IEEE Open Journal of the Computer Society

Received 3 April 2024; revised 5 September 2024; accepted 15 September 2024. Date of publication 18 September 2024; date of current version 17 October 2024. The review of this article was arranged by Associate Editor Lin Gu.

Digital Object Identifier 10.1109/OJCS.2024.3463173

Artificial Intelligence of Behavior for Human Emotion Recognition in Closed Environments

GONZALO-ALBERTO ALVAREZ-GARCIA [1] (Member, IEEE), CLAUDIA ZÚÑIGA-CAÑÓN [1] (Senior Member, IEEE), ANTONIO-JAVIER GARCIA-SANCHEZ [2], JOAN GARCIA-HARO [2] (Member, IEEE), MILTON SARRIA-PAJA [3] (Senior Member, IEEE), AND RAFAEL ASOREY-CACHEDA [2] (Member, IEEE)

[1]Research Group COMBA I+D, Universidad Santiago de Cali, Cali 760035, Colombia
[2]Department of Information and Communications Technologies, ETSIT, Universidad Politécnica de Cartagena (UPCT), 30202 Cartagena, Spain
[3]Research Group GIEIAM, Universidad Santiago de Cali, Cali 760035, Colombia

CORRESPONDING AUTHOR: RAFAEL ASOREY-CACHEDA (e-mail: rafael.asorey@upct.es).

This work was supported in part by MICIU/AEI/10.13039/501100011033, in part by European Union NextGenerationEU/PRTR under Grant PID2023-148214OB-C21 and Grant TED2021-129336B-I00, in part by Fundación Séneca under Grant 22236/PDC/23, in part by the Ministerio de Ciencia, Innovación y Universidades, in part by European Union NextGenerationEU/PRTR-C17.I1, in part by Fundación Séneca with funding from Comunidad Autónoma Región de Murcia (CARM) through ThinkInAzul and AgroAINext Programmes, in part by DAIMon, a Cascade funding action deriving from the Horizon Europe Project aerOS, through European Commission under Grant 101069732, in part by the Asociación Universitaria Iberoamericana de Postgrado (AIUP), and in part by Dirección General de Investigaciones of Universidad Santiago de Cali through Project "Crowdsourcing Optimized Wireless Sensor Network Deployment (CRoWD)" under Grant 613-621119-852.

ABSTRACT Understanding human emotions and behavior in closed environments is essential for creating more empathetic and humane spaces. Environmental factors, such as temperature, noise, and light, play a crucial role in influencing behavior, but individuals' emotional states are equally important and often go unnoticed. Artificial Intelligence of Behavior (AIoB) offers a novel approach that integrates environmental measurements with human emotions to create spatially adaptive processes that can influence behavior. In this article, we present a new human emotion sensor developed using video cameras and implemented on a System on Chip (SoC) development board. Our approach uses Convolutional Neural Networks (CNNs) to recognize the presence of emotions in enclosed spaces and generate parameters that can influence emotional states and behavior within an AIoB system. The research successfully integrates advanced CNN technology into a System on Chip (SoC) platform, allowing for real-time processing of video data. The versatility of utilizing an energy-efficient SoC extends its application to smart environments aimed at improving mental health. By employing algorithms capable of detecting emotional states across various individuals, the study enhances its effectiveness. Additionally, it identifies the best CNN operations tailored to the technical specifications of the devices involved. Thus, The development involves a three-step process: (i) collecting enough data to build a robust model, (ii) training the model and evaluating its performance using test values, and (iii) applying the model on the development board. Our study demonstrates the feasibility of using AIoB to recognize and respond to human emotions in closed areas. By integrating emotional cues with environmental measurements, our system can create more personalized and empathetic spaces that cater to the needs of individuals. Our approach could have significant implications for designing public spaces to promote well-being and emotional satisfaction.

INDEX TERMS Convolutional neural networks, compressive sensing, emotions, image classification, instrumentation, system on chip.

I. INTRODUCTION

Recognizing human emotions has become increasingly important in recent years, as it has the potential to enhance user experience and improve efficiency in various domains, such as advertising, psychology, and human resource management [1], [2]. However, identifying human emotions is a major technological challenge since it involves changing conditions and subjective parameters [3], [4], [5]. Further complexity

results from developing a system that can establish objective parameters in non-intrusive and non-invasive systems [5], [6].

Algorithms based on neural networks have emerged as a promising approach to identifying human emotions given their ability to learn patterns and relationships in large datasets. Developing these algorithms and using real-time images involve high-performance computing systems. Artificial Intelligence of Behavior (AIoB) is a novel term that integrates Artificial Intelligence (AI) and Internet of Things (IoT) technologies to monitor human spaces and improve people's comfort or induce desired responses like activity or tranquility [7]. Environmental factors, such as temperature, noise, and light, may significantly affect human behavior. Therefore, acting on these parameters could create a general response among a group of people. Our proposal consists of an implementation intended to operate on top of an SoC with limited capabilities. An SoC is a complete computer system, but it is not reconfigurable [6]. This allows a user interface to visualize and intuitively configure the system without the need for advanced knowledge [8].

Controlling personal environments by, for example, opening or closing a window, may seem like simple actions, but they involve making complex decisions based on a range of parameters, like noise, temperature, and air quality [9], [10]. In closed social environments, such as airports, hospitals, and government buildings, individuals are subject to rules and restrictions that can affect their moods and well-being [11]. For instance, long waiting times, crowded spaces, and loud noises may all contribute to increased stress levels [12].

Managing these complex environments is a challenge since the factors influencing human behavior are multifaceted and interdependent. For example, increasing air circulation may improve air quality, but this could also lead to more noise from outside, which may negatively impact people's concentration and mood. Similarly, reducing noise levels may improve concentration but could also increase the perceived temperature, which can affect comfort levels [4].

To provoke the desired response in a given environment, it is crucial to understand the emotional state of the individuals in that environment. However, traditional methods of measuring emotions can be intrusive and invade privacy [13]. To address these challenges, this article presents a novel AIoB module that identifies emotions in a way that is non-invasive and preserves people's privacy. Our proposed module is based on image recognition techniques that analyze individuals' facial expressions to identify their predominant emotions [14].

Understanding the emotional state of the individuals involved is crucial for achieving the desired response in a particular setting. Nevertheless, conventional approaches to assessing emotions are often invasive. To tackle these obstacles, this article introduces an innovative AIoB module designed to identify emotions in a way that is both non-invasive and respects people's privacy. Our proposed module uses cutting-edge image recognition techniques to analyze facial expressions, thereby enabling the prevailing emotion to be identified [14].

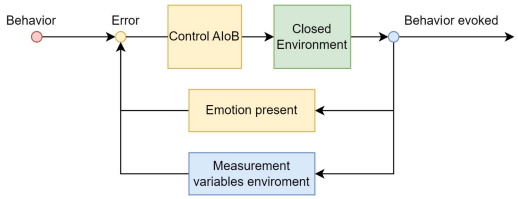

FIGURE 1. Closed-loop representation of an AIoB System.

