Article

A Survey on Physiological Signal-Based Emotion Recognition

Zeeshan Ahmad * and Naimul Khan

Department of Electrical, Computer and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada
* Correspondence: z1ahmad@ryerson.ca

Abstract: Physiological signals are the most reliable form of signals for emotion recognition, as they cannot be controlled deliberately by the subject. Existing review papers on emotion recognition based on physiological signals surveyed only the regular steps involved in the workflow of emotion recognition such as pre-processing, feature extraction, and classification. While these are important steps, such steps are required for any signal processing application. Emotion recognition poses its own set of challenges that are very important to address for a robust system. Thus, to bridge the gap in the existing literature, in this paper, we review the effect of inter-subject data variance on emotion recognition, important data annotation techniques for emotion recognition and their comparison, data pre-processing techniques for each physiological signal, data splitting techniques for improving the generalization of emotion recognition models and different multimodal fusion techniques and their comparison. Finally, we discuss key challenges and future directions in this field.

Keywords: data annotation; physiological signals; data variability; emotion models; challenges; review

1. Introduction

Emotion is a psychological response to some external stimulus and internal cognitive processes, supported by a series of physiological activities going on in human body. Thus, emotion recognition is a promising and challenging work area which enables us to recognize the emotions of a person for stress detection and management, risk prevention, mental health and interpersonal relations.

Emotional response also depends on the age. Socio-cognitive approaches suggest that the ability to understand emotions should be well maintained in adult aging. However, neuropsychological evidence suggests potential impairments in processing emotions in older adults [1]. However, long term depression in any age can lead to chronic diseases [2]. The pandemic of COVID-19 affected the emotions of people across the globe. The prevalence of a high suicide risk increased from pre-pandemic to during the pandemic, appearing to be largely influenced by social determinants, in conjunction with the implications of the COVID-19 pandemic [3].

Emotion recognition is an emerging research area due to its numerous applications in our daily life. Applications include areas such as developing models for inspecting driver emotions [4], health care [5,6], software engineering [7] and entertainment [8].

Different modalities of data can be used for emotion recognition. They are commonly divided into behavioral and physiological modalities. Behavioral modalities includes emotion recognition from facial expressions [9–13], from gestures [14–16] and from speech [17–19], while physiological modalities include emotion recognition from physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), galvanic skin response (GSR), electrodermal activity (EDA) and so on [20–24].

Behavioral modalities can be effectively controlled by user and the quality of expressing them may be significantly influenced by personality of the user and the current

environment of the subject [25]. On the other hand, physiological signals are continuously available and cannot be controlled intentionally or consciously. The commonly used physiological signals are ECG, EEG and GSR signals.

Many papers exist in literature where review on emotion recognition using physiological signals is presented.

In [26], a comprehensive review on physiological signal-based emotion recognition was presented that includes emotion models, emotion elicitation methods, the published emotional physiological datasets, features, classifiers and the frameworks for emotion recognition based on the physiological signal. In [27], a literature review provides a concise analysis of physiological signals and instruments used to monitor them, emotion recognition, emotion models and emotional stimulation approaches. The authors also discuss selected works on emotional recognition by physiological signals in wearable devices. In [28], different emotion recognition methods using physiological signals by machine learning techniques were explained. This paper also reviewed different stages such as data collection, data processing, feature extraction and classification models for each recognition method. In [29], recent advancements in emotion recognition research using physiological signals, including emotion models and stimulation, pre-processing, feature extraction and classification methodologies were presented. In [30], the current state-of-the art of emotion recognition was presented. The paper also presented the main challenges and future opportunities that lie ahead, in particular for the development of novel machine learning (ML) algorithms in the context of emotion recognition using physiological signals.

In [31], the different stages of Facial Emotion Recognition (FER) such as pre-processing, feature extraction and classification using various methods and state-of-the-art CNN models are discussed. Comparison between different deep learning models, their benchmark accuracy and their architectural details are also discussed for model selection based on application and dataset. In [32], the authors present a systematic literature review of scientific studies investigating automatic emotion recognition in a clinical population composed of at least a sample of people with a disease diagnosis. Based on the findings it is revealed that most clinical applications involved neuro-developmental, neurological and psychiatric disorders with the aims of diagnosing, monitoring, or treating emotional symptoms. In [33], emotion recognition for everyday life using physiological signals from wearables is presented. The authors observed that deep learning architectures provide new opportunities to solve complex tasks in this field of study. However, the limitation is that the study presents classification of binary or few class problem. In [34], the current situation in the EEG-based emotion recognition research along with the tutorial is presented to guide the researchers to start from a very beginning, as well as illustrate the theoretical basis and the research motivation. EEG data pre-processing, feature engineering, selection of classical and deep learning models for EEG-based emotion recognition are discussed.

In [35–37], the neural network, deep learning models and transfer learning-based models were discussed for different physiological signals and physiological signal-based datasets.

Existing review papers on emotion recognition based on physiological signals looked for only the regular steps involved in the workflow of emotion recognition such as pre-processing, feature extraction and classification but did not discuss the set of important challenges that are particular to emotion recognition.

The existing review papers on emotion recognition based on physiological signals provide only the following information:

1. feature extraction and selection techniques;
2. generic data pre-processing techniques for physiological signals;
3. different types of classifiers and machine learning techniques used for emotion recognition;
4. databases for emotion recognition;
5. assessment and performance evaluation parameters for ML models such as calculation of accuracy, recall, precision and F_1 score from confusion matrix.

Although the above information is useful and important for emotion recognition, it is established for any signal processing application and thus missing those challenging factors that are specific for emotion recognition using physiological signals. The following challenging factors are not discussed in the existing literature.

1. The problems faced during data annotation of physiological signals are not elaborated.
2. The data pre-processing techniques are bundled together and presented as generic techniques for any physiological signal. In our opinion, each physiological signal presents its own unique set of challenges when it comes to pre-processing, and the pre-processing steps should be discussed separately.
3. Inter-subject data variability has a huge impact on emotion recognition. The existing reviews neither discuss this effect nor provide recommendations to reduce inter-subject variability.
4. A comparison of data splitting techniques, such as subject-independent and subject-dependent, is not provided for better emotion recognition and generalization of the trained classification models.
5. The comparison and advantages of different multimodal fusion methods are not provided in these review papers.

The aforementioned challenging factors for emotion recognition that are missed in the existing literature are considered as the research gaps of the existing work. Thus, to bridge the above gap and to address the shortcomings of the existing reviews, in this paper, we are focussed on the missing but essential elements of emotion recognition. These are:

1. different kinds of data annotation methods for emotion recognition and advantages/disadvantages of each method;
2. data pre-processing techniques that are specific for each physiological signal;
3. effect of inter-subject data variance on emotion recognition;
4. data splitting techniques for improving the generalization of emotion recognition models;
5. different multimodal fusion techniques and their comparison.

This paper is organised as follows: in Section 2, the emotion models are explained. In Section 3, databases for emotion recognition using physiological signal are presented. In Section 4, the data annotation techniques for emotion recognition are illustrated. Section 5 demonstrates the data pre-processing technique for each physiological signal. In Section 6, the effect of inter-subject data variance on emotion recognition is described. In Section 7, data splitting techniques for improved generalization of emotion recognition models are analyzed. Section 8 provides the details of different multimodal fusion techniques for emotion recognition. Section 9 summarizes future challenges of emotion recognition using physiological signals and finally we conclude the paper in Section 10.

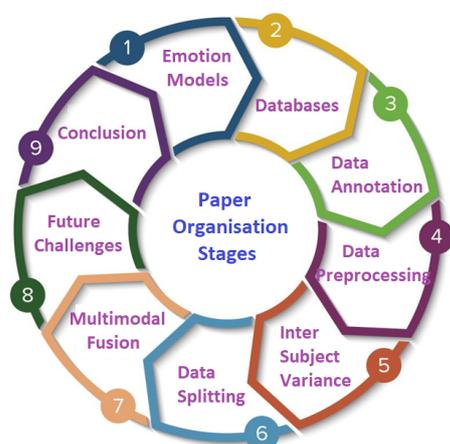The organisation of the paper is also shown in Figure 1.

Figure 1. Paper organisation stages.

2. Emotion Models

Over the past few decades, different emotion models have been proposed by researchers based on the quantitative analysis and the emergence of new emotion categories. Based on recent research, emotional states can be represented with two models: discrete and multidimensional.

2.1. Discrete or Categorical Models

Discrete models are the most commonly used models because they contain a list of distinct emotion classes that are easy to recognize and understand. Ekman [38] and Plutchik [39] are amongst those scientists that present the concept of discrete emotion states. Ekman briefed that there are six basic emotions—happy, sad, anger, fear, surprise, and disgust—and all the other emotions are derived from these six basic emotions. Plutchik presented a famous wheel model to describe eight discrete emotions. These emotions are joy, trust, fear, surprise, sadness, disgust, anger and anticipation, as shown in Figure 2. The model describes the relations between emotion concepts, which are analogous to the colors on a color wheel. The cone's vertical dimension represents intensity, and the circle represents degrees of similarity among the emotions. The eight sectors are designed to indicate that there are eight primary emotion dimensions and rest of the motion dimensions are derived from them.
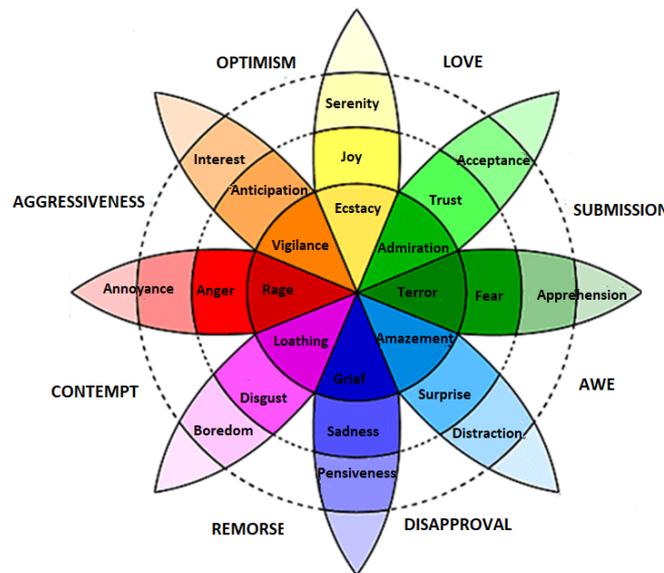
Figure 2. Plutchik wheel of discrete emotions [39].

