International Journal of Innovative Research and Scientific Studies, 8(3) 2025, pages: 5100-5116

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Lightweight 1D CNN model for emotion classification using GSR signal

Amita Dessai[1*], Hassanali Virani[2]

[1,2]*Goa College of Engineering, Affiliated with Goa University, Goa, India.*

Corresponding author: Amita Dessai (*Email: amitachari@gec.ac.in*)

Abstract

Emotions can be detected through facial expressions, voice signals, and physiological signals such as Galvanic Skin Response (GSR), Electroencephalogram (EEG), and Electrocardiogram (ECG). However, there have been limited studies on using GSR signals for emotion detection. Emotion recognition systems (ERS) use artificial intelligence to diagnose emotions and trigger appropriate actions accurately. This study introduces a novel lightweight deep Convolutional Neural Network (LWDCNN) for emotion classification using GSR data. The LWDCNN model reduces computational complexity while improving the speed and accuracy of the classification. GSR data is normalized based on suitable segmentation. A seven-fold cross-validation technique is employed to classify emotions using LWDCNN based on the arousal and valence dimensions. Emotions are classified based on the social and non-social context of the subjects using GSR data from the AMIGOS (A Dataset for Affect, Personality, and Mood Research on Individuals and Groups) database in both individual (non-social) and group (social) settings. The accuracy of valence classification is 78.83% for individual settings and 79.96% for group settings. The accuracy of arousal classification is 80.62% for individual settings and 84.42% for group settings. This model can effectively be used to diagnose the subjects' mental health by detecting the appropriate emotions.
Goa College of Engineering, Affiliated with Goa University, Goa, India.

Keywords: AMIGOS, Emotions, GSR, Lightweight 1D CNN, Social context.

DOI: 10.53894/ijirss.v8i3.7709
Funding: This study received no specific financial support.
History: Received: 23 April 2025 / **Revised:** 26 May 2025 / **Accepted:** 30 May 2025 / **Published:** 10 June 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors have contributed to the work involved in this manuscript. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Acknowledgments: The authors are thankful to the Goa College of Engineering affiliated with Goa University for supporting the work carried out in this research.
Publisher: Innovative Research Publishing

1. Introduction

Brain-computer interaction (BCI) systems allow users to trigger computer actions without physical movement [1]. Affective BCI (aBCI) systems, which originated from the field of affective computing (AC), focus on enabling actions to be

triggered by emotions [2]. The use of AC is vast, such as analyzing movie reviews for sentiment detection [3] and monitoring the impact of online games on students' grades [4]. Within medical applications, Affective BCI (aBCI) systems are particularly valuable for aiding in rehabilitation, movement, and interaction to mitigate the effects of conditions such as stroke, cerebral palsy, and spinal cord injury [1]. Apart from medical applications, a BCI is also in demand for analyzing cognitive states during marketing contexts, driver assistance, detecting students' emotions in an online classroom scenario, and advertising products. The use of affective BCI systems can effectively detect the mental health of individuals and prevent suicidal attempts, promoting societal balance. The emotions are classified into disgust, sadness, anger, surprise, happiness, and fear states [5, 6] using the discrete model as suggested by Paul Ekman or into valence and arousal states using the two-dimensional model suggested by Egger et al. [6] as revealed in Figure 1. The intensity of emotions is measured by arousal, whereas pleasantness is measured by valence. The HVHA (High Valence and High Arousal) quadrant represents happiness, while the LVLA (Low Valence and Low Arousal) quadrant indicates sadness. HVLA (High Valence and Low Arousal) emotions indicate calmness and relaxation, whereas LVHA (Low Valence and High Arousal) emotions are associated with fear and anger [7].

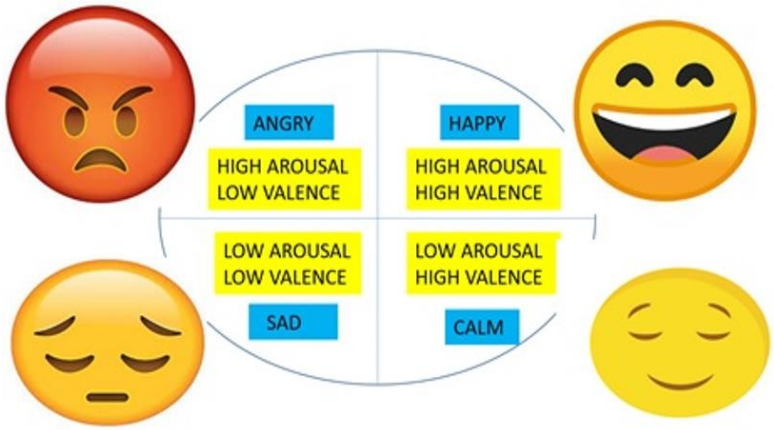

Figure 1.
Russel's Model.

According to a study by Ismail et al. [5], approximately 53% of researchers use the dimensional model, while 38% use the discrete model, and 9% use a hybrid model that combines both. Emotion Recognition Systems (ERS) mostly use audiovisuals to elicit emotions effectively, however, some studies incorporate visual, game, and audio elements as well [5].

Affective BCI systems are not limited to neurophysiological signals such as electroencephalogram (EEG); other physiological signals such as photoplethysmogram (PPG), galvanic skin response (GSR), and electromyography (EMG) also have great potential in many applications. Facial signals, text signals, and biological signals such as EEG and Electrocardiogram (ECG) have been extensively researched for emotion recognition. Moreover, PPG [8], GSR, skin temperature [9], and rate of respiration can also be used for emotion detection.

The fusion of several modalities, such as audio-visual, speech-text, and multiple biological signals has been explored by a few researchers for emotion recognition [10, 11]. ERS usually involves numerous steps, comprising preprocessing, feature extraction [12], feature selection, and classification. Machine learning classifiers such as Support Vector Machine (SVM), Ada Boost, K Nearest Neighbor (KNN), Naive Bayes, Random Forest (RF), Decision Tree (DT) are utilized for emotion classification by extracting time and frequency features from GSR data. SVM classifiers are the most accurate for small training sizes, while RF and DT classifiers can also provide reasonable accuracy. Some researchers believe that KNN classifiers are faster and simpler than SVM classifiers, and they outperform them. The handcrafted feature extraction method is used in machine learning techniques on preprocessed datasets, but deep learning techniques based on CNN perform automatic feature extraction and classification [13]. Affective computing focuses on developing novel algorithms for feature extraction and classification for the detection of emotions. Artificial Intelligence products powered by deep learning are expected to replace humans in various tasks, such as transportation, medical diagnosis, driving, and emotion detection. 1D CNN is particularly suitable for emotion classification using GSR data. However, the architecture of 1D CNN can be modified by tuning hyperparameters.

Human experiences are shaped by social emotions, which are complex feelings that require individuals to be aware of their own mental states as well as those of others. These emotional experiences naturally occur during spontaneous, dynamic, and responsive interactions [14]. The social context significantly influences how emotions are perceived and rated in terms of their valence (positive or negative) and arousal (intensity). It is crucial to examine emotion detection in group settings, where participants experience emotions collectively, as most emotions are felt in groups. While researchers have typically evaluated emotional models in individual contexts, there is a clear need to assess their performance in both individual and group settings using deep learning techniques. In a classroom environment, smart bands can capture students' emotions, such as sadness or happiness. By analyzing this data, educators can develop tailored teaching plans that enhance the learning experience while considering the diverse needs of children. Similarly, in a cinema, smart bands can be used to monitor audience emotions, allowing for relevant adjustments to improve the entertainment experience for the target audience.

Recent literature has primarily concentrated on developing emotion recognition models in non-social contexts, where participants experience emotions individually. However, it is crucial to examine how social context influences emotion

detection, as this aspect has not been thoroughly investigated. This study aims to fill this gap by classifying emotions in group settings using a 1D convolutional neural network (CNN) model. Past studies have typically utilized complex CNN architectures with numerous layers, which increase hardware complexity and reduce processing speed. In contrast, we incorporate the LWDCNN architecture to enhance accuracy while minimizing complexity by using fewer parameters. LWDCNN is designed to facilitate the implementation of CNNs on small devices, such as mobile phones.