By leveraging real-time data from cameras strategically placed in enclosed spaces, our module can accurately detect a wide range of emotions, including but not limited to Happiness, Sadness, Anger, and Fear. This information can then be used to dynamically adapt environmental factors like temperature, lighting, and noise, thereby fostering a more personalized and comfortable experience.

The potential applications of our proposed AIoB module extend across fields including advertising, healthcare, and public services. For instance, in a hospital setting, this module could prove instrumental in identifying patients' emotional states and subsequently tailoring their environment to promote healing and provide optimum comfort. In a retail store, the module could play a pivotal role in optimizing product display arrangements and store layouts to enhance customers' overall experience and increase sales.

By facilitating a more personalized and empathetic environment, our AIoB module has the potential to enhance people's well-being, productivity, and overall quality of life. An immersive experience is fostered by assessing individuals' emotional states within a given environment and integrating them with corresponding environmental parameters [15], positively influencing individual and collective behavior.

This approach allows us to pinpoint specific environmental factors that can be modified to optimize well-being. Consequently, the AIoB system can be conceptualized as a closed-loop control system, as illustrated in Fig. 1, where an emotional or behavioral state is the variable to be regulated.

In this context, an emotional or behavioral state is understood as a constellation of sensations experienced by an individual, as outlined by Ajzen's model [16]. Therefore, it is evident that these states are malleable and can be influenced by controllable environmental factors.

The present emotion module plays a crucial role in the AIoB system as it identifies people's emotions which, in turn, allows behavior parameters to be modified in response to environmental conditions.

This is illustrated in Fig. 1, which depicts the closed-loop control system underpinning the AIoB system. By analyzing facial expressions through image recognition techniques, the system can accurately determine people's predominant emotions in a given environment. This non-invasive and privacy-sensitive approach ensures that individuals' emotional states can be evaluated discreetly and respectfully [17].

Recognizing the emotions of a population group is a technological challenge requiring advanced AI techniques and the analysis of vast quantities of data [16], [18], [19].

To this end, the widely accepted Azjen behavioral model is employed as a reference framework. This model identifies three fundamental parameters: attitude toward behavior, subjective norm, and perceived behavioral control, which facilitate the understanding and modification of human behavior.

The design of the emotion measurement module is based on a dynamic system analysis, taking into account various factors, including the fluctuation of environmental conditions and comfort levels in a closed space [12].

The elements in spaces where people connect with each other, such as microclimatic parameters, play a major role in determining individuals' emotional states [20]. However, people are not equally sensitive to physical factors like light, temperature, and noise, and these may influence emotions differently [21]. For instance, the amount of light could favor concentration or calmness, while the tone, intensity, or color of artificial light may elicit individual or collective attitudes [22]. Similarly, noise and the effect of the primary frequencies of the audible spectrum could affect people's comfort levels [23]. Controlled sounds can significantly impact the perception of space, thereby affecting the way people feel and behave in that environment [24].

This article presents a novel implementation of a critical module within the AIoB architecture that accurately determines the emotional states present in a room. Therefore, the objective of this research is to develop an integral system capable of analyzing human emotions through video, delivering a parameterized value locally and in the cloud [25].

Given the limitations implied by using an SoC, the two-dimensional structure of the images discretely acquired, and the time invariance of the generated information, the proposed approach employs a CNN training methodology to develop a robust model, which is then implemented on an SoC platform for efficient processing.

A small device that is barely noticeable and is non-invasive without sacrificing accuracy has been developed. This device can easily be implemented as an integrated sensor [26]. The resulting system provides real-time emotion detection capabilities and represents a substantial advance in the field of artificial intelligence and environmental sensing.

Among the possible architectures, CNN, MobileNet, Inception v3, RestNet-50, and VGG19 were analyzed [14], [27], [28]. The training process was carried out using 4 open databases of Face Emotion Recognition (FER). They are cataloged as: 2 casual or spontaneous (FER 2013, AM-FED), 1 acted (RaFD), and 1 synthetic (FERG) [1].

AIoB technology prioritizes non-invasive methods and protects privacy as a fundamental feature. To achieve this, the system uses video cameras to detect emotional states by processing information within the system itself, that is, at the edge [29]. Only the final results, such as categories, are transmitted, ensuring the anonymity of the people in the room. This approach also offers a significant advantage in terms of resource consumption as there is no need to transmit video data and all the processing is distributed. This more efficiently uses computational resources and generates a categorical database without affecting individuals privacy [8], [30].

The proposed system identifies 7 emotions with an accuracy of between 80% and 53%. It identifies and recognizes the emotions of 100 people at a speed of 2 fps and attains the best performance in design and implementation for the use of VGG19.

To summarize, the main contributions of this paper are:

1) Introduction of a new implementation of a critical module within the AIoB architecture.
2) Development of a system capable of accurately determining emotional states in a room.
3) Creation of an integral system capable of analyzing human emotions through video.
4) Delivery of parameterized values locally and in the cloud for enhanced accessibility and usability.

The rest of the article is structured as follows: In Section II, we provide a comprehensive background on AIoB technology, image analysis, and emotion classification. In Section III, we describe the sensor architecture of the AIoB system. Section IV discusses the results of tests conducted in different environments. Finally, in Section V, we conclude by summarizing the contributions of the article.

II. RELATED WORK

Identifying emotions in individuals or groups is a complex technical challenge that has been extensively researched. Several theories have been proposed based on criteria such as microexpressions or the correlation of specific facial points and their distances to generic emotions [14], [19]. However, accurately identifying emotions manually requires an expert to compare the design of facial triangulation with sample tables indicating the emotion being expressed [25], [31].

To automate this task, a deterministic function based on image analysis [32] was developed. However, the results of this analysis may change depending on several variables, such as the subject's biotype [29], age, sex [19] and the technical location of the face in the image (for example, the results varied if the face was slightly tilted) [33]. Other studies have concluded that human facial expression provides a visual way to understand underlying emotions and is widely used in applications such as robotics and security. The use of binary synaptic neural networks (BSN) is an alternative to the analysis of the phenomenon. In addition, based on neuromorphic computing (NC) to recognize human facial emotions (FER). This implementation using hardware description with Verilog enabled 67.5% accuracy under controlled light and data collection conditions.

The use of CNN is an effective technique to address the problem of emotion recognition. Unlike traditional methods that require frontal images of the face, neural networks can establish features based on a collection of training photos that include three-dimensional rotations, allowing emotions to be recognized without faces needing to be in a predetermined position [34], [35].

This approach leverages stochastic processes to produce results within a range of probabilities [13], [24], [36], which can lead to false positives or false negatives [37].

AI-based systems can effectively detect the presence of individuals in an image through facial segmentation techniques [19], [31], [35]. These systems can also count the number of people in the captured image [38].

One of the benefits of these systems is their ability to adapt to varying environmental conditions [39], [40], [41]. However, the use of filters or comparison matrices has a significant impact on the performance of the [42] system.

To achieve effective training, CNNs require a substantial volume of data [27]. Considering the potentially thousands or even millions of features involved in each model, understanding the purpose and representation of each parameter becomes a challenging task [43].