2.2. Continuous or Multidimensional Models

All emotional states listed in discrete sets of emotions cannot be confined by a single word and they need a range of intensities for description. For instance, a person may feel less or more excited, or become less or more afraid in response to a particular stimulus. Thus, to cover the range of intensities in emotions, multidimensional models such as 2D and 3D are proposed. Amongst 2D models, the model presented in [40] is the famous model that describes the emotion along the dimensions of High Arousal (HA) and Low Arousal (LA) and High Valence (HV) and Low Valence (LV) as shown in Figure 3. The 2D model classifies emotions based on two dimensional data consisting of valence and arousal value. On the other hand, the 3D model deals with valence, arousal and dominance. Valence indicates the level of pleasure, Arousal indicates the level of excitation and Dominance indicates the level of controlling or dominating emotion. The 3D model is shown in Figure 4.
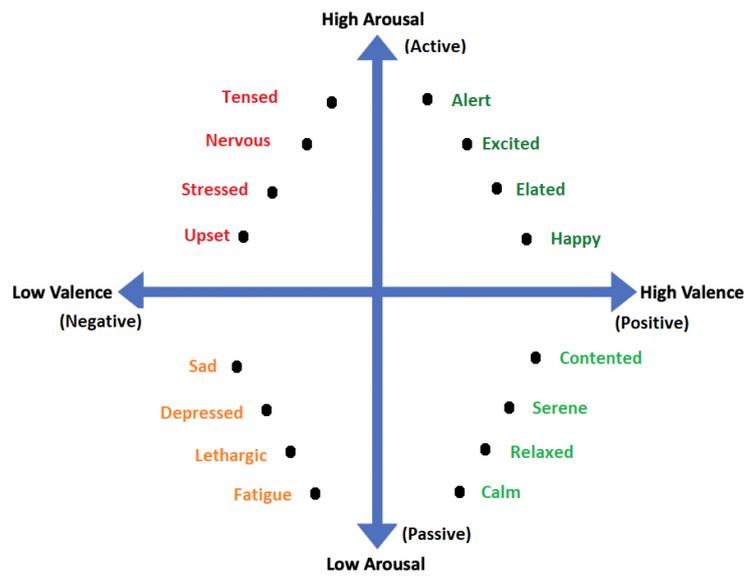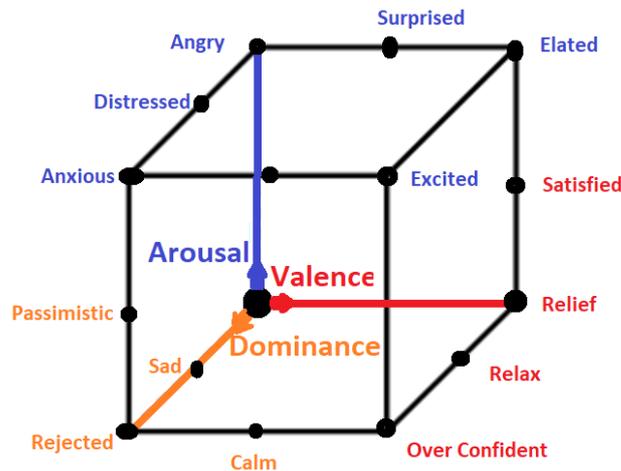
Figure 3. 2D Valence-arousal Model.

Figure 4. 3D Emotion Model.

3. Databases

Several datasets for physiological-based emotion recognition are publicly available, such as AMIGOS [41], ASCERTAIN [42], BIO VID EMO DB [43], DEAP [44], DREAMER [45], MAHNOB-HCI [46], MPED [47], SEED [48] as shown in Table 1. All datasets, other than SEED, are multimodal and provide more than one modality of physiological signals. As can be seen, some of these datasets utilize the continuous emotion model, while the others use the discrete model. Most of these datasets are unfortunately limited to a small number of subjects due to the elaborate process of data collection. Some details of these datasets such as annotation and pre-processing are discussed in the following sections.

Table 1. Publicly available datasets in alphabetical order for physiological signals-based emotion recognition.

Database	Modalities	Stimulai	Subjects	Sampling Rate	Emotion States
AMIGOS [41]	EEG, ECG, EDA	Video clips	40	Not Provided	Anger, Fear, Sadness, Disgust, Neutrality, Surprise, Happiness

Table 1. *Cont.*

Database	Modalities	Stimulai	Subjects	Sampling Rate	Emotion States
ASCERTAIN [42]	EEG, ECG, EDA	Movie clips	58	EDA at 100 Hz EEG at 32 Hz	Valence (–3–3), Arousal (0–6)
BIO-VID-EMO DB [43]	ECG, EMG,SC	Film clips	86	512 Hz	Valence, Arousal, Amusement, Sadness, Anger, Disgust, Fear
DEAP [44]	EEG, EDA, EMG, PPG, EOG, RSP	Music Videos	32	512 Hz	Valence (1–9), Arousal (1–9), Dominance (1–9), Liking (1–9), Familiarity (1–5)
DREAMER [45]	EEG, ECG	Film clips	23	Not Provided	Anger, Fear, Sadness, Disgust, Calmness, Surprise, Amusement, Happiness, Excitement
MAHNOB-HCI [46]	EEG, ECG, EDA, RSP, SKT	Video clips	27	256 Hz	Anger, Anxiety, Fear, Sadness, Disgust, Neutrality, Surprise, Amusement, Joy
MPED [47]	EEG, ECG, EDA, RSP	Video clips	23	1000 Hz	Anger, Fear, Sadness, Disgust. Neutrality, Funny, Joy
SEED [48]	EEG	Film clips	15	1000 Hz	Negativity, Neutrality, Positivity

4. Data Annotation

Data annotation is amongst the most important steps after data acquisition and eventually for emotion recognition. However, this subject is not addressed in detail in existing literature. For better discussion on this matter, we divide the data annotation procedure into discrete and continuous data annotation.

In discrete annotation, participants are given questionnaire after the stimuli and they are asked to rate their feelings between some scores ranging from 1 to 5 or from 0 to 9 and then the final annotations are decided as shown in Figure 5.

(a) Stimuli

Please rate your emotion category according to the stimulai which you just saw.

weak   strong

Positive ○ ○ ○ ○ ○ ○

Negative ○ ○ ○ ○ ○ ○

(b) Questionaire and self report

Angry

Sad

.
.
.

Happy
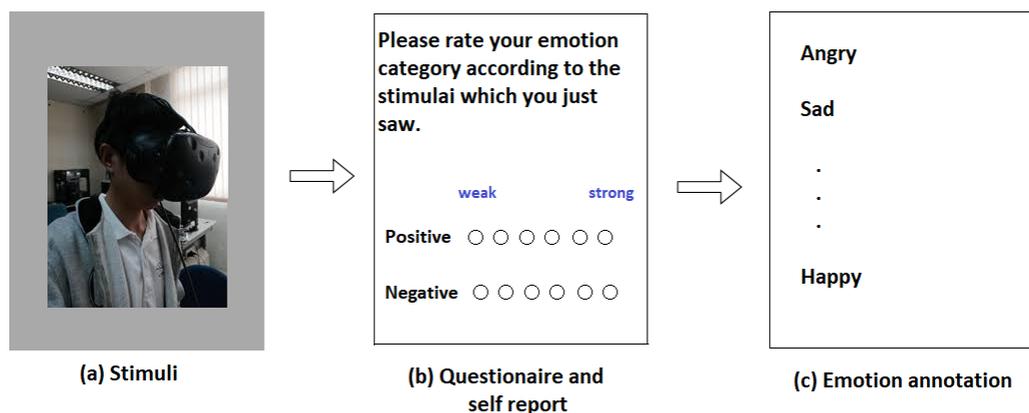
(c) Emotion annotation

Figure 5. Steps involving discrete annotation.

On the other hand, in continuous annotations, the participants are required to continuously annotate the data in real time using some human computer interface (HCI) mechanism as shown in Figure 6.

Figure 6. Continuous annotation using HCI mechanism [49].

In Sections 4.1 and 4.2, we provide a review on discrete and continuous annotation.

4.1. Discrete Annotation

In [47], three psychological questionnaires—PANAS [50], SAM [51] and DES [52]—were used for data annotation of seven emotion categories such as joy, funny, anger, sad, disgust, fear and neutrality. Besides questionnaires, a T-test was also conducted to evaluate the effectiveness of the selected emotion categories.

In [42,53], participants were briefed on details of the experiments including the principal experimental tasks and later on the self-report questionnaires were filled by the participants according to their experiences during stimuli. These questionnaires were finally used for data annotation. In [22], for automatic ECG-based emotion recognition, music is used as stimuli and in the self-assessment stages, the participants were asked to indicate how strongly they had experienced the corresponding feeling during stimulus presentation and then participants gave marks ranging from 1 to 5 for each of the nine emotion categories.