The LWDCNN model is evaluated on two datasets: one capturing galvanic skin response (GSR) data from group settings (social context) and another from individual settings (non-social context). The main research contributions of this study are as follows:

- Developing a novel LWDCNN model for emotion detection using Galvanic Skin Response (GSR) data. This proposed model incorporates fewer layers for feature extraction and classification, which enhances speed and reduces task complexity.
- Developing the LWDCNN model with GSR data for emotion classification in two contexts: when subjects watched short-duration videos alone (non-social context) and when they watched long-duration videos in groups (social context).

The novelty of this study lies in the development of a lightweight deep learning convolutional neural network (LWDCNN) model that utilizes fewer layers for feature extraction and classification. This optimization enhances the speed of processing while simplifying the task. Moreover, the model is tested in both individual and group settings, making it the first study to apply a 1D CNN specifically for group classification in social contexts using galvanic skin response (GSR) data.

2. Related Works

Several publicly available databases such as DEAP [15], DREAMER [16], ASCERTAIN [17], and AMIGOS [18] have been investigated by researchers for emotion classification through physiological signals such as ECG, EEG, GSR, PPG, ST. Machine learning techniques, such as SVM, KNN, RF, and Decision tree classifiers, have been employed by researchers to classify emotions using GSR signals. However, the features have to be extracted manually [19, 20] and the classification accuracies reported are lesser. 1 D CNN techniques for emotion classification with GSR signals used by researchers are described below. Santamaria-Granados et al. [21] preprocess GSR data from the AMIGOS database, segment it, and normalize using standard deviation and the mean. This helps to standardize the data and make it more readily classified. Four convolutional layers and three max-pooling layers perform feature extraction, and four fully connected layers are used for classification. The model is trained using two hundred epochs, 90% of the data, and the leftover 10% of the data is used for testing. The data is classified into four categories: HALV, LVHA, LVLA, and HVHA, and achieved 71% arousal accuracy and 75% valence accuracy. Al Machot et al. [22] have classified emotions using the MAHNOB and DEAP datasets for subject-independent classification. The normalized data is passed through the convolutional layers for classification. Three convolutional layers, 3 max-pooling layers, and 3 fully connected layers are used. The last layer uses the Softmax function for classification. The grid search optimization technique optimizes the number of layers and filters. Additionally, ten-fold cross-validation is performed using the categorical cross-entropy loss function to minimize errors, while the Adam optimizer is employed to achieve faster convergence. The data is classified into four classes: HALV, LVHA, LVLA, and HVHA. For the subject-dependent classification, an accuracy of 85% for DEAP and 81% for MAHNOB is obtained. Dar et al. [23] normalize the GSR data using the Z-score normalization. Two 1D Convolutional layers, 2 max-pooling layers, and a flattened layer are followed by an LSTM layer. This is followed by three fully connected layers with a softmax function for classification. Adam optimizer is used and a batch size of two hundred and forty is used for training the classifier. 70% of the data is used for training and 30% for testing. The output is classified into four classes: HALV, LVHA, LVLA, and HVHA, with an accuracy of 63.67%. Kang and Kim [24] used PPG and GSR data for classifying the emotions. They achieved an accuracy of 81.33% in the arousal domain and 80.25% in the valence domain. The noise in the signal is reduced using the Butterworth filter, and for feature extraction, 1 D CNN model with two convolution and pooling layers, and the encoder and decoder are used. The features from both PPG and GSR are concatenated before classification. Two fully connected layers are used for classification. In contrast, Dessai and Virani [25] classified emotions using GSR data with 2D CNN models by converting the CWT of the GSR signals into scalograms and then into images that were classified using various pretrained 2D CNN models, MobileNet, DenseNet 201, InceptionResnetV2, NASNetMobile, and EfficientNetB7. However, the 2D CNN techniques require specialized hardware for training due to their computational complexity. Furthermore, CNN techniques are used in various studies to classify emotions using ECG data [26-28], PPG [29], and EEG data [30-34]. Applied novel data augmentation CNN to classify emotions using ECG data to address the issue of unbalanced data. They obtained an accuracy of 85.56% for arousal and 95.16% for valence on the DREAMER database using the seven-layer CNN model. Fan et al. [35] have used the attention mechanism in CNN. The weight information is added to the ECG features extracted using the deep CNN [35].

3. Methodology

The block diagram for classifying emotions captured through GSR data is shown in Figure 2. The data is segmented, normalized, and fed to the CNN model for classification. The emotions are classified into valence and arousal states.

The flowchart illustrated in Figure 3 depicts the methodology employed to classify emotions built on the GSR data of the AMIGOS dataset [18]. The GSR signal is split into two thousand-point segments using peak-based fiducial points, which are then normalized using Z-score normalization. Then, the dataset is split into training and test sets, and the K-fold cross-validation procedure is applied for better results. A 1D CNN model is designed and

trained with a batch size of ten and one hundred eighty epochs to classify the test dataset based on valence and arousal. The precision, recall, accuracy, and F1 score are computed, and a confusion matrix, accuracy, and loss plot are obtained for better visualization. Further details are explained below.

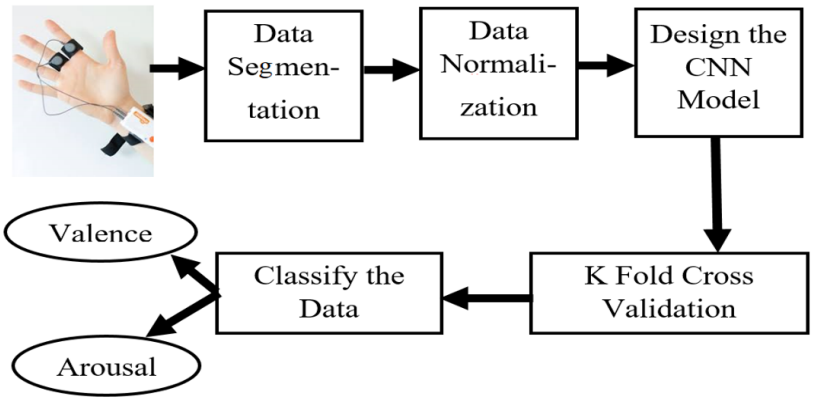

Figure 2.
Block diagram.

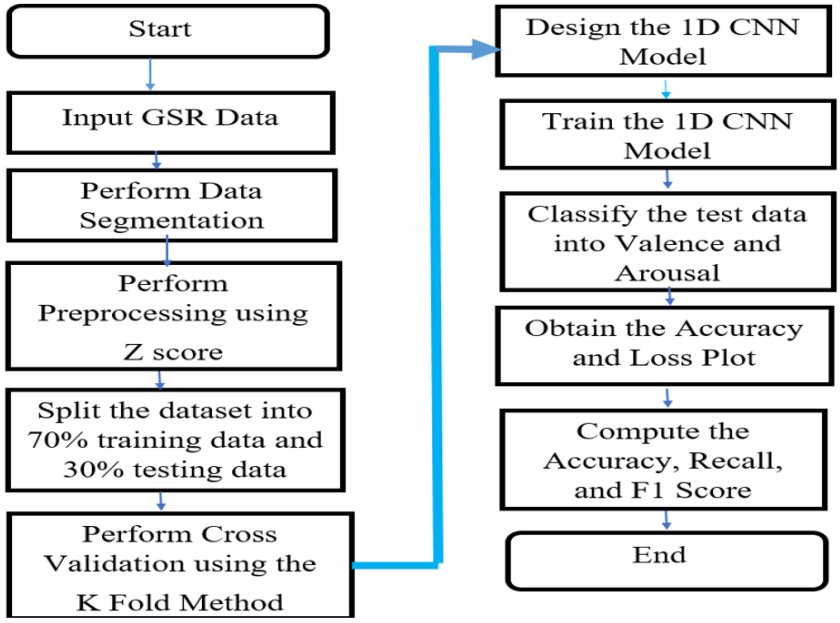

Figure 3.
Flowchart.

3.1. Dataset

To validate the proposed technique, the AMIGOS dataset [18] is used which contains several physiological signals such as EEG, GSR, and ECG.

The database consists of recordings of forty participants with diverse ages and equal gender distribution. The self-assessment is done using the Self-Assessment Manikins (SAM) questionnaire.

