electronics

MDPI

Article

Emotion Classification Based on CWT of ECG and GSR Signals Using Various CNN Models

Amita Dessai * and Hassanali Virani

Department of Electronics and Telecommunication Engineering, Goa College of Engineering, Goa University, Ponda 403401, Goa, India

* Correspondence: amitachari@gec.ac.in

Abstract: Emotions expressed by humans can be identified from facial expressions, speech signals, or physiological signals. Among them, the use of physiological signals for emotion classification is a notable emerging area of research. In emotion recognition, a person's electrocardiogram (ECG) and galvanic skin response (GSR) signals cannot be manipulated, unlike facial and voice signals. Moreover, wearables such as smartwatches and wristbands enable the detection of emotions in people's naturalistic environment. During the COVID-19 pandemic, it was necessary to detect people's emotions in order to ensure that appropriate actions were taken according to the prevailing situation and achieve societal balance. Experimentally, the duration of the emotion stimulus period and the social and non-social contexts of participants influence the emotion classification process. Hence, classification of emotions when participants are exposed to the elicitation process for a longer duration and taking into consideration the social context needs to be explored. This work explores the classification of emotions using five pretrained convolutional neural network (CNN) models: MobileNet, NASNetMobile, DenseNet 201, InceptionResnetV2, and EfficientNetB7. The continuous wavelet transform (CWT) coefficients were detected from ECG and GSR recordings from the AMIGOS database with suitable filtering. Scalograms of the sum of frequency coefficients versus time were obtained and converted into images. Emotions were classified using the pre-trained CNN models. The valence and arousal emotion classification accuracy obtained using ECG and GSR data were, respectively, 91.27% and 91.45% using the InceptionResnetV2 CNN classifier and 99.19% and 98.39% using the MobileNet CNN classifier. Other studies have not explored the use of scalograms to represent ECG and GSR CWT features for emotion classification using deep learning models. Additionally, this study provides a novel classification of emotions built on individual and group settings using ECG data. When the participants watched long-duration emotion elicitation videos individually and in groups, the accuracy was around 99.8%. MobileNet had the highest accuracy and shortest execution time. These subject-independent classification methods enable emotion classification independent of varying human behavior.

Keywords: emotions; AMIGOS; ECG; GSR; CWT; scalogram; CNN

Citation: Dessai, A.; Virani, H. Emotion Classification Based on CWT of ECG and GSR Signals Using Various CNN Models. *Electronics* **2023**, *12*, 2795. https://doi.org/10.3390/electronics12132795

Academic Editor: Elena Carlotta Olivetti

Received: 31 March 2023
Revised: 3 June 2023
Accepted: 14 June 2023
Published: 24 June 2023

1. Introduction

Emotion intelligence is a trending area of research that explores the detection of emotions using suitable machine-learning techniques. Emotion recognition based on facial expressions and speech signals does not indicate genuine emotions, since people can misinterpret such signals. However, biological parameters such as ECG, electroencephalogram (EEG), GSR, respiration rate, and skin temperature do not lead to false emotion recognition since they cannot be misinterpreted. Much research is carried out on emotion classification using EEG, but it is more suitable for clinical applications. Using ECG and GSR for emotion classification is the most suitable for the naturalistic environment of participants. Emotion elicitation can be done using pictures, sound, and videos, and virtual reality has become a new technique.

Self-assessment of emotions can be done by participants using a set of questionnaires and self-assessment reports. Using Russell's circumplex model, emotions can be separated into two dimensions, valence and arousal [1], and four discrete classes: high valence, low arousal (HVLA), low valence, high arousal (LVHA), high valence, high arousal (HVHA), and low valence, low arousal (LVLA). Emotions such as happiness exhibit HVHA, anger exhibits LVHA, sadness indicates LVLA, and calmness indicates HVLA [2–4]. The COVID-19 pandemic increased negative emotions such as fear, anxiety, and stress in people worldwide. A study proved that the lockdown period increased the recognition of negative emotions and decreased the identification of positive emotions, such as happiness [5]. It was necessary to detect people's emotions during the pandemic in order to trigger the necessary actions to achieve a proper societal balance. ECG and GSR signals enable the classification of emotions using deep learning or machine learning classifiers using non-intrusive techniques [6,7]. Moreover, databases are available for people to conduct research on emotion classification based on physiological signals [8–11]. However, studies have yet to consider factors such as ethnicity, age, the use of virtual reality to elicit emotions, and the naturalistic environment of the subjects. A new database could be made available for research on emotion classification using ECG and GSR data considering the following factors:

- Emotions can be elicited using virtual reality, which resembles the real environment.
- Participants with diverse cultural backgrounds and in different age groups who are not familiar with one another can be evaluated.
- ECG and GSR recordings can be captured when participants are in a naturalistic environment using smart bands.

Biological signals such as ECG, GSR, and respiration rate aid in deriving information about people's emotions. ECG is the most commonly used approach since smart devices can quickly obtain heartbeat information. Once emotions are detected, proper counseling can be recommended. ECG and GSR signals represent the time domain variation of voltage [12,13]. The amplitude and time duration changes of ECG and GSR waveforms enable us to discriminate among the various emotions [13]. Emotions depend on people's personality and moods. Subject-independent classification models can enable the detection of emotions regardless of the subject [14]. Raw ECG and GSR signals are contaminated with noise and need to be suitably filtered. Signal processing and feature extraction techniques extract the appropriate data from ECG and GSR signals for emotion classification. Continuous wavelet transform (CWT) analyzes the frequency content of these signals and transforms time-domain signals into the time-frequency domain, and 1D signals can be transformed into a 2D matrix [15]. Various researchers have used machine learning techniques for emotion classification, but deep learning techniques can improve classification accuracy. Various pretrained CNN models, such as MobileNet, NASNetMobile, DenseNet 201, InceptionResnetV2, and EfficientNetB7, can be used for emotion classification based on the transfer learning approach.

The main research contributions of this work are described below.

The duration of the emotional stimulus period and the social and non-social context of the participants influence the classification of emotions, hence these factors need to be considered. Other studies on emotion classification using ECG signals focused on short-duration emotion stimulus periods and participants experiencing emotions individually. This work explores a novel classification of emotions considering the social context of individuals, using deep learning techniques and a longer duration of emotional stimulus. We also propose a filtering technique to eliminate the noise from ECG and GSR signals and a continuous wavelet transform technique to convert the signals into time–frequency scalograms, and further classify emotions using several pretrained CNN models: MobileNet, NASNetMobile, DenseNet 201, InceptionResnetV2, and EfficientNetB7. Deep learning models with automatic feature extraction have proved to be effective at increasing the accuracy of classification. The architectural advancements of the latest CNN models considerably help to improve the accuracy. Other studies have extracted CWT coefficients from ECG data and classified emotions using traditional machine learning techniques [16],

but they have yet to explore deep learning techniques for classifying emotions. Thus, this work proposes a novel algorithm for emotion classification in two cases:

- ECG and GSR recordings were obtained while participants watched short videos. Emotions were classified using the novel approach of obtaining CWT coefficients and classifying based on deep learning models.
- ECG recordings were obtained while participants watched long-duration videos individually and in groups, and emotions were classified using the novel technique.

The primary objective of this study was to increase the accuracy of emotion classification based on ECG and GSR data, as well as to explore emotion classification based on groups of participants undergoing longer-duration emotion elicitation using 2D CNN models. However, challenges arose when using the 2D CNN models on the numerical ECG and GSR recordings of the AMIGOS dataset. To address this problem, scalograms of CWT coefficients of biosignals were derived and converted into images to enable classification using the 2D CNN models. This research provides a valuable contribution to the field of human–computer interaction based on emotional intelligence.

The paper is structured as follows: Section 2 provides a review of the literature. Section 3 gives the methodology followed in this work and describes the various techniques used. Section 4 presents the results, and Section 5 provides the conclusion.

2. Related Works

This section gives an overview of the work done by other researchers using ECG, GSR, CWT, machine learning, and deep learning techniques. In some studies, the time and frequency domain features were extracted from the signals, and machine learning classifiers such as support vector machine (SVM), random forest (RF), and k-nearest neighbor (KNN) were used for classifying emotions [17–19]. However, peaks from ECG and GSR signals need to be detected from pre-processed signals, and handcrafted features need to be extracted. Using a CNN-based deep learning approach with automatic feature extraction simplifies the classification task. Dessai et al. analyzed studies on emotion classification using ECG and GSR signals [20]. They concluded that deep learning techniques using automatic extraction of features increased the classification accuracy compared to traditional machine learning techniques.

Deep learning models such as CNN, long short-term memory (LSTM), and recurrent neural network (RNN) are also used for classifying emotions [20]. Al Machot et al. obtained classification accuracy of 78% using the MAHNOB database and 82% accuracy using the DEAP database on GSR data using a CNN classifier [14]. Dar et al. used a CNN-LSTM model to classify emotions based on ECG and GSR data from the DREAMER and AMIGOS databases. They achieved accuracy of 98.73% for ECG data and 63.67% for GSR data from the AMIGOS dataset [21]. Santamaria-Granados et al. pre-processed ECG signals from the AMIGOS dataset and segmented them to a length of 200 R peaks, then used a 1D CNN for classification. They obtained accuracy of 76% and 75%, respectively, for arousal and valence using ECG data and 71% and 75% using GSR data [22]. After suitable pre-processing, the 1D CNN was used to classify photoplethysmography (PPG) data from the DEAP database, resulting in 75.3% and 76.2% accuracy for valence and arousal. A normalization technique was used to normalize the data from PPG pulses to avoid variations in the data from different people [2]. Hammad et al. extracted spatial and temporal features in parallel. The features were fused, and the grid search technique was used for optimization to achieve higher classification accuracy using CNN, resulting in accuracy of 97.56% for valence and 96.34% for arousal [23]. Lee et al. used Pearson's correlation method to select statistical features combined with automatically mined features from CNN and improved the classification accuracy [24]. Harper et al. used a probabilistic procedure to select the output, and achieved peak classification accuracy of 90% [25]. Numerical ECG data were converted into images, and supervised machine learning techniques were used to classify emotions [26]. Sepúlveda et al. extracted features from ECG data of the AMIGOS database using CWT and classified emotions using machine learning techniques. They

determined that wavelet scattering with an ensemble resulted in accuracy of 89%, and the KNN classifier attained 82.7% for valence and 81.2% for arousal [16].