In [54], the questionnaire, divided into three parts, one for each activity, in each of which the participants had to select the emotions they felt before, during and after the activity, was given to the participants. Based on the participants marking, 28 emotions are arranged on a 2-dimensional plane with Valence and Arousal at each axis as shown in Figure 3, according to the model proposed in [55]. In [56], the data were collected in a VR environment. The data annotation was based on the subjects' responses to the two questions regarding level of excitement and pleasantness which were graphed on a 9 square grid, a discretized version of the arousal-valence model [57]. Each grid shows how many subjects reported feeling the corresponding level of "pleasantness" and "excitement" for each VR session. The valence scale was measured as "Unpleasant", "Neutral", or "Pleasant" (from left to right) and the arousal scale was measured as "Relaxing", "Neutral", or "Exiting" (from bottom to top). The results are very asymmetrical with the majority of the sessions rated as "Exciting" and "Pleasant".

4.2. Continuous Annotation

In [49], participants viewed a 360° videos as stimuli and they rated emotional states (valence and arousal) continuously using the joystick. It is observed that collecting continuous annotations can be used to evaluate the performance of fine-grained emotion recognition algorithms such as weakly supervised learning or regression. In [58], the authors claim that the lack of continuous annotations is the reason why they failed to validate their weakly-supervised algorithm for fine-grained emotion recognition since

the continuous labels allow the algorithms to learn the precise mappings between the dynamic emotional changes and input signals. Moreover, if only discrete annotations are available, ML algorithms can overfit because the discrete labels represent only the most salient or recent emotion rather than the dynamic emotional changes that may occur within stimuli [59]. A software for continuous emotion annotation is introduced in [60] and is called EMuJoy (Emotion measurement with Music by using a Joystick). It is better than Schubert's 2DES software [61]. It helps subjects to generate self reports for different media in real time with a computer mouse, joystick, or any other human–computer interface. Using a joystick has an advantage, because joysticks have a return spring to automatically realign them in the middle of the space. This contrasts with Schubert's software, where mouse is used. In [62], stimuli videos were also continuously annotated along valence and arousal dimensions. Long-short-term-memory recurrent neural networks (LSTM-RNN) and continuous conditional random fields (CCRF) were utilized in detecting emotions automatically and continuously. Furthermore, the effect of lag on continuous emotion detection is also studied. The database is annotated using a joystick and delays from 250 ms up to 4 s in annotations. It is observed that 250 ms increased the detection performance whereas longer delays deteriorated it. This shows the advantage of using joystick over mouse. The authors of [63] analyzed the effect of lag on continuous emotion detection on SEMAINE database [64]. They found a delay of 2 s will improve their emotion detection results. SEMAINE database is annotated by Feeltrace [65] using a mouse as annotation interface. Thus, the observations in [63] also show that the response time of joystick is less than mouse. Another software named DARMA is introduced in [66] for continuous measurement system that synchronizes media playback and the continuous recording of two-dimensional measurements. These measurements can be observational or self-reported and are provided in real-time through the manipulation of a computer joystick.

Touch events are dynamic and can be bidirectional, i.e., 2D along valence-arousal plane. During continuous annotation, participants are instructed to annotate their emotion experience by moving the joystick head into one of the four quadrants. The movement of the joystick head maps the emotions into a 2D valence-arousal plane, in which the x axis indicates valence while the y axis indicates arousal. Since the nature of emotions is time-varying, during annotation, participants could lose the control over the speed, i.e., the movement of a joystick which could lead collection of less precise ground truth labels. Thus, the training of participants on HCI mechanism and study of lag for continuous emotion detection are two important factors for precise continuous annotation.

4.3. Summary of the Section

There is no definite or clear-cut advantage of discrete annotation on continuous annotation and vice versa. However, the choice of annotation method depends on the application. For example, in biofeedback systems, participants learn how to control physical and psychological effects of stress using feed back signals. Biofeedback is a mind–body technique that involves using visual or auditory feedback to teach people to recognize the physical signs and symptoms of stress and anxiety, such as increased heart rate, body temperature and muscle tension. Since stress level changes during biofeedback mechanism, real time continuous annotation is more important in this case [67]. The choice of annotation also depends on length of the data, the number of emotion categories, number of participants, feature extraction method and the choice of classifier.

5. Data Pre-Processing

Physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG) and galvanic skin response (GSR) are time-series and delicate signals in the raw form. During acquisition, these physiological signals are contaminated by numerous factors such as line interference of 50 Hz or 60 Hz, electromagnetic interference, noise and baseline drifts and different artifacts due to body movements and different responses of participants to different stimuli. The bandwidths of the physiological signals are different. For instance,

EEG signals lie in the range of 0.5 Hz to 35 Hz, ECG signals are intensive in the range of 0 to 40 Hz and GSR signals are mainly concentrated in the range of 0 to 2 Hz. Since the strength of physiological signals is rich in different frequency ranges, different physiological signals required different pre-processing techniques. In this section, we review the pre-processing techniques used for physiological signals.

5.1. EEG Pre-Processing

In addition to noise, the EEG signals are affected by Electrooculography (EOG) with frequency below 4 Hz. This artifact is caused by the facial muscle movement or eye movement [68,69]. Different methods that have been used for EEG pre-processing or cleaning include the rejection method, linear filtering, statistical methods such as linear regression and Independent Component Analysis (ICA).

5.1.1. Rejection Method

The rejection method involves both manual and automatic rejection of epochs contaminated with artifacts. Manual rejection needs less computations but it is more laborious than automatic rejection. In automatic rejection pre-determined threshold is used to remove the artifact-contaminated trials from the data. The common disadvantages associated with the rejection method are the sampling bias [70] and loss of valuable information [71].

5.1.2. Linear Filtering

Linear filtering is simple and easy to apply and is beneficial mostly when artifacts located in certain frequency bands do not overlap with the signal of interest. For example, low-pass filtering can be used to remove EMG artifacts and high-pass filtering can be used to remove EOG artifacts [48]. However, linear filtering flops when EEG signal and the EMG or EOG artifacts lie in the same frequency band.

5.1.3. Linear Regression

Linear regression using least square method has been used for EEG signal processing for removing EOG-based artifacts. In linear regression, using least square criteria, residual is calculated by subtracting the EOG signal from the EEG signal and then this residual is optimized to get the cleaned EEG signal [72]. This method is not suitable for EMG-based artifact removal because EMG data are collected from multiple muscles groups and therefore EMG data has no single reference.

5.1.4. Independent Component Analysis

ICA is amongst those methods that do not require reference artifacts for cleaning the EEG signal and thus making the components independent [73]. The disadvantage associated with ICA is that it requires prior visual inspection to identify the artifact part.

5.2. ECG Pre-Processing

ECG signal is usually corrupted by various noise sources and artifacts. These noises and artifacts deteriorate the signal quality and affect the detection of QRS.

5.2.1. Filtering

Filtering is the simplest method of removing noise and artifacts from the ECG signal and to improve the signal to noise ratio (SNR) of the QRS complex. In ECG signal, the three critical frequency regions include the very low frequency band below 0.04 Hz, the low frequency band (0.04 to 0.15 Hz), and the high frequency band (0.15 to 0.5 Hz) [74]. Thus the commonly used filters for ECG are high pass filter, low pass filter and band pass filter.

High-pass filters allow only higher frequencies to pass through them. They are used to suppress low-frequency components the ECG signal. Low frequency components include motion artifact, respiratory variations and baseline wander. In [75], high pass filters with cut-off frequency of 0.5 Hz was selected to waive off the baseline wander.

Low-pass filters are usually employed to eliminate high-frequency components from the ECG signal. These components include noise, muscle artifacts and powerline interference [76].

Since bandpass filters remove most of the unwanted components from the ECG signal, it is extensively used to pre-process ECG. The structure of Band-pass filter comprises of both a high-pass and low-pass filter. In [77], bandpass filter is used to remove muscle noise and baseline wander from ECG.

5.2.2. Normalization

In ECG, the amplitude of QRS complex amplitude increase from birth to adolescence and then begins to decrease afterward [78]. Due to this, ECG signals suffer from inter-subject variability. To overcome this problem, amplitude normalization is performed. In [79], a method of normalizing ECG signals to a standard heart rate to lower the false rate detection, was introduced. The commonly used normalization techniques are min-max normalization [80], maximum division normalization [81] and Z-score normalization [82].

5.3. GSR Pre-Processing

There is still no standard methods for GSR pre-processing. In this section, we are discussing a few methods to deal with noise and artifacts present in raw GSR signal.

5.3.1. Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is the technique that best addresses the non-linear and nonstationary nature of the signal while removing noise and artifact and was introduced in [83] as a tool to adaptively decompose a signal into a collection of AM–FM components. The EMD relies on a fully data-driven mechanism that does not require any a priori known basis, like Fourier and wavelet-based mechanisms. EMD decomposes the signal into a sum of intrinsic mode functions (IMFs). An IMF is defined as a function with equal number of extrema and zero crossings (or at most differed by one) with its envelopes, as defined by all the local maxima and minima, being symmetric with respect to zero.

After extraction of IMFs from a time series signal the residue tends to become a monotonic function, such that no more IMFs can be extracted. Finally, after the iterative process, the input signal is decomposed into a sum of IMF functions and a residue.

