Implementation of Facial Emotion Recognition System using CNN and IoT

Arul P [1], R Sudha [2] , S Padmapriya[3], Abirami N [4], Manishankar R[5]

[1*] *Dept. of Electronics and Communication Engineering, R.M.D. Engineering College Chennai, India*

[2,3,5] *Dept. of Computer Science and Engineering, AVC College of Engineering, Mayiladuthurai, India*

[4] *Dept. of Information Technology, R.M.D. Engineering College Chennai, India*

Keywords: ***ECG, GSR, CNN, Normalization, Facial Expressions, IoT.***

Abstract. A lot of effort has been paid to emotion modelling and recognition by fields including psychology, cognitive science, and, more recently, engineering. While behavioral modalities have been the subject of extensive investigation, physiological signals have received less attention. Electrocardiograph (ECG) signals can vary depending on the emotion, and different emotions can be identified by different changes in ECG signals. The goal of this study is to use ECG signals to recognize emotions. Four different emotions are represented by the data: happy, thrilling, tranquil, and tense. A finite impulse filter is then used to de-noise the raw data. To improve the accuracy of emotion recognition, we utilize the Discrete Cosine Transform (DCT) to extract characteristics from the collected data. Electrocardiograms (ECGs) and GSR are used in this project's emotion recognition research as both a unimodal and multimodal approach to emotion recognition systems. There are critical observations made of the following processes: pre-processing, validation, dimensionality reduction, feature extraction, feature selection, and data collecting. Additionally, this project showcases architectures with accuracy levels greater than 90%. Also evaluated are the existing ECG and GSR inclusive emotional databases, and a popularity analysis is provided. This review also covers the advantages of emotion recognition technologies for healthcare systems. We conclude with a full discussion of the topic and recommendations for future work based on the evaluated literature. The results offered here are helpful for aspiring researchers looking to review the overview of earlier studies on ECG and GSR -based emotion recognition systems, identify knowledge gaps, and develop and design future applications of emotion recognition systems, particularly for enhancing healthcare.

Introduction

Emotion recognition and classification in real time are essential for productive human-robot interaction. Emotions in humans are relatively permeable to events, ideas, and physical qualities. It is crucial that no data is wasted during the gathering process, hence a compact model configuration that can swiftly deliver findings is required. Although real-time research in this field has been extensively pursued, it remains impossible to rely on facial expressions and vocal messages since they are under the control of the human somatic nervous system (SNS). Since human control over the signals of the autonomic nerve system (ANS) is restricted, the ANS can be used to correctly determine internal emotional states [1, 2]. Human- robot interaction can be studied, designed, and implemented with the help of emotional intelligence.

Complex linked emotions, which develop in response to one's unique set of internal and external contexts and traits, have an effect on the efficacy of emotion identification and categorization systems. Therefore, using the fusion technology of bio-signals (ANS), facial expressions, and voice data (SNS), one may build a reliable emotion recognition model by improving the internal and external emotion recognition and classification performance [3]. For instance, the medical field can employ internal emotion recognition technology for patient rehabilitation [5] and the aviation industry can use it for training pilots on complex aircraft [4] by analysing their emotional bio-signals. Traditional methods of ANS monitoring and evaluation have involved the use of physiological indicators such as electrocardiograms, galvanic skin reactions, blood pressure, and respiration rates. The electrocardiogram (ECG) and the galvanic skin response

(GSR) stand out as particularly useful tools for evaluating a wide range of clinical and psychophysiological conditions. Both the electrocardiogram (ECG) and the galvanic skin response (GSR) can be used as useful indicators of autonomic nervous system (ANS) activity [1]. These assessments are simple to implement, inexpensive, harmless, and continuous in nature. However, intuitive interpretation is necessary for pinpointing the patterns you seek among the various mental and physical states.Since stress affects people of all ages, backgrounds, and professions today, it is often discussed. Workplace cultural changes, increased stress, alterations in personal priorities, and technological advancements are all possible explanations for this development.

