based intelligent face recognition in IoT-cloud environment. *Comput. Commun.* **2020**, *152*, 215–222. [[CrossRef](#)]
43. Medapati, P.K.; Tejo Murthy, P.H.S.; Sridhar, K.P. LAMSTAR: For IoT-based face recognition system to manage the safety factor in smart cities. *Trans. Emerg. Telecommun. Technol.* **2020**, *31*, e3843. [[CrossRef](#)]
44. Viola, P.; Jones, M. 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, 8–14 December 2001; Volume 1, p. I.
45. Arriaga, O.; Valdenegro-Toro, M.; Plöger, P.G. Real-time convolutional neural networks for emotion and gender classification. In Proceedings of the 27th European Symposium on Artificial Neural Networks, ESANN 2019, Computational Intelligence and Machine Learning, Brügge, Belgium, 24–26 April 2019; pp. 221–226.
46. Blood Pressure Sensor—Serial Output. Available online: <https://www.sunrom.com/p/blood-pressure-sensor-serial-output> (accessed on 17 May 2021).