In [84], an algorithm, modified from EMD, called Empirical iterative algorithm is proposed for denoising the GSR from motion artifacts and quantization noise as shown in Figure 7 [84]. The algorithm does not rely on the shifting process of EMD, and provides the filtered signal directly as an output of each iteration.
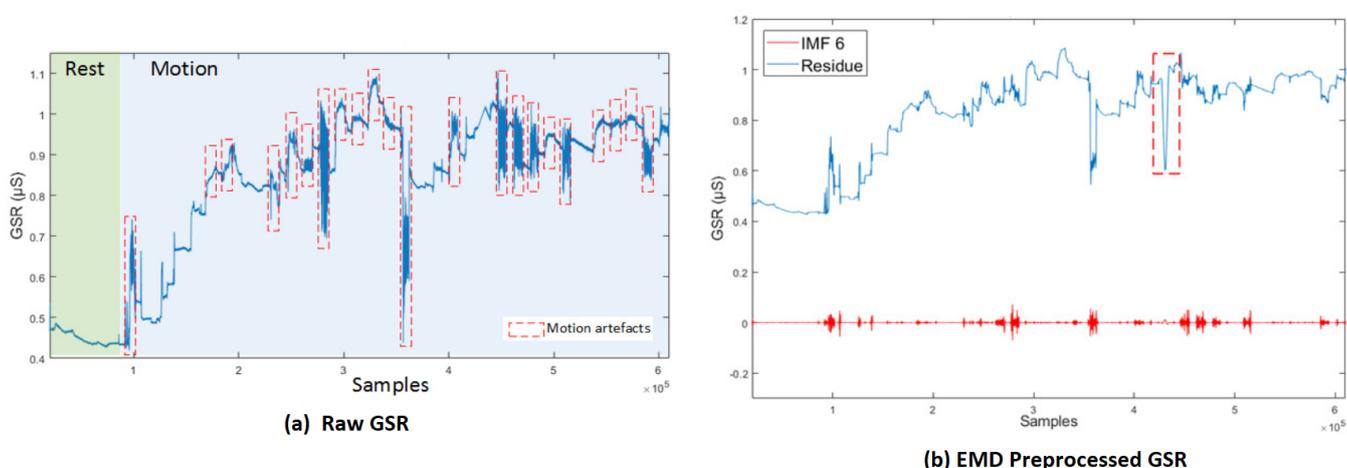

Figure 7. (**a**) Raw GSR: rest and motion phases. Signals corresponding to the movements involving the right hand are delimited by red lines. (**b**) GSR decomposition based on EMD, IMF6 and its respective residue [84].

5.3.2. Kalman Filtering

Low pass and moving average filters have been used for GSR pre-processing [29]; however, the Kalman filter is the better choice. Kalman filter is a model-based filter and a state estimator. It employs a mathematical model of the process producing the signal to be filtered [85]. In [86], an extended Kalman filter is used to remove noise and artifact from GSR signal in real time. The Kalman filter used in this article is comparable to a third order Butterworth low pass filter with a similar frequency response. However, the Kalman filter is more robust than its counterpart Butterworth low pass filter while suppressing the noise and artifacts.

5.4. 1D to 2D Conversion

Physiological signals such as EEG, ECG and GSR are 1D signal in raw form. In [87,88], raw ECG signal was converted into 2D form, i.e., into three statistical images namely Gramian Angular Field (GAF) images, Recurrence Plot (RP) images and Markov Transition Field (MTF) images as shown in Figure 8. Experimental results show the superiority of 1D to 2D pre-processing. In [89], ECG and GSR are converted in 2D scalogram for better emotion recognition. Furthermore, in [90], ECG and GSR signals are converted in 2D RP images for improved emotion recognition as compared to 1D form signals.

Thus, the transformation from 1D to 2D has been shown to be an important pre-processing step for physiological signals.
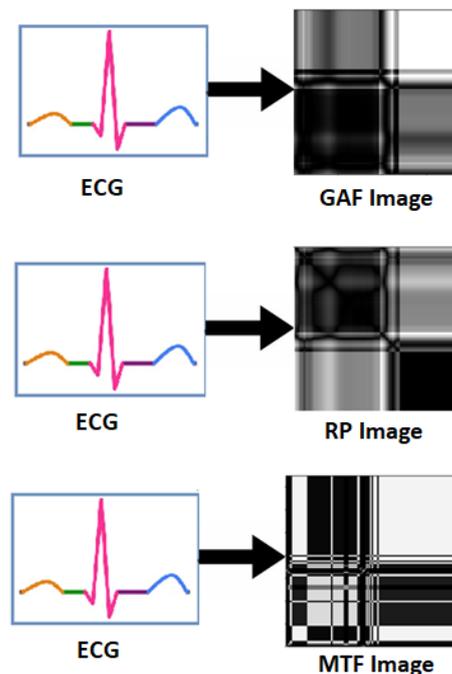

Figure 8. Transformation of ECG signal into GAF, RP and MTF Images.

5.5. Summary of the Section

The pre-processing for the EEG data is mainly to remove EOG artifacts with a frequency less than 4 Hz that caused by eye blink, EMG artifacts with a frequency more than 30 Hz, power frequency artifacts in the environment with a frequency between 50 to 60 Hz and so on. Since the noise and artifacts present at different frequency ranges, therefore linear filtering, linear regression and ICA are more useful pre-processing methods for EEG. The other methods such as EMD decomposes the signal into a sum of intrinsic mode functions (IMFs). This may supress the information from EEG data. Similarly, the limitations of Kalman filtering are that it is like a third order Butterworth low pass filter and thus cannot remove noise from all the frequency bands of EEG.

6. Inter-Subject Data Variance

In this section, we will review the effect of inter-subject data variance on emotion recognition. We will also discuss the reasons of inter-subject data variance and recommend solutions for the problem.

There exists few papers [42,53,91–94], which only reported the problem of inter-subject data variance and its effect on degradation of emotion recognition accuracy, but they did not shed light on the possible reasons behind inter-subject data variability.

To derive our point home, we performed experiments on ECG data of two datasets, WESAD dataset [95] and our own Ryerson Multimedia Research Laboratory (RML) dataset [96]. For WESAD dataset, we use three categories. These categories are Amusement, Baseline and Stress. For RML dataset, three categories of stress are low stress, medium stress and high stress. The issue of inter-subject data variability was particularly apparent in the RML dataset, since we controlled the data collection process.

To observe the high inter-subject variability, we utilize leave-one-subject-out cross-validation (LOSOCV) for all our experiments, where the models are trained with data from all but one subject, and tested on the held out subject. The average results across all subjects are presented in Tables 2–5. Tables 2 and 4 show the results of experiments conducted with 1D ECG for both datasets. To see the affect of data pre-processing on inter-subject data variability, we transform the raw ECG data into spectrograms. ResNet-18 [97] was trained on these spectrograms and results are shown in Tables 3 and 5.

Table 2. Accuracy, precision, recall and F_1 score using 1D raw WESAD data with 1D CNN.

Testing Sub	Accuracy	Precision	Recall	F_1 Score
2	44.2	47.3	44.2	40.8
3	47.5	47.9	47.5	45.6
4	46.8	55.4	46.8	43.8
5	51.5	51.3	51.5	49
6	46	45.7	46	43.9
7	51.3	47	51.3	46
8	51	51.4	51	50.6
9	42.8	41.5	42.8	41
10	57.6	56.1	57.6	53.4
11	56.1	57.5	56.1	57.4
13	52.1	51.6	52.1	50.6
14	61.6	60	61.6	55.4
15	52.9	52.1	52.9	50.6
16	56.4	52.5	56.4	52
17	53.1	50.7	53.1	50.9
Average	**51.4**	**51.2**	**51.4**	**48.7**

Table 3. Accuracy, precision, recall and F_1 score using 2D spectrograms of WESAD data with ResNet-18.

Testing Sub	Accuracy	Precision	Recall	F_1 Score
2	75.1	75.3	75.1	75.2
3	73.3	75.6	73.3	73.8
4	73.4	76.7	73.4	73.8
5	73.8	77.4	73.8	74.4
6	65.3	67.7	65.3	65.7

Table 3. *Cont.*

Testing Sub	Accuracy	Precision	Recall	F_1 Score
7	71.3	76	71.3	72
8	66.8	70.4	66.8	67.7
9	67.9	72.9	67.9	68.4
10	69.5	71.8	69.5	70
11	75.3	78.4	75.3	75.7
13	68.8	73	68.8	69.3
14	78.1	80.1	78.1	78.4
15	67.8	70.6	67.8	68.7
16	75.8	79	75.8	76.1
17	78.1	81.6	78.1	78.2
Average	**72**	**75.1**	**72**	**72.5**

Table 4. Accuracy, precision, recall and F_1 score using 1D raw RML data with 1D CNN.

Testing Sub	Accuracy	Precision	Recall	F_1 Score
2	58.9	60.9	58.9	58.6
3	65.4	66.4	65.4	65.4
4	59.3	59.4	59.3	58.9
5	64.7	64.2	64.7	63.2
11	45.3	48.8	45.3	38
12	69.6	72.6	69.6	69.1
13	70.6	76.3	70.6	70.3
14	66.7	69	66.7	67.2
16	64.8	67.2	64.8	64.1
Average	**62.8**	**62.5**	**62.8**	**61.6**

In Tables 2–5, we observe the inter-subject data variance across all the metrics, i.e., accuracy, precision, recall and F_1 score. We also observe that by transforming 1D data to 2D using spectrograms have increased the accuracy of the affective state but did not reduce the data variance significantly.

Table 5. Accuracy, precision, recall and F_1 score using 2D spectrograms of RML data with ResNet-18.

Testing Sub	Accuracy	Precision	Recall	F_1 Score
2	72.1	74.87	72.1	72.25
3	71.3	75.6	71.25	71.17
4	60.7	58.65	60.7	56.06
5	62.5	63.3	62.5	62.54
11	49.3	37.6	49.3	39.35
12	61.11	63.48	61.11	61.25
13	76.1	83.7	76.15	76.22
14	64	66.13	64	63.59
16	70.3	73.28	70.3	70.8
Average	**65.34**	**66.2**	**65.34**	**63.7**

Looking at the WESAD results, we can see that the average accuracy is only 72%. This is significantly lower than the average accuracy achieved if we would use k-fold validation. However, using k-fold cross validation means that a segment of data for each subject was always present in the training set. This is an unrealistic scenario for practical applications. Therefore, dealing with inter-subject variability, conducting experiments in a LOSOCV setting is very important. We elaborate on this point further in Section 7.

6.1. Statistical Analysis

We perform statistical analysis by plotting error-bars on 1D and 2D WESAD data to notice the variability in accuracy of each individual subject across the overall mean value of accuracy as shown in Figures 9 and 10. The error-bars show the spread or variability in the data. Figures 9 and 10 can also be used to detect the outlier subjects from the dataset that could be removed later on for reducing the variability amongst the remaining subjects.