SAM assisted in annotating the valence and arousal levels corresponding to participants' emotions. The participants rated the emotions on a scale of one to ten [18]. The GSR signal is captured using the Shimmer sensor by placing two electrodes in the middle of the left hand's index and middle fingers [25]. The Shimmer 2R sensor captures the GSR signal at a sampling frequency of 128 Hz, with a 12-bit resolution [25].

In this study, the emotion classification is done using the GSR data of the AMIGOS dataset for the two scenarios as described below. AMIGOS database captures GSR data in consideration of the individual and group settings of the participants [18].

Scenario 1: Individual settings (Non-Social Context)

The subjects watched the videos individually, exploring the non-social context. The emotion elicitation for individual settings was done using the sixteen short-duration videos of less than 250 seconds when the participants watched the videos individually [18]. The low valence data corresponding to video 6 and video 8, and high valence data corresponding to video 1 and video 12 are obtained for individual valence classification [18]. To classify the level of arousal, the low arousal data from video 1, video 6, and the high arousal data from video 8 and video 12 were considered [18].

Scenario 2: Group Settings (Social context)

The subjects watched the videos together, and the social context of their experiences was explored. Additionally, for the group settings, the emotion elicitation was done using the four long-duration videos of greater than fourteen minutes when five groups of four participants watched the videos in groups [18].

To classify the group valence, the data related to video number 18 is compared with the data related to video number 19 [18]. Similarly, for group arousal classification, the data corresponding to video number 17 is compared to the data corresponding to video number 20 [18].

3.2. Segmentation

To avoid degradation of a lengthy GSR signal when passed through a deep learning structure, it needs to be suitably segmented [25]. This encompasses extracting a complete cycle of the GSR signal from the GSR waveform. The GSR signal waveform is indicated below in Figure 4.

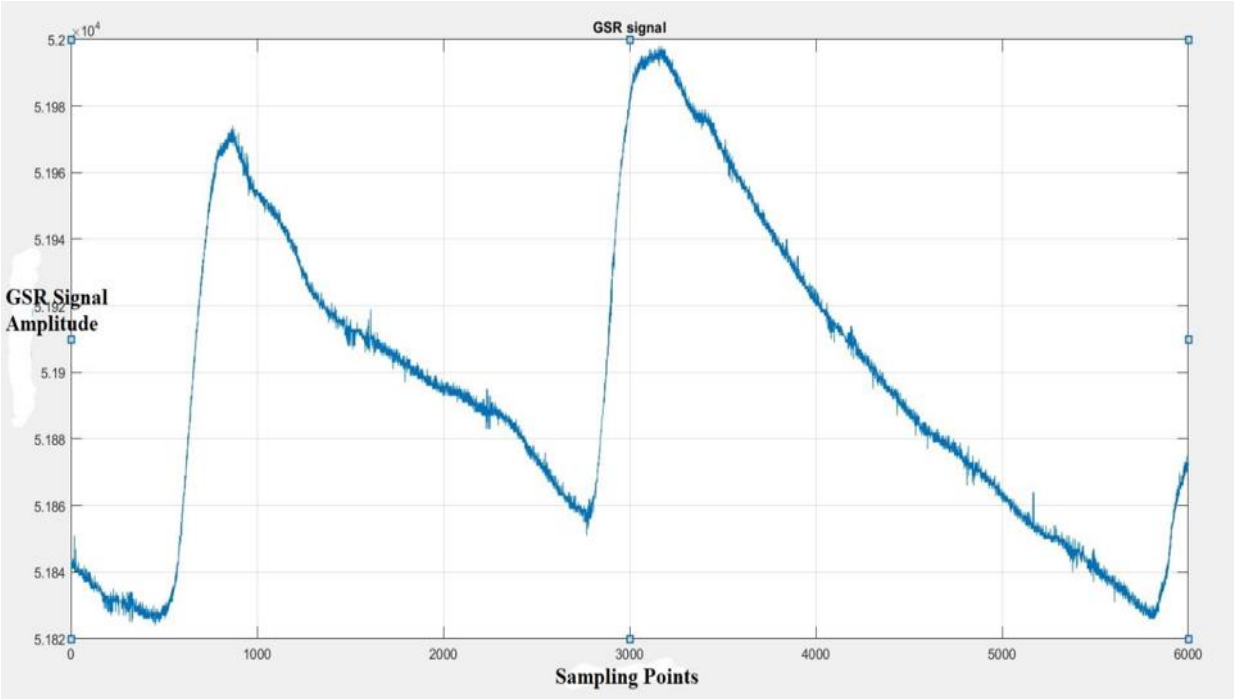

Figure 4.
GSR waveform.

The GSR signal is sampled at 128 Hz [18]. Considering the peaks as fiducial points, the signal is sampled at an interval of 2000 sampling points.

Scenario 1: For short video classification, each recording of the participant with 8000 sampling points is fragmented into four parts, and a total of 624 samples are obtained, with 312 samples of one class.

Scenario 2: For group classification, each recording of 14000 length is split into seven parts of length 2000 each. Hence, 119 samples per class and 238 samples for two classes are obtained.

3.3. Preprocessing Based on Z Score Normalization

The data is preprocessed using the Z-score normalization technique as mentioned in (1).

$$z = \frac{(x-\mu)}{s} \quad (1)$$

Where, $'x'$, is the sample, $'\mu'$ is the mean, and $'s'$ specifies the standard deviation. Z-score value indicates the amount by which a value deviates from the mean. This transformation is crucial for attaining a zero mean and unit variance of the features. Reducing the variation in the training data helps the CNN model to converge faster and improves its accuracy.

3.4. K-Fold Cross-Validation

Data is split into 70% training and 30% test. The training dataset is further distributed into two parts - 80% for training and 20% for validation. The training data undergoes a seven-fold cross-validation process, where the average accuracy of all seven folds is computed [36]. Finally, the model is tested on the 30% unseen test data, ensuring subject-independent classification. For each training and validation, a different dataset is used. This helps us to understand the best-case and worst-case scenarios for valence and arousal classification. K-fold cross-validation alleviates the overfitting risk and provides a reliable model performance assessment on unseen data.

3.5. 1D CNN Model

The configuration of a 1D-CNN is determined by several hyper-parameters, including the number of hidden CNN and Multilayer perceptron(MLP) layers/neurons, filter (kernel) size in each CNN layer, subsampling factor in each CNN layer, and the choice of pooling and activation functions [13]. In the hidden CNN layer, each neuron performs a sequence of convolutions, the sum of which is passed through the activation function, followed by the sub-sampling operation [13]. The input layer is a passive layer that receives the raw 1D signal, and the output layer is an MLP layer with the number of neurons equal to the number of classes [13]. The CNN layers process the raw 1D data and learn to extract features that are used in the classification task performed by the MLP-layers [13]. Both feature extraction and classification operations are fused into one process that can be optimized to maximize classification performance. 1 D CNN model automatically extracts features using the convolutional layers and pooling layers [13]. CNN layers filter the noise in the signal and also detect the relevant information from the GSR signals. The convolutional operation is performed in the CNN layer on the input data. The maximum pooling layer determines the maximum value of the data across the applied filter size. This helps in reducing the dimension of the data. The flattened layer converts the data into a dimension vector to create a single long feature vector. The output of the flattened layer is served as input to the fully connected layers. In these layers, all the neurons are connected. The fully connected layer performs the classification task based on the activation function and softmax classifier.

Dropout is a technique used to randomly deactivate a certain number of neurons during training, meaning that the output from these neurons is not included in the computations. This effectively creates a smaller network and helps reduce overfitting. By using different subsets of neurons for training each time, the network is better able to perform on unseen data. To further prevent overfitting, a dropout value is applied as a form of regularization. The ReLU activation function helps to avoid the vanishing gradient problem and provides better computation performance. The softmax function for classification transforms the raw outputs of the neural network into a vector of probabilities, enabling the prediction of the most probable class based on the input features. The model has been trained for 180 epochs with a batch size of ten. Figure 5 indicates the layout of the proposed model.