In addition, the intricate nature of the neural network architecture, with its multitude of paths and weights, presents difficulties in manually optimizing the numerous interlocking weights that correspond to the model parameters by altering the values of the [10] features.

Some factors that can affect emotion identification systems using images are lighting conditions, camera white balance, and people's ethnic characteristics. One of the proposals is the one made by Cruz Albarran [4]. It makes use of thermographic images for the identification of 5 kinds of emotions. In this study they reached an accuracy of 89.9% this analysis is done in a particular way and in a suitable place for that purpose. Another way to approach the research problem is a non-contact emotion recognition method based on complexion and heart rate (HR). The method they develop Du et al. It uses complexion and heart rate as indicators of non-contact emotion recognition [5]. This uses a common device a Kinect camera as data input. The method combines CNN and bidirectional random fields of short-term and conditional memory (Bi-LSTM-CRF) to extract complexion and heart rate characteristics for contactless emotion recognition [43]. They analyze 4 categories and obtain an accuracy of recognition of emotions greater than 80% in the detection of anger, depression, doubt and indignation [5].

III. DEVELOPMENT

The main part of the system is an emotion sensor that uses artificial intelligence to detect and understand human emotions. This sensor includes several important elements, such as parts that collect data, units that process signals, and machine learning algorithms that analyze the data to accurately identify emotional expressions.

The first step in the development process is the design of the hardware, where the necessary physical parts, like sensors, processors, and communication devices, are chosen and arranged to ensure they can handle data reliably. After this, a suitable CNN model is selected because of its ability to find patterns in the data that are linked to different emotions, which is essential for achieving accurate and efficient results. The final step is a proof-of-concept, in which the entire system is tested in real-world situations to check its performance and confirm that the chosen model works correctly.

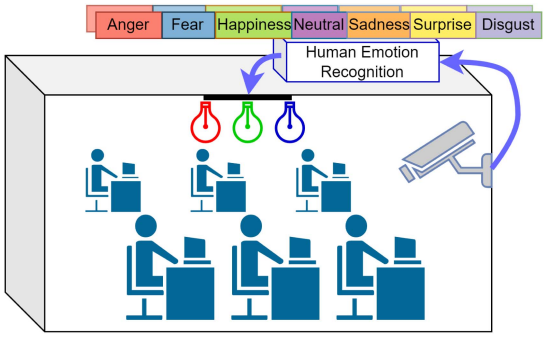

FIGURE 2. Components of system for detected and analyzed emotion in images of video by and categories of classification.

The system that is developed involves lines of components through the proposed methodology. The emotion sensor using artificial intelligence involves several critical elements.

A. HARDWARE DESIGN

An emotion recognition system using images that uses artificial intelligence requires two types of systems. First, a computer system with the conditions to perform the training and validation of the neural network model. Suitable hardware that can process large amounts of data and run complex algorithms.

Then the system where it will be implemented and compatibility with the image capture sensor is required. In this order, some of the required components are:

- A powerful processor, high-performance graphics card, high-speed RAM, and a high-capacity hard drive are required for CNN's training system to store large sets of image data and process them efficiently [14].
- A high-quality camera capable of capturing high-resolution, high-quality images is necessary for accurate recognition of emotions. This camera must have automatic white balance and chroma adjustment systems [27].
- A SOC that supports image processing and deep learning software, such as TensorFlow, Keras, or PyTorch, to run deep learning models using images compatible with external capture devices as input. In addition, the system must have the ability to process and transmit the characteristics using time flags and the statistical identification of the predominant emotion [17].

It is also necessary an emotion monitoring system, in an emotion database with a set of algorithms capable of identifying patterns in images and determining the emotions that are shown. Creating the information of the emotions present and their statistical weight.

From a hardware perspective, a computer, a webcam, and a minicomputer are the basic elements needed for system implementation (as shown in Fig. 2). Processing the information

locally and transmitting the results to a database is possible with this configuration.

These emotions are connected to the images produced during the facial response and have characteristics related to the expressions or microexpressions (facial expressions of short duration) of the face [44].

To train the neural network, we use 4 open datasets that contain among their samples the necessary ones for the 7 emotions. The databases used are: FER 2013, AM-FED, RaFD and FERG. Together they collect a set of 123,795 images labeled in one of the 7 emotions: anger 13,7%, fear 12,6%, happiness 23,5%, neutral 15,8%, sadness 14,8%, surprise 10,4%, disgust 9,2%. For the realization of the CNN model, 80% of the images were used for training and the remaining 20% for validation. Despite the use of a large amount of data, the distribution of training photos for each emotion is not uniform. After completing the training, we evaluate the performance of the system by performing new video captures using open images.

B. NEURAL NETWORK MODEL SELECTION

CNNs are highly effective at processing and analyzing images due to their specialized architecture, which allows them to detect specific patterns and features in images. In the field of Facial Emotion Recognition (FER), a novel technique gaining prominence involves the use of Vision Transformers (ViT). Specifically, two ViT models, ViT token-to-token and ViT Mobile, have shown promising findings [45]. However, successful utilization of these models hinges on balanced training data. In this sense, research results indicate that when dealing with imbalanced databases, ViT Mobile achieved a maximum precision of approximately 63%, whereas with balanced data, the precision improved to 75%. Notably, model validation occurred on high-performance computing systems, which presents challenges for implementation in SoC architectures.

This study sheds light on the trade-offs between model performance and data balance, emphasizing the need for careful dataset curation when deploying CNN-based FER systems. Thus, several CNN models have been analyzed to identify features in the images, including:

- ResNet-50: A deep neural network model that uses residual layers to improve the accuracy of image classification. ResNet-50 has a 50-layer architecture that provides better feature detection [46].
- Inception v3: A CNN model developed by Google that employs a multi-path architecture to improve the accuracy of image classification. Inception v3 uses convolution filters of various sizes to identify features in images [47].
- MobileNet: A deep neural network model designed for mobile and low-power applications. MobileNet uses a simpler network architecture that reduces model size and improves processing speed without compromising accuracy [48].
- VGG19: A highly efficient deep convolutional neural network for image classification, popular due to its deep architecture, small convolution filters, and training with large datasets [49].

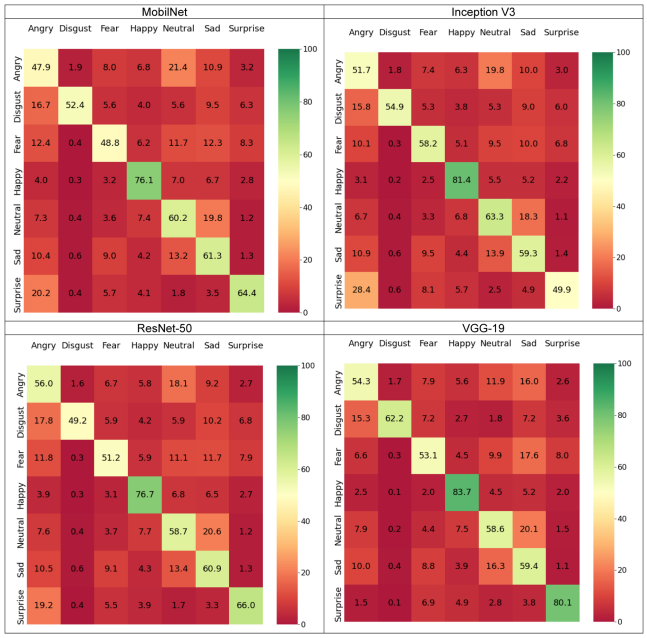

FIGURE 3. Convolution matrix for Face Emotion Recognition in ResNet, Inception, Mobilenet and VGG19 models.

