Voice-Based Smart System for Emotion Recognition and Regulation

M. Parvathi[1], V. Pranathi[1], Maithreyi Varma[1], and B. V. V. Satyanarayana[2(✉)]

[1] ECE Department, BVRIT Hyderabad College of Engineering for Women, Telangana, India
[2] ECE Department, Vishnu Institute of Technology, Bhimavaram 534202, India
satyanarayana.b@vishnu.edu.in

Abstract. Recognizing and supporting children's emotional growth can be challenging, often leading to upsets in their emotional regulation since they lack an internal alarm system to regulate their emotions. It is important to observe their continuous mood swings and emotions to assess their psychological state. In general, younger people are faster in expressing their emotions rather than elders. Emotions that are naturally exhibited if suppressed would lead to psychological traumas. Also, emotional deregulation has a detrimental effect on brain function and interpersonal abilities may likely to lead personality disorder issues. Young adults can address the impacts of childhood trauma and lead fulfilling lives with the help of efficient, evidence-based treatment or by any means of smart emotion recognition alert systems. This paper proposes a smart emotion recognition alert system based on real emotion features that will be collected in real-time. The major concerns of emotions are sadness, fear, crying, screaming, shouting, anger and frustration which are collected from the supported sensors built around the design, and will get processed by ML algorithms to get back with voice command notes for regulation. The proposed model is tested using a dataset with emotions captured by children ages 4 to elders 50. The clustering algorithm applied in assessing the accuracy along with voice modules and face detection support, observed the accuracy level reached a maximum of 72.5%. The emotion detection system alone will not be adequate in this regard as the outcomes demonstrated that, to return to normal while lessening the strength of the emotion, a voice alarm system is also necessary.

Keywords: Raspberry Pi · Pulse rate Sensor · Microphone · Emotion Recognition · Speech Recognition · Face Recognition

1 Introduction

Emotion is commonly defined in psychology as a complex emotional state that can alter thought and behaviour as well as cause physical and psychological changes. According to some theoretical stances, emotions are not the cause but rather the result of a

P. Pareek et al. (Eds.): IC4S 2024, LNICST 599, pp. 474–488, 2025.
https://doi.org/10.1007/978-3-031-77081-4_37

confluence of various elements such as behaviour, motivation, feelings, and physiological alterations [1, 2]. Other psychological ideas like temperament, mood, personality, motivation, and disposition are frequently entwined with emotion. Human emotions are essential for promoting social interactions because they allow people to relate to one another, communicate their own emotions, and give helpful criticism. Numerous studies have shown how emotions have a significant influence on human interactions. Because of this, a brand-new area of research known as automatic emotion detection has arisen to precisely recognize and comprehend emotions. Numerous modalities, including physiological signals, speech patterns, facial expressions, and more, have been studied in this subject in previous studies [3, 4]. The ultimate objective is to provide trustworthy techniques for identifying and analyzing emotional states.

Recent years have seen a fast advancement in sensor and information technology, opening the door for machines to understand and interpret human emotions. In many different domains, emotion recognition is a significant area of research. Emotions in people can take many different forms. As a result, emotion recognition can be achieved by the examination of physiological signals, speech, behaviour, and facial expressions. Various sensors are gathering these signals. Accurately identifying human emotions can help affective computing advance.

Visual, auditory, radar, and other physiological signal sensors are the primary sensors used in emotion detection. These sensors may gather signals of various dimensions and, using certain algorithms, do emotional analysis. Applications for various sensors vary when it comes to emotion identification.

1.1 Visual Sensor

One of the most popular techniques for recognizing emotions is based on visual sensors. Its benefits include easy data collection and minimal cost. Currently, facial expression recognition (FER) is the primary method utilized by visual sensors to identify emotions [5]. It is challenging for machines to replicate human expression detail. Emotion recognition mistakes arise from the ease with which facial expressions can be concealed. For instance, even when they are not in a good mood, people typically smile pleasantly during certain social events. The same emotion can have distinct facial features, and subtle variations in an individual's feelings are often imperceptible. The camera's ability to recognize facial expressions presents a classification difficulty because of the huge intra-class distance and modest inter-class distance in emotion detection. Additionally, it might be challenging to accurately identify emotions when a mask is worn or when the subject is being photographed from a variety of perspectives.

1.2 Audio Sensor

One of the most crucial aspects of human culture is language. Language is a means by which people may interact with one another and express themselves. Rich information included in human speech can be utilized to identify emotions.

It takes skill to discern the emotion in speech. Acoustic variability is caused by the various speaking styles of individuals, and this has an immediate impact on speech feature extraction and categorization. Different emotions may be expressed in the same statement, and certain specific emotional variances may rely on the speaker's upbringing or local culture, which presents further difficulties for Speech emotion recognition (SER) [6].