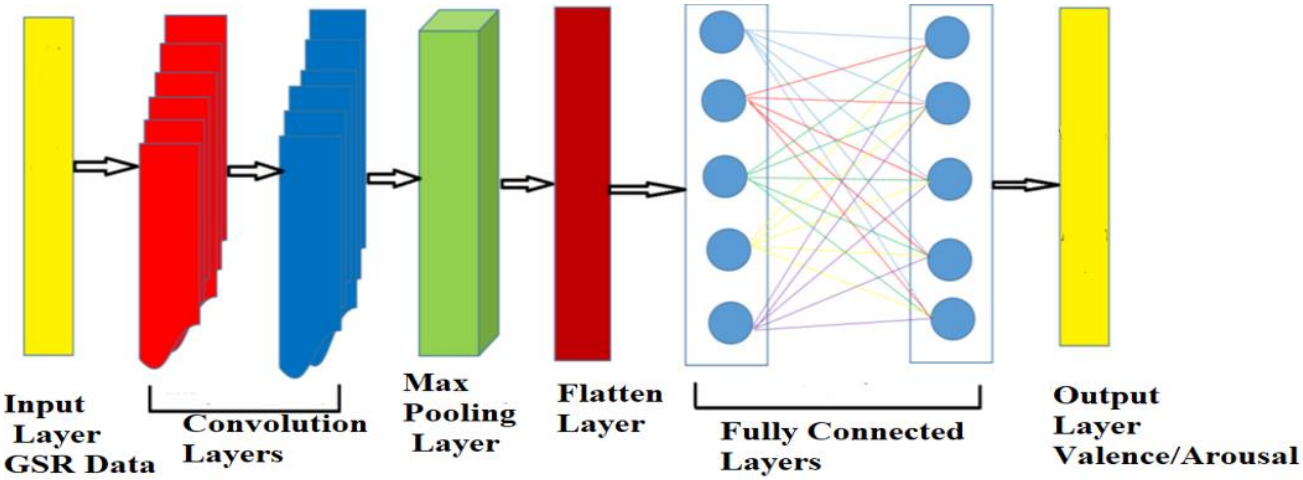

Figure 5.
1 D CNN Model.

The input signal, which has a size of 2000, is processed by the first convolutional layer. This layer has 64 filters, a kernel size of 3, a ReLU activation function, and a dropout rate of 0.3. The processed data is then passed to the second convolutional layer, which has 150 filters, a kernel size of 10, a ReLU activation function, and a dropout rate of 0.1. During the training process, a batch normalization layer is used to normalize the output of intermediate layers for a batch size of 10. This helps improve the speed of the model. Max pooling with a size of 2 and a stride of 2 is applied to reduce the dimensionality of the GSR data, allowing the convolutional layers to extract relevant features. The flattened layer is then used to convert the 2D array into a 1D vector, which is then fed into a dense layer. The output is classified into two classes using a softmax activation function in the dense layer. To make deep learning models converge faster and improve their accuracy and speed, an Adam optimizer with a learning rate of 0.001 is used. This helps minimize the categorical cross-entropy loss function that quantifies the dissimilarity between the predicted probabilities and the true categorical labels. Figure 6 shows the architectural details of the proposed CNN model.

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d_2 (Conv1D)           (None, 2000, 64)          256

 dropout_1 (Dropout)         (None, 2000, 64)          0

 conv1d_3 (Conv1D)           (None, 2000, 150)         96150

 batch_normalization_1       (None, 2000, 150)         600
 (Batch Normalization)

 max_pooling1d_1             (None, 1000, 150)         0
 (Max Pooling 1D)