C. PROOF OF CONCEPT

After training the network and its implementation in the system, a proof of concept is necessary. This test aims to verify the conditions of the system and determine its application in the context of the problem analyzed. The system is assembled and the variance parameters are established using the European technical standard for ambient lighting. A group of volunteers performs the proof of concept using as a stimulus particular situations that can denote any of the 7 categories of analysis. In the test, volunteers respond spontaneously under a premise of a suggested emotion.

IV. RESULTS

1) SELECT CNN

The training of all four models was carried out using the same datasets. In the case of ResNet-50 and VGG19, pre-trained models were loaded and fine-tuned for the specific tasks or domains of interest. This approach leverages the pre-trained weights and learned features from these models on a large-scale dataset (often ImageNet) to provide a strong starting point for the target tasks, saving computational resources and time.

Fine-tuning allows these models to adapt their knowledge to the specific dataset and problem at hand, making them effective choices for various computer vision tasks.

The results of the confusion matrices for the four training processes are shown in Fig. 3. Despite the similarities observed in the training results of the models, it is important

TABLE 1. Summary of Test Process Results in the Implementation CNN Model

CNN	ResNet-50	Inception v3	MobileNet	VGG19
Accuracy	76.7 %	81.4 %	76.1 %	83.7 %
PyBench	965 ms	96 ms	28 ms	175 ms
Size	79 MB	45 Mb	102 MBk	457 MB
Resolution	4k	2k	1k	4k
Speed	1 fps	10 fps	30 fps	5 fps
Pre-training	Y	N	N	Y

to note that superior performance has been achieved in two of them. In particular, the Inception model achieved an 81.4% accuracy, while VGG19 with an 83.7%.

These results highlight the effectiveness of Inception and VGG19in the specific tasks for which they were trained and underscore the importance of carefully selecting the convolutional neural network architecture for each computer vision problem.

Furthermore, VGG19 has demonstrated impressive performance in image classification, object detection, and semantic segmentation tasks, making it a robust choice for a variety of computer vision and deep learning applications.

2) DEVICE DEVELOPMENT

To develop the system, three main tasks were carried out: obtaining the neural network model, generating the model, and implementing the system. The results In the first task, four preset models were used to train the neural networks. The main results are summarized in Table 1.

The CNNs selected for the training process were chosen for their ability to analyze and process images, identify patterns, and extract features.

ResNet-50, in particular, can extract up to 1000 features from images, and its deep architecture is helpful for classification tasks. However, ResNet-50 is computationally expensive to implement. In the PyBench tests the result was 965ms per image evaluation cycle with detection of 100 identified people. The main reason for the execution time is the lack of FPU on the Raspberry Pi. This slows down the estimation of the residuals between layers and makes the time between cycles longer than in other CNN models. Its advanced layers are designed to detect more complex features, which are less extended across the image than the features extracted by earlier layers.

The developed system can be used on any device that supports Python code processing, machine learning libraries, and data manipulation. Raspberry Pi 3B+ was chosen due to its capability for processing these algorithms. Additionally, being a pocket-sized computer allows for low energy consumption and its compact size makes it a non-invasive element.

It is worth noting that Inception v3 and MobileNet share similar structures but differ in the number of intermediate and output layers, resulting in a reduced number of classification

Algorithm 1: Convolution Neural Network.

CNN based on VGG 19 load
Initialization :
1: Epochs for training
LOOP 50 times
2: **for** Layers in model **do**
3: training
4: **if** (model is better) **then**
5: update model
6: **end if**
7: **end for**
8: **return**Last model
Export Model type. h5 save

features. However, to analyze images across seven distinct groups, these features are sufficient. Pushing the resolution beyond a certain threshold may impose limitations on network capacity and limit the efficiency of the feature extraction and retrieval processes.

VGG19 (as depicted in Table 1) has the highest precision among CNNs, has the essential capabilities required for the current problem. Is efficient classifies the seven characteristics and effectively filters the environment. Ambient noise, such as variations in lighting and facial noises positioning.

The foundation of the CNN model employed in this study traces back to the influential VGG16 model [28], which established its prominence in the realm of computer vision. As computational capabilities advanced, the model's architecture was expanded, leading to the development of the VGG19 model.

It is worth noting that augmenting the number of layers at this stage did not yield a substantial improvement in model accuracy. However, it did result in an increased number of features, subsequently lengthening processing time and resource consumption [14].

The training process involved categorizing images into seven different emotions. The final model for implementation in the system was trained using Algorithm 1, which includes image resizing to 48×48 pixels from the database. The picture size can be adjusted automatically based on the needs of the system.

The resulting neural network achieved a training mean accuracy of 64.6% and a validation mean accuracy of 63.2%, which can be considered satisfactory given the classification of seven distinct classes.

The highest error probability, accounting for false positives and detection errors across all classes except the identified one, stands at 5%. Detailed accuracy and recall results are presented in Table 2.

Notably, the emotion "Happiness" shows the highest accuracy, approximately 85%, while the accuracy for "Sadness" is lower, at around 47%. Furthermore, the emotion "Fear" has the lowest recall value, indicating more likelihood of false positives than true positives for this particular emotion.

TABLE 2. Training and Validation Test Model CNN VGG19 Summary

Emotion	Precision	Recall	F1-score	Support
Angry	54.3%	54.2%	56.5%	3392
Disgust	62.2%	60.1%	60.9%	2278
Fear	53.1%	48.7%	52.3%	3120
Happy	83.7%	84.2%	816%	5818
Neutral	58.6%	59.1%	54.9%	3912
Sad	59.4%	57.8%	52.3%	3664
Surprise	80.1%	80.5%	79.1%	2575

Although the recall value for "Fear" is less than 50%, the F1-Score is higher than 49%, indicating that the identification of this emotion is adequate. For the remaining emotions, the values are above 50%, making the model statistically acceptable.

The training phase is crucial and entails data processing to ensure the completeness, relevance, appropriate size, and correct labeling of the data.

The model construction summary includes information on the total number of extracted features to be included in the system, which in our case is 2,346,183 after building the CNN and assembling the model configuration. The model is then exported in h5 format and validated before being used on a Raspberry Pi 3 B+ mini-computer. The main features in the system of support are: Broadcom BCM2837B0 Cortex-A53 64-bit SoC @ 1.4GHz processor, RAM: 1GB LPDDR2 SDRAM, Wireless connections Wi-Fi Dual Band b/g/n/ac, H.264 1080p30 decoding and encoding, Supports PXE. The system consumes 260mA (1.4W) without a graphical interface and reaches 350mA (1.9W) with the graphical interface. Consumption is measured when the system is connected and transmitting data, and it decreases by 10mA when the system is offline.