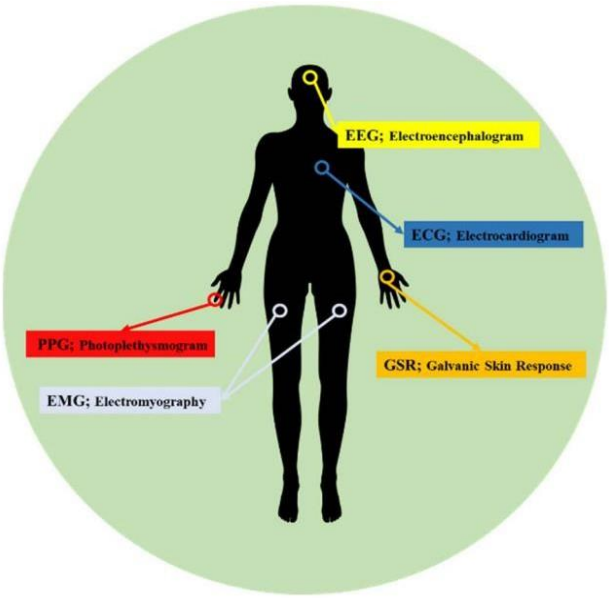

Fig 1. Bio-signals of an ANS based on measurement location and type

That's why, unlike decades ago, stress detection systems are crucial today. Because stress is always present, safeguarding the world's human capital from its growing impact is essential. In order to improve people's mental and physical health, it is crucial that stress be diagnosed and addressed early on. Despite their transient nature, stress and emotions share important physiological underpinnings. When stress levels are really high, many negative emotions might surface.

Given the inextricable link between emotion detection systems and stress detection systems, this paper gives a complete overview of both, with a particular focus on the role that machine learning algorithms play in both.Several different bio-signals can be acquired without any quality degradation using either invasive or non- invasive methods. The invasive variety includes inserting a sensor and needle into the person's body for a precise signal, but this raises ethical concerns regarding the potential for harm to the victim.

Consequently, the non-invasive method [6, 7], in which electrodes and sensors are only glued to the skin of the individual, is the most used technique. Figure 1 shows how ANS signals obtained from different parts of the human body and using different methods can seem very different from one another. Fig 1. Shows photoplethysmogram (PPG) monitors the amount of light reflected from blood vessels, while an electrocardiogram (ECG) examines electrical activity in the heart. Galvanic skin response (GSR) detects perspiration and thermal changes. Electromyography (EMG) is a technique used to detect and record electrical activity in skeletal muscles.

Literature Review

The simplicity of use, possibility for commercialization, and unique regularity of a bio-signal all seem different depending on the measurement method utilised, as stated above. There is also the possibility that the number and placement of electrodes used to collect various bio-signals has a significant bearing on the usability and efficiency gained by users. Different experiments and examinations have been done from diverse views due to the broad diversity of information offered by these signals and the strategies of utilising this information. EEG readings require a large number of electrodes (up to 64 channels) and complex data. The power spectral density (PSD) of the EEG signals was used by Mantini et al. [6] to extract emotional information; furthermore, Topic et al. [4] used the topography of the EEG data to extract features of persistent emotional information. The electrical activity of the heart is recorded via an ECG signal, which also includes the patient's blood pressure and heart rate. Cai et al. and Puurtinen et al. [10], [12] employed DWT and TS algorithms to classify emotions, while Golgowski et al. [10] investigated wavelet modification for faulty heart detection. The PPG signal captures information on blood volume changes, blood flow, and a patient's regular heartbeat by using light reflected from the blood vessels. Rakshit et al. [13] successfully used HRV extracted from PPG signals to categorise subjects' emotions. Lee et al. [14] performed rapid emotional classification by simply pre-processing short-length PPG data using a 1D convolutional neural network (CNN) model. Since the GSR signal includes information on perspiration and body temperature shifts in response to changes in mood, it is strongly linked to the physical domain [15]. In [11], Ganapathy et al. created a CNN model to extract 38 characteristics via Fourier transformations; in [10], Susanto et al. used a 1D CNN and residual bidirectional GRU to categorise emotions.

A technique for matching statistical appearance models with photographs was proposed by Cootes et al. [15]. By using the learnt connection between mistakes in model parameters and those in results, the technique develops a robust iterative matching process that controls the shape and gray-level changes learned in the training set. To classify human facial expressions, Phavish et al. looked into applying deep learning with CNN models. Facial landmarks and contours were identified using an edge detection framework, and facial expressions were classified using a stochastic classifier [08]. Collecting facial expression data via images and videos is expensive, and some individuals worry that their privacy may be invaded in the process.