Emotion classification has also been done by using textual data [27], speech signals [28], and facial expressions [29–31]. However, physiological parameters cannot be falsified, and this represents a new dimension for research in human–computer interaction. This is mainly due to the development of advanced sensors [32]. Researchers have extracted CWT coefficients of ECG data for applications such as arrhythmia and biometric classification. Wang et al. extracted CWT coefficients from the MIT-BIH arrhythmia database to detect arrhythmia based on ECG data. They attained accuracy of 98.74% using a CNN classifier [33]. Byeon et al. obtained CWT coefficients for classifying biometric parameters using ECG data. Images based on scalograms were used as inputs for AlexNet, GoogLeNet, and ResNet deep machine learning classifiers, and their performance was tested [34]. Mashrur et al. carried out arrhythmia detection by obtaining CWT coefficients from ECG data using the pre-trained AlexNet model for classification [35]. Aslan et al. converted electroencephalogram (EEG) data from a database for emotion recognition based on EEG signals and various computer games (GAMEEMO) into frequency domain features using CWT. They plotted scalograms and converted them into images, used the pre-trained GoogLeNet deep learning model to classify emotions, and achieved 93.31% accuracy [36]. Garg et al. classified emotions based on EEG data using the SelectKBest feature selection technique and traditional machine learning techniques [37]. Alsubai et.al.classified emotions using a deep normalized attention-based neural network for feature extraction using EEG data [38].

The impact of user engagement on purchase intentions has also been analyzed based on EEG data [39]. In addition, emotion classification based on the EEG data of participants while listening to music is an emerging area of research [40].

We found that researchers have used various machine learning techniques for emotion classification. Researchers have also used the scalogram method based on CWT coefficients using ECG data for other applications, such as biometrics and arrhythmia classification, with supervised and deep machine learning classifiers proving to be efficient tools for ECG data. Moreover, studies in the literature explored emotion classification with a shorter duration of emotion elicitation and when the emotions are experienced individually, using traditional machine learning and 1 D CNN techniques. However, people experience emotions in groups while watching movies or playing games, or in the classroom environment. Hence, the emotions of participants need to be classified considering the group setting. This study explored emotion classification considering the group setting using the proposed model.

3. Methodology

In this study, emotions are classified based on Russell's two-dimensional circumplex model, as illustrated in Figure 1. Valence indicates the pleasantness of emotions and arousal indicates the intensity of emotions. For example, calmness indicates low arousal and high valence [41].

The methodology for classifying emotions using the CWT and deep learning framework is shown in Figure 2. ECG recordings corresponding to individuals and groups watching short-duration and long-duration videos from the AMIGOS dataset were used for emotion classification [11], and GSR recordings corresponding to short-duration videos were used to classify emotions. The scalogram obtained from the CWT coefficients was converted into an image and fed to the CNN classifier.

3.1. Data Description

The ECG signals were recorded using the Shimmer 2R5 platform by attaching the electrodes to the person's right and left arm and the reference electrode to the left ankle. Signals were sampled at 256 Hz sampling frequency with 12-bit resolution [11]. Figure 3 indicates the ECG signal waveform [12].

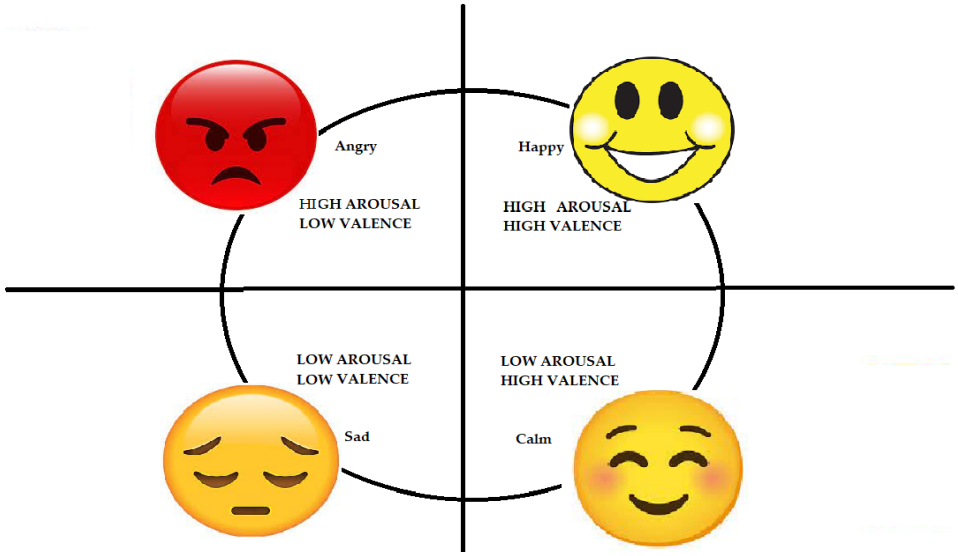

Figure 1. Russell's circumplex model.

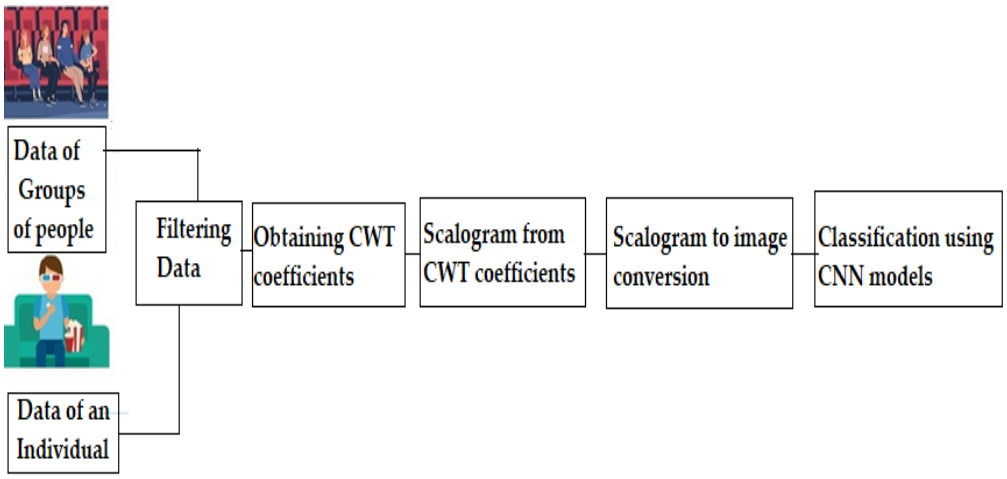

Figure 2. Block diagram.

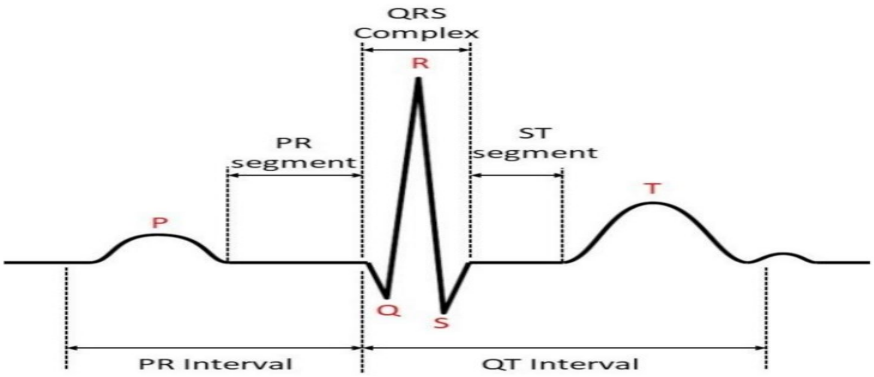

Figure 3. ECG signal [12].

GSR signals were captured using the Shimmer 2R with a sampling frequency of 128 Hz and 12-bit resolution, with two electrodes placed at the middle of the left middle and index fingers, as indicated in Figure 4 [11,20].

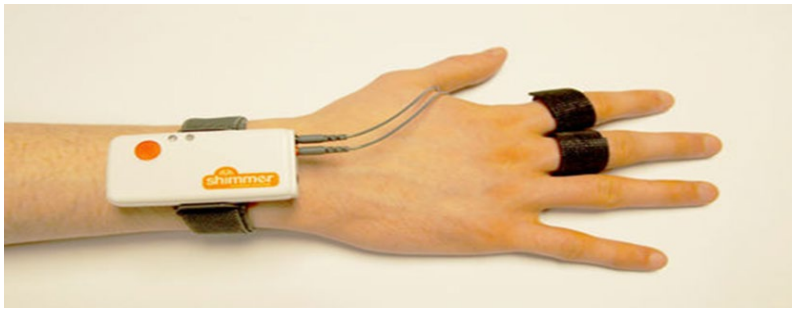

Figure 4. GSR sensor [20].

The AMIGOS database is a publicly available database that considers the individual or group setting of participants and emotion elicitation for longer duration in classifying emotions [11]. ECG and GSR recordings were acquired of 40 healthy participants aged 21 to 40 years (13 female) while watching 16 short videos (less than 250 s) and 4 long videos (16 min) [11]. Studies on emotion classification are based on ECG recordings when participants are watching short-duration videos for a few minutes, whereas emotions are experienced mainly in groups. In addition, playing games, watching movies, and classroom activities, among other activities, are also done in groups.

In this work, we utilized ECG and GSR recordings from the AMIGOS database for two cases, as follows:

3.1.1. Case 1: Valence and Arousal Classification with Short Videos

A multimodal emotion classification model based on ECG and GSR signals is more reliable than a model based on a single sensor [42].

While subjects are watching short videos, the ECG and GSR recordings can be used to classify their emotions based on the valence arousal scale. The ECG and GSR recordings of 40 participants watching short videos were obtained [11]. However, only the recordings of the first 17 participants were selected in this work to establish a comparison with case 2 below. Table 1 shows details of the short-duration videos from the AMIGOS dataset.

Table 1. Short video details.

Video Number	Category	Source Dataset	Source Movie	Video Duration (Minutes)
1	HVLA	DECAF	August Rush	01:30
6	LVLA	DECAF	My Girl	01:00
8	LVHA	MAHNOB	Silent Hill	01:10
12	HVHA	DECAF	Airplane	01:25

The short-duration videos were selected to elicit particular affective states in the subjects. In the table, the video number is listed in the first column, and the quadrant of Russell's model is given in the second column, indicating the emotional arousal and valence levels. For ECG valence classification, data corresponding to videos 6 and 12 were considered. For ECG arousal classification, data corresponding to videos 1 and 12 were considered. For GSR arousal classification, recordings of 39 participants were considered. For low arousal, data corresponding to videos 1 and 6 were considered, and for high arousal, videos 8 and 12 were considered. Similarly, for GSR valence classification, low valence data corresponding to videos 6 and 8, and high valence data corresponding to videos 1 and 12 were considered.