1.3 Physiological Sensor

It is thought that intense emotion may have an impact on how well important organs operate. Aristotle thought that physiological conditions like a fast heartbeat, elevated body temperature, or hunger loss are indicative of the impact of emotions on physiology. The autonomic nervous system would become active in reaction to external stimuli, causing the brain to go into physiological response mode. People laugh when they're joyful, for instance; they stand on end when they're terrified, and they cry when they're sad. A real-time facial emotion identification system with deep learning-based emotion detection is confined to identifying emotions; it cannot provide alarms or real-time feedback.

The development of children depends on their emotional health, yet it can be difficult for parents to recognize and control their feelings. The inadequacy of current emotion-detection technologies to offer caregivers real-time alerts and instruction leaves a vacuum in safeguarding children's emotional well-being. Young adults who have experienced childhood trauma are affected psychologically, physically, and cognitively, among other aspects. Studies reveal that those who went through traumatic experiences as children had a threefold increased risk of developing mental health issues and a fifteen-fold increased risk of developing borderline personality disorder [7, 8]. Relationship abilities and brain function might be adversely affected by adverse childhood experiences (ACES). Young adults can address the impacts of childhood trauma and lead fulfilling lives with the help of efficient, evidence-based treatment. With the use of a cutting-edge Emotion Alert System (EAS), the proposed system seeks to transform emotional well-being at a time of growing human-machine connections. This system reliably detects and interprets human emotions by utilizing state-of-the-art technology to smoothly combine inputs from a variety of sources, such as voice modulation, facial expressions, and pulse rate.

The suggested work is based on the utilization of biometric data, the extraction of physiological indications in real-time, and the application of sophisticated algorithms for the detection of emotions. The proposed Emotion Alert System offers a thorough picture of an individual's emotional state through a mix of wearable biometric sensors and advanced voice and facial analysis technologies. An intelligent alert system that reacts to changes in the user's emotional baseline is the fundamental functionality. The technology is designed to facilitate proactive interventions and support by triggering timely notifications upon identification of substantial shifts in emotional states. This helps the user become more self-aware and makes it possible for outside caregivers or support systems to adequately attend to the user's emotional needs.

One of the Emotion Recognition System's primary benefits is its capacity to be tailored to the unique emotional expression of each user through user-customizable thresholds. Furthermore, the suggested effort placed a greater emphasis on data security, user privacy, and employing strong encryption techniques to protect sensitive data. The goal of the proposed smart device development is to use technology to understand and actively promote people's mental well-being. At the vanguard of this vision is the

Emotion Intelligent Smart System, which will usher in a new era of tailored emotional intelligence, resilience, and healthy interactions between people and technology. The remaining paper deals with a review of related work with limitations in Sect. 2. The proposed smart system with architectural details and working is dealt with in Sect. 3.

2 Review of Related Work

Speech signals are an attractive source for affective computing due to several intrinsic benefits. For example, speech signals are typically easier and less expensive to capture than many other biological signals (such as an ECG). Speech emotion recognition (SER) [9, 10] has attracted the interest of most researchers because of this. SER uses a speaker's voice to infer her underlying emotional state. Over the past few years, there has been a growing interest in this field of research. Information regarding students' emotional states can be used to focus on improving the quality of instruction in the classroom or through online learning platforms. For instance, a teacher can utilize SER to determine which subjects are appropriate to teach and has to be able to create plans for handling different emotions in the classroom. The emotional state of the students should therefore be taken into account in the classroom.

For a SER system to be successful, three important difficulties must be resolved: (1) selecting an appropriate emotional speech database, (2) collecting useful characteristics, and (3) utilizing machine learning algorithms to create trustworthy classifiers. In everyday interpersonal relationships, emotion is a major factor. By expressing one's emotions and providing feedback to others, it aids in matching and comprehending the feelings of others. Emotion has a significant influence on how people engage with one another, according to research [11]. Emotional expressions reveal a great deal about a person's mental health. This has led to the emergence of a novel field of research called automated emotion recognition, whose primary goal is to understand and retrieve pleasant emotions. Previous research has investigated many modalities, including facial expressions [12], speech [13], physiological signals [14], and others, to identify emotional states. One of the primary problems with the SER system is emotional feature extraction. Important speech elements that convey emotion have been proposed by numerous researchers [15], including modulation spectral features (MSFs) [16], energy, pitch, formant frequency, linear prediction cepstrum coefficients (LPCC), mel-frequency cepstrum coefficients (MFCC), and pitch. As a result, the majority of researchers favour combining feature sets, which are made up of many feature types and incorporate additional emotional information [17].