 flatten_1 (Flatten)         (None, 150000)            0

 dense_1 (Dense)             (None, 2)                 300002

=================================================================
Total params: 397008 (1.51 MB)
Trainable params: 396708 (1.51 MB)
Non-trainable params: 300 (1.17 KB)
```

Figure 6.
Model Summary of 1D CNN Model

In Figure 6,' none' indicates the batch size, which is initially set to none since the batch size is not known in advance, 2000 size of input data, 64 number of filters. The next CNN layer uses 150 filters for convolution. The batch normalization layer normalizes data, allowing for higher learning rates and accelerating convergence. This leads to better and more efficient neural network training, improving the network's speed. The feature map reduces to 1000 after passing through the max pooling layer. The flattened layer generates a vector of dimension 150,000. The dense layer classifies the output into two classes using the softmax activation function. Trainable parameters are updated with backpropagation, while non-trainable parameters remain static during the training period. The hyperparameters used in designing the CNN model include the number of CNN layers, max pooling layers, filter sizes, and the number of fully connected layers, which are selected based on optimizing the model's accuracy.

4. Results

The proposed LWDCNN model is validated on the two different scenarios mentioned above. The recall, accuracy, precision, F1 score, and confusion matrix are obtained for the validation of the results. The K-fold cross-validation accuracies, corresponding to seven folds for Valence Classification, are indicated in Table I, Figure 7, and Figure 8.

Table 1.
K-fold cross-validation accuracy for valence classification

K-Fold Cross Validation Number	Individual Settings, Valence Accuracy %	Group Settings, Valence Accuracy %
Fold 1	68.98	56.94
Fold 2	71	79.16
Fold 3	71.12	80.55
Fold 4	81.81	83
Fold 5	82.35	83.33
Fold 6	83.42	86.11
Fold 7	93.04	90.27

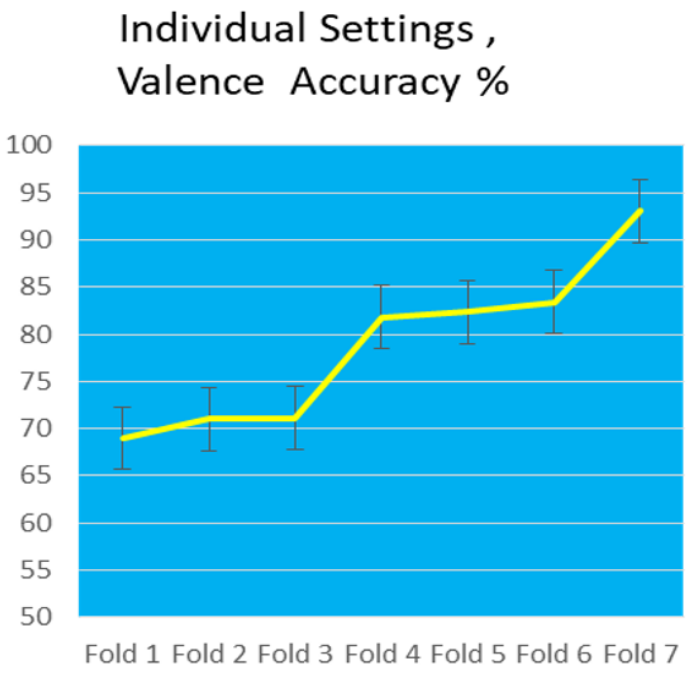

Figure 7.
K-fold Accuracy for Individual Settings.

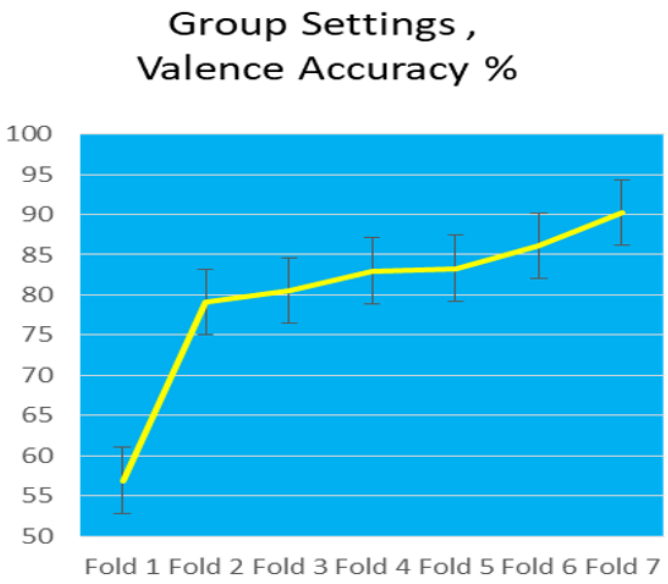

Figure 8.
K-fold Accuracy for Group settings.

To determine the best and worst-case scenarios for valence classification, a different fold is used for validation each time. The k-fold cross-validation results help in achieving this understanding. As per Table 1, Figure 7, and Figure 8, the worst-case accuracy obtained for individual settings is 68.98%, while for group settings, it is 56.94%. On the other hand, the best-case accuracy obtained for individual settings is 93.04%, and for group settings, it is 90.27%. Thus, the model demonstrates reasonably good performance for valence classification.

Table 2 and Figure 9 display the Precision, Recall, Accuracy, and F1 score for valence classification in two scenarios.

Table 2.
Performance evaluation for valence classification.

Valence	Accuracy	Precis-ion	Recall	F1 Score
Individual Settings	78.83 %	0.93	0.93	0.93
Group Settings	79.96 %	0.86	0.85	0.83

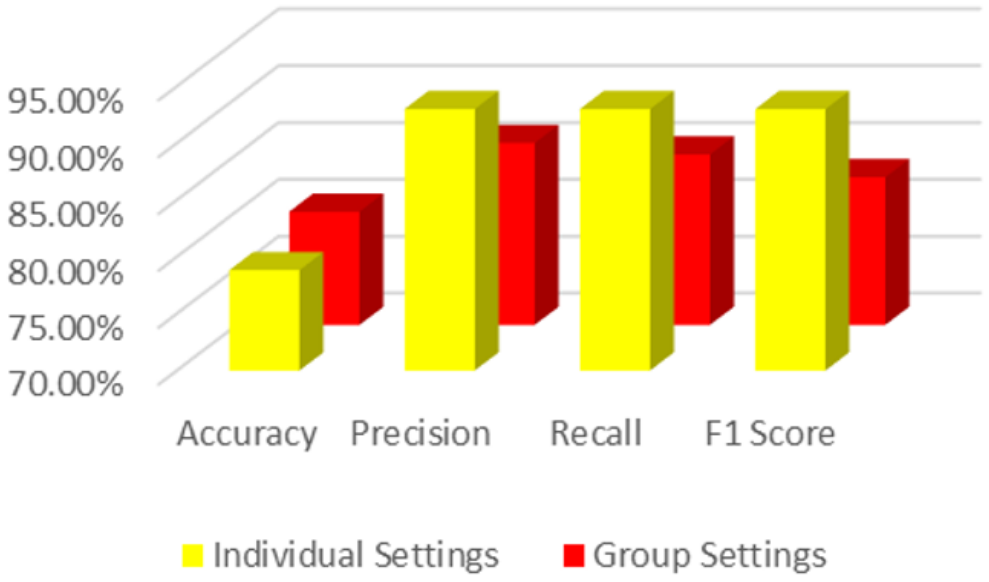

Figure 9.
Performance evaluation for valence classification.

The model gives a better performance for Valence classification in both individual and group settings. Similarly, the results obtained for Arousal Classification are indicated below in Table 3.

Table 3.
K-fold cross-validation accuracy for arousal classification.

K-Fold Cross Validation Number	Individual Settings, Arousal Accuracy %	Group Settings, Arousal Accuracy% %
Fold 1	65.95	73.61
Fold 2	66.48	75.0
Fold 3	67.02	83.33
Fold 4	85.10	88
Fold 5	91.48	88.88
Fold 6	92.02	90.27
Fold 7	96.27	94.44

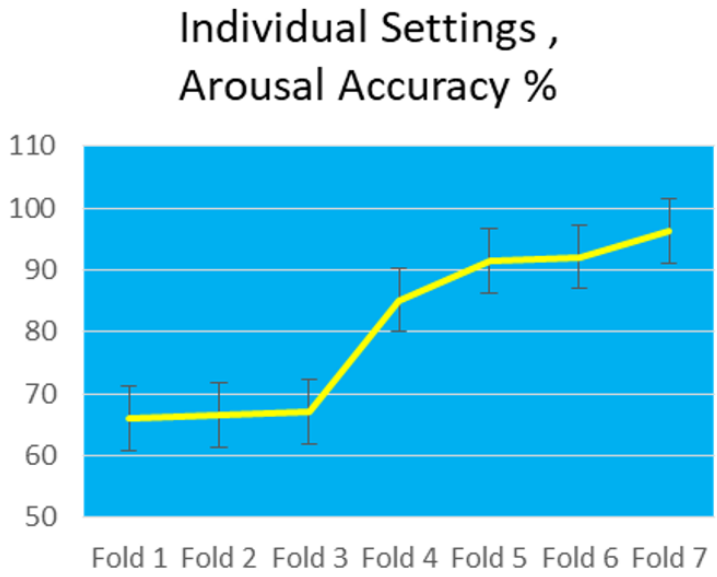

Figure 10.
K-fold Accuracy for Individual Settings

For each training and validation, a different dataset is used. This is why the k-fold cross-validation is important. They show the best-case and worst-case scenarios for arousal classification. Table 3, Figure 10, and Figure 11 specify that the worst-case accuracy obtained for individual settings is 65.95% and for group settings is 73.61%, respectively. Individual settings achieved 96.27% accuracy, and group settings achieved 94.44% accuracy in the best-case scenario. Table 4 provides

the accuracy, recall, precision, and F1 score for the arousal classification. Overall, the model performs better for arousal classification in both scenarios.

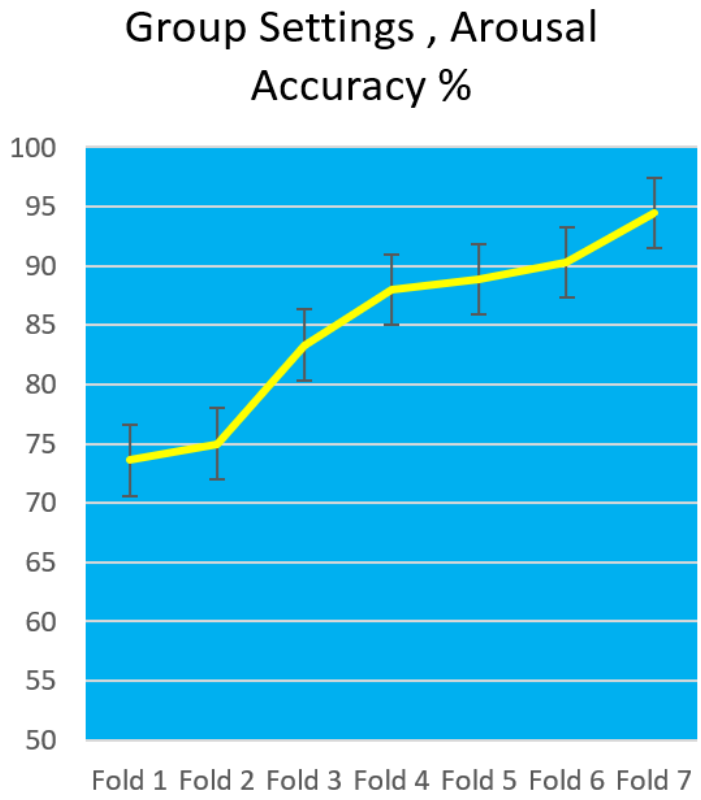

Figure 11.
K-fold Accuracy for Group settings.

Table 4.
Performance evaluation for arousal classification.

Mode	Accuracy %	Precision	Recall	F1 Score
Individual Settings	80.62	0.92	0.92	0.92
Group Settings	84.92	0.90	0.89	0.89

Furthermore, the accuracy of the Valence Arousal classification for both scenarios is presented in Table 5.

Table 5.
Valence Arousal classification accuracy.

Mode	Valence Accuracy	Arousal Accuracy
Individual Settings	78.83 %	80.62%
Group Settings	79.96 %	84.92%

The average accuracy obtained for individual settings is 78.83 % for Valence classification and 80.62% for arousal classification. The average accuracy obtained for group settings is 79.96 % for Valence classification and 84.92% for arousal classification. Figure 12 indicates the valence arousal classification accuracies.

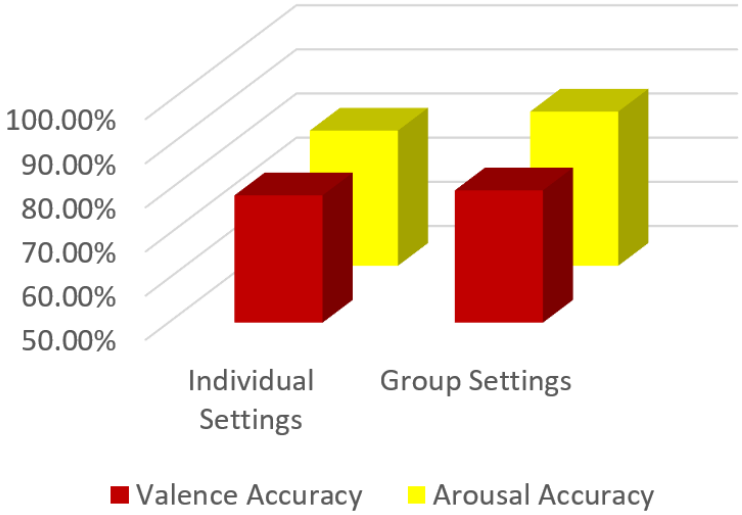

Figure 12.