3.1.2. Case 2: Valence and Arousal Classification with Long Videos

ECG recordings of 17 participants watching long videos individually and in groups were obtained. Table 2 shows details of the long-duration videos from the AMIGOS dataset [11].

Table 2. Long video details.

Video Number	Category
17	LVHA
18	HVLA
19	LVLA
20	LVLA

The long videos ran for more than 14 min, and they provoked emotions according to the diverse quadrant of Russell's model. The long-duration videos could elicit various affective states over time, resulting in a context for the affective states [11]. For individual valence classification, data corresponding to videos 18 and 19 were considered. For individual arousal classification, data corresponding to videos 17 and 20 were considered.

ECG recordings of participants watching long videos in groups were also captured. Groups of people familiar with each other and with the same cultural background were formed [11]. Five groups with four individuals each watched the long videos in order to explore the group setting. ECG recordings of only 17 participants were taken to maintain uniformity in the number of samples for both cases. For group arousal classification, data corresponding to videos 17 and 20 were obtained. For group valence classification, data corresponding to videos 18 and 19 were used.

3.2. Pre-Processing

The raw ECG signal amplitude measured from the electrodes on the right arm and left leg versus time was plotted using MATLAB software, as shown in Figure 5.

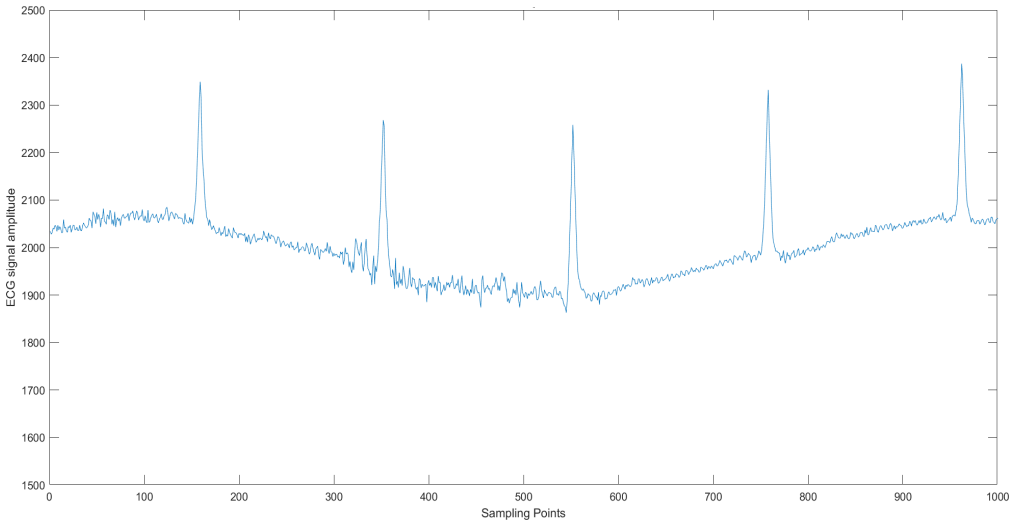

Figure 5. ECG signal.

Similarly, the raw GSR signal amplitude measured from the electrodes placed on fingers versus time was plotted using MATLAB software, as shown in Figure 6. Increased sweat content results in a large flow of current, thus increasing the conductance [20].

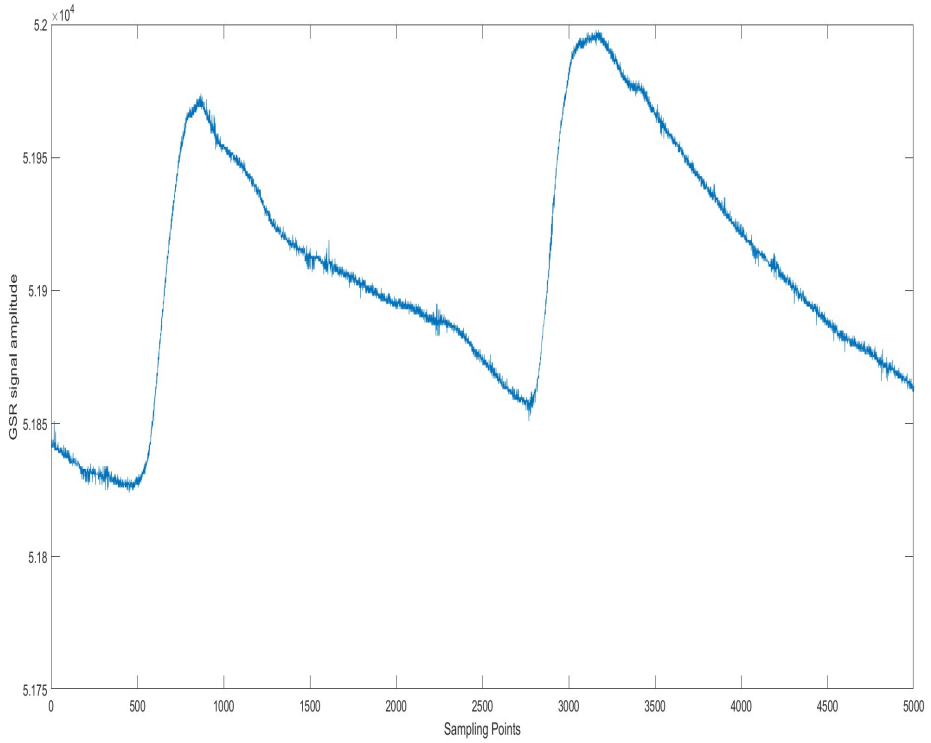

Figure 6. GSR signal.

3.2.1. Filtering

Raw ECG and GSR signals are prone to various noises, such as the motion of the electrodes, muscle artifacts, baseline drift, and power line interference, hence it becomes necessary to eliminate such noises using suitable filtering techniques. Missing data of the participants were eliminated, and the signals were suitably filtered with the sixth-order low-pass Butterworth filter with 15 Hz cut-off frequency and high-pass filter with 0.5 Hz cut-off frequency [13].

3.2.2. Segmentation

Each ECG recording used in our work had 16,200 sampling points. If a lengthy signal is passed through the deep learning framework, it could result in degradation, hence the ECG recording has to be suitably segmented. The segmentation requires extracting a complete cycle of the ECG signal from the ECG waveform. In our study, no algorithm was used for detecting the QRS complex; instead, the annotated R-peak location was the reference point to segment the signal. As shown in Figure 5, R peaks are predominantly noticed in the signal, separated by a duration of fewer than 200 sampling points. Hence, the signal is segmented using a fixed window size of 200, which would generate around 81 segments per participant.

GSR signals with 8096 sampling points were sampled at intervals of 2024 samples. The peaks in the signal are the reference points, as observed in Figure 6. A total of 4 segments per participant were obtained.

3.3. Continuous Wavelet Transform

ECG and GSR signals consist of various frequency components, hence CWT based on a series of wavelets needs to be performed in order to convert the signals into the time–frequency domain so that the appropriate features can be extracted. CWT of ECG and GSR data was obtained, and the signal was compared with the shifted version of the

wavelet. The CWT equation is given as Equation (1), where the inner products identify the similarities between the waveform and the analyzing function ψ (mainly a wavelet):

$$C(a, b; f(t), \varphi(t)) = \int [f(t).a\varphi * (t - ba)dt] \quad (1)$$

We used the analytic Morlet (Gabor) wavelet with 12 wavelet bandpass filter banks, called the Amor wavelet, with equal variance in time and frequency. Variations in scale parameter a and position parameter b give CWT in time t; time, scale, and coefficients define the wavelet. The fast-changing details of the signal are captured at a lower scale since the wavelet is compressed and its frequency is high, and vice versa [15,34].

3.4. Scalogram

All CWT coefficients were arranged to form a scalogram. The scalogram was characterized on a jet type color map with 128 colors and transformed into an image in RGB color format, and the images were classified using the various CNN models (Figures 7 and 8).

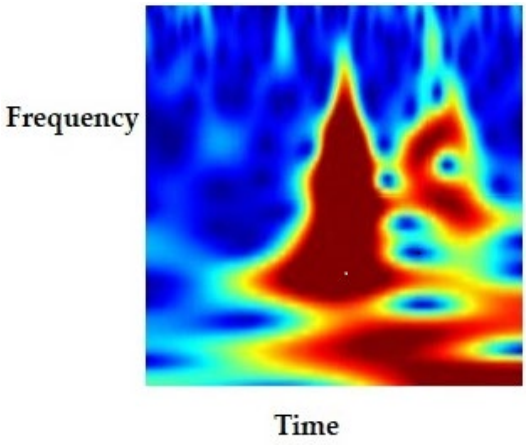

Figure 7. Scalogram of ECG signal.

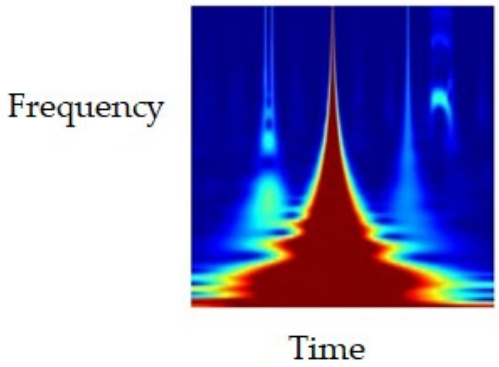

Figure 8. Scalogram of GSR signal.

3.5. Convolutional Neural Network

A CNN model learns directly from the data without human intervention. The automatic feature extraction technique in CNN models reduces the complexity of the task as compared to traditional machine learning techniques.

The objective of this study was to classify emotions using 2D pretrained CNN models, including MobileNet, which can be used on devices and in the naturalistic environment of participants, where emotions can be detected easily using smart bands and mobile phones.

Other recent pretrained CNN models—DenseNet 201, NASNetMobile, Inception-ResnetV2, and Efficient-NetB7—were chosen to validate the performance of MobileNet

in terms of accuracy and speed compared to other models. Khan et al. used seven machine learning classifiers for emotion classification and proved that the performance of a particular classifier will vary based on the dataset [43].

The CNN architectures of DenseNet 201, MobileNet, NASNetMobile, InceptionResnetV2, and EfficientNetB7 are explained in this section.