A combined feature set, however, may result in high dimension and redundancy of speech features, complicating the learning process for the majority of machine learning algorithms and raising the risk of over fitting. As a result, feature selection is essential for minimizing feature redundancy in dimensions. Enhancing learning efficiency, reducing computing complexity, creating more generalizable models, and requiring less storage are all possible with feature extraction and feature selection.

Classification is the final stage in speech emotion recognition. It entails using attributes taken from the data to categorize the raw data—which is a speech or frame of an utterance—into a specific class of emotion. Many classification techniques, including the Gaussian mixture model (GMM), hidden Markov model (HMM), support vector

Table 1. Overview of Various Emotion Recognition Techniques.

Approach type	Technology involved	Examples	Limitations
Feature Extraction	Key characteristics are manually selected and extracted from facial photos using geometric landmarks, further to be used in classification	Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and eigenfaces	Restricted by the level of detail in the features chosen hence less accurate in emotion recognition task
ML algorithms	These approaches use set of labeled data for the automatic classification of emotions from facial expressions	Popular are Support Vector Machines (SVMs), k-nearest Neighbors (kNN), or Random Forests (RF) etc. based on accuracy interest	The advantage of ML models is that they can recognize intricate patterns in facial expressions; but, to train them, a substantial amount of labeled data is needed
Deep Learning	These approaches use training neural networks with numerous layers for the classification of emotions from facial expressions	Popular are Convolutional Neural Networks (CNNs), DNN	Complexity lies with large amounts of labeled data hence computationally expensive

machine (SVM), neural networks (NN), and recurrent neural networks (RNN), have been presented by academics in recent years for speech emotion recognition [18–30]. Researchers have also developed alternative classifiers, such as the modified brain emotional learning model (BEL), which combines the multilayer perceptron (MLP) and adaptive neuro-fuzzy inference system (ANFIS) to recognize speech emotions.

Another approach that has been suggested is the multiple kernel Gaussian process (GP) classification, wherein the linear kernel and radial basis function (RBF) kernel are combined to present two conceptions that are comparable in the learning procedure. In contrast to the conventional approach, the Voiced Segment Selection (VSS) algorithm, which was previously proposed in [31], deals with the voiced signal segment as the texture image processing feature. To classify the data, the voiced and unvoiced characteristics are extracted from the spectrogram using the Log-Gabor filters. In addition, a feature selection method is evaluated to extract high-quality features from the extracted collection of features; its limitations are displayed in Table 1.

Table 1 shows that facial expressions are the primary means of capturing emotions. There are two methods for analyzing facial expressions: automated and manual. The manual analysis involves the classification of expressions and observations that need to be completed by skilled humans. The 1970s saw the development of the Facial Action Coding System (FACS), one of the most extensive and conventional systems for characterizing facial motions based on the underlying muscle actions, by psychologists

Wallace Friesen and Paul Ekman. In automated analysis, on the other hand, computer algorithms play a significant role in the detection and classification of observed facial expressions, with further classification relying more on facial features. Additionally, the accuracy level of the selected classifier can be confirmed by using the most appropriate machine-learning algorithms [32, 33]. Research on emotion recognition generally relies heavily on facial expression analysis, and advancements in this area may improve both the precision of algorithms and the understanding of human emotion. Although the corrective mechanism has not yet been designed, emotion identification algorithms have been applied in several research studies. Of the most popular techniques for recognizing emotions is based on visual sensors. Its benefits include easy data collection and minimal cost. Currently, facial expression recognition (FER) is the primary method utilized by visual sensors to identify emotions.

3 The Architecture of the Proposed System and Working

The proposed system's core components are biometric data extraction, real-time physiological signal extraction, and the use of sophisticated emotion identification algorithms. The proposed Emotion Alert System offers a thorough picture of an individual's emotional state through a mix of wearable biometric sensors and advanced voice and facial analysis technologies. The Emotion Alert System continuously scans for minute signs that are frequently missed, serving as a tailored emotional intelligence partner. The system dynamically adapts to the user's baseline by evaluating changes in speech modulation, facial expressions, and pulse rate. This enables individualized and sophisticated emotion recognition.

The intelligent alert system's primary function is to react when the user's emotional baseline has deviated from normal. The system initiates timely warnings upon identification of notable changes in emotional states, so enabling proactive interventions and assistance. This improves self-awareness and makes it possible for support systems or outside caregivers to adequately attend to the user's emotional needs. User-customizable thresholds, which guarantee flexibility to individual variability in emotional expression, are a key component of the proposed Emotion Alert System.