Performance evaluation for Arousal classification.

The confusion matrix for the classification of Valence is displayed in Figures 13 and 14.

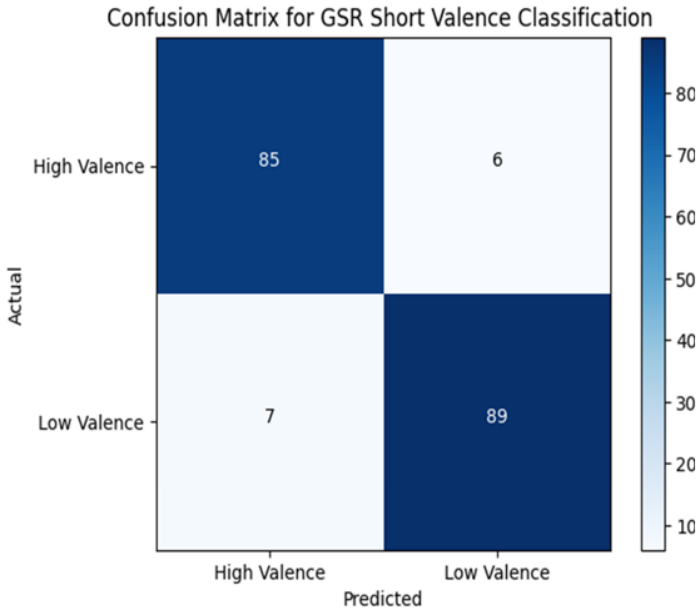

Figure 13.
Confusion matrix for individual setting for Valence Classification.

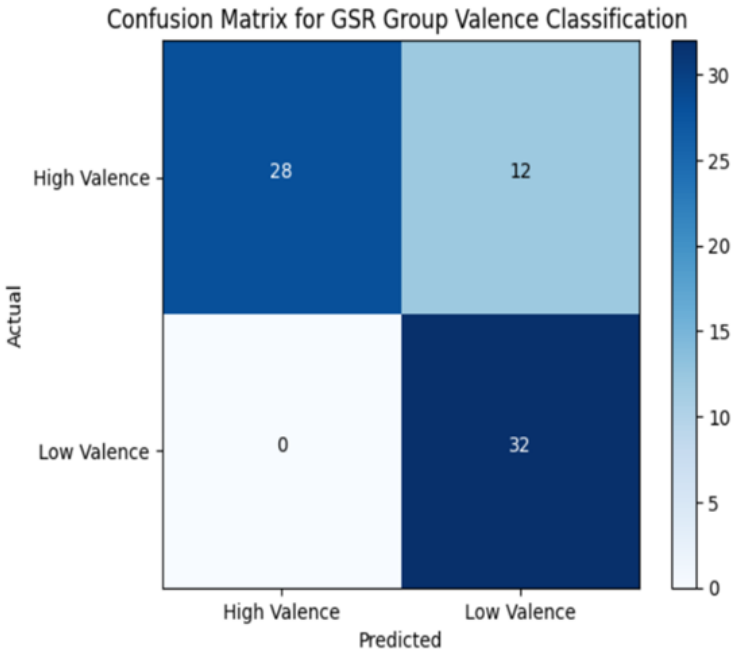

Figure 14.
Confusion matrix for group setting for Valence Classification.

.

The confusion matrix for the classification of Arousal is displayed in Figures 15 and 16.

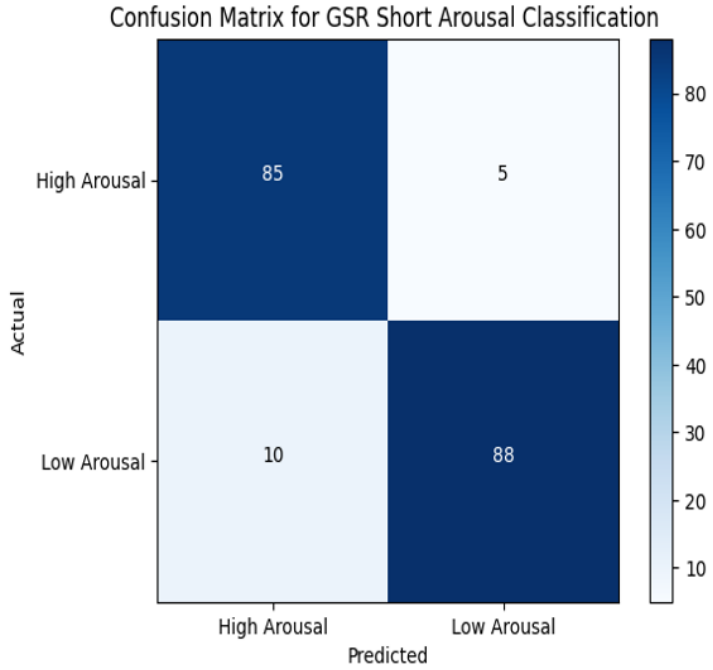

Figure 15.
Confusion matrix for individual setting for Arousal Classification.

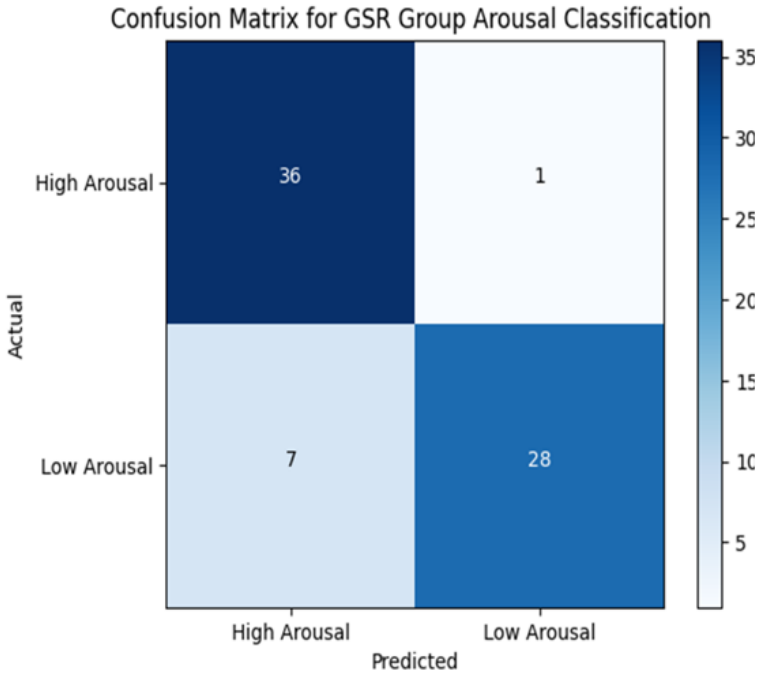

Figure 16.
Confusion matrix for group setting for Arousal Classification.

The Accuracy, Precision, Recall and F1 scores, as indicated in Table 2 and Table 4, are obtained from the confusion matrices using (2), (3), (4), and (5) as mentioned below. True Positive (TP) occurs when the predicted class and the actual class of a data point are both positive. True Negative (TN) occurs when the predicted class and the actual class are both negative. False Positive (FP) occurs when the predicted class is positive, but the actual class isnegative. False Negative (FN) occurs when the predicted class is negative, but the actual classis positive.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1\ Score = 2\ \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

The graphs in Figure 17 and Figure 18 show training and validation, accuracy, and loss, respectively, for classifying valence based on short video data.

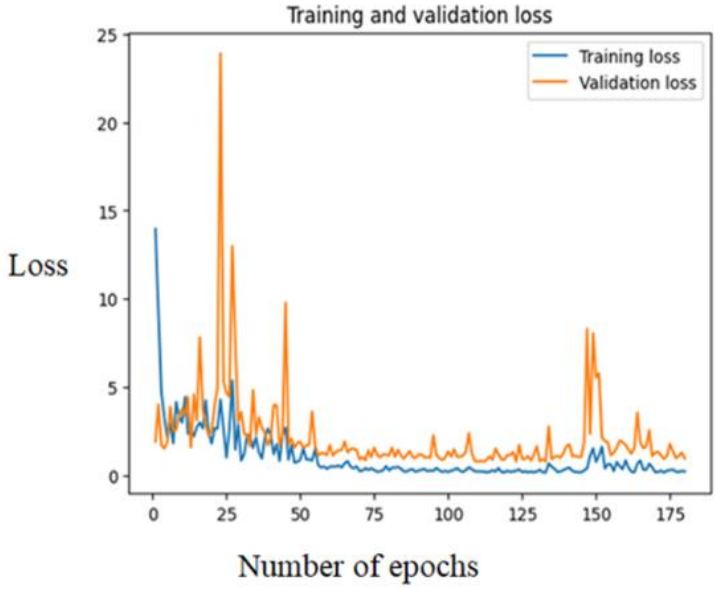

Figure 17.
Loss Plot.

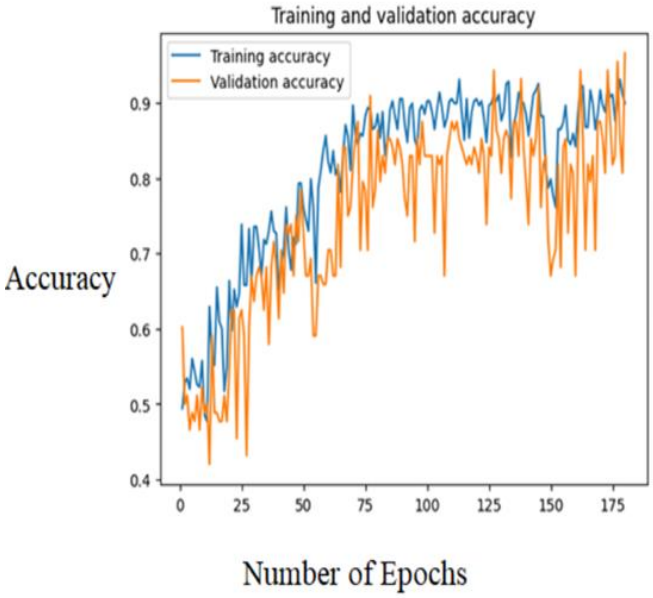

Figure 18.
Accuracy Plot.

5. Discussion

Manual feature extraction and machine learning classifiers have often been used for emotion classification from GSR data. However, Dessai and Virani [25] have employed pretrained CNN models to achieve improved accuracy. Nonetheless, the size, complexity, and cost of these 2D CNN models pose significant challenges. Moreover, the additional task of conversion of 1D data to 2D images or scalograms using continuous wavelets needs to be undertaken. A one-dimensional CNN has a lower computational complexity when compared to a two-dimensional CNN. An image with $N \times N$ dimensions convolving with KxK kernel will have a computational complexity ~ O (N^2K^2), while in the corresponding

1D convolution (with the same dimensions, N and K) is ~ O(NK) [13]. Millions of parameters are involved in the 2D CNNs as compared to thousands of parameters in 1D CNNs.

Table 6.
Comparison table.

Sr. No.	Reference Number	Physiological signals	Database	Methodology	Convolutional layers	Max Pooling layers	Fully connected layers	Dropout	Number of Classes	Classification Accuracy
1	Kang and Kim [24]	PPG AND GSR	DEAP	1D CNN	2	2	2	0.2	2	Valence: 80.25% Arousal: 81.33%
2	Al Machot, et al. [22]	GSR	DEAP, MAHNOB	1D CNN	3	3	3	0.2	4	a)MAHNOB:78% b)DEAP:82%
3	Dar, et al. [23]	GSR	DREAMER, AMIGOS	1D CNN-LSTM	2	2	3	0.5	4	a)DREAMER: 89.25% b)AMIGOS :63.67%
4	Santamaria-Granados, et al. [21]	GSR	AMIGOS	1D CNN	4	3	4	0.5	2	Valence:75% Arousal:71%
5	Proposed	GSR: Individual Settings	AMIGOS	1D CNN	2	1	1	0.3	2	Valence:78.83 % Arousal: 80.62%

The shallow architectures of 1D CNNs make it much easier to implement and train. Moreover, training 2D CNNs requires special graphical processing units. However, a CPU on a standard computer is realistic and comparatively faster for training 1D CNNs. Low computational requirements make 1D CNN models effective for emotion detection using GSR in real-time environments, particularly on mobile phones with low processing capability and battery life. Although only a limited number of studies have used 1D CNN models for emotion classification using GSR data, they have shown promising results. The hyperparameters that are typically involved in the design of 1D CNN models include the number of convolution layers, fully connected layers, number of filters, and choice of activation and pooling functions. Table 6 displays the architectural specifications of 1D CNN models utilized by other researchers, as well as their corresponding classification accuracy. Additionally, the proposed model is included.

Kang and Kim [24] concatenated the features extracted from the PPG and GSR signals and passed them to the fully connected layers for classification. They achieved an accuracy of 81.33% in the arousal domain and 80.25% in the valence domain. However, they utilized both PPG and GSR signals from the DEAP dataset, and the model is not suitable exclusively for the GSR data.

Al Machot et al. [22] have classified emotions using the MAHNOB and DEAP datasets for subject-independent classification. The data is classified into four classes: HALV, LVHA, LVLA, and HVHA. For the subject-dependent classification, an accuracy of 85% for DEAP and 81% for MAHNOB is obtained. However, they designed a 1D CNN model for four-class classification.

Dar et al. [23] categorized emotions into four classes using GSR data from the AMIGOS dataset, achieving an accuracy of 63.67%. The output is classified into four classes: HALV, LVHA, LVLA, and HVHA, with an accuracy of 63.67%. However, the additional LSTM layer increases the complexity of the task.