3.5.1. DenseNet CNN

In CNN architecture, the features of one layer act as the input to the next layer. The output totally depends on the features of the last layer. Hence, as the path of information increases, it can cause certain information to "vanish" or get lost, which reduces the network's ability to train effectively. However, in the DenseNet CNN, the features of one layer act as the input to every other layer, hence the vanishing gradient problem does not occur. Figure 9 shows the architecture of the DenseNet 121 CNN.

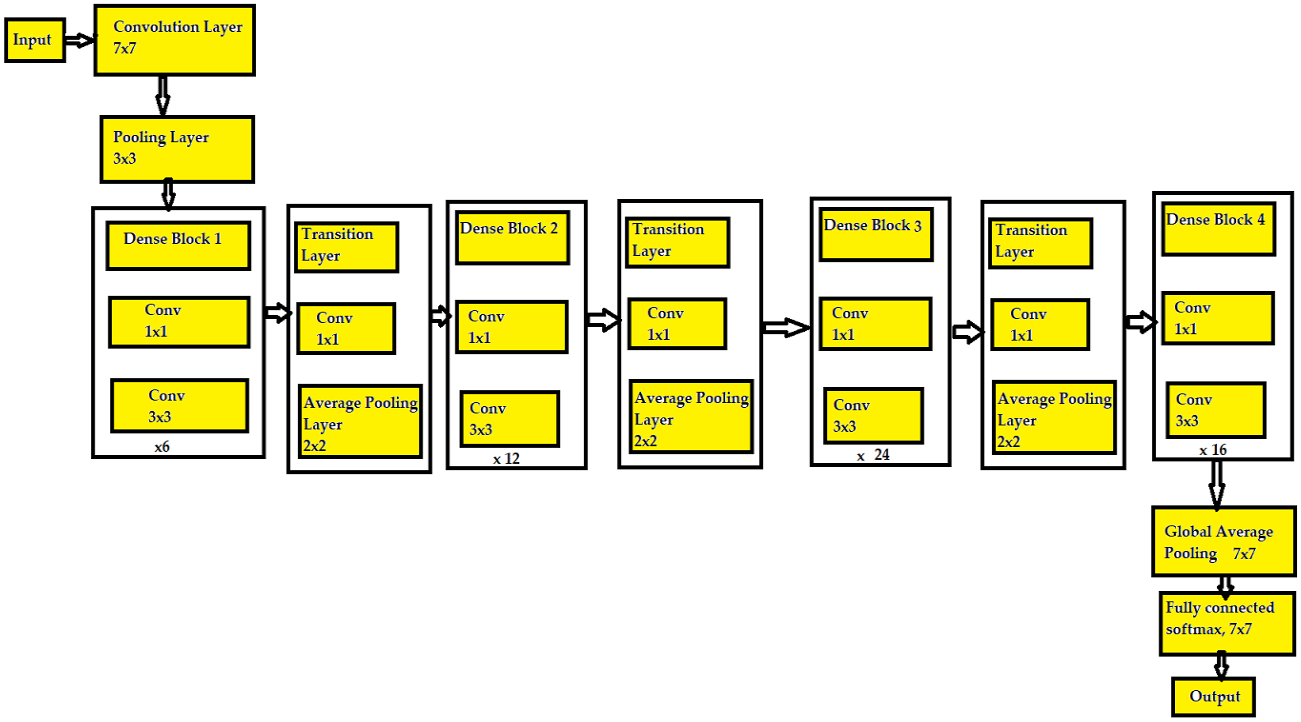

Figure 9. Architecture of DenseNet 121 CNN.

In DenseNet 121, there 121 layers with 4 dense blocks, each with a different number of convolutional layers. Dense block 1 has 6 convolutional layers, dense block 2 has 12 layers, dense block 3 has 24 layers, and dense block 4 has 16 layers [44,45]. Input is given to the CNN network with a filter size of 7×7 and a stride of 2, then it is given to a pooling layer with a filter size of 3×3 and a stride of 2.

After every dense block, there is a transition layer. DenseNet overcomes the vanishing gradient problem and provides high accuracy by connecting every layer directly. Hence, each layer receives the feature maps from every other layer, which enables DenseNet to strengthen the feature propagation. The features learned by layer 1 are directly accessible by layers 2 to 5. One layer has already learned something, and all the other layers can directly access the information through concatenation. Errors are calculated and passed to all the layers, including the early layers of the network. The last layer is the fully connected layer and uses the softmax activation function [44,45].

Similarly, the DenseNet 201 model has 201 layers, hence it is larger compared to DenseNet 121, and can enhance the classification accuracy.

3.5.2. MobileNet CNN

MobileNet uses depthwise separable convolution, in which an input image with three dimensions is split into three channels. Each channel is separately convolved with a filter. Finally, the output of all channels is combined to get the entire image. In a standard CNN, a 3 × 3 layer is followed by batch normalization and the ReLU activation function. MobileNet uses a depthwise separable convolutional layer comprising a 3 × 3 depthwise convolution with batch normalization and ReLU followed by a 1 × 1 pointwise convolution with batch normalization and ReLU, as shown in Figure 10 [46]. Owing to this architecture, MobileNet is much smaller and faster than other CNN architectures, making it suitable for implementation on mobile phones.

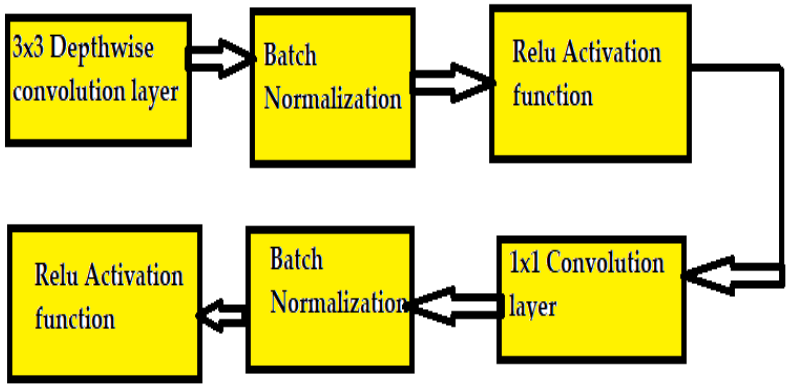

Figure 10. MobileNet CNN architecture.

3.5.3. NASNet Mobile CNN

In the Neural Architecture Search Network (NASNet) CNN, the blocks are not predefined but are detected using the reinforcement learning search method. The NAS has three basic strategies: search space, search strategy, and performance estimation strategy (Figure 11).

Figure 11. NASNet mobile CNN architecture.

The search space looks for the best possible architecture based on the type of application. The global search space allows any architecture the freedom to set hyperparameters, and the cell-based search space searches a particular architecture based on the cells. Architectures built on the cell-based search space are smaller and more effective and can be built into larger networks. The recurrent neural network controller searches within normal and reduction cells, thus the best combinations to form the cells are searched to improve the performance. The search strategy helps to find the best architecture for a given performance. The estimation strategy searches the architecture based on the best performance, such as the accuracy of the model [47,48].

3.5.4. InceptionResnetV2 CNN

The multiple-sized convolutional filters in the inception architecture are combined with the residual connections. This helps avoid the degradation problem due to the deep architecture and reduces the training time [49,50].

3.5.5. EfficientNetB7 CNN

This network was introduced in 2021 to optimize both parameter efficiency and training speed. The model was developed using a training-aware neural architecture search and scaling. A combination of scaling in width, depth, resolution, and neural architecture search was used to improve its performance [51].

Table 3 lists the architectural details of the pretrained CNN models. Depth refers to the number of layers in the model, excluding the input layer.

Table 3. Comparison of model architectures.

CNN MODEL	Architectural Specifications	Size	Depth	Total Parameters	Trainable Parameters	Non-Trainable Parameters
MobileNet	Depthwise separable convolution; smaller and faster	16 Mb	88	3,228,864	3,206,976	21,888
NASNetMobile	Blocks are not predefined but are detected using reinforcement learning search method	23 Mb	_	4,269,716	4,232,978	36,738
DenseNet201	Features of one layer act as input to every other layer; resolves vanishing gradient problem	80 Mb	201	18,321,984	18,092,928	229,056
InceptionResnetV2	Multiple-sized convolutional filters in inception architecture are combined with residual connections	215 Mb	572	54,336,736	54,276,192	60,544
EfficientNetB7	Uses a combination of training-aware neural architecture search and scaling		813	64,097,687	63,786,960	310,727

Table 3 also shows a comparison of total parameters of the CNN models and the number of trainable and non-trainable parameters. The total parameters are obtained by adding the parameters from all layers. The number of weights that get optimized indicates the trainable parameters. The weights that are not updated during training with backpropagation indicate the non-trainable parameters. MobileNet has the smallest number of parameters and EfficientNetB7 has the highest. MobileNet has the smallest size and depth.

In this work, CWT- and scalogram-based images were obtained using the MATLAB tool, and Python software on the Google Colab platform was used for CNN classification. The training used 70% of images from the dataset, validation used 20%, and testing used the remaining 10%. For ECG data, of the 2754 samples, 1926 samples were used for training, 550 samples for validation, and 278 for testing. For the GSR data, of the 624 samples, 436 samples were used for training, 124 samples for validation, and 64 for testing. The accuracy of valence and arousal classification was obtained.

4. Results and Discussion

The results obtained in the two cases described above are presented below. The model's performance was evaluated based on accuracy, precision, recall, F1 score, and confusion matrix parameters. Accuracy is calculated by the ratio of the number of correct predictions over the total number of predictions. Precision indicates the correctness of the number of optimistic predictions. Recall or sensitivity indicates the number of positive cases predicted correctly out of all positive cases. F1 score is the harmonic mean of precision and recall [52].

4.1. Case 1: Arousal and Valence Classification Using ECG and GSR Data while Participants Watched Short Videos

ECG and GSR recordings of the 17 participants watching short videos were captured [11], and arousal and valence classification was performed. The results obtained are described below.

4.1.1. ECG Arousal Classification

Accuracy, F1 score, precision, and recall for arousal classification were obtained using the pretrained CNN models based on the transfer learning approach. Table 4 compares these parameters using ECG data when participants were watching short-duration videos using MobileNet, NASNetMobile, DenseNet 201, InceptionResnetV2, and EfficientNetB7. The time for data execution by each CNN is also indicated in Table 4.

Table 4. Performance evaluation of arousal classification using ECG data.