3.1 Emotion Recognition System

The emotion alert sensor uses a camera for facial identification, a microphone for voice modulation analysis, and a sensor for pulse rate monitoring. The technology interprets a person's emotional state by examining their voice patterns, facial expressions, and pulse rate fluctuations. The sensor offers a thorough and multifaceted approach to emotional assessment when consistent emotional patterns are found across these modalities. In response, the sensor generates alerts specific to the indicated emotion.

3.2 Facial Recognition Using Camera

The person in focus's face was captured by a camera that employed computer vision algorithms to scan the image and assess facial traits like mouth, eye, and brow movements. This determines the person's emotional state by comparing these characteristics with pre-established emotional patterns.

3.3 Voice Modulation Analysis

A microphone is used by the suggested setup to record the user's voice. The main function of this module is to examine the voice's modulation, pitch, and tone. To determine the person's emotional state, compare these voice traits with recognized emotional patterns.

3.4 Pulse Rate Sensor

To determine the user's heart rate, the device uses a heart rate sensor. Different emotional states are associated with variations in pulse rate. The sensor determines which of the preset emotional profiles matches the collected pulse rate data.

3.5 Alert Generation

The results of voice modulation analysis, facial recognition, and pulse rate detection are integrated into the suggested smart system. An alert is generated if a recurring emotional pattern is found using all of these modalities. The recognized emotion, such as happiness, stress, or sadness, might be used as feedback to the user in the form of an alert. The suggested system includes a voice-based message alert that functions similarly to an alarm. If the threshold of any sub-module is crossed, the alert system will sound the alarm until the person's emotions return to normal. Figure 1 shows the overall sequence of operation in the process of the proposed emotion detection feedback alert system.

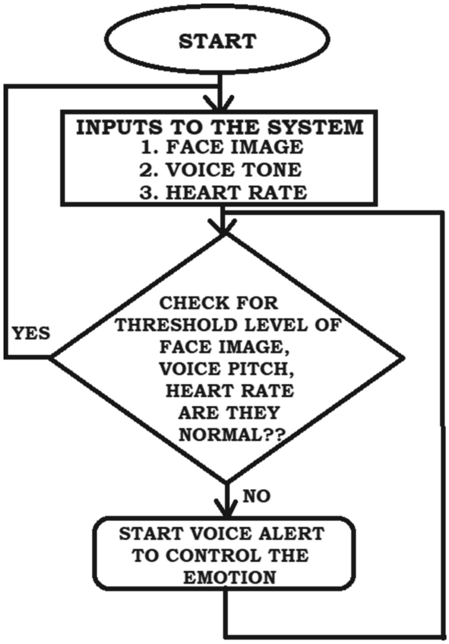

Fig. 1. A Flowchart for the sequence of operations in the process of emotion detection

3.6 Hardware Requirement

The next phase involves designing and implementing a smart system for emotion recognition and regulation. It is divided into three main sections: input, control, and output. The prototype design for the output component uses a Raspberry Pi with voice, camera, and heart rate sensor modules. The voice module of the Raspberry Pi's input section uses a microphone to record the user's voice and examines the voice's modulation, pitch, and tone. To determine the person's emotional state, compare the voice traits of the input voice with recognized emotional patterns.

On the other hand, the individual's heart rate is measured by a heart rate or pulse rate sensor. Different emotional states are associated with variations in pulse rate. To determine the present emotional state, the sensor compares the acquired pulse rate data with pre-established emotional profiles. The camera module for facial recognition is the third module in the input area. It records the focused person's facial expressions and analyzes face characteristics like mouth, eye, and brow movements using facial expression recognition algorithms like SVM or KNN further to determine the person's emotional state by comparing these characteristics with pre-established emotional patterns. The block diagram and hardware setup of the proposed design are shown in Fig. 2 and Fig. 3

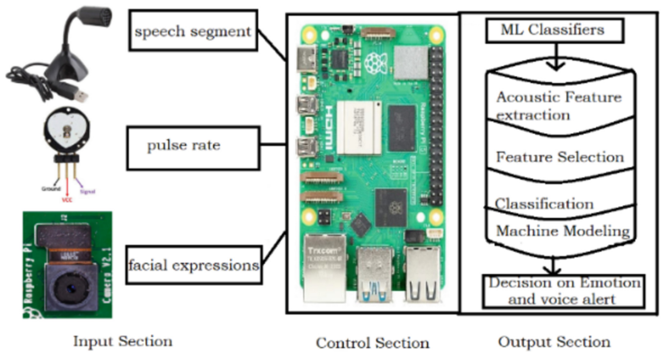

Fig. 2. Block diagram of proposed smart system for emotion recognition and correction

3.7 Dataset Requirement

The datasets are gathered from many sources, including Figshare.com and Kaggle.com, to validate the suggested design while taking into account a range of emotional factors. In addition to voice inputs, the combined emotional qualities taken into consideration are stress, rage, sobbing, yelling, and dullness. The microphone that records regular conversations provides the voice inputs, but when the threshold level is raised in response to a change in pitch, the alert system is triggered to produce a warning message. Figure 4 displays the dataset that was used for the suggested work.