Data complexity increases owing to age, geography, and language differences in how voice signals are conveyed. Attributes need to be standardised and pre-processed before feature extraction may proceed. Speech-based emotion classification using acoustic features was performed, speech signals' prosodic and spectral characteristics were analysed, and a two-stage hierarchical classifier was used for speech recognition and classification in a study by Albornoz et al. [6]. Eyben et al. developed an LSTM-RNN model for continuous audio data that does not require emotion recognition or window slicing. Using feature extraction, we combined linguistic and acoustic data in our previous work [15]. When compared to their visual counterparts, voice-based emotion recognition and classification provide significantly fewer privacy concerns. But this doesn't help those who have problems communicating verbally or in settings where people's movements are limited, including libraries, construction sites, or underwater habitats.

Since the ANS's bio-signals are electrical impulses transmitted by the body's muscles and bones, they are relatively untouched by their surroundings. Therefore, problems with facial expression and voice signals can be avoided, and technology that incorporates extra bio-signals shows strong emotion recognition and classification performance and carries out a variety of functions [7], [7]. In contrast to the user-controllable SNS, the ANS is able to accurately detect and classify internal emotions by means of uncontrollable impulses. However, the volume and quality

of the learning data are constrained when a single biosignal is employed for internal emotion recognition.

Therefore, it has been suggested that a multimodal emotion recognition and classification system based on several bio- signals be used to improve accuracy and efficiency. Hui et al. presented a decision recognition fusion model that incorporate emotion detection consistency across dimensions. Preprocessing and feature extraction for non-invasive EEG and ECG signals were described in [6]. The signals were merged, and then emotion recognition was carried out using a probabilistic neural network (PNN). The results of the experiments showed that the combined signals were more effective than the individual ones. Yang et al. extracted EEG and PPG data into 11 statistical time zones and 5 frequency zones using a one-dimensional convolutional neural network, and then classified the signals based on their emotional content. In a previous study, we calculated the pulse transit time (PTT) between the ECG and PPG data to further improve the final classification accuracy [7]. A signal latency of 10 s was required for feature extraction and 60 s was maximum for emotion recognition. A signal latency of more than 10 s makes studies that mix several inputs (images, audio, etc.) inappropriate for real-time emotion recognition systems.

The categorization and regression of PPG and GSR signals was investigated by Ayata et al. using a variety of techniques, including ensemble learning, random forest, k-nearest neighbours (KNN), support-vector machines, and others. Using GSR and PPG signals lasting only 3 and 8 seconds, respectively, they were able to extract features with an arousal accuracy of 72.06 percent and a valence accuracy of 71.05 percent [10].It was also found that PPG and GSR signals are linked when people are in different emotional states. Heart rate changes in PPG signals were analysed by Lee et al. using both post- normalization frequency domain features and normal to normal (NN) temporal domain characteristics. Arousal and valence were extracted from a 10-second PPG signal with 82.1% and 80% accuracy, respectively [2].Kim and Andre [1] proposed a method for identifying emotions from electrical activity in the muscles, the heart, the breath, and the skin. Mathematical approaches like as entropy, time-frequency indices, spectral measurements, and geometric analysis were used to determine the most salient features. There were two types of data classification used by them: by subject and by actual data.

In the two examples provided, the highest possible categorization rates were 95% and 70%, respectively. The methods of k nearest neighbour (KNN), support vector machine (SVM), and least squares were evaluated for their ability to identify emotions based on EEG data by Duan et al. [5]. Features were collected from the smoothed EEG power spectrum, and then dimension reduction was done using minimal redundancy maximal relevance (MRMR) and principal component analysis (PCA).For a long time, researchers in the behavioural and social sciences have struggled with how to define and categorise different types of emotional states. Both the discrete model proposed by Ekman[8] and the two-dimensional continuous model proposed by Lang [9] are used often in research on emotion recognition.