Implementing emotion detection hardware on a Raspberry Pi 3 can produce different results based on the deep learning technique used and the quality of the input data. However, in general, the Raspberry Pi 3 can process and analyze input data in real-time and produce an output based on emotion detection.

Initially, we used the Raspberry Pi camera with 1080p resolution and 30 fps speed to develop the proposal. The Algorithm 2 runs smoothly on the Raspberry Pi, but the camera does not provide sufficient information to obtain the minimum parameters required for face identification in large spaces. To improve the capture quality, a camera with native 4k resolution was chosen. The characteristics of sensor by video capture are: 4K/60 fps (up to 4096 x 2160pixels), Camera mega pixel: 13, Focus type: Autofocus, Diagonal field of view (dFoV): 90°/78°/65°, Digital zoom: 5x, RightLight 3 with HDR, Sensor with infrared technology.

The implementation of emotion detection hardware on a Raspberry Pi 3 aims to adapt to different scenarios and environments. This can be achieved by using a 4K camera that obtains accurate features and ensures adequate processing.

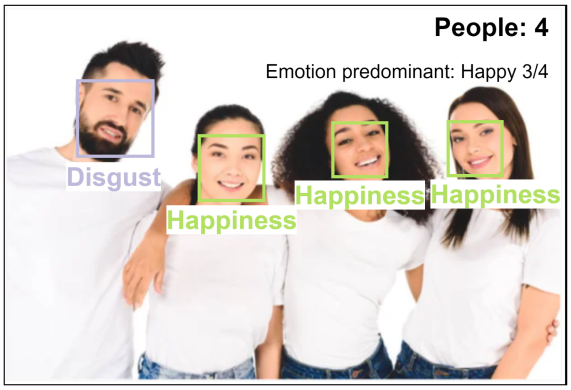

FIGURE 4. Results of system testing showing the predominant emotion detected as "Happiness".

Algorithm 2: Emotion Identification.

Image Capture from videoload
Initialization :
if (Identification face) **then**
  Save face
  Inc Count
end if
Emotion identifier
emotion=[[Anger],[Fear],[Happiness],[Neutral], +[Sadness],[Surprise],[Disgust]]
for face count **do**
  for i **do**
    if Emotion identifier==emotion[i] **then**
      $emotion_count[i] \leftarrow emotion_count[i] + 1$
      break
    end if
  end for
end for
$emotionPredominat \leftarrow Max(emotion_count)$
Show (face count)
Show (emotionPredominant/face count)
return Last model
New Capture save

With a speed of 5 fps, images can be processed with robust enough information to handle variations in lighting conditions, camera position, and changes in peolple's appearance, such as the use of accessories.

Using the trained neural network, the people in the room can be identified and the emotion of each person can be detected, as shown in Algorithm 2.

Fig. 4 shows an example of the system's behavior, where four people are in the image, and the identified predominant emotion is "Happiness." The selection criterion for the predominant emotion is based on the most frequently displayed emotions.

Similarly, in the test depicted in Fig. 5, five individuals are detected, and the system recognizes "Disgust" as the

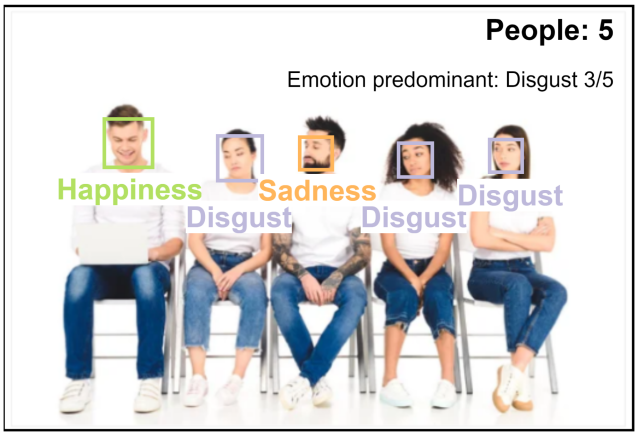

FIGURE 5. Results of system testing showing the predominant emotion detected as "Disgust".

FIGURE 6. Emotion detection test in a waiting area, with the predominant emotion detected as "Neutral".

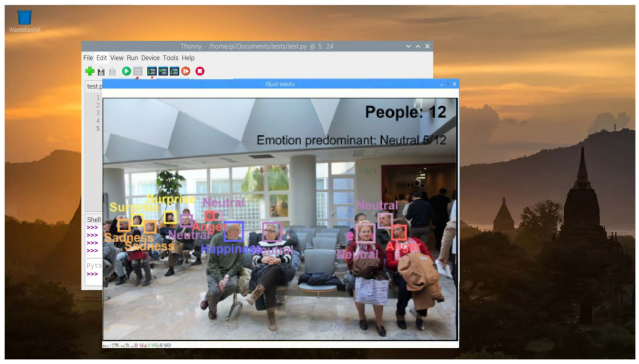

FIGURE 7. Graphical user interface (GUI) of Raspbian for emotion detection development and implementation.

TABLE 3. Ethnographic and Age Characteristics of the Proof-of-Concept Group

Age	Participants	Female	Male	Caucasian	Indigenous	Mestizo	Afrodescendants
18	3	1	2	2	0	1	0
19	3	2	1	0	2	1	0
20	4	1	3	2	0	2	0
21	3	0	3	3	0	0	0
22	2	0	2	1	0	0	1
23	2	1	1	0	1	0	1
24	0	0	0	0	0	0	0
25	1	1	0	1	0	0	0
26	1	0	1	0	0	0	1
27	1	0	1	1	0	0	0
Total	20	6	14	10	3	4	3

predominant emotion, while also detecting "Happiness" and "Sadness."

In the last example, represented in Fig. 6, the system was unable to identify some people as their full face was not visible, resulting in the detection of only 12 people. The predominant emotion detected was "Neutral," although the system also identified emotions such as "Sadness," "Happiness," and "Anger" in some individuals. This highlights the limitation of the system in detecting emotions when facial features are not fully visible, such as, in this case, in a waiting room scenario. One of the advantages of implementing emotion detection on a Raspberry Pi 3 is its power efficiency, which allows for longer battery life and increased portability.

In addition, the implementation is expected to be able to handle a high processing load without significant delays in response time. The implementation of emotion detection hardware on a Raspberry Pi 3 should be able to handle a higher processing load without excesive delays in response time, providing users with a seamless experience.

The emotion detection sensor is implemented on a Raspberry Pi Linux distribution running the user-friendly Raspbian graphical desktop environment. This interface enables seamless monitoring of system operations both online and offline, enhancing user convenience, as illustrated in Fig. 7. However, it is important to note that offline monitoring is not a necessity for the system, as data transmission to the cloud allows for online querying and analysis.

Furthermore, the identified emotions and associated data are transmitted to a cloud where they can be accessed, without transmitting images or video signals. This offers an added layer of privacy and security to the system.