Before applying machine learning methods, 638 samples of voice and image data are examined. These datasets are then reframed with relevant data using threshold values. Speech recognition modules and the sentiment analysis pipeline were started. The

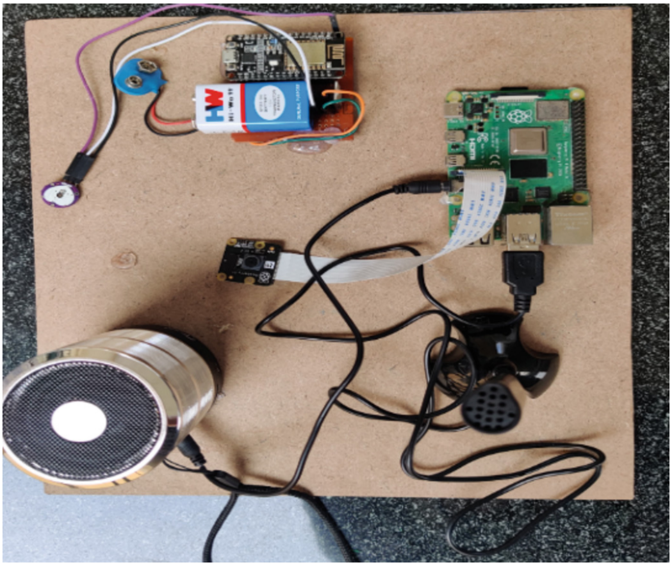

Fig. 3. Hardware setup for the proto-typed design

Age	feeling.nervous	panic	anger	over.react	feeling.tired	introvert	stressful	crying	screaming	shouting	staring for long	self beating	teath biting	repetitive.behaviour	throw out things	Disorder
23	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	MDD
14	1	0	1	1	0	1	0	1	1	0	0	0	0	1	0	ASD
25	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	Loneliness
29	1	0	0	1	0	0	0	0	0	0	1	1	0	0	1	bipolar
32	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	anexiety
40	1	1	1	1	0	0	1	1	0	0	0	1	0	1	0	PTSD
18	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	sleeping disorder
32	1	0	0	0	1	1	1	0	1	1	1	0	1	0	0	psychotic deprission
24	1	1	0	1	0	1	0	0	1	0	1	0	0	1	0	ASD
37	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	eating disorder
4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	ADHD
24	0	1	1	0	0	1	0	1	1	0	0	0	0	1	0	ASD
18	1	0	1	1	0	1	0	1	1	0	0	0	0	1	0	ASD
27	1	0	0	0	0	0	0	0	0	0	1	0	0	1	1	ADHD
15	0	1	1	1	1	0	1	1	0	1	1	0	0	0	0	MDD
19	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	PDD
40	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	PTSD
12	1	0	0	0	1	0	1	0	0	0	0	0	0	1	1	OCD
21	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	eating disorder
19	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	bipolar
35	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	sleeping disorder
40	1	0	0	0	1	0	1	1	0	0	1	0	0	0	0	PTSD

Fig. 4. Dataset used for validation

suggested system only has one audio input device connected, however, in actual use; several audio inputs can be linked to pre-trained audio modules for analysis.

A heart rate pulse reading module is also utilized. Before determining if the sensor is attached to the body through the device, the heart rate pulse reading module will determine this. Boolean parameter "bool = true," which denotes that the device is still attached to the body, is used for this. Apart from the modules for voice and heart rate, there is also an image module that utilizes a webcam to continuously read postures. While the prototyped design employs a pi-camera with a resolution of five megapixels per inch intact to the board itself, the suggested design used a webcam to capture the photographs. To do the emotion comparisons with the designated threshold levels, the entire hardware consists of an integrated arrangement of controller, voice, and camera modules that operate in addition to accessing and executing the code in RealVNC viewer operating on a fast GUI processor.