6.2. Reasons of Inter-Subject Data Variance

The following reasons could explain the inter-subject variability:

- Some users feel uncomfortable when wearing a sensor on their bodies and behave unnaturally and unexpectedly during data collection.
- Body movements during data collection adversely effect the quality of recorded data. For instance, head and eyeball movement during EEG data collection and hand movement during EMG data collection are the sources of degradation of these data.
- An intra-variance subject is also caused due to the placement of electrodes on different areas of the body. For example, during chest-based ECG collection, electrodes are placed on different areas of the chest and during collection of EEG data, the electrodes are placed on all over the scalp, i.e., from left side to right side through the center of the head.
- Different users respond differently to the same stimuli. This difference in their response depends on their mental strength and physical stability. For instance, a movie clip can instill fear in some users while other users may feel excitement while watching it.
- The length of the collected data is also amongst the major reasons of creating subject diversity. A long video is more likely to elicit diverse emotional states [98], while the data with limited samples will create problem during the training of the model and classification of the affective states.
- Experimental environment and quality of sensors/equipment are also the reasons of variance in data. For instance, uncalibrated sensors may create errors at different stages of data collection.

6.3. Reducing Inter-Subject Data Variance

The following actions could be taken to remove the inter-subject variability:

- The collected sample size for each subject should be same. The length of the data should be set carefully to avoid variance in the data.
- The users or subjects should be trained to remain calm and behave normally during data collection or wearing a sensor.
- Different channels of the data collecting equipment should be calibrated against the same standards so that the intra-variance in the data for the subject can be reduced.
- In [99], statistics-based unsupervised domain adaptation algorithm was proposed for reducing the inter-subject variance. A new baseline model was introduced to tackle the classification task, and then two novel domain adaptation objective functions, cluster-aligning loss and cluster-separating loss were presented to mitigate the negative impacts caused by domain shifts.
- A state-of-the-art attention-based parallelizable deep learning model called transformer was introduced in [100] for language modelling. This model uses multihead attention to focus more on relevant features for improved language modelling and

translational task. A similar kinds of attention-based deep learning models should be developed for physiological signals for reducing intersubject variability.
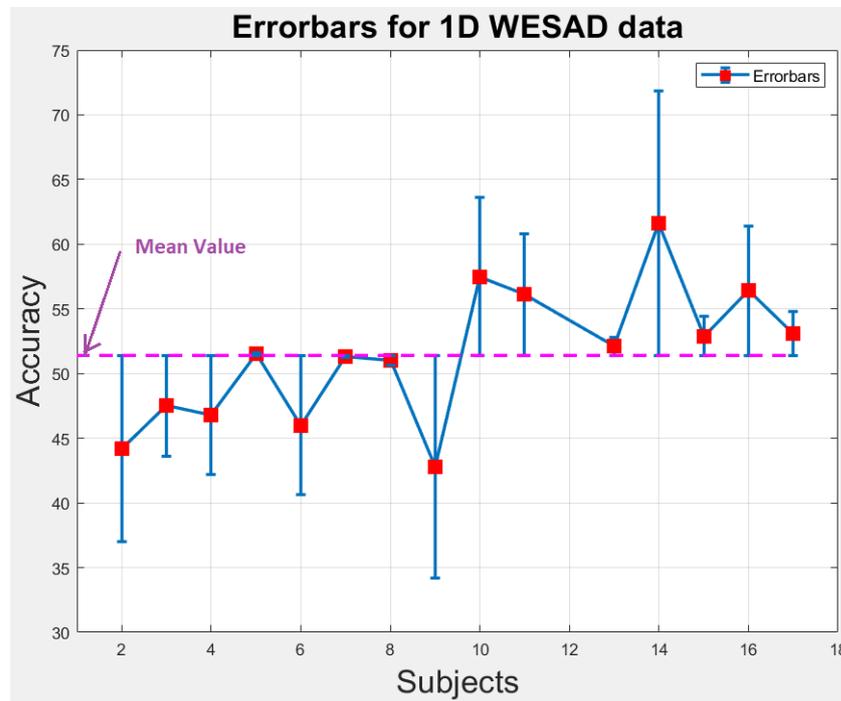

Figure 9. Error-bars showing inter-subject variability in terms of accuracy across the mean value for 1D WESAD data.
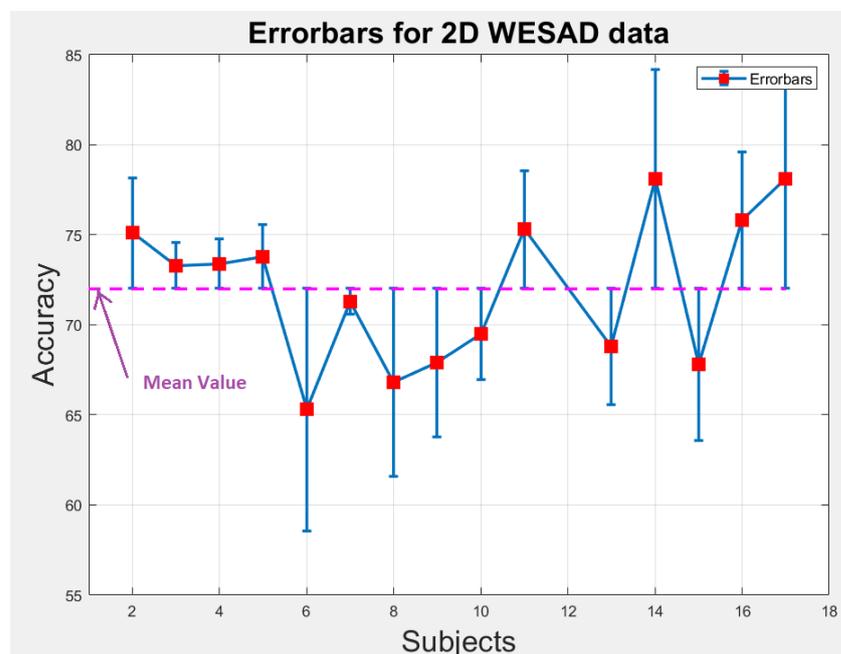

Figure 10. Error-bars showing inter-subject variability in terms of accuracy across the mean value for 2D WESAD data.

6.4. Summary of the Section

The following points are the key findings of the section:

1. Inter-subject data variance depends on the behavior of the user during data collection, calibration of the sensors and experimental environment.

2. By performing experiments, we prove that inter-subject data variance could be reduced by data pre-processing, i.e., by transforming the raw ECG data (1D) into spectrograms (2D). Transforming data from 1D to 2D enables ResNet-18 to extract those relevant features that are more specific to emotions.
3. We believe more research in the area of domain adaptation and developing attention-based models could help tackle inter-subject data variability.

7. Data Splitting

There are two major types of data splitting techniques that are being used for emotion recognition using physiological signals. These two methods of splitting are called subject-independent and subject-dependent.

In the subject-independent case, the model is trained from data of various subjects and then tested on data of new subjects (that are not included in the training part). In the subject-dependent case, the training and testing part is randomly distributed among all the subjects.

7.1. Subject-Independent Splitting

The most commonly used subject-independent data splitting method is leave-one-subject-out (LOSO) cross-validation. In LOSO, the model is trained on (k −1) subjects and leave one subject for testing and this process is repeated for every subject of the dataset.

In [101,102], LOSO cross validation technique is used for DEAP dataset containing 32 subjects. Thus, the experiment is repeated for every subject and the final performance is considered as the average of all experiments. In [103], leave-one-subject out method is conducted to evaluate the performance, and the final performance is calculated by the averaging method. In [104], EEG-based emotion classification was conducted by a leave-one-subject out validation (LOSOV). One subject among the 20 subjects is used for testing, while the others were used for the training set. The experiments were repeated for all subjects and the mean and standard deviation of the accuracies were used to evaluate the model performance.

7.2. Subject-Dependent Splitting

In subject-dependent splitting, various strategies are used. In one of the strategies, training and testing parts are taken from a single subject with a particular splitting percentage. In another setting, few trials of same subject are taken as a training part and rest of them are used as testing. Furthermore, in another method of subject-dependent splitting, testing and training parts are randomly selected across all the subjects either by particular percentage or by k-fold cross validation. For k = 5, the concept of cross validation is shown in Figure 11. The final performance is the average of all the five evaluations.

In [47], physiological signals of 7 trials out of 28 were selected as testing for each subject and rest of the trials are used for training. This was repeated for every subject. In [105], models are evaluated using 10-fold cross validation performed on each subject. This was repeated on all 32 subjects of DEAP datasets and the final evaluation about accuracies and F1 scores are based on averaging method. In [106], physiological samples of all subjects of DEAP dataset are used to evaluate the model. The 10-fold cross validation technique is used to analyze the emotion recognition performance. All samples are randomly divided into ten subsets. One subset is regarded as a test set, and another subset as a validation set. The remaining subsets are regarded as a training set. The above process is repeated ten times until all subsets are tested.