No.	CNN Model	Accuracy (%)	Precision	Recall	F1 Score	Data Execution Time (s)
1	InceptionResNetV2	91.45	0.91	0.90	0.90	920.778
2	MobileNet	90.55	0.90	0.90	0.90	215.212
3	DenseNet201	90.55	0.92	0.92	0.92	850.947
4	NASNetMobile	89.27	0.90	0.90	0.90	297.214
5	EfficientNetB7	80.18	0.82	0.80	0.80	1743.611

Figure 12 shows a comparison of the accuracy of these CNN models. InceptionResnetV2 had the highest classification accuracy of 91.45%. In this CNN, multiple-sized convolutional filters in the inception architecture are combined with residual connections, which improves accuracy. MobileNet had the shortest execution time of 215.212 s owing to its smaller size and depth, as indicated in Figure 13. EfficientNetB7 had the lowest accuracy and longest execution time owing to its greater size and depth.

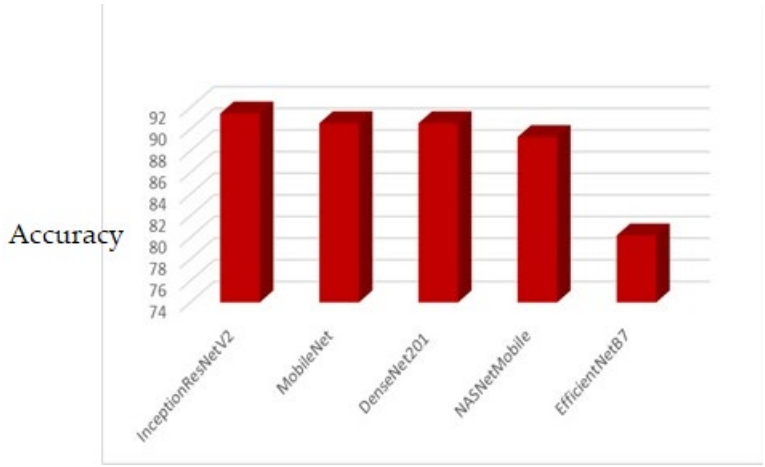

Figure 12. Accuracy comparison of short video arousal classification using ECG.

The precision, recall, and F1 score were the highest for the DenseNet201 CNN model, as shown in Figure 14.

4.1.2. ECG Valence Classification

ECG recordings of the 17 participants while watching short videos were captured for valence classification. Accuracy, F1 score, precision, and recall were obtained using the pretrained CNN models. Table 5 shows a comparison of these parameters for valence classification of participants watching short-duration videos based on the transfer learning approach using MobileNet, NASNetMobile, DenseNet201, InceptionResNetV2, and EfficientNetB7.

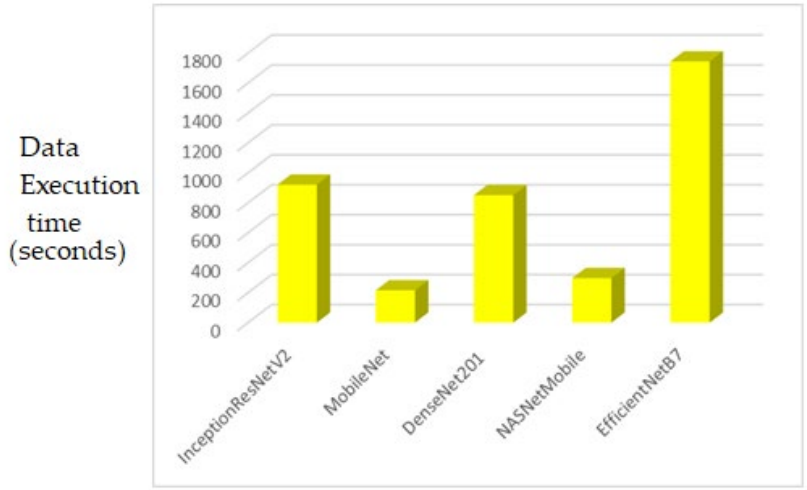

Figure 13. Comparison of data execution time (in seconds) for short video arousal classification using ECG.

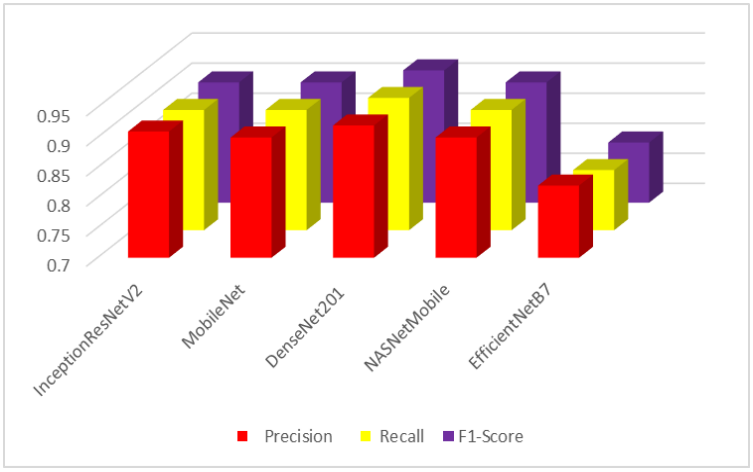

Figure 14. Precision, recall, and F1 score for short video arousal classification using ECG.

Table 5. Performance evaluation of valence classification using ECG data.

No.	CNN Model	Accuracy (%)	Precision	Recall	F1 Score	Data Execution Time (s)
1	InceptionResNetV2	91.27	0.9	0.9	0.9	892.63
2	MobileNet	90.55	0.93	0.93	0.93	183.081
3	DenseNet201	92	0.93	0.93	0.93	895.516
4	NASNetMobile	89.09	0.92	0.92	0.92	303.057
5	EfficientNetB7	71.64	0.78	0.72	0.70	17,772.865

DenseNet201 had the highest classification accuracy of 92%. Figure 15 shows the classification accuracy and loss plots for DenseNet201 in ECG valence classification. MobileNet had the shortest execution time owing to its smaller size and depth. DenseNet201 had better accuracy since it resolves the vanishing gradient problem. EfficientNetB7 had the lowest accuracy, precision, recall, and F1 score and the longest execution time owing to its greater depth.

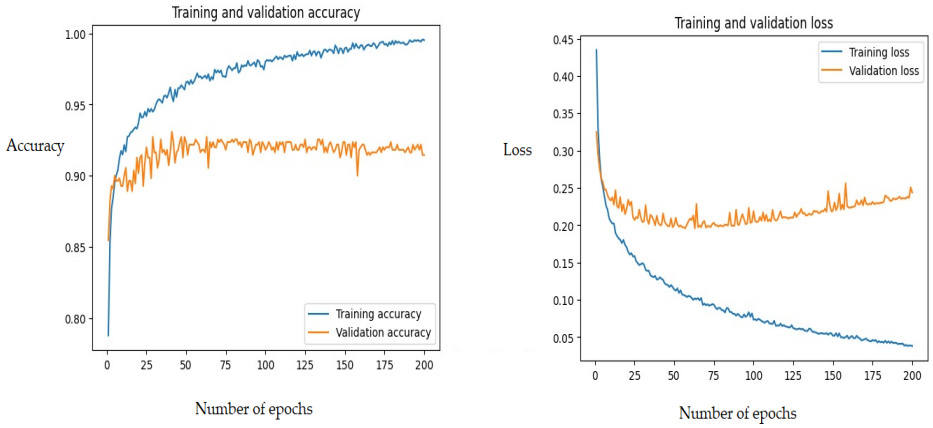

Figure 15. Classification accuracy and loss for DenseNet201 in valence classification.

The training and validation accuracy and loss do not deviate much below 90% accuracy, indicating better model performance. In total, 126 samples were correctly classified as high valence, and 133 samples as low valence.

4.1.3. GSR Arousal Classification

GSR recordings of the 17 participants watching short videos were captured, and arousal and valence classification were carried out. The results obtained are presented below. Table 6 shows a comparison of accuracy, precision, recall, and F1 score for arousal classification using GSR data.

Table 6. Performance evaluation of arousal classification using GSR data.

No.	CNN Model	Accuracy (%)	Precision	Recall	F1 Score	Data Execution Time (s)
1	InceptionResNetV2	97.58	0.98	0.98	0.98	205.267
2	MobileNet	98.39	0.98	0.98	0.98	62.214
3	DenseNet201	99.19	0.98	0.98	0.98	163.155
4	NASNetMobile	96.77	0.97	0.97	0.97	76.611
5	EfficientNetB7	76.61	0.75	0.75	0.75	486.086

DenseNet201 had the highest classification accuracy. Figure 16 shows the confusion matrix for the arousal classification of DenseNet201 using GSR data. In total, 32 samples were correctly classified as high arousal and 31 samples as low arousal.

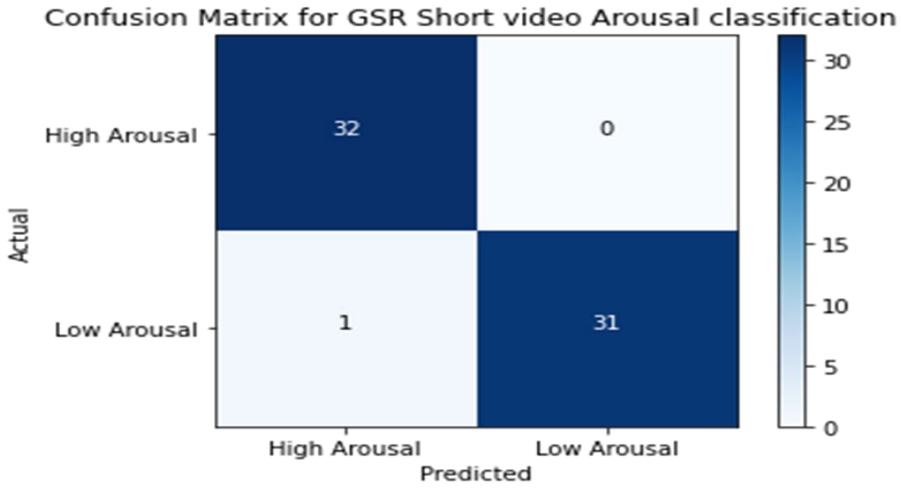

Figure 16. Confusion matrix for GSR arousal classification.

Figure 17 shows the training and validation classification accuracy and loss for the MobileNet CNN classifier with GSR data.