4 Results and Observations

To validate the prototype proposed system, KNN algorithm is used. The proposed design aims to emphasize voice alert messages over the observed negative moods of the user. In this respect, using previously collected dataset with emotions is used for training the model. The number of neighbors taken are 5 as a choice. Then, the algorithm used the input data from the camera, with a function using 'faces = face_detector.detectMultiScale(gray_frame, scaleFactor = 1.1, minNeighbors = 5)'. This function further converted each frame to resized frame using function 'frame = cv2.resize(frame, (1280, 720))', and then converted to gray frame using function 'gray_frame = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)'. Finally the identified reshaped face with frame is given as input to the classifier to predict the emotion on the face, using the function 'emotion_prediction = emotion_model.predict(cropped_img)'. The flattened and reshaped face is then passed to the KNN classifier's 'predict()'

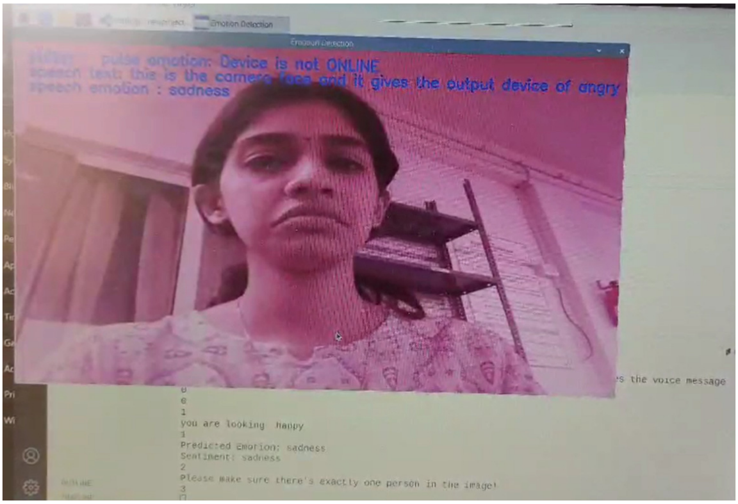

Fig. 5. Emotion recognition for sad mood

method to predict the label of the face. Further the voice module is initiated with speech_recognition library by giving recorded voices as input data. Further, the heart rate monitoring is carried out using the opencv library, and the estimation of heart rate is done using the frequency filtering method.

A selection of the output observations, which were obtained in the moods of the user's happy, neutral, and sad states, are displayed in Fig. 5, Fig. 6 and Fig. 7. Instead of being there in a negative mood for a long time, this would enable the user or users to get rid of and return to normal conditions a little faster. Mood swings are fairly prevalent, especially in younger individuals, therefore it will be beneficial for them to learn to let go of control over their lives.

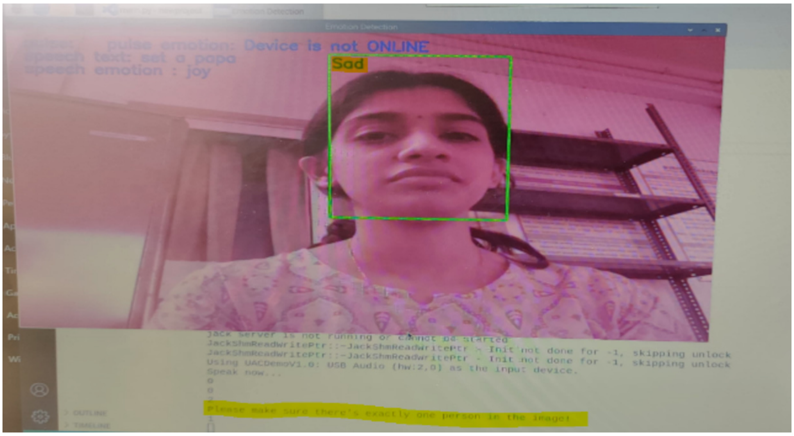

Fig. 6. Emotion recognition for neutral mood

The heart rate threshold is selected as 65 pulses per minute as normal, beyond this level is considered abnormal and to be controlled. The voice alert will start cautioning on noticing the level and continue till it reaches back to normal. The user will hear an audio message containing caution or alarm when the system detects an abnormality in any of the emotions. An alert voice message such as "you are looking sad, everything is going to be OK" will be heard if the user's mood is noticed to be depressed. This will warn the user until their face or mood returns to normal.