3) EXPERIMENTAL TESTS

The emotion sensor proof of concept using images was conducted on a group of 20 people. The objective of the test was to evaluate the ability of the sensor to detect different participants' emotions.

The test group used in the emotion sensor proof concept consisted of 20 people (see Table 3). All participants were between the ages of 18 and 27.

The selection of this age group was because young people may be more receptive to emotions and, therefore, might show

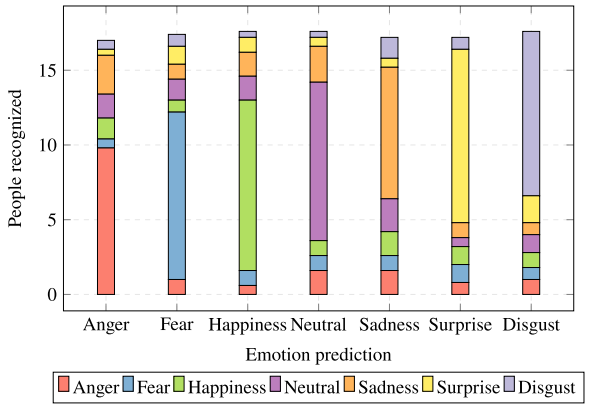

FIGURE 8. Proof of concept test results for each analysis category.

FIGURE 9. Indoor environment conditions and proof-of-concept participants.

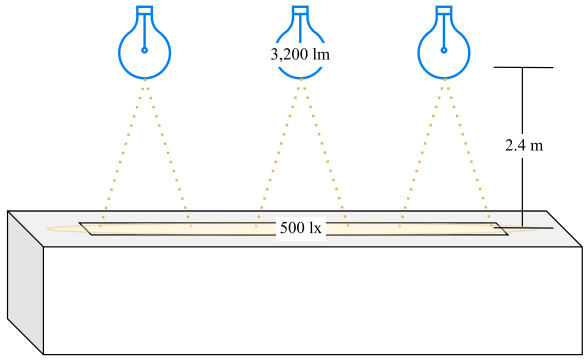

FIGURE 10. Design of maximum radiation power and luminous flux over the average distance from the target.

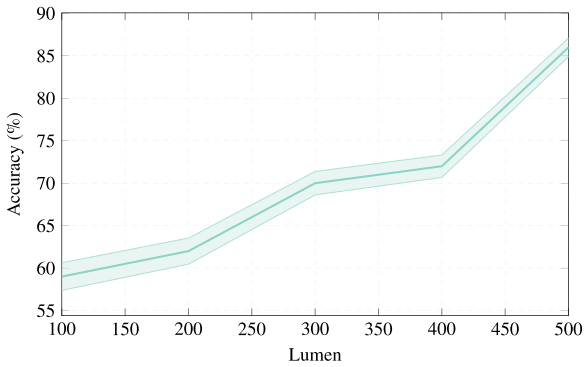

FIGURE 11. Sensor accuracy depending on lighting conditions.

greater variability in their emotional expressions. This age range was also chosen because participants were expected to have a variety of cultural and ethnic backgrounds.

Fig. 8 shows the results of the test, demonstrating the sensor's ability to detect the predominant emotion in a group setting. The test was conducted with 20 participants, and the system was able to accurately identify an average of 17 to 18 individuals in the room.

a) Individual tests: Each participant completed an individual and group testing session (see Fig. 9), during which they were presented with 35 images randomly selected from a database of emotional images.

The images featured faces with expressions of anger, disgust, fear, happiness, neutrality, sadness, and surprise. Participants were instructed to replicate the facial expression depicted in each image, providing the sensor with a reference for detecting the emotion.

It should be noted that participants had no prior knowledge of the emotions being tested, to avoid bias.

The test involved modifying only the lighting conditions in the test environment, while keeping the temperature and humidity constant. Initially, the level of light was set at 20% of the total available light intensity. The minimum 100 lx (20%) and maximum values 500 lx (100%) of the measurement were established using the European technical standard for lighting public spaces UNE 12464.1 as a parameter. This is used so that indoor lighting values meet visual needs and guarantee comfort conditions.

Subsequently, the intensity was gradually increased by 20% until reaching a maximum of 500 lx. The aim of this test was to assess the sensor's performance across diverse lighting conditions, simulating real-world scenarios where lighting can vary significantly (Fig. 10).

The test results were carefully analyzed and used to optimize the sensor's performance in different lighting environments. For future investigations, it may be worthwhile to explore independent modifications of environmental conditions (such as temperature and/or humidity) as well as lighting variations, or even consider their combined effects.

b) Group tests: Consistency is essential for testing the emotion sensor's performance. Therefore, during each session, the distance between the sensor and the participants, as well as the position of the presented images, was kept constant. To evaluate the sensor's ability to detect and classify emotions under varying lighting conditions, the lighting was modified in each session, as depicted in Fig. 11.

The test results revealed that the emotion sensor successfully detected and classified emotions in all the lighting conditions set, including the lowest condition of 20%. Nonetheless, the accuracy of the sensor appeared to improve gradually with more intense light, ultimately reaching its highest accuracy level at 3,200 lm.

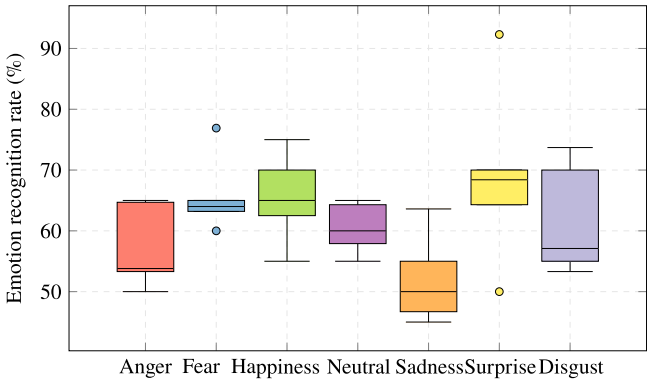

FIGURE 12. Average emotion recognition rate based on experimental data.

c) Variation in illumination tests: The emotion sensor employed sophisticated image processing techniques to analyze the facial features of individuals in the images, such as the shape of their mouth, eyes, and eyebrows. Based on these features, the sensor was capable of evaluating the intensity of the emotion in each image and assigning it to one of seven pre-established emotional categories.

As seen in Fig. 12, the test results indicate that the emotion sensor achieved a maximum accuracy rate of 74% for the measurement of the emotion Happiness. The most accurate data was recorded for the detection of the expression of Surprise, with 92%, although it is an atypical value. The least accuracy was observed for the detection of Sadness, with 42%. This last data belongs to the lowest value, without being an outlier.

The results of the systems developed and the proof-of-concept tests show that the hypotheses concerning the system and the different models obtained comply with the parameters and establish references similar to the modeling requirements.

V. CONCLUSION

Emotion sensing is a rapidly growing field with numerous practical applications in various domains, such as security, entertainment, health, and customer service. Raspberry Pi, an affordable and versatile hardware platform, can be used to implement emotion detection systems in a privacy-preserving and non-intrusive way. The human emotion recognition module is a crucial component of the AIoB system that enables emotions in a room to be identified.