7.3. Summary of the Section

Subject-dependent classification is usually performed to investigate the individual variability between subjects on emotion recognition from their physiological signals.
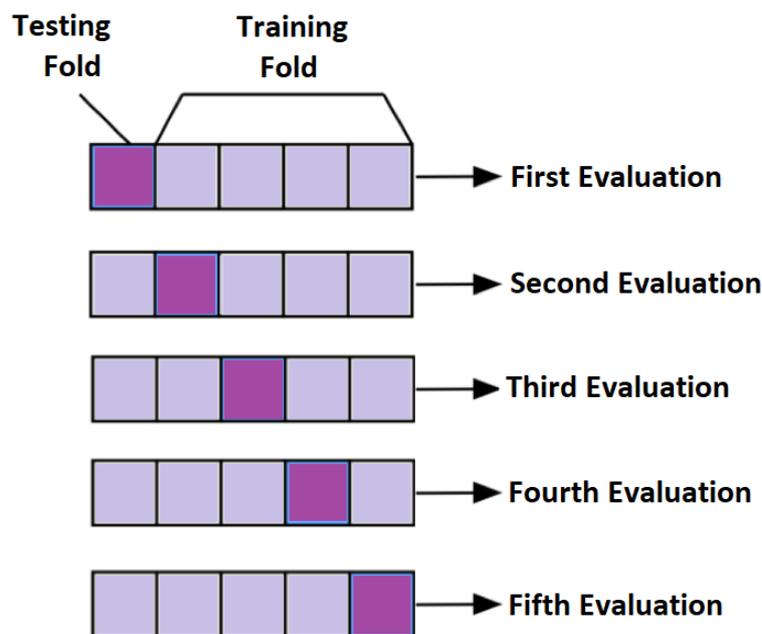
Figure 11. Conceptual Visualization of 5-fold cross validation.

In real application scenarios, emotion recognition system with subject-independent classification is considered more practical as the system has to predict emotions of each subject or patient separately when implemented for emotion recognition task.

In [94], comparison between subject-dependent and subject-independent setting is provided. it is observed that subject-independent splitting shows 3% lower accuracy because of high inter-subject variability. However, the generalized capability of subject-independent model on unseen data is more as these models learn inter-subject variability.

Thus, it is more valuable to develop a good subject-independent emotional recognition model so that they can generalize well on unseen data or new patients.

8. Multimodal Fusion

The purpose of multimodal fusion is to obtain complementary information from different physiological modalities to improve the emotion recognition performance. Fusing different modalities alleviates the weaknesses of individual modalities by integrating complementary information from the modalities.

In [107], new emotional discriminative space is constructed utilizing discriminative canonical correlation analysis (DCCA) is used from physiological signals with the assistance of EEG signals. Finally, machine learning techniques are utilized to build emotion recognizer. The authors of [108] trained feedforward neural networks using both the fused and non-fused signals. Experiments show that the fused method outperforms each individual signal across all emotions tested. In [109], feature level fusion is performed between the features from two modalities, i.e., facial expression images and EEG signals. A feature map of size 128×128 is constructed by concatenating the facial images and their corresponding EEG feature maps. Emotion recognition model is finally trained on these feature maps. Experimental results show the improved performance of multimodal features over single modality features. In [110], hybrid fusion of face data, EEG and GSR is performed. First, features from EEG and GSR modalities are integrated for estimating the level of arousal. Final fusion was the late fusion of EEG, GSR and face modalities.

The authors of [111] presented a new emotion recognition framework based on decision fusion, where three separate classifiers were trained on ECG, EMG and skin conductance level (SCL). The majority voting principle was used to determine a final classification result on the three outputs of the separate classifiers.

In [112], multidomain features such as features from time domain, frequency domain and wavelet domain are fused with different combinations to identify stable features which would best represent the underlying EEG signal characteristics and performed better classification. Multidomain feature fusion is performed using concatenation to obtained final feature vector. In [113], weighted average fusion of physiological signal was conducted to classify valence and arousal. Overall performance was quantified as the percentage of correctly classified emotional states per video. It is also observed that the classification performance obtained was slightly better for arousal than valence. The authors of [114] provide the study about combining the EEG data and audio-visual features of the video. Then, PCA algorithm was applied to reduce the feature dimensions. In [115], feature level fusion of sensor data from EEG and EDA sensors was performed for human emotion detection. Nine features (delta, theta, low alpha, high alpha, low beta, high beta, low gamma, high gamma and galvanic skin response) are selected for training a neural network model. The classification performance of each modality was also evaluated separately. It is observed that multimodal emotion recognition is better than using a single modality. In [116], the authors developed two late multi-modal data fusion methods with deep CNN models to estimate positive and negative affects (PA and NA) scores and explored the effect of the two late fusion methods and different combination of the physiological signals on the performance. In the first late fusion method, a separate CNN is trained for each modality and in the second late fusion method, the average of the three classes' (baseline, stress and amusement) probabilities across the pretrained CNNs is calculated, and the emotion class with the highest average probability is selected. In the case of the PA or NA estimation, the average of the estimated scores across the pretrained CNNs is provided as the estimated affect score. In [96], multidomain fusion of ECG signal is performed for multilevel stress assessment. First, the ECG signal was converted into image and then images are made multimodal and multidomain by transforming them into frequency and time-spectrum domain using Discrete Fourier Transform (DFT) and Gabor wavelet transform (GWT) respectively. Features in different domains are extracted by CNNs and then decision level fusion is performed for improving the classification accuracy as shown in Figure 12.
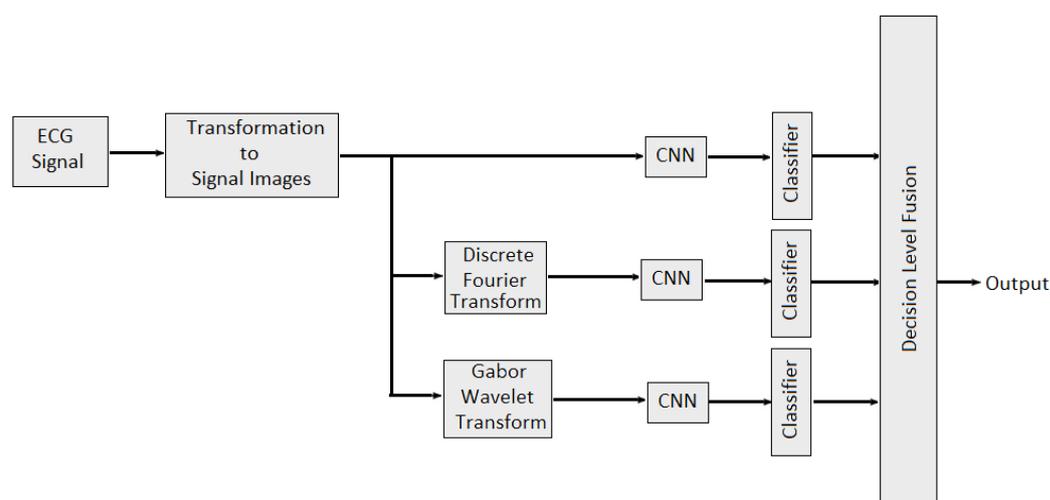

Figure 12. Multimodal Multidomain Fusion of ECG Signal.

Summary of the Section

The above review shows that the fusion of the physiological modalities exhibits better performance than the single modality in terms of classification accuracy for both arousal and valence dimensions. Furthermore, it is observed from the review that the two major fusion methods practiced for physiological data features are feature fusion and decision fusion; however, feature fusion is adopted by the researcher more than decision level. The greatest advantage of feature level fusion is that it utilizes the correlation among the

modalities at an early stage. Furthermore, only one classifier is required to perform a task, making the training process less tedious. The disadvantage with decision level fusion is the use of more than one classifier. This makes the task time consuming.

9. Future Challenges

The workflow for emotion recognition consists of many steps such as data acquisition, data annotation, data pre-processing, feature extraction and selection and recognition. The problems and challenges are in each step of emotion recognition and are explained in this section.

9.1. Data Acquisition

One of the significant challenges in acquiring physiological signals is to deal with noise, baseline drifts, different artifacts due to body movements, different responses of participants to different stimuli and low graded signal. The data acquiring devices carry noise which corrupts the signal and induced artifacts that superimposed on the signal. The second challenge is the setting of stable and noiseless lab and selection of stimuli so that genuine emotions are induced which are closed to real world feelings. Another challenge in acquiring a high quality data is the various responses of participants or subjects to the same stimuli. This causes large inter-subject variability in the data and is also the reason that most of the available emotion recognition datasets are of short duration.

To overcome the above challenges, well-designed labs and proper selection of stimuli are necessary. Furthermore, subjects must be properly trained to avoid large variance in the data and possibility of gathering datasets with long duration and less inter-subject variability.

9.2. Data Annotation

High volume research has been conducted on emotion recognition, even then there is no uniform standards to annotate data. Discrete and continuous data annotation techniques are commonly used; however, due to unavailability of uniform annotation standard, there is no compatibility between the datasets. Furthermore, all emotions can neither be listed as discrete nor continuous as different emotions need different range of intensities for description. The problem with existing datasets is that some of them are annotated with four emotions, some are with basic six emotions or with eight emotions. Due to this different emotion labels, datasets are not compatible and emotion recognition algorithms do not work equally well on these datasets.

To face the aforementioned challenges of data annotation, hybrid data annotation technique should be introduced where each emotion is labelled according to its range of energy. However, care must be taken while selecting new standards of data annotation because hybrid annotation technique may lead to imbalance data.

9.3. Feature Extraction and Fusion

Features for emotion recognition are categorized into time domain features and transform domain features. Time domain features mostly include statistical features of first and higher orders. Transform domain features are further divided into frequency domain and time-frequency domain such as feature extracted from wavelet transform. Still, there is no set of features that guarantees to work for all models and situations.

One of the solution of the above problem is to design adaptive intelligent systems that can automatically select the best features for the classifier model. Furthermore, another solution is to perform multimodal fusion. In Section 8, it is explained that commonly practiced fusion techniques for emotion recognition are feature level and decision level fusion. However, for feature level fusion, concatenation is mostly used and for decision level fusion, majority voting is mostly practiced. Thus, a need for more improved and intelligent-based fusion methods is arising because the simple feature level and decision level fusion cannot counter the non-stationary and nonlinear nature of features.

9.4. Generalization of Models

One of the biggest challenges in emotion recognition is to design the models that can generalize well on unseen data or new datasets. The main obstacles are the limited samples in the datasets, non-standard data splitting techniques and large inter-subject data variability. The commonly used data splitting techniques are subject-dependent and subject-independent techniques. It is explained in Section 7 that the models trained on subject-independent settings are more likely to generalize well. However, existing subject-independent recognition models are not intelligent enough to perform well in realistic and real-time applications.