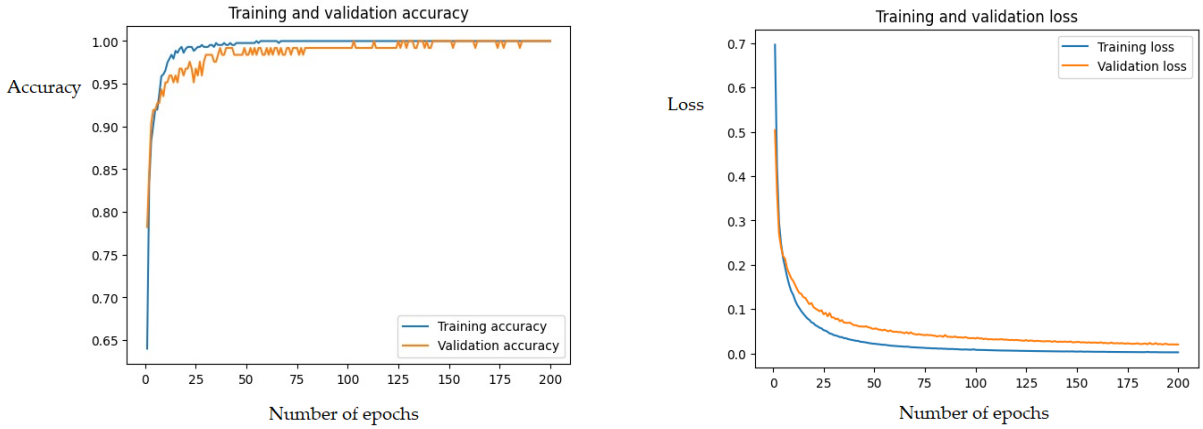

Figure 17. MobileNet CNN arousal classification accuracy and loss using GSR.

The training and validation accuracy and loss do not deviate much, indicating better model performance.

4.1.4. GSR Valence Classification

Table 7 shows a comparison of the accuracy, precision, recall, and F1 score for valence classification using GSR data.

Table 7. Performance evaluation of valence classification using GSR data.

No.	CNN Model	Accuracy (%)	Precision	Recall	F1 Score	Data Execution Time (s)
1	InceptionResNetV2	91.13	0.95	0.95	0.95	202.349
2	MobileNet	99.19	0.97	0.97	0.97	43.428
3	DenseNet201	96.77	0.97	0.97	0.97	186.739
4	NASNetMobile	95.97	0.93	0.92	0.92	79.452
5	EfficientNetB7	79.84	0.79	0.78	0.78	479.478

MobileNet had the highest valence classification accuracy.

We also compared the accuracy of the models using the ECG and GSR recordings of participants. Table 8 shows the arousal and valence classification using the ECG and GSR data of the participants while watching short videos.

Table 8. Arousal and valence classification using ECG and GSR data.

No.	CNN Model	ECG Arousal Accuracy (%)	GSR Arousal Accuracy (%)	ECG Valence Accuracy (%)	GSR Valence Accuracy (%)
1	InceptionResNetV2	91.45	97.58	91.27	91.13
2	MobileNet	90.55	98.39	90.55	99.19
3	DenseNet201	90.55	99.19	92	96.77
4	NASNetMobile	89.27	96.77	89.09	95.97
5	EfficientNetB7	80.18	76.61	71.64	79.84

Table 8 shows that using GSR data led to better model performance than ECG data based on short videos in arousal and valence classification. Hence, GSR is a more suitable modality for emotion classification using data corresponding to short-duration emotion elicitation.

Filtering and CWT were used to pre-process and extract the relevant features from the data before emotion classification. Time–frequency scalograms were obtained and converted to images, enabling the use of deep learning architectures such as 2D CNN

for emotion classification. The performance was validated using the diverse pretrained CNN models.

Researchers have used various deep-learning models for emotion classification based on short-duration emotion elicitation. Sepulveda et al. used traditional machine learning techniques on CWT coefficients obtained from ECG data and reported a classification accuracy of 90.2% for arousal and 88.8% for valence using short video data from the AMIGOS database [17]. Table 9 lists the work done by other researchers using deep learning models for emotion classification using ECG and GSR data.

Table 9. Comparison with other related works.

Ref. No.	Databases	Biosignals	Methods	Accuracy
[2]	DEAP	ECG	1D CNN	Valence: 75.3%; arousal: 76.2%
[14]	DEAP, MAHNOB	GSR	1D CNN	MAHNOB: valence and arousal: 78% DEAP: valence and arousal: 82%
[21]	DREAMER, AMIGOS	ECG, GSR	1D CNN, LSTM	DREAMER: valence and arousal: 89.25% AMIGOS ECG valence and arousal: 98.73%; GSR valence and arousal: 63.67%
[22]	AMIGOS	ECG, GSR	1D CNN	ECG valence: 75%; arousal: 76% GSR valence: 75%; arousal: 71%
[24]	DEAP	PPG	1D CNN	Valence: 82.1%; arousal: 80.9%
[25]	DREAMER, AMIGOS	ECG	CNN-LSTM (Bayseian)	DREAMER: valence: 86%; arousal: 83% AMIGOS: valence: 90%; arousal: 88%
[53]	DREAMER	ECG	1D CNN	Self-supervised CNN: valence: 85%; arousal: 85.9% Fully supervised CNN: valence: 66.6%; arousal: 70.7%
Proposed Work	AMIGOS	ECG (short video data)	InceptionResNetV2 CNN	Valence: 91.27% Arousal: 91.45%
		GSR (short video data)	MobileNet CNN	Valence: 99.19% Arousal: 98.39%

It can be seen that other researchers specifically used 1D CNN models on numerical ECG and GSR datasets. In our work, we used 2D CNN models to improve classification accuracy.

4.2. Case 2: Arousal and Valence Classification of Participants while Watching Long Videos in Individual and Group Settings

In this work, we carried out a novel classification of emotions using deep learning techniques to explore the social context of participants using ECG data corresponding to long videos. The ECG recordings of the 17 participants watching long videos individually and in groups were captured and classified using the arousal and valence dimensions. The model performance was evaluated using classification accuracy, F1 score, precision, and recall, as described below.

4.2.1. Individual Arousal Classification

For the arousal classification, ECG recordings of the 17 participants watching long videos individually were captured [11]. Table 10 compares the accuracy, precision, recall, and F1 score for arousal classification of participants while watching long-duration videos individually based on the transfer learning approach using the pretrained MobileNet, NASNetMobile, DenseNet 201, and InceptionResnetV2 deep learning CNN classifiers. All models had similar accuracy, precision, recall, and F1 score. MobileNet had the shortest execution time owing to its smaller size and depth. Figure 18 shows classification accuracy and loss plots.

Table 10. Performance evaluation of individual arousal classification.

No.	CNN Model	Accuracy (%)	Precision	Recall	F1 Score	Data Execution Time (s)
1	InceptionResNetV2	99.8	1	1	1	857.976
2	MobileNet	99.8	1	1	1	211.122
3	DenseNet201	99.8	1	1	1	881.271
4	NASNetMobile	99.8	1	1	1	300.944

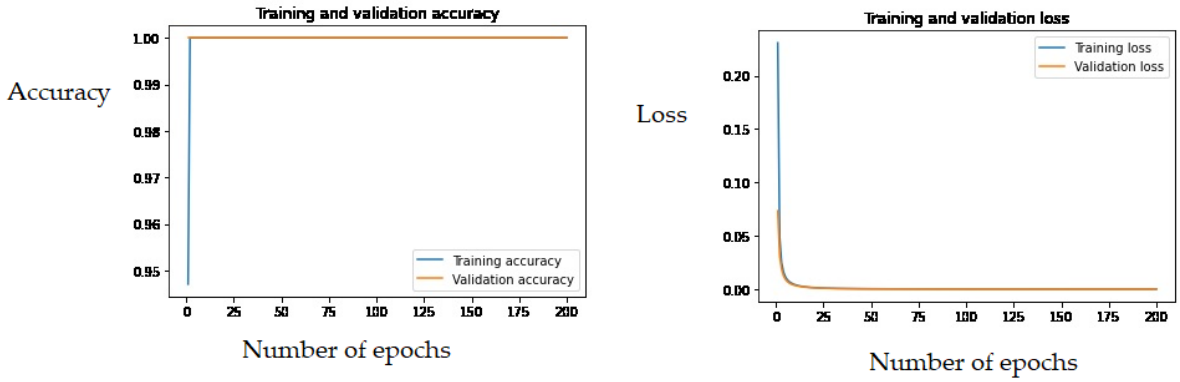

Figure 18. Individual arousal classification accuracy and loss for NASNetMobile CNN.

The training and validation accuracy plots do not indicate any discrepancy, hence the model is the most suitable for classification.

4.2.2. Individual Valence Classification

For the valence classification, ECG recordings of the 17 participants watching long videos individually were captured [11]. Accuracy, F1 score, precision, and recall were obtained using the pretrained CNN models. Table 11 compares these parameters for valence classification of participants watching long videos individually using MobileNet, NASNetMobile, DenseNet 201, InceptionResnetV2. All models attained the highest classification accuracy and had similar performance in terms of accuracy, precision, recall, and F1 score, and MobileNet had the shortest execution time, as indicated in Table 11.

Table 11. Performance evaluation of individual valence classification.

No.	CNN Model	Accuracy (%)	Precision	Recall	F1 Score	Data Execution Time (s)
1	InceptionResNetV2	99.8	1	0.9	1	314.81
2	MobileNet	99.8	1	1	1	174.88
3	DenseNet201	99.8	1	1	1	932.799
4	NASNetMobile	99.8	1	1	1	561.291

In total, 139 samples were correctly classified as high valence, and 139 as low valence. No deviation was observed between training and validation accuracy.

4.2.3. Group Arousal Classification

For the classification based on social context, ECG recordings of the 17 participants watching long videos in groups were captured [11]. Accuracy, F1 score, precision, and recall were obtained using the pretrained CNN models. Table 12 compares these parameters for arousal classification of participants watching long videos in groups based on the transfer learning approach using MobileNet, NASNetMobile, DenseNet 201, InceptionResnetV2. MobileNet had the highest classification accuracy, as shown in Figure 19.

Table 12. Performance evaluation of group arousal classification.