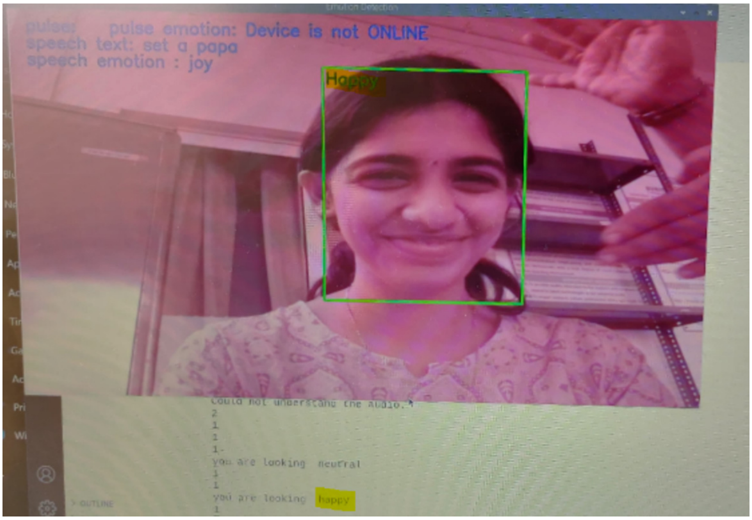

Fig. 7. Emotion recognition for a happy mood

Pseudo code for heart rate monitoring is as follows:

```
    if(data &lt;=10)://check whether sensor is on the body
    pulsetext = &quot;Device is not in contact with body&quot;
    elif(data&gt;10 and int(data) &lt;=65):
    pulsetext = &quot;HeartBeat is normal&quot;
    elif(data&gt;=65):
    pulsetext = &quot;HeartBeat is Abnormal control heart beat&quot;
    speaktext = pulsetext
speakbool = True
```

Pseudo code for voice based feedback system:

```
ifemotion_dict[maxindex].lower() in &quot;sad&quot;:
 speaktext = &quot;videomode,you are looking sad, Everything is going to be OK
 &quot;
speakbool = True
elif(emotion_dict[maxindex].lower() in &quot;disgust&quot;):
speaktext = &quot;video mode dont get disgusted, Its temporary&quot;
speakbool = True
elif(emotion_dict[maxindex].lower() in &quot;fearful&quot;):
speaktext = &quot;videomode,you are looking fearful, dont get scared please seek
help&quot;
speakbool = True
elif(emotion_dict[maxindex].lower() in &quot;angry&quot;):
speaktext = &quot;videomode,you are looking angry, Take a deep breath&quot;
```

5 Conclusions

To sum up, a new system has been suggested. It analyzes and detects children's emotions by combining a heart rate sensor, camera, and microphone. This device not only recognizes emotions but also notifies caretakers in real-time. It gives parents and caregivers peace of mind by assisting in the maintenance of emotional equilibrium. This system is an important tool for developing young brains since it will be tested on a dataset of children ages 3 to 18. Its accuracy and dependability will be assessed on this dataset, the real-time expected labels and rectangles surrounding the faces that were detected. The accuracy with the selected image test data observed using KNN algorithms is 72.5%. The entire system operates in real-time and can be configured to accept user mood swings in three different ways: speech, picture, and heart rate. Based on these inputs, if any one of them indicates a breach, the user will automatically hear a warning voice until their mood returns to normal.

References

1. Cabanac, M.: What is emotion? Behav. Proc. **60**(2), 69–83 (2002)
2. Schacter, D.L., Gilbert, D.T., Wegner, D.M.: Introducing Psychology, 2nd edn. Macmillan, New York (2011)
3. Barrett, L.F. and Russell, J.A.: An introduction to psychological construction. In: Barrett, L.F., Russell, J.A., (eds.), The Psychological Construction of Emotion, pp. 1–17. The Guilford Press (2015)
4. James, W.: II - What is an emotion? Mind **9**(34), 188–205 (1884)
5. Kerkeni, L., Serrestou, Y., Mbarki, M., Raoof, K. and Mahjoub, M.A.: A review on speech emotion recognition: case of pedagogical interaction in classroom. In: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 1–7. IEEE, Morocco (2017)
6. Ekman, P.: An argument for basic emotions. Cogn. Emot. **6**(3/4), 169–200 (1992)
7. Matilda, S.: Emotion recognition: a survey. Int. J. Adv. Comput. Res. **3**(1), 14–19 (2015)
8. Koolagudi, S.G., Rao, K.S.: Emotion recognition from speech: a review. Int. J. Speech Technol. **15**, 99–117 (2012)
9. Boekaerts, M.: The crucial role of motivation and emotion in classroom learning. In: The Nature of Learning: Using Research to Inspire Practice, pp. 91–111. OECD Publishing, Paris (2010)
10. Schirmer, A., Adolphs, R.: Emotion perception from face, voice, and touch: comparisons and convergence. Trends Cogn. Sci. **21**(3), 216–228 (2017)
11. He, C., Yao, Y.J., Ye, X.S.: An emotion recognition system based on physiological signals obtained by wearable sensors. In: Yang, C., Virk, G., Yang, H. (eds.) Wearable Sensors and Robots. Lecture Notes in Electrical Engineering, vol. 399. Springer, Singapore, pp. 15−25 (2017). https://doi.org/10.1007/978-981-10-2404-7_2
12. Ali, H., Hariharan, M., Yaacob, S., Adom, A.H.: Facial emotion recognition using empirical mode decomposition. Expert Syst. Appl. **42**(3), 1261–1277 (2015)
13. Liu, Z.T., Wu, M., Cao, W.H., Mao, J.W., Xu, J.P., Tan, G.Z.: Speech emotion recognition based on feature selection and extreme learning machine decision tree. Neuro-Comput. **273**, 271–280 (2018)
14. Sree, G.D., Chandrasekhar, P., Venkatesshulu, B.: SVM based speech emotion recognition compared with GMM-UBM and NN. Int. J. Eng. Sci. 3293 (2016)