Since, while testing, an emotion recognition model has to face an unseen subject, one solution is to train and validate the model on large number of subjects so that it can generalize well on testing. The problem of a limited dataset can be solved by carefully applying data augmentation techniques because that techniques could lead to data imbalance and overfitting of the model.

9.5. Modern Machine Learning Methods

Advanced machine learning tools need to be developed to mitigate the challenges of emotion recognition using physiological signals. For instance, one of the modern techniques which is being used is transfer learning. In transfer learning, the knowledge learned while solving one task can be applied for different but related work. For example, the rich features learned by a machine learning model while conducting emotion recognition based on EEG data from a large dataset could also be applicable for EEG data of another dataset. This can easily solve the problem of a small dataset where fewer data samples are available. Transfer learning methods are getting success but still lot of research work is required to be conducted for physiological signal-based transfer learning methods for emotion recognition.

10. Conclusions

In this paper, we provide a review on the physiological signal-based emotion recognition. Existing reviews on the physiological signal-based emotion recognition presented only the generic steps of emotion recognition such as combined techniques for data pre-processing, feature extraction and selection methods, selection of machine learning techniques and classifiers, but did not elaborate on the most important factors that are crucial for the performance of emotion recognition systems and their generalization. These important factors include the challenges during data annotation, specific data pre-processing techniques for each physiological signal, effect of inter-subject data variance, data splitting methods and multimodal fusion. Thus, in this paper, we address these all challenging factors to bridge the gap in the existing literature. In this research, we provide comprehensive review on these factors and report our key findings about each factor. We also discuss the future challenges about physiological signal-based emotion recognition based on this research.

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

References

1. Phillips, L.H.; MacLean, R.D.; Allen, R. Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* **2002**, *57*, P526–P530. [CrossRef] [PubMed]
2. Giordano, M.; Tirelli, P.; Ciarambino, T.; Gambardella, A.; Ferrara, N.; Signoriello, G.; Paolisso, G.; Varricchio, M. Screening of Depressive Symptoms in Young–Old Hemodialysis Patients: Relationship between Beck Depression Inventory and 15-Item Geriatric Depression Scale. *Nephron Clin. Pract.* **2007**, *106*, c187–c192. [CrossRef] [PubMed]
3. Schluter, P.J.; Généreux, M.; Hung, K.K.; Landaverde, E.; Law, R.P.; Mok, C.P.Y.; Murray, V.; O'Sullivan, T.; Qadar, Z.; Roy, M. Patterns of suicide ideation across eight countries in four continents during the COVID-19 pandemic era: Repeated cross-sectional study. *JMIR Public Health Surveill.* **2022**, *8*, e32140. [CrossRef] [PubMed]
4. De Nadai, S.; D'Incà, M.; Parodi, F.; Benza, M.; Trotta, A.; Zero, E.; Zero, L.; Sacile, R. Enhancing safety of transport by road by on-line monitoring of driver emotions. In Proceedings of the 2016 11th System of Systems Engineering Conference (SoSE), Kongsberg, Norway, 12–16 June 2016; pp. 1–4.
5. Ertin, E.; Stohs, N.; Kumar, S.; Raij, A.; Al'Absi, M.; Shah, S. AutoSense: Unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, Seattle, WA, USA, 1–4 November 2011; pp. 274–287.
6. Kołakowska, A. Towards detecting programmers' stress on the basis of keystroke dynamics. In Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), Gdańsk, Poland, 11–14 September 2016; pp. 1621–1626.
7. Kołakowska, A.; Landowska, A.; Szwoch, M.; Szwoch, W.; Wróbel, M.R. Emotion recognition and its application in software engineering. In Proceedings of the 2013 6th International Conference on Human System Interactions (HSI), Piscataway, NJ, USA, 6–8 June 2013; IEEE: Manhattan, NY, USA 2013; pp. 532–539.
8. Szwoch, M.; Szwoch, W. Using Different Information Channels for Affect-Aware Video Games-A Case Study. In *Proceedings of the International Conference on Image Processing and Communications*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 104–113.
9. Muhammad, G.; Hossain, M.S. Emotion recognition for cognitive edge computing using deep learning. *IEEE Internet Things J.* **2021**, *8*, 16894–16901. [CrossRef]
10. Siddharth, S.; Jung, T.P.; Sejnowski, T.J. Impact of affective multimedia content on the electroencephalogram and facial expressions. *Sci. Rep.* **2019**, *9*, 16295. [CrossRef] [PubMed]
11. Wang, S.H.; Phillips, P.; Dong, Z.C.; Zhang, Y.D. Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. *Neurocomputing* **2018**, *272*, 668–676. [CrossRef]
12. Yang, D.; Alsadoon, A.; Prasad, P.C.; Singh, A.K.; Elchouemi, A. An emotion recognition model based on facial recognition in virtual learning environment. *Procedia Comput. Sci.* **2018**, *125*, 2–10. [CrossRef]
13. Zhang, Y.D.; Yang, Z.J.; Lu, H.M.; Zhou, X.X.; Phillips, P.; Liu, Q.M.; Wang, S.H. Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access* **2016**, *4*, 8375–8385. [CrossRef]
14. Gunes, H.; Shan, C.; Chen, S.; Tian, Y. Bodily expression for automatic affect recognition. In *Emotion Recognition: A Pattern Analysis Approach*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2015; pp. 343–377.
15. Piana, S.; Staglianò, A.; Odone, F.; Camurri, A. Adaptive body gesture representation for automatic emotion recognition. *ACM Trans. Interact. Intell. Syst. TiiS* **2016**, *6*, 1–31. [CrossRef]
16. Noroozi, F.; Corneanu, C.A.; Kamińska, D.; Sapiński, T.; Escalera, S.; Anbarjafari, G. Survey on emotional body gesture recognition. *IEEE Trans. Affect. Comput.* **2018**, *12*, 505–523. [CrossRef]
17. Zheng, H.; Yang, Y. An improved speech emotion recognition algorithm based on deep belief network. In Proceedings of the 2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, 12–14 July 2019; IEEE: Manhattan, NY, USA 2019; pp. 493–497.
18. Latif, S.; Rana, R.; Khalifa, S.; Jurdak, R.; Qadir, J.; Schuller, B.W. Survey of deep representation learning for speech emotion recognition. *IEEE Trans. Affect. Comput.* **2021**. [CrossRef]
19. Akçay, M.B.; Oğuz, K. Speech emotion recognition: Emotional models, databases, features, pre-processing methods, supporting modalities, and classifiers. *Speech Commun.* **2020**, *116*, 56–76. [CrossRef]
20. Li, W.; Zhang, Z.; Song, A. Physiological-signal-based emotion recognition: An odyssey from methodology to philosophy. *Measurement* **2021**, *172*, 108747. [CrossRef]
21. Suhaimi, N.S.; Mountstephens, J.; Teo, J. EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Comput. Intell. Neurosci.* **2020**, *2020*, 8875426. [CrossRef] [PubMed]
22. Hsu, Y.L.; Wang, J.S.; Chiang, W.C.; Hung, C.H. Automatic ECG-based emotion recognition in music listening. *IEEE Trans. Affect. Comput.* **2017**, *11*, 85–99. [CrossRef]
23. Sarkar, P.; Etemad, A. Self-supervised learning for ecg-based emotion recognition. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 3217–3221.
24. Zhang, Q.; Lai, X.; Liu, G. Emotion recognition of GSR based on an improved quantum neural network. In Proceedings of the 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 27–28 August 2016; Volume 1, pp. 488–492.
25. Pentland, A.; Heibeck, T. *Honest Signals*; MIT Press: Cambridge, MA, USA, 2008.

26. Shu, L.; Xie, J.; Yang, M.; Li, Z.; Li, Z.; Liao, D.; Xu, X.; Yang, X. A review of emotion recognition using physiological signals. *Sensors* **2018**, *18*, 2074. [CrossRef]

27. Wijasena, H.Z.; Ferdiana, R.; Wibirama, S. A Survey of Emotion Recognition using Physiological Signal in Wearable Devices. In Proceedings of the 2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), Delft, The Netherlands, 12–16 July 2021; pp. 1–6.

28. Joy, E.; Joseph, R.B.; Lakshmi, M.; Joseph, W.; Rajeswari, M. Recent survey on emotion recognition using physiological signals. In Proceedings of the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 19–20 March 2021; Volume 1, pp. 1858–1863.

29. Fan, X.; Yan, Y.; Wang, X.; Yan, H.; Li, Y.; Xie, L.; Yin, E. Emotion Recognition Measurement based on Physiological Signals. In Proceedings of the 2020 13th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 12–13 December 2020; pp. 81–86.

30. Bota, P.J.; Wang, C.; Fred, A.L.; Da Silva, H.P. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE Access* **2019**, *7*, 140990–141020. [CrossRef]

31. Dalvi, C.; Rathod, M.; Patil, S.; Gite, S.; Kotecha, K. A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions. *IEEE Access* **2021**, *9*, 165806–165840.