No.	CNN Model	Accuracy (%)	Precision	Recall	F1 Score	Data Execution Time (s)
1	MobileNet	99.82	1	1	1	223.211
2	DenseNet201	99.10	1	1	1	960.337
3	InceptionResNetV2	98.74	1	1	1	871.891
4	NASNetMobile	98.74	0.99	0.99	0.99	301.785

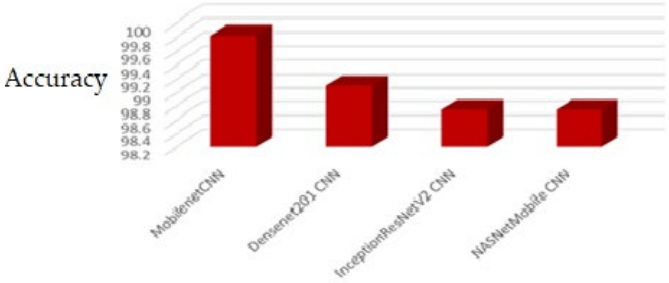

Figure 19. Classification accuracy comparison.

4.2.4. Group Valence Classification

For group valence classification, ECG recordings of the 17 participants watching long videos in groups were captured [11]. Accuracy, F1 score, precision, and recall were obtained using the pretrained CNN models. Table 13 compares these parameters for valence classification of participants watching long videos in groups based on the transfer learning approach using MobileNet, NASNetMobile, DenseNet 201, InceptionResnetV2. MobileNet and DenseNet 201 had the highest accuracy, precision, recall, and F1 score, and MobileNet had the shortest execution time.

Table 13. Performance evaluation of valence classification.

No.	CNN Model	Accuracy (%)	Precision	Recall	F1 Score	Data Execution Time (s)
1	MobileNet	99.82	1	1	1	191.844
2	DenseNet201	99.82	1	1	1	849.547
3	NASNetMobile	99.45	1	1	1	300.732
4	InceptionResNetV2	99.27	1	1	1	872.207

Figure 20 shows the confusion matrix for valence classification using InceptioResNetV2, and Figure 21 shows the classification accuracy and loss plots. All samples were correctly classified, with no misclassification, indicating the highest precision, recall, and F1 score values.

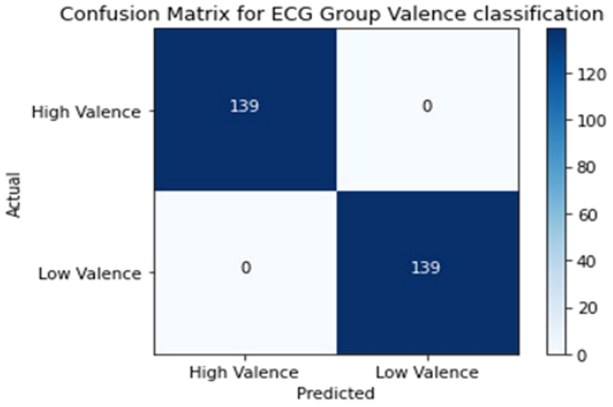

Figure 20. Confusion matrix.

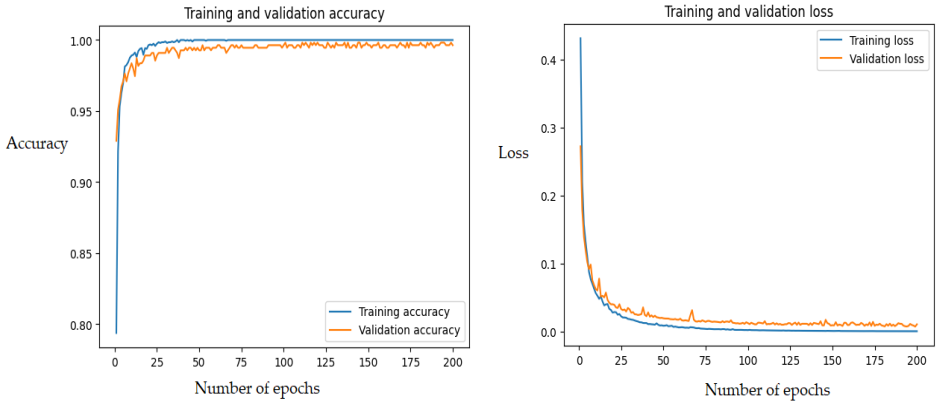

Figure 21. Group valence classification accuracy and loss for InceptionResNetV2 CNN.

Finally, we compared classification accuracy when the participants experienced emotions while watching long-duration videos individually and in groups, and while watching short videos, for ECG data as indicated in Table 14 and Figure 22.

Table 14. Valence arousal classification accuracy.

No.	CNN Model	ECG Short Video Arousal Accuracy (%)	ECG Short Video Valence Accuracy (%)	ECG Individual Long Video Arousal Accuracy (%)	ECG Individual Long Video Valence Accuracy (%)	ECG Group Long Video Arousal Accuracy (%)	ECG Group Long Video Valence Accuracy (%)
1	InceptionResNetV2	91.45	91.27	99.8	99.8	98.74	99.27
2	MobileNet	90.55	90.55	99.8	99.8	99.82	99.82
3	DenseNet201	90.55	92	99.8	99.8	99.10	99.82
4	NASNetMobile	89.27	89.09	99.8	99.8	98.74	99.45

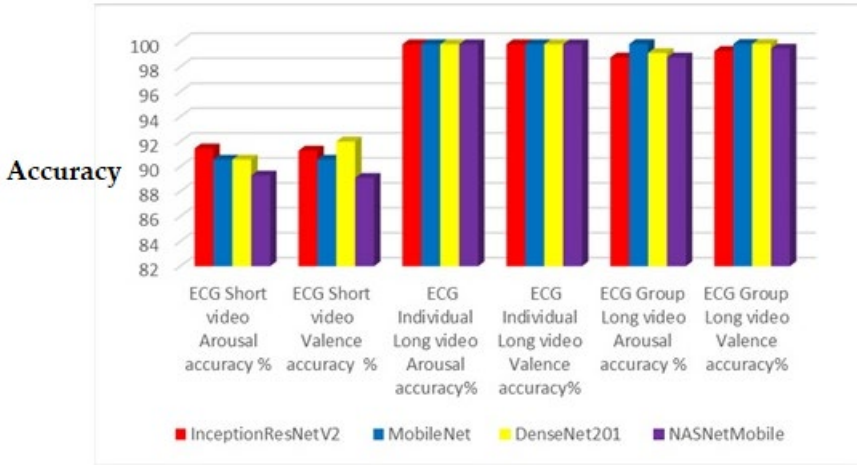

Figure 22. Accuracy comparison chart.

Figure 22 indicates that experimentally emotion classification accuracy using ECG data is improved when the emotion elicitation occurs over a longer duration and in groups.

5. Conclusions

The novelty of this study is in the CWT-based deep learning classification for ECG and GSR data using pretrained CNN models. The valence arousal classification accuracy of 99.19% and 98.39% obtained using GSR data outperforms the accuracy obtained using

ECG data. Furthermore, this novel work explored emotion classification considering long-duration emotion elicitation and the social context of participants based on a deep learning framework, resulting in an accuracy of around 99.80%, and when the participants watched long-duration videos in groups, accuracy was around 99.82%. Emotions are experienced in a better way when participants are engaged in groups. The architectural advancements in five pretrained CNN models, MobileNet, NASNetMobile, DenseNet 201, InceptionResnetV2, and EfficientNetB7, help improve the accuracy considerably. Additionally, automatic feature extraction makes the task less complex than traditional machine learning techniques with manual feature extraction.

The performance of five pretrained CNN models, trained on millions of datasets, was validated in terms of accuracy, precision, recall, F1 score, and execution speed. MobileNet, DenseNet201, and InceptionResnetV2 exhibited similar performance in terms of accuracy in both cases. MobileNet had the shortest execution time owing to its depthwise convolution architecture, followed by NASNetMobile. MobileNet is suitable for implementation on devices such as mobile phones. MobileNet outperformed the other models and can be suitably used in the naturalistic environment where people's emotions can be detected easily using smart bands and mobile phones. The subject-independent classification was explored, since different users react differently to the same stimuli based on their mental strength and physical stability, in order to ensure that the classification was independent of personality and mental status.

Some potential applications of the proposed model are mentioned below. The social isolation due to the COVID-19 pandemic led to sadness and depression, resulting in affective and behavioral problems among the population. This model could help to identify emotions such as depression and calmness and thus enable appropriate action. In healthcare scenarios, robots could be used to assess the emotions of patients and provide appropriate assistance. The emotions of the driver such as anger, and calmness can be captured and passengers may be alerted. The emotions of students, such as happy and sad, could be accurately captured using wearable sensors, and suitable changes could be adopted in teaching methodologies. Companies could develop appropriate marketing strategies based on customers' emotions to advertise products.

This model can accurately classify emotions based on two discrete classes, and could be further extended to classify emotions in the four dimensions of Russell's model. The model could also be applied to classify emotions based on EEG recordings of subjects and self-assessment reports. However, wearable devices such as Shimmer ECG and GSR sensors are required to be in contact with the subject's body, leading to some discomfort; in addition, people might be reluctant to be monitored throughout the day. The size and weight of the sensors could lead to discomfort for subjects, which could affect the accuracy of the model. To overcome these limitations, lightweight sensors could be developed using the latest technologies to capture ECG and GSR recordings.

Author Contributions: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing—original draft preparation, Writing—review and editing, Visualization, Supervision, Project administration, A.D. and H.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are available upon approval at http://www.eecs.qmul.ac.uk/mmv/datasets/amigos/index.html (publicly available database for research).

Acknowledgments: The authors are thankful to the Goa College of Engineering, affiliated with Goa University, for supporting the work carried out in this research.