To improve emotion categorization algorithms, the discrete emotional model has been widely used in this study. The discrete emotional model recognises the individuality of basic emotions and treats them accordingly. These four emotions encompass every other emotion. Our study took into account the six fundamental emotions recognised by Ekman et al. [8]: joy, sadness, surprise, anger, disgust, and fear. convolutional neural networks (CNN)

Materials and methods

A. *Electrocardiogram (ECG)*

This work presents the Asian Affective and Emotional State (A2ES) Database, which includes electrocardiogram (ECG) and polygraphic (PPG) readings from 47 people of different Asian backgrounds. Reported problems with ECG and PPG recordings due to subject's skin tone exist [31,33]. To monitor the heart's electrical activity, doctors utilise electrocardiograms [34]. This activity begins in the sinoatrial node and leads to muscular contractions that pump blood throughout the body. As indicated in Figure 1, the ECG has three basic components: P wave, QRS wave, and T wave. In contrast, PPG is an inexpensive and noninvasive method of measuring alterations in blood volume in response to cardiac activity in humans.

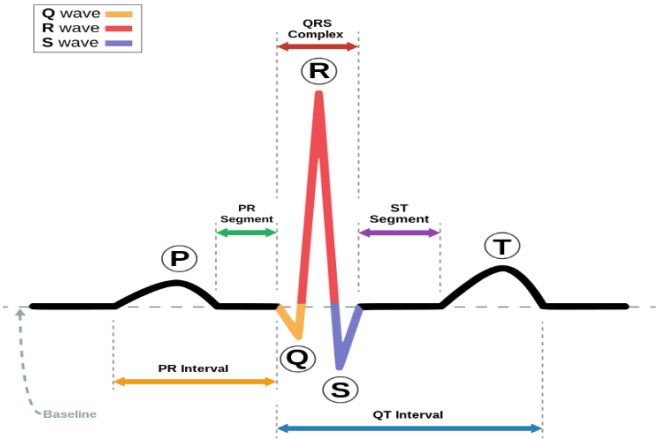

Fig. 2. PPG signal element

An incoherent light source and a photoreceiver are the two fundamental parts of a PPG system Figure 2 and Figure 3 depicts a typical PPG signal element, complete with systolic and diastolic periods (representing the influx and relaxation of blood, respectively) and a dicrotic notch (representing the reflection of the pulse) Twenty-five audio-visual stimuli were played for the data subjects, each designed to elicit a different emotion.

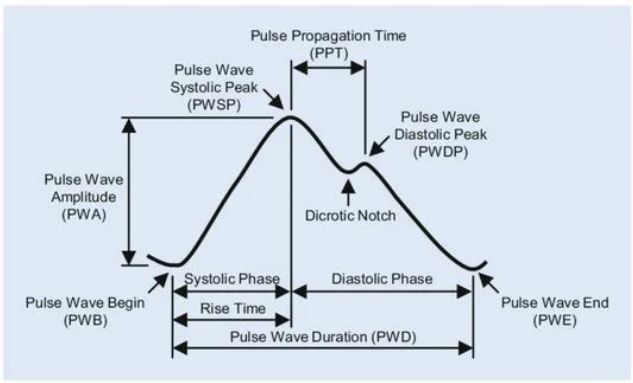

Fig.3. Typical transmission of a PPG signal with the systolic period

B. *GSR*

Time-frequency domain signal feature extraction and data pre-processing are common steps in traditional bio-signal- based emotion recognition and classification methods. Conventional methods, on the other hand, are too complex to be used with a real-time system and depend on extra manual feature extraction. In this study, Figure 4 we provide a real- time emotion categorization system that uses PPG and GSR data for user convenience, along with a labelling strategy that takes into account the bio-signal features. The proposed method consists of three main parts: weight-reduced model learning for emotion classification, training data labelling,

and data pre-processing using window slicing and waveform duplication and loss avoidance. Essentially, a multimodal neural network model uses the window of PPG and GSR data split into a single pulse unit as input, with no further manual processing involved.