Santamaria-Granados et al. [21] classified emotions into Valence and Arousal and obtained an accuracy of 71% and 75%, respectively, using 1 D CNN model. The data is classified into four categories: HALV, LVHA, LVLA, and HVHA, and achieved 71% arousal accuracy and 75% valence accuracy. This being the first study to implement the CNN model for emotion detection using the GSR data, however, the complexity of the model is greater.

As observed in Table 6, the literature studies have utilized 1D CNN models with a greater number of layers, which increases the model's complexity. In addition to this, Dar et al. have incorporated LSTM layers. The proposed model aims to reduce the system's complexity while improving the accuracy of classification.

In this study, the emotions are classified using the proposed LWDCNN model on the Valence Arousal scale specifically for GSR data.

The study reduced the number of convolutional and max pooling layers needed for feature extraction. Furthermore, only one fully connected layer for classification is used, thus reducing the intricacy of the architecture and hardware requirements. Moreover, a novel classification of emotions to explore the group settings and individual settings of the participants using GSR data is performed. This is the first study to use a 1D CNN model on GSR data to classify emotions based on this concept. The accuracy of the model could be enhanced further through the optimization of the hyperparameters, addressing the limitations of the study.

6. Conclusion

A substantial amount of scientific interest exists in the use of GSR in affective brain-computer interaction studies. The 1D CNN model automatically extracts and classifies features, reducing the time required for feature extraction in time series classification. The novel LWCNN architecture has fewer layers, which reduces complexity and hardware requirements. Furthermore, this research introduces a novel approach to ERS by developing an LWCNN model that explores both group settings (social context) and individual settings (non-social context) of subjects using GSR data. An accuracy of valence classification of 78.83% for individual settings and 79.96% for group settings is obtained. The accuracy of arousal classification is 80.62% for individual settings and 84.42% for the group settings. The model achieved reasonable valence-arousal classification accuracy with improved precision, F1 score, and recall. Modifying the architecture of the CNN model in terms of the hyperparameters can further improve the accuracy of emotion classification.

Notably, the accuracy of group settings outperforms the accuracy obtained from individual settings. Furthermore, the LWDCNN model was tested on unseen data, enabling subject-independent classification. Thus, this study makes a significant contribution to the field of affective computing using GSR signals.

In the future, baseline removal methods can be utilized to remove noise, thereby enhancing classification accuracy. The LWDCNN model can also be adapted for other physiological signals such as ECG, EEG, and skin temperature by appropriately segmenting the signals and optimizing the model's hyperparameters.

References

[1] I. Daly, L. Chen, S. Zhou, and J. Jin, "An investigation into the use of six facially encoded emotions in brain-computer interfacing," *Brain-Computer Interfaces*, vol. 3, no. 1, pp. 59-73, 2016. https://doi.org/10.1080/2326263X.2016.1149360

[2] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: Principles, state-of-the-art, and challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66-84, 2014.

[3] O. G. Horsa and K. K. Tune, "Aspect-based sentiment analysis for afaan oromoo movie reviews using machine learning techniques," *Applied Computational Intelligence and Soft Computing*, vol. 2023, no. 1, p. 3462691, 2023. https://doi.org/10.1155/2023/3462691

[4] M. Alshraideh, A. A.-J. Abu-Zayed, M. Leiner, and I. M. AlDajani, "Beyond the scoreboard: A machine learning investigation of online games' influence on Jordanian university students' grades," *Applied Computational Intelligence and Soft Computing*, vol. 2024, no. 1, p. 1337725, 2024. https://doi.org/10.1155/2024/1337725

[5] S. N. M. S. Ismail, N. A. A. Aziz, S. Z. Ibrahim, and M. S. Mohamad, "A systematic review of emotion recognition using cardio-based signals," *ICT Express*, vol. 10, no. 1, pp. 156-183, 2024. https://doi.org/10.1016/j.icte.2023.09.001

[6] M. Egger, M. Ley, and S. Hanke, "Emotion recognition from physiological signal analysis: A review," *Electronic Notes in Theoretical Computer Science*, vol. 343, pp. 35-55, 2019. https://doi.org/10.1016/j.entcs.2019.04.009

[7] A. F. Bulagang, N. G. Weng, J. Mountstephens, and J. Teo, "A review of recent approaches for emotion classification using electrocardiography and electrodermography signals," *Informatics in Medicine Unlocked*, vol. 20, p. 100363, 2020. https://doi.org/10.1016/j.imu.2020.100363

[8] W. Mellouk and W. Handouzi, "CNN-LSTM for automatic emotion recognition using contactless photoplythesmographic signals," *Biomedical Signal Processing and Control*, vol. 85, p. 104907, 2023. https://doi.org/10.1016/j.bspc.2023.104907

[9] S. Pant, H.-J. Yang, E. Lim, S.-H. Kim, and S.-B. Yoo, "PhyMER: Physiological dataset for multimodal emotion recognition with personality as a context," *IEEE Access*, vol. 11, pp. 107638-107656, 2023. https://doi.org/10.1109/ACCESS.2023.3320053