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

References

1. Egger, M.; Ley, M.; Hanke, S. Emotion recognition from physiological signal analysis: A review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55. [CrossRef]
2. Lee, M.S.; Lee, Y.K.; Pae, D.S.; Lim, M.T.; Kim, D.W.; Kang, T.K. Fast emotion recognition based on single pulse PPG signal with convolutional neural network. *Appl. Sci.* **2019**, *9*, 3355. [CrossRef]
3. Bulagang, A.F.; Weng, N.G.; Mountstephens, J.; Teo, J. A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. *Inform. Med. Unlocked* **2020**, *20*, 100363. [CrossRef]
4. Dessai, A.; Virani, H. Emotion Classification using Physiological Signals: A Recent Survey. In Proceedings of the 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Thiruvananthapuram, India, 10–12 March 2022; IEEE: Piscataway, NJ, USA; Volume 1, p. 333.
5. Meléndez, J.C.; Satorres, E.; Reyes-Olmedo, M.; Delhom, I.; Real, E.; Lora, Y. Emotion recognition changes in a confinement situation due to COVID-19. *J. Environ. Psychol.* **2020**, *72*, 101518. [CrossRef] [PubMed]
6. Goshvarpour, A.; Abbasi, A.; Goshvarpour, A. An accurate emotion recognition system using ECG and GSR signals and matching pursuit method. *Biomed. J.* **2017**, *40*, 355–368. [CrossRef]
7. Hsu, Y.; Wang, J.; Chiang, W.; Hung, C. Automatic ECG-Based Emotion Recognition in Music Listening. *IEEE Trans. Affect. Comput.* **2020**, *11*, 85–99. [CrossRef]
8. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31. [CrossRef]
9. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A Multimodal Database for Affect Recognition and Implicit Tag-ging. *IEEE Trans. Affect. Comput.* **2012**, *3*, 42–55. [CrossRef]
10. Subramanian, R.; Wache, J.; Abadi, M.K.; Vieriu, R.L.; Winkler, S.; Sebe, N. Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Trans. Affect. Comput.* **2018**, *9*, 147–160. [CrossRef]
11. Miranda-Correa, J.A.; Abadi, M.K.; Sebe, N.; Patras, I. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Trans. Affect. Comput.* **2021**, *12*, 479–493. [CrossRef]
12. Maggio, A.V.; Bonomini, M.P.; Leber, E.L.; Arini, P.D. Quantification of ventricular repolarization dispersion using digital processing of the surface ECG. In *Advances in Electrocardiograms—Methods and Analysis*; IntechOpen: Rijeka, Croatia, 2012.
13. Dessai, A.; Virani, H. Emotion Detection Using Physiological Signals. In Proceedings of the 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 9–10 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–4.
14. Al Machot, F.; Elmachot, A.; Ali, M.; Al Machot, E.; Kyamakya, K. A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors. *Sensors* **2019**, *19*, 1659. [CrossRef]
15. Available online: https://en.wikipedia.org/wiki/Continuous_wavelet_transform (accessed on 20 March 2023).
16. Sepúlveda, A.; Castillo, F.; Palma, C.; Rodriguez-Fernandez, M. Emotion recognition from ECG signals using wavelet scattering and machine learning. *Appl. Sci.* **2021**, *11*, 4945. [CrossRef]
17. Shukla, J.; Barreda-Angeles, M.; Oliver, J.; Nandi, G.C.; Puig, D. Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity. *IEEE Trans. Affect. Comput.* **2019**, *12*, 857–869. [CrossRef]
18. Romeo, L.; Cavallo, A.; Pepa, L.; Bianchi-Berthouze, N.; Pontil, M. Multiple Instance Learning for Emotion Recognition Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2022**, *13*, 389–407. [CrossRef]
19. Bulagang, A.F.; Mountstephens, J.; Teo, J. Multiclass emotion prediction using heart rate and virtual reality stimuli. *J. Big Data* **2021**, *8*, 12. [CrossRef]
20. Dessai, A.U.; Virani, H.G. Emotion Detection and Classification Using Machine Learning Techniques. In *Multidisciplinary Applications of Deep Learning-Based Artificial Emotional Intelligence*; IGI Global: Hershey, PA, USA, 2023; pp. 11–31.
21. Dar, M.N.; Akram, M.U.; Khawaja, S.G.; Pujari, A.N. CNN and LSTM-based emotion charting using physiological signals. *Sensors* **2020**, *20*, 4551. [CrossRef] [PubMed]
22. Santamaria-Granados, L.; Munoz-Organero, M.; Ramirez-Gonzalez, G.; Abdulhay, E.; Arunkumar, N.J.I.A. Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS). *IEEE Access* **2018**, *7*, 57–67. [CrossRef]
23. Hammad, D.S.; Monkaresi, H. ECG-Based Emotion Detection via Parallel-Extraction of Temporal and Spatial Features Using Convolutional Neural Network. *Traitement Du Signal* **2022**, *39*, 43–57. [CrossRef]
24. Lee, M.; Lee, Y.K.; Lim, M.T.; Kang, T.K. Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features. *Appl. Sci.* **2020**, *10*, 3501. [CrossRef]
25. Harper, R.; Southern, J. A bayesian deep learning framework for end-to- end prediction of emotion from heartbeat. *IEEE Trans. Affect. Comput.* **2020**, *13*, 985–991. [CrossRef]
26. Ismail, S.N.M.S.; Aziz, N.A.A.; Ibrahim, S.Z.; Nawawi, S.W.; Alelyani, S.; Mohana, M.; Chun, L.C. Evaluation of electrocardiogram: Numerical vs. image data for emotion recognition system. *F1000Research* **2021**, *10*, 1114.
27. Pólya, T.; Csertő, I. Emotion Recognition Based on the Structure of Narratives. *Electronics* **2023**, *12*, 919. [CrossRef]
28. Bhangale, K.; Kothandaraman, M. Speech Emotion Recognition Based on Multiple Acoustic Features and Deep Convolutional Neural Network. *Electronics* **2023**, *12*, 839. [CrossRef]

29. Dukić, D.; Sovic Krzic, A. Real-Time Facial Expression Recognition Using Deep Learning with Application in the Active Classroom Environment. *Electronics* **2022**, *11*, 1240. [CrossRef]
30. Kamińska, D.; Aktas, K.; Rizhinashvili, D.; Kuklyanov, D.; Sham, A.H.; Escalera, S.; Nasrollahi, K.; Moeslund, T.B.; Anbarjafari, G. Two-Stage Recognition and beyond for Compound Facial Emotion Recognition. *Electronics* **2021**, *10*, 2847. [CrossRef]
31. Gaddam, D.K.R.; Ansari, M.D.; Vuppala, S.; Gunjan, V.K.; Sati, M.M. Human facial emotion detection using deep learning. In *ICDSMLA 2020: Proceedings of the 2nd International Conference on Data Science, Machine Learning and Applications*; Springer: Singapore, 2022; pp. 1417–1427.
32. Stanisław, S. Bringing Emotion Recognition Out of the Lab into Real Life: Recent Advances in Sensors and Machine Learning. *Electronics* **2022**, *11*, 496. [CrossRef]
33. Wang, T.; Lu, C.; Sun, Y.; Yang, M.; Liu, C.; Ou, C. Automatic ECG classification using continuous wavelet transform and convolutional neural network. *Entropy* **2021**, *23*, 119. [CrossRef]
34. Byeon, Y.-H.; Pan, S.-B.; Kwak, K.-C. Intelligent deep models based on scalograms of electrocardiogram signals for biometrics. *Sensors* **2019**, *19*, 935. [CrossRef]
35. Mashrur, F.R.; Roy, A.D.; Saha, D.K. Automatic identification of arrhythmia from ECG using AlexNet convolutional neural network. In Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 20–22 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.
36. Muzaffe, A. CNN based efficient approach for emotion recognition. *J. King Saud Univ.-Comput. Inf. Sci.* **2021**, *34*, 7335–7346.
37. Garg, N.; Garg, R.; Anand, A.; Baths, V. Decoding the neural signatures of valence and arousal from portable EEG headset. *Front. Hum. Neurosci.* **2022**, *16*, 808. [CrossRef]
38. Alsubai, S. Emotion Detection Using Deep Normalized Attention-Based Neural Network and Modified-Random Forest. *Sensors* **2023**, *23*, 225. [CrossRef] [PubMed]
39. Castiblanco Jimenez, I.A.; Gomez Acevedo, J.S.; Olivetti, E.C.; Marcolin, F.; Ulrich, L.; Moos, S.; Vezzetti, E. User Engagement Comparison between Advergames and Traditional Advertising Using EEG: Does the User's Engagement Influence Purchase Intention? *Electronics* **2022**, *12*, 122. [CrossRef]
40. Cui, X.; Wu, Y.; Wu, J.; You, Z.; Xiahou, J.; Ouyang, M. A review: Music-emotion recognition and analysis based on EEG signals. *Front. Neuroinform.* **2022**, *16*, 997282. [CrossRef] [PubMed]
41. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [CrossRef]
42. Cai, Y.; Li, X.; Li, J. Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review. *Sensors* **2023**, *23*, 2455. [CrossRef]
43. Khan, C.M.T.; Ab Aziz, N.A.; Raja, J.E.; Nawawi, S.W.B.; Rani, P. Evaluation of Machine Learning Algorithms for Emotions Recognition using Electrocardiogram. *Emerg. Sci. J.* **2022**, *7*, 147–161. [CrossRef]
44. DenseNet | Densely Connected Convolutional Networks, YouTube. 2022. Available online: https://www.youtube.com/watch?v=hCg9bolMeJM (accessed on 29 March 2023).
45. Ahmed, A. Architecture of Densenet-121, OpenGenus IQ: Computing Expertise & Legacy. OpenGenus IQ: Computing Expertise & Legacy. 2021. Available online: https://iq.opengenus.org/architecture-of-densenet121/ (accessed on 29 March 2023).
46. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. Available online: https://arxiv.org/abs/1704.04861 (accessed on 29 March 2023).
47. Available online: https://sh-tsang.medium.com/review-nasnet-neural-architecture-search-network-image-classification (accessed on 29 March 2023).
48. Nair, D. NASNet—A Brief Overview. OpenGenus IQ: Computing Expertise & Legacy. 2020. Available online: https://iq.opengenus.org/nasnet (accessed on 29 March 2023).
49. Elhamraoui, Z. INCEPTIONRESNETV2 Simple Introduction, Medium. 2020. Available online: https://medium.com/@zahraelhamraoui1997/inceptionresnetv2-simple-introduction-9a2000edcdb6 (accessed on 29 March 2023).
50. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-V4, inception-resnet and the impact of residual connections on learning. *arXiv* **2016**, arXiv:1602.07261. Available online: https://arxiv.org/abs/1602.07261 (accessed on 29 March 2023).
51. Tan, M.; Le, Q. EFFICIENTNETV2: Smaller Models and Faster Training. *arXiv* **2021**, arXiv:2104.00298v3. Available online: https://arxiv.org/pdf/2104.00298 (accessed on 29 March 2023).
52. Available online: https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score (accessed on 29 March 2023).
53. Sarkar, P.; Etemad, A. Self-supervised ECG representation learning for emotion recognition. *IEEE Trans. Affect. Comput.* **2020**, *13*, 1541–1554. [CrossRef]