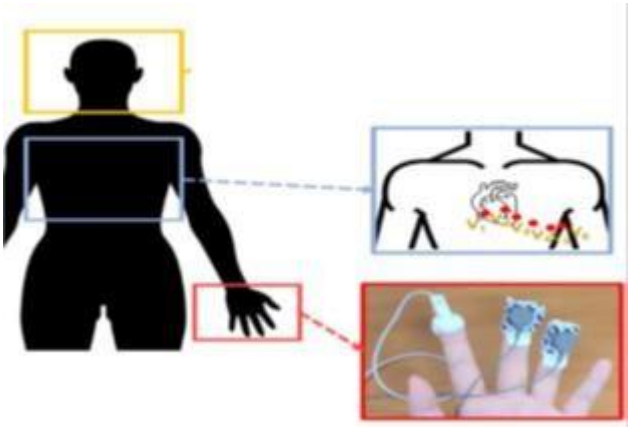

Fig 4. Actual acquisition method and location according to bio-signal with ECG and GSR

Convolutional neural networks, often known as CNNs or ConvNets, are a particular kind of neural network that excels in processing input that has a grid-like design, such as images. A digital image is a binary representation of visual data. It consists of a grid-like arrangement of pixels, with a pixel value assigned to each one to specify its colour and brightness. The human brain processes a vast amount of information as soon as it detects a picture. Each neuron has a unique receptive field and is connected to other neurons to form a network that spans the whole visual field.

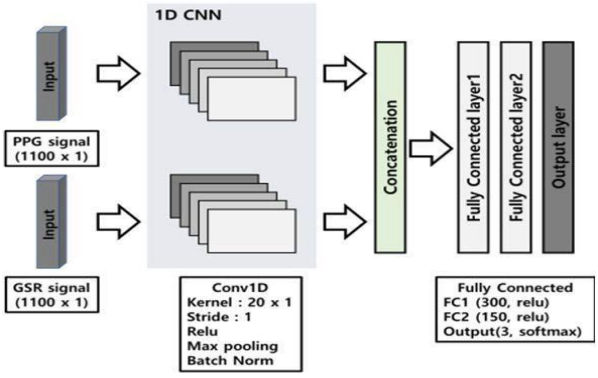

Fig 5. 1D CNN MODEL

Similar to how each neuron in the biological vision system responds to stimuli only in the limited region of the visual field known as the receptive field, each neuron in a CNN processes data only in its receptive field. The layers first identify basic patterns, such as lines and curves, then more complex patterns, such as faces and objects. Using a CNN shows in Fig 5, one can grant computers sight.

Architectural Design

Using the Shimmer sensors, the human body's natural ECG signals may be read out. The signals are scaled down to two-dimensional Scalograms, which are frequency spectra. These scalograms served as input to the deep learning algorithm (CNN), which classifies based on the input scalograms into seven basic human emotions (happy, neutral, sad, fear, anger, surprise and disgust). The process flow for the suggested approach is depicted in Fig. 6.

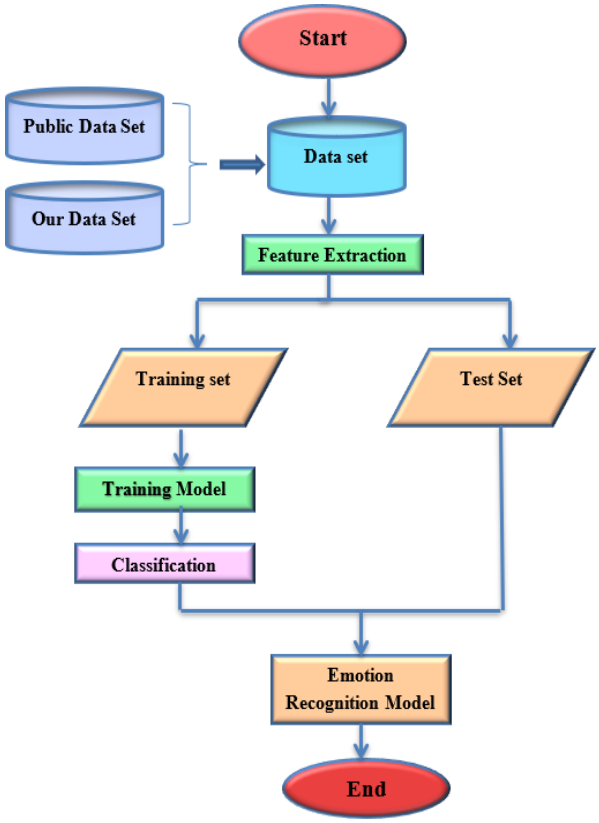

Fig 6. Architectural Design

Data Collection and preprocessing

A. Real Time Data collection

Fig.7 shows that the GSR sensor connected and sample data set collected from real time environment.