32. Pepa, L.; Spalazzi, L.; Capecci, M.; Ceravolo, M.G. Automatic emotion recognition in clinical scenario: A systematic review of methods. *IEEE Trans. Affect. Comput.* **2021**. [CrossRef]

33. Saganowski, S.; Perz, B.; Polak, A.; Kazienko, P. Emotion Recognition for Everyday Life Using Physiological Signals from Wearables: A Systematic Literature Review. *IEEE Trans. Affect. Comput.* **2022**, *1*, 1–21. [CrossRef]

34. Li, X.; Zhang, Y.; Tiwari, P.; Song, D.; Hu, B.; Yang, M.; Zhao, Z.; Kumar, N.; Marttinen, P. EEG based Emotion Recognition: A Tutorial and Review. *ACM Comput. Surv. CSUR* **2022.** doi: 10.1145/3524499. [CrossRef]

35. Danala, G.; Maryada, S.K.; Islam, W.; Faiz, R.; Jones, M.; Qiu, Y.; Zheng, B. Comparison of Computer-Aided Diagnosis Schemes Optimized Using Radiomics and Deep Transfer Learning Methods. *Bioengineering* **2022**, *9*, 256. [CrossRef] [PubMed]

36. Ponsiglione, A.M.; Amato, F.; Romano, M. Multiparametric investigation of dynamics in fetal heart rate signals. *Bioengineering* **2021**, *9*, 8. [CrossRef] [PubMed]

37. Bizzego, A.; Gabrieli, G.; Esposito, G. Deep neural networks and transfer learning on a multivariate physiological signal Dataset. *Bioengineering* **2021**, *8*, 35. [CrossRef]

38. Ekman, P. An argument for basic emotions. *Cogn. Emot.* **1992**, *6*, 169–200. [CrossRef]

39. Plutchik, R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am. Sci.* **2001**, *89*, 344–350. [CrossRef]

40. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [CrossRef]

41. 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.* **2018**, *12*, 479–493. [CrossRef]

42. 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.* **2016**, *9*, 147–160. [CrossRef]

43. Zhang, L.; Walter, S.; Ma, X.; Werner, P.; Al-Hamadi, A.; Traue, H.C.; Gruss, S. "BioVid Emo DB": A multimodal database for emotion analyses validated by subjective ratings. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; pp. 1–6.

44. 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.* **2011**, *3*, 18–31. [CrossRef]

45. Katsigiannis, S.; Ramzan, N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 98–107. [CrossRef] [PubMed]

46. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* **2011**, *3*, 42–55. [CrossRef]

47. Song, T.; Zheng, W.; Lu, C.; Zong, Y.; Zhang, X.; Cui, Z. MPED: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access* **2019**, *7*, 12177–12191. [CrossRef]

48. Zheng, W.L.; Lu, B.L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [CrossRef]

49. Xue, T.; El Ali, A.; Zhang, T.; Ding, G.; Cesar, P. CEAP-360VR: A Continuous Physiological and Behavioral Emotion Annotation Dataset for 360 VR Videos. *IEEE Trans. Multimed.* **2021**, *14*. [CrossRef]

50. Watson, D.; Clark, L.A.; Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personal. Soc. Psychol.* **1988**, *54*, 1063. [CrossRef]

51. Bradley, M.M.; Lang, P.J. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **1994**, *25*, 49–59. [CrossRef]

52. Gross, J.J.; Levenson, R.W. Emotion elicitation using films. *Cogn. Emot.* **1995**, *9*, 87–108. [CrossRef]

53. Yang, K.; Wang, C.; Gu, Y.; Sarsenbayeva, Z.; Tag, B.; Dingler, T.; Wadley, G.; Goncalves, J. Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition. In *IEEE Trans. Affect. Comput.* **2021**. [CrossRef]

54. Althobaiti, T.; Katsigiannis, S.; West, D.; Ramzan, N. Examining human-horse interaction by means of affect recognition via physiological signals. *IEEE Access* **2019**, *7*, 77857–77867. [CrossRef]

55. Russell, J.A.; Mehrabian, A. Evidence for a three-factor theory of emotions. *J. Res. Personal.* **1977**, *11*, 273–294. [CrossRef]

56. Hinkle, L.B.; Roudposhti, K.K.; Metsis, V. Physiological measurement for emotion recognition in virtual reality. In Proceedings of the 2019 2nd International Conference on Data Intelligence and Security (ICDIS), South Padre Island, TX, USA, 28–30 June 2019; pp. 136–143.

57. Posner, J.; Russell, J.A.; Peterson, B.S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **2005**, *17*, 715–734. [CrossRef] [PubMed]

58. Romeo, L.; Cavallo, A.; Pepa, L.; Berthouze, N.; Pontil, M. Multiple instance learning for emotion recognition using physiological signals. *IEEE Trans. Affect. Comput.* **2019**, *13*, 389–407. [CrossRef]

59. Fredrickson, B.L.; Kahneman, D. Duration neglect in retrospective evaluations of affective episodes. *J. Personal. Soc. Psychol.* **1993**, *65*, 45. [CrossRef]

60. Nagel, F.; Kopiez, R.; Grewe, O.; Altenmüller, E. EMuJoy: Software for continuous measurement of perceived emotions in music. *Behav. Res. Methods* **2007**, *39*, 283–290. [CrossRef]

61. Schubert, E. Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Aust. J. Psychol.* **1999**, *51*, 154–165. [CrossRef]

62. Soleymani, M.; Asghari-Esfeden, S.; Fu, Y.; Pantic, M. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans. Affect. Comput.* **2015**, *7*, 17–28. [CrossRef]

63. Mariooryad, S.; Busso, C. Analysis and compensation of the reaction lag of evaluators in continuous emotional annotations. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; pp. 85–90.

64. McKeown, G.; Valstar, M.; Cowie, R.; Pantic, M.; Schroder, M. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Trans. Affect. Comput.* **2011**, *3*, 5–17. [CrossRef]

65. Cowie, R.; Douglas-Cowie, E.; Savvidou, S.; McMahon, E.; Sawey, M.; Schröder, M. 'FEELTRACE': An instrument for recording perceived emotion in real time. In Proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, Newcastle, UK, 5–7 September 2000.

66. Girard, J.M.; C Wright, A.G. DARMA: Software for dual axis rating and media annotation. *Behav. Res. Methods* **2018**, *50*, 902–909. [CrossRef]

67. Xu, W.; Chen, Y.; Sundaram, H.; Rikakis, T. Multimodal archiving, real-time annotation and information visualization in a biofeedback system for stroke patient rehabilitation. In Proceedings of the 3rd ACM workshop on Continuous Archival and Retrival of Personal Experences, New York, NY, USA, 28 October 2006; pp. 3–12.

68. Muthukumaraswamy, S. High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Front. Hum. Neurosci.* **2013**, *7*, 138. [CrossRef]

69. Fatourechi, M.; Bashashati, A.; Ward, R.K.; Birch, G.E. EMG and EOG artifacts in brain computer interface systems: A survey. *Clin. Neurophysiol.* **2007**, *118*, 480–494. [CrossRef] [PubMed]

70. Gratton, G. Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behav. Res. Methods Instrum. Comput.* **1998**, *30*, 44–53. [CrossRef]

71. Ramoser, H.; Muller-Gerking, J.; Pfurtscheller, G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 441–446. [CrossRef] [PubMed]

72. Croft, R.J.; Chandler, J.S.; Barry, R.J.; Cooper, N.R.; Clarke, A.R. EOG correction: A comparison of four methods. *Psychophysiology* **2005**, *42*, 16–24. [CrossRef]

73. Bigirimana, A.D.; Siddique, N.; Coyle, D. A hybrid ICA-wavelet transform for automated artefact removal in EEG-based emotion recognition. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 004429–004434.

74. Venkatesan, C.; Karthigaikumar, P.; Paul, A.; Satheeskumaran, S.; Kumar, R. ECG signal pre-processing and SVM classifier-based abnormality detection in remote healthcare applications. *IEEE Access* **2018**, *6*, 9767–9773. [CrossRef]

75. Alcaraz, R.; Sandberg, F.; Sörnmo, L.; Rieta, J.J. Classification of paroxysmal and persistent atrial fibrillation in ambulatory ECG recordings. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 1441–1449. [CrossRef]

76. Patro, K.K.; Jaya Prakash, A.; Jayamanmadha Rao, M.; Rajesh Kumar, P. An efficient optimized feature selection with machine learning approach for ECG biometric recognition. *IETE J. Res.* **2022**, *68*, 2743–2754. [CrossRef]

77. Cordeiro, R.; Gajaria, D.; Limaye, A.; Adegbija, T.; Karimian, N.; Tehranipoor, F. Ecg-based authentication using timing-aware domain-specific architecture. *IEEE Trans.-Comput.-Aided Des. Integr. Syst.* **2020**, *39*, 3373–3384. [CrossRef]

78. Surawicz, B.; Knilans, T. *Chou's Electrocardiography in Clinical Practice: Adult and Pediatric*; Elsevier Health Sciences: Amsterdam, The Netherlands, 2008.

79. Saechia, S.; Koseeyaporn, J.; Wardkein, P. Human identification system based ECG signal. In Proceedings of the TENCON 2005—2005 IEEE Region 10 Conference, Melbourne, Australia, 21–24 November 2005; pp. 1–4.

80. Wei, J.J.; Chang, C.J.; Chou, N.K.; Jan, G.J. ECG data compression using truncated singular value decomposition. *IEEE Trans. Inf. Technol. Biomed.* **2001**, *5*, 290–299.

81. Tawfik, M.M.; Selim, H.; Kamal, T. Human identification using time normalized QT signal and the QRS complex of the ECG. In Proceedings of the 2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010), Newcastle Upon Tyne, UK, 21–23 July 2010; pp. 755–759.

82. Odinaka, I.; Lai, P.H.; Kaplan, A.D.; O'Sullivan, J.A.; Sirevaag, E.J.; Kristjansson, S.D.; Sheffield, A.K.; Rohrbaugh, J.W. ECG biometrics: A robust short-time frequency analysis. In Proceedings of the 2010 IEEE International Workshop on Information Forensics and Security, Seattle, WA, USA, 12–15 December 2010; pp. 1–6.

83. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]

84. Gautam, A.; Sim oes-Capela, N.; Schiavone, G.; Acharyya, A.; De Raedt, W.; Van Hoof, C. A data driven empirical iterative algorithm for GSR signal pre-processing. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 1162–1166.

85. Haug, A.J. *Bayesian Estimation and Tracking: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2