Moreover, implementing emotion detection hardware on a Raspberry Pi 3 can generate reliable and accurate real-time results while being energy-efficient, adaptable to various environments, easily customizable, and scalable. By improving the performance of computer systems, it is possible to solve new problems and run AI-based applications on devices with limited computing resources.

The raspberry pi 3 B+ mini pc does not have FPU (Floating Process Unity), this makes the processing time of the residuals in the ResNet model slower layer by layer without having deep capabilities. The VGG19 model despite being more complex and heavier. Its millions of feature segmentations without loading residuals between capable offer a more efficient step with the hardware's own restrictions, being more efficient and with better precision.

Although VGG19 may have worse performance than other models, its use can result in a feasible emotion sensing module that can be adapted to the implemented computational system with an appropriate speed for the given problem. Additionally, the Raspbian graphical environment has numerous benefits, such as a user-friendly interface, useful development tools, customization and configuration, and a large community of users and online resources.

Using emotion detection hardware on a Raspberry Pi 3 offers a cost-effective solution for both individuals and businesses. Enterprises can leverage emotion sensing to enhance customer experience and optimize business decision-making, while individuals can develop applications for entertainment and social space control.

AI and IoT are the essential tools required for monitoring and modeling social spaces. The AIoB system builds on these technologies to modify social space conditions and achieve a desired social state response. By controlling ambient sound, lighting, and temperature, among other aspects, and determining the dominant emotion, it is possible to intelligently act on the environment to achieve and maintain the desired social state.

The results of our study indicate that the emotion sensor has promising potential in detecting and classifying emotions through imaging. However, further testing and adjustments to the technology are needed to improve the accuracy in detecting certain emotions, such as sadness.

Our findings also suggest that the emotion sensor is capable of performing in different lighting conditions, but the accuracy of its emotion detection may be affected by light intensity. Therefore, further optimization of the emotion sensor technology may be required to improve its ability to detect and classify emotions in low or unfavorable lighting conditions.

REFERENCES

[1] M. Karnati, A. Seal, D. Bhattacharjee, A. Yazidi, and O. Krejcar, "Understanding deep learning techniques for recognition of human emotions using facial expressions: A comprehensive survey," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 5006631.

[2] A. Kołakowska, W. Szwoch, and M. Szwoch, "A review of emotion recognition methods based on data acquired via smartphone sensors," *Sensors*, vol. 20, no. 21, 2020, Art. no. 6367.

[3] H. Xiao and Z. Hu, "Feature-similarity network via soft-label training for infrared facial emotional classification in human-robot interaction," *Infrared Phys. Technol.*, vol. 117, 2021, Art. no. 103823.

[4] I. A. Cruz-Albarran, J. P. Benitez-Rangel, R. A. Osornio-Rios, and L. A. Morales-Hernandez, "Human emotions detection based on a smart-thermal system of thermographic images," *Infrared Phys. Technol.*, vol. 81, pp. 250–261, 2017.

[5] G. Du, Q. Tan, C. Li, X. Wang, S. Teng, and P. X. Liu, "A noncontact emotion recognition method based on complexion and heart rate," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 5018614.

[6] A. B. S and M. Rao, "Design of emotion recognition system using neuromorphic computing technique," in *Proc. 18th Int. SoC Des. Conf.*, 2021, pp. 355–356.

[7] Y. Cai, X. Li, and J. Li, "Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review," *Sensors*, vol. 23, no. 5, 2023, Art. no. 2455, doi: 10.3390/s23052455. [Online]. Available: https://www.mdpi.com/1424-8220/23/5/2455

[8] D. Parikh, *Raspberry Pi and MQTT Essentials: A Complete Guide to Helping You Build Innovative Full-Scale Prototype Projects Using Raspberry Pi and MQTT Protocol*. Birmingham, U.K.: Packt Publishing Ltd., 2022.

[9] D. Soemantri, C. Herrera, and A. Riquelme, "Measuring the educational environment in health professions studies: A systematic review," *Med. Teacher*, vol. 32, no. 12, pp. 947–952, 2010.

[10] W.-H. Tang, W.-H. Ho, and Y. J. Chen, "Data assimilation and multisource decision-making in systems biology based on unobtrusive internet-of-things devices," *Biomed. Eng. Online*, vol. 17, no. Suppl 2, 2018, Art. no. 147.

[11] D. T. Hettich, E. Bolinger, T. Matuz, N. Birbaumer, W. Rosenstiel, and M. Spüler, "EEG responses to auditory stimuli for automatic affect recognition," *Front. Neurosci.*, vol. 10, 2016, Art. no. 244.

[12] S. Ablameyko, *Neural Networks for Instrumentation, Measurement and Related Industrial Applications* (NATO Science Series. Series III, Computer and Systems Sciences), vol. 185. Birmingham, AL, USA: IOS and EBSCO Industries, Inc, 2003.

[13] V. Braun, J. Blackmore, R. O. Cleveland, and C. R. Butler, "Transcranial ultrasound stimulation in humans is associated with an auditory confound that can be effectively masked," *Brain Stimulation*, vol. 13, no. 6, pp. 1527–1534, 2020.

[14] G. Cao, Y. Ma, X. Meng, Y. Gao, and M. Meng, "Emotion recognition based on CNN," in *Proc. Chin. Control Conf.*, 2019, pp. 8627–8630.

[15] U. Ali et al., "EEG emotion signal of artificial neural network by using capsule network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 434–446, 2020.

[16] I. Ajzen, "The theory of planned behavior: Frequently asked questions," *Hum. Behav. Emerg. Technol.*, vol. 2, no. 4, pp. 314–324, 2020.

[17] C.-S. Lee et al., "BCI-based hit-loop agent for human and AI robot colearning with AIoT application," *J. Ambient Intell. Humanized Comput.*, vol. 14, pp. 3583–3607, 2021.

[18] Y. Christina and N. N. K. Yasa, "Application of theory of planned behavior to study online booking behavior," *Int. J. Data Netw. Sci.*, vol. 5, pp. 331–340, 2021.

[19] R. Pathar, A. Adivarekar, A. Mishra, and A. Deshmukh, "Human emotion recognition using convolutional neural network in real time," in *Proc. 1st Int. Conf. Innov. Inf. Commun. Technol.*, 2019, pp. 1–7.

[20] B. M. Franco, L. Hernández-Callejo, and L. M. Navas-Gracia, "Virtual weather stations for meteorological data estimations," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12801–12812, 2020.

[21] D. R. Brown and J. F. Cavanagh, "The sound and the fury: Late positive potential is sensitive to sound affect," *Psychophysiology*, vol. 54, no. 12, pp. 1812–1825, 2017.

[22] U. Hernandez-Jayo and J. Garcia-Zubia, "Remote measurement and instrumentation laboratory for training in real analog electronic experiments," *Measurement*, vol. 82, pp. 123–134, 2016.