Fig 7. AVC Data Set –ECG and GSR

B. Data Preprocessing

Even though raw PPG signal data is more practical and usable than ECG readings, it still requires cleaning to remove high- frequency noise from the power source and low-frequency noise from the capillaries. The linear frequency-domain filter with the simplest mathematical expressions, the

Butterworth filter, was chosen for this study. The dynamic noise was reduced using a moving average filter, a noise in the data baseline, and a high-order polynomial fitting. The GSR signal went through the same preliminary processing steps as before.

Butterworth Filter

The basic formula for filter approximation is

$$H(j\omega) = \frac{K}{\sqrt{1 + \varepsilon^2 (\omega^2)}} \quad (1)$$

POLYNOMIAL CURVE FITTING

The Butterworth filter is a frequency filter with a smooth transition band and a flat passband amplitude spectrum. The transition band is smaller and more abrupt with higher order N in the simple linear frequency filter, which is stated as follows.

Nth-order polynomial equation:

$$y(x, w) = \sum_{j=0}^{N} W_j x^j \quad (2)$$

Table 1 shows PPG and GSR signal segment label distribution.

Table 1. PPG and GSR Signal Segment

		Annotation labeling	Self- assessment labeling
Number of Data extracted		32000 Segments	
Label (Arousal)	high	14,489	2
	neutral	11,190	2
	low	6,321	1
Label (Valence)	positive	6,716	1
	Neutral	5,289	1
	negative	14,227	3

Result And Discussion

A. Performance Measure

The performance of a model is evaluated using a variety of metrics, including accuracy, loss, precision, recall, and F- score shown in the Fig.9. The following will define these measures.

Accuracy

$$\frac{A + B}{A + B + C + D} \quad (3)$$

Precision

$$\frac{A+B}{A+B} \tag{4}$$

Recall

$$\frac{A}{A+C} \tag{5}$$

F1 Score

$$F1\ Score = \frac{2 * precision * recall}{precision + recall} \tag{6}$$

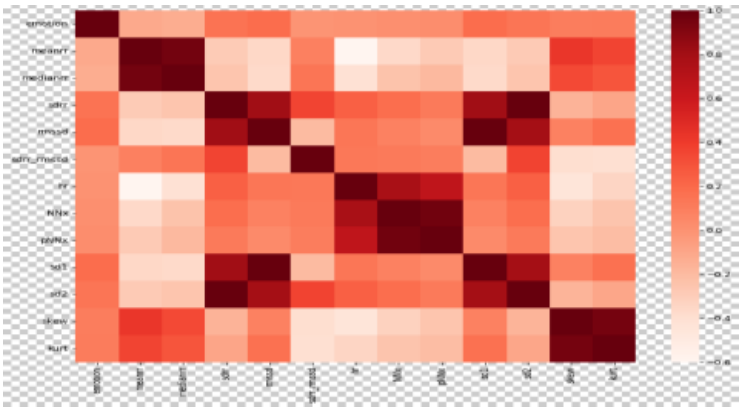

Fig. 9. Performance distribution

B. Output

Table.2 shows the real time result for emotion recognition and Figure 10 shows graphical representation of signals.

TABLE II. Real Time Result

Set	Target Emotion	Happy	Sad	Anger	Surprise	Disgust	Neutral	Fear
1	Happy/Surprise/Neutral	0.01	0.04	0.01	89.1	0.00	10.8	0.00
2	Sad/Anger/Disgust	0.58	89.6	0.5	0.15	0.33	0.19	1.6
3	Sad/Anger/Disgust	0.43	0.22	99.2	0.00	0.00	0.05	0.01
4	Sad/Fear/Disgust	1.5	0.04	0.01	89.1	0.8	93.4	0.00
5	Happy/Neutral	50.0	49.3	0.03	0.50	0.04	0.17	0.00
6	Neutral/Anger	1.5	1.34	72.9	0.9	0.01	0.3	22.7
7	Sad/Disgust/Anger	5.92	0.29	42.4	0.02	0.01	51.3	0.00
8	Sad/Surprise	11.8	46.9	0.83	22.2	1.17	17.0	0.69
9	Disgust/Anger	22.7	1.4	72.3	0.00	0.04	0.7	2.7