[10] S. Dutta, B. K. Mishra, A. Mitra, and A. Chakraborty, "A multi-modal approach for emotion recognition through the quadrants of valence–arousal plane," *SN Computer Science*, vol. 4, no. 5, p. 460, 2023. https://doi.org/10.1007/s42979-023-01925-8

[11] B. Pan, K. Hirota, Z. Jia, and Y. Dai, "A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods," *Neurocomputing*, vol. 561, p. 126866, 2023. https://doi.org/10.1016/j.neucom.2023.126866

[12] A. Dessai and H. Virani, "Emotion detection using physiological signals," presented at the International Conference on Electrical, Computer, and Energy Technologies, ICECET. https://doi.org/10.1109/ICECET52533.2021.9698729, 2021.

[13] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, 2021.

[14] G. Gilam and T. Hendler, "With love, from me to you: embedding social interactions in affective neuroscience," *Neuroscience & Biobehavioral Reviews*, vol. 68, pp. 590-601, 2016. https://doi.org/10.1016/j.neubiorev.2016.06.027

[15] S. Koelstra *et al.*, "Deap: A database for emotion analysis; Using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, 2011. https://doi.org/10.1109/T-AFFC.2011.15

[16] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42-55, 2011. https://doi.org/10.1109/T-AFFC.2011.25

[17] R. Subramanian, "ASCERTAIN: Emotion and personality recognition using commercial sensors," Retrieved: http://mhug.disi.unitn.it/index.php/datasets/ascertain/, 2016.

[18] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality and mood research on individuals and groups," Retrieved: http://arxiv.org/abs/1702.02510, 2017.

[19] J. Shukla, M. Barreda-Angeles, J. Oliver, G. C. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Transactions on Affective Computing*, vol. 12, no. 4, pp. 857-869, 2019. https://doi.org/10.1109/TAFFC.2019.2901673

[20] A. Dessai and H. Virani, "Multimodal and multidomain feature fusion for emotion classification based on electrocardiogram and galvanic skin response signals," *Sci*, vol. 6, no. 1, p. 10, 2024. https://doi.org/10.3390/sci6010010

[21] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 57-67, 2018. https://doi.org/10.1109/ACCESS.2018.2883213

[22] F. Al Machot, A. Elmachot, M. Ali, E. Al Machot, and K. Kyamakya, "A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors," *Sensors*, vol. 19, no. 7, p. 1659, 2019. https://doi.org/10.3390/s19071659

[23] M. N. Dar, M. U. Akram, S. G. Khawaja, and A. N. Pujari, "CNN and LSTM-based emotion charting using physiological signals," *Sensors*, vol. 20, no. 16, p. 4551, 2020. https://doi.org/10.3390/s20164551

[24] D.-H. Kang and D.-H. Kim, "1D convolutional autoencoder-based PPG and GSR signals for real-time emotion classification," *IEEE Access*, vol. 10, pp. 91332-91345, 2022. https://doi.org/10.1109/ACCESS.2022.3201342

[25] A. Dessai and H. Virani, "Emotion classification based on CWT of ECG and GSR signals using various CNN models," *Electronics*, vol. 12, no. 13, p. 2795, 2023. https://doi.org/10.3390/electronics12132795

[26] D. S. Hammad and H. Monkaresi, "Ecg-based emotion detection via parallel-extraction of temporal and spatial features using convolutional neural network," *Traitement du Signal*, vol. 39, no. 1, pp. 43-57, 2022. https://doi.org/10.18280/ts.390105

[27] R. Harper and J. Southern, "A bayesian deep learning framework for end-to-end prediction of emotion from heartbeat," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 985-991, 2020. https://doi.org/10.1109/TAFFC.2020.2981610

[28] A. U. Dessai and H. G. Virani, "Emotion Detection Using ECG Signals and a Lightweight CNN Model," *Computer Systems Science & Engineering*, vol. 48, no. 5, 2024.

[29] M. Lee, Y. K. Lee, M.-T. Lim, and T.-K. Kang, "Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features," *Applied Sciences*, vol. 10, no. 10, p. 3501, 2020. https://doi.org/10.3390/app10103501

[30] W. Zheng and B. Pan, "A spatiotemporal symmetrical transformer structure for EEG emotion recognition," *Biomedical Signal Processing and Control*, vol. 87, p. 105487, 2024. https://doi.org/10.1016/j.bspc.2023.105487

[31] J. Chen, X. Lin, W. Ma, Y. Wang, and W. Tang, "EEG-based emotion recognition for road accidents in a simulated driving environment," *Biomedical Signal Processing and Control*, vol. 87, p. 105411, 2024. https://doi.org/10.1016/j.bspc.2023.105411

[32] M. K. Chowdary, J. Anitha, and D. J. Hemanth, "Emotion recognition from EEG signals using recurrent neural networks," *Electronics*, vol. 11, no. 15, p. 2387, 2022. https://doi.org/10.3390/electronics11152387

[33] S. Nita, S. Bitam, M. Heidet, and A. Mellouk, "A new data augmentation convolutional neural network for human emotion recognition based on ECG signals," *Biomedical Signal Processing and Control*, vol. 75, p. 103580, 2022.

[34] Z. Huang, Y. Ma, R. Wang, W. Li, and Y. Dai, "A model for EEG-based emotion recognition: CNN-BI-LSTM with attention mechanism," *Electronics*, vol. 12, no. 14, p. 3188, 2023. https://doi.org/10.3390/electronics12143188

[35] T. Fan *et al.*, "A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition," *Computers in Biology and Medicine*, vol. 159, p. 106938, 2023.

[36] P. Gupta, "Cross-validation in machine learning towards data science," Retrieved: https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f, 2017.