15. Chen, S., Jin, Q.: Multi-modal dimensional emotion recognition using recurrent neural networks. In: Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge, pp. 49–56. ACM, Brisbane (2015)
16. Motamed, S., Setayeshi, S., Rabiee, A.: Speech emotion recognition based on a modified brain emotional learning model. Biologically Inspired Cogn. Archit. **19**, 32–38 (2017)
17. Lim, W., Jang, D., Lee, T.: Speech emotion recognition using Convolutional and recurrent neural networks. In: 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 1–4. IEEE, Korea (2016)
18. Kerkeni, L., Serrestou, Y., Mbarki, M., Raoof, K. and Mahjoub, M.A.: Speech emotion recognition: methods and cases study. In: International Conference on Agents and Artificial Intelligence (ICAART), pp. 175–182. Springer, Portugal (2018)
19. Ragot, M., Martin, N., Em, S., Pallamin, N., Diverrez, J.M.: Emotion recognition using physiological signals: laboratory vs. wearable sensors. In: Ahram, T., Falcão, C. (eds.) Advances in Human Factors in Wearable Technologies and Game Design. AHFE 2017. Advances in Intelligent Systems and Computing, vol. 608, pp. 15–22. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-60639-2_2
20. Fernández-Caballero, A., et al.: Smart environment architecture for emotion detection and regulation. J. Biomed. Inform. **64**, 55–73 (2016)
21. Wu, S., Falk, T.H., Chan, W.Y.: Automatic speech emotion recognition using modulation spectral features. Speech Commun. **53**(5), 768–785 (2011)
22. Wagner, J., André, E., Jung, F.: Smart sensor integration: a framework for multimodal emotion recognition in real-time. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 1–8. IEEE, Netherlands (2009)
23. Tang, J., Alelyani, S., Liu, H.: Feature selection for classification: a review. In: Data classification: Algorithms and applications, pp. 37–64. CRC Press (2014)
24. Vondra, M., Vích, R.: Recognition of emotions in German speech using gaussian mixture models. In: Esposito, A., Hussain, A., Marinaro, M., Martone, R. (eds.) Multimodal Signals: Cognitive and Algorithmic Issues. Lecture Notes in Computer Science(), vol. 5398, pp. 256–263. Springer, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00525-1_26
25. Ingale, A.B., Chaudhari, D.S.: Speech emotion recognition using hidden Markov model and support vector machine. Int. J. Adv. Eng. Res. Stud. **1**(3), 316–318 (2012)
26. Milton, A., Roy, S.S., Selvi, S.T.: SVM scheme for speech emotion recognition using MFCC feature. Int. J. Comput. Appl. **69**(9), 34–39 (2013)
27. Pan, Y., Shen, P., Shen, L.: Speech emotion recognition using support vector machine. Int. J. Smart Home **6**(2), 101–108 (2012)
28. Shen, P., Changjun, Z., Chen, X.: Automatic speech emotion recognition using support vector machine. In: Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, pp. 621–625. IEEE, China (2011)
29. Graves, A., Jaitly, N.: Towards end-to-end speech recognition with recurrent neural networks. In: 31st International Conference on Machine Learning (ICML 2014), pp. 1764–1772. Springer, China (2014)
30. Prasomphan, S.: Improvement of speech emotion recognition with neural network classifier by using speech spectrogram. In: 2015 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 73–76. IEEE London (2015)
31. Gu, Y., Postma, E., Lin, H.X., Herik, J.V.D.: Speech emotion recognition using voiced segment selection algorithm. In: Proceedings of the Twenty-Second European Conference on Artificial Intelligence, pp. 1682–1683. IOS Press, Netherlands (2016)

32. Ngo, D., Nguyen, A., Dang, B. and Ngo, H.: Facial expression recognition for examining emotional regulation in synchronous online collaborative learning. Int. J. Artif. Intell. Educ. 1–20 (2024)
33. Huang, Z.Y., et al.: A study on computer vision for facial emotion recognition. Sci. Rep. **13**(1), 8425 (2023)