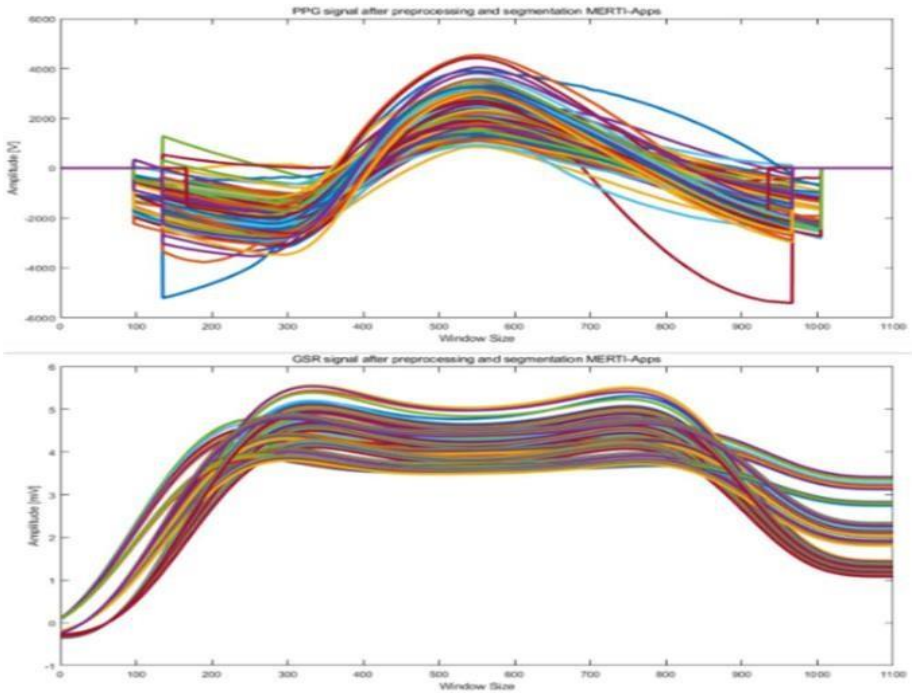

Fig. 10 . Final PPG, GSR signal output (sample size 1100); (a) Results of the PPG signal; and (b) results of GSR signal

TABLE III. AVC-GSR Dataset Result

Sample No.	GSR Sensor Value	Conductive Voltage	Emotion Expression
Sample 1	500-598	1.72-1.9	**Smile**
	400-490	3.27-3.74	**Neutral**
	700-798	4.10-4.25	**Excitement**
Sample 2	300-350	1.72-1.9	**Sad**
	727-789	3.27-3.74	**Neutral**
	855-900	4.10-4.25	**Happy**

Table 3 shows result of Figure 7 real time dataset collected. Our proposed CNN has achieved 80% of accuracy compared to other algorithms shown in Fig.11.