[23] D. Pizzagalli, T. Koenig, M. Regard, and D. Lehmann, "Affective attitudes to face images associated with intracerebral EEG source location before face viewing," *Cogn. Brain Res.*, vol. 7, no. 3, pp. 371–377, 1999.

[24] C. C. Berger and H. H. Ehrsson, "The content of imagined sounds changes visual motion perception in the cross-bounce illusion," *Sci. Rep.*, vol. 7, 2017, Art. no. 40123.

[25] P. B. Chopade and N. Prabhakar, "Human emotion recognition based on block patterns of image and wavelet transform," *Int. J. Adv. Technol. Eng. Exploration*, vol. 8, no. 83, 2021, Art. no. 1394.

[26] C. Pattamadilok and M. Sato, "How are visemes and graphemes integrated with speech sounds during spoken word recognition? ERP evidence for supra-additive responses during audiovisual compared to auditory speech processing," *Brain Lang.*, vol. 225, 2022, Art. no. 105058.

[27] T. Debnath, M. M. Reza, A. Rahman, A. Beheshti, S. S. Band, and H. Alinejad-Rokny, "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 6991.

[28] W. N. Ismail, M. M. Hassan, H. A. Alsalamah, and G. Fortino, "CNN-Based health model for regular health factors analysis in internet-of-medical things environment," *IEEE Access*, vol. 8, pp. 52541–52549, 2020.

[29] J. Liang, Y. Li, Z. Zhang, and W. Luo, "Sound gaps boost emotional audiovisual integration independent of attention: Evidence from an ERP study," *Biol. Psychol.*, vol. 168, 2022, Art. no. 108246.

[30] P. Kumar, H. Bagga, B. S. Netam, and V. Uduthalapally, "SAD-IoT: Security analysis of DDoS attacks in IoT networks," *Wireless Pers. Commun.*, vol. 122, pp. 87–108, 2022.

[31] N. Kulishova, "Emotion recognition using Sigma-Pi neural network," in *2016 IEEE 1st Int. Conf. Data Stream Mining Process.*, 2016, pp. 327–331.

[32] F. González-Hernández, R. Zatarain-Cabada, M. L. Barrón-Estrada, and H. Rodríguez-Rangel, "Recognition of learning-centered emotions using a convolutional neural network," *J. Intell. Fuzzy Syst.*, vol. 34, no. 5, pp. 3325–3336, 2018.

[33] K. N. Spreckelmeyer, M. Kutas, T. P. Urbach, E. Altenmüller, and T. F. Münte, "Combined perception of emotion in pictures and musical sounds," *Brain Res.*, vol. 1070, no. 1, pp. 160–170, 2006.

[34] G. Cao, Y. Ma, X. Meng, Y. Gao, and M. Meng, "Emotion recognition based on CNN," in *Proc. 2019 Chin. Control Conf. (CCC)*, Guangzhou, China, 2019, pp. 8627–8630, doi: 10.23919/ChiCC.2019.8866540.

[35] J. G. Razuri, D. Sundgren, R. Rahmani, and A. M. Cardenas, "Automatic emotion recognition through facial expression analysis in merged images based on an artificial neural network," in *Proc. 12th Mex. Int. Conf. Artif. Intell.*, 2013, pp. 85–96.

[36] M. Bruchmann, S. Schindler, J. Heinemann, R. Moeck, and T. Straube, "Increased early and late neuronal responses to aversively conditioned faces across different attentional conditions," *Cortex: J. Devoted Study Nervous Syst. Behav.*, vol. 142, pp. 332–341, 2021.

[37] X.-L. Chou et al., "Contextual and cross-modality modulation of auditory cortical processing through pulvinar mediated suppression," *eLife*, vol. 9, 2020, Art. no. e54157.

[38] L. M. Alonso-Valerdi, "Python executable script for estimating two effective parameters to individualize brain-computer interfaces: Individual alpha frequency and neurophysiological predictor," *Front. Neuroinform.*, vol. 10, 2016, Art. no. 22.

[39] M. Bekisz, W. Bogdan, A. Ghazaryan, W. J. Waleszczyk, E. Kublik, and A. Wróbel, "The primary visual cortex is differentially modulated by stimulus-driven and top-down attention," *PLoS One*, vol. 11, no. 1, 2016, Art. no. e0145379.

[40] O. Vynokurova and D. Peleshko, Eds., *DSMP 2016: Proceedings of the 2016 IEEE First International Conference on Data Stream Mining & Processing: Lviv, Ukraine, Aug. 23-27, 2016.* Piscataway, NJ, USA: IEEE, 2016.

[41] D. Schön and M. Besson, "Visually induced auditory expectancy in music reading: A behavioral and electrophysiological study," *J. Cogn. Neurosci.*, vol. 17, no. 4, pp. 694–705, 2005.

[42] L. Rodrigo-Salazar, I. González-Carrasco, and A. R. Garcia-Ramirez, "An IoT-based contribution to improve mobility of the visually impaired in smart cities," *Computing*, vol. 103, no. 6, pp. 1233–1254, 2021.

[43] S. Márquez-Sánchez, I. Campero-Jurado, D. Robles-Camarillo, S. Rodríguez, and J. M. Corchado-Rodríguez, "BeSafe B2.0 smart multisensory platform for safety in workplaces," *Sensors*, vol. 21, no. 10, 2021, Art. no. 3372.

[44] J. Domínguez-Borràs, S. W. Rieger, C. Corradi-Dell'Acqua, R. Neveu, and P. Vuilleumier, "Fear spreading across senses: Visual emotional events alter cortical responses to touch, audition, and vision," *Cereb. Cortex*, vol. 27, no. 1, pp. 68–82, 2017.

[45] S. Bobojanov, B. M. Kim, M. Arabboev, and S. Begmatov, "Comparative analysis of vision transformer models for facial emotion recognition using augmented balanced datasets," *Appl. Sci.*, vol. 13, no. 22, 2023, Art. no. 12271, doi: 10.3390/app132212271. [Online]. Available: https://www.mdpi.com/2076-3417/13/22/12271

[46] C. Zhang et al., "ResNet or DenseNet? Introducing dense shortcuts to ResNet," in *2021 IEEE Winter Conf. Appl. Comput. Vis.*, 2021, pp. 3549–3558, doi: 10.1109/WACV48630.2021.00359.

[47] Y. Zahid, M. A. Tahir, and M. N. Durrani, "Ensemble learning using bagging and inception-V3 for anomaly detection in surveillance videos," in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 588–592, doi: 10.1109/ICIP40778.2020.9190673.

[48] D. Sinha and M. El-Sharkawy, "Thin MobileNet: An enhanced mobilenet architecture," in *Proc. IEEE 10th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf.*, 2019, pp. 0280–0285, doi: 10.1109/UEMCON47517.2019.8993089.

[49] A. Bagaskara and M. Suryanegara, "Evaluation of VGG-16 and VGG-19 deep learning architecture for classifying dementia people," in *Proc. 4th Int. Conf. Comput. Informat. Eng.*, 2021, pp. 1–4, doi: 10.1109/IC2IE53219.2021.9649132.