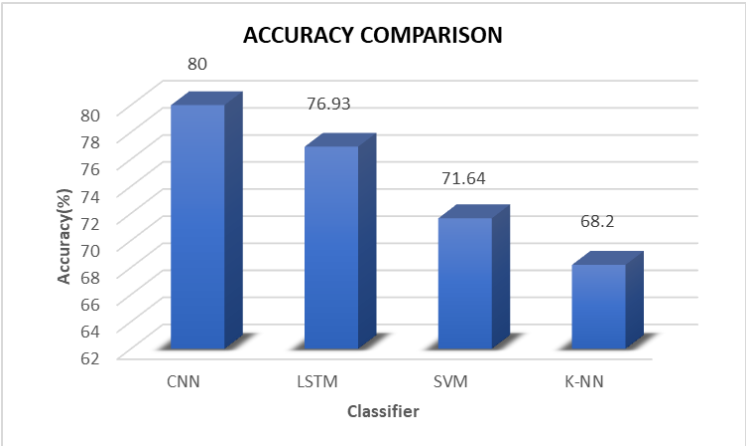

Fig. 11 . Accuracy comparison

Conclusion

Physiological signals represent a person's internal emotional state in a way that cannot be easily altered, the use of physiological signals as a method for detecting emotions is regarded as the method that is the most dependable and authentic. This is because it cannot be easily manipulated. In this thesis, we provided a solution that is based on ECG physiological signals, and it is designed to recognise seven main emotions: joyful, fearful, sad, angry, disgusted, and surprised. The categorization was carried out with the assistance of a Convolutional Neural Network utilising an AlexNet architecture. The signals were first transformed into a scalogram before being fed into the Convolutional Neural Network (CNN). The overall accuracy that we have achieved with the AMIGOS dataset is 93%, and the accuracy that we have achieved with the real-time dataset is 68.5%. In the future, with more advanced sensors for the collecting of physiological signals, the application of a variety of other deep learning algorithms could potentially lead to improvements in the obtained results.

References

[1] R.BA (ed) (1995).” Substance Abuse and Mental Health Statistics Sourcebook” (DHHS Publication No. SMA 95-3064). Washington, DC, U.S. Government Printing Office

[2] B.Ahmed, “Depression and anxiety: a snapshot of the situation in Pakistan,” International Journal of Neuroscience and Behavioral Science, February 2016

[3] N.Bach,”World Mental Health Day 2017: Illness in the Workplace Is More Common Than You Might Think”,[Online].Available:

[4] A.Greco, “Affective computing in virtual reality:emotion recognition from brain and heartbeat dynamics using wearable sensors,” SCIENTIFIC REPORTS ,September 2018.

[5] E.Andre, “Emotion Recognition Based on Physiological Changes in Music Listening,” IEEE Transactions on Pattern Analysis and Machine Intelligence,January 2009

[6] P.Tzirakis, G.Trigeorgis, M.A.Nicolaou, B. W. Schuller, and S. Zafeiriou,”End-to-end multimodal emotion recognition using deep neural networks”, IEEE Journal of Selected Topics in Signal Processing, 11(8):1301–1309, 2017

[7] H.Ferdinando, T.Seppanen, and E.Alasaarela,” Enhancing emotion recognition from ecg signals using supervised dimensionality reduction”, In ICPRAM, p.112–118, 2017

[8] Y.Xu and G.Yuan Liu,”A method of emotion recognition based on ecg signal. In Computational Intelligence and Natural Computing”, 2009.CINC'09, International Conference on, vol. 1, p. 202-205, IEEE,2009

[9] S.Katsigiannis and N.Ramzan,”Dreamer: a database for emotion recognition through EEG and ECG signals from wireless low-cost o_- the-shelf devices”, IEEE journal of biomedical and health informatics, 2018.

[10] Busso, C.; Deng, Z.; Yildirim, S.; Bulut, M.; Lee, C.M.; Kazemzadeh, A.; Lee, S.; Neumann, U.; Narayanan, S. Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information. In Proceedings of the 6th International Conference on Multimodal Interfaces, Sorrento, Italy, 25–29 November 2012; pp. 205–211

[11] Tivatansakul, S.; Ohkura, M. Improvement of emotional healthcare system with stress detection from ECG signal. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Milan, Italy, 25–29 August 2015.

[12] Minhad, K.N.; Ali, S.H.M.D.; Reaz, M.B.I. A design framework for human emotion recognition using electrocardiogram and skin conductance response signals. J. Eng. Sci. Technol. 2017, 12, 3102–3119

[13] Alemi, O.; Li, W.; Pasquier, P. Affect-expressive movement generation with factored conditional Restricted Boltzmann Machines. In Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction, ACII, Xi'an, China, 21–24 September 2015

[14] Lang, P.J.; Bradley, M.M.; Cuthbert, B.N. International Affective Picture System (IAPS): Instruction Manual and Affective Ratings; University of Florida: Gainesville, FL, USA, 2005.

[15] Mahmoodabadi, S.Z.; Ahmadian, A.; Abolhasani, M.D.; Eslami, M.; Bidgoli, J.H. ECG feature extraction based on multiresolution wavelet transform. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology, Shanghai, China, 17–18 January 2006; pp. 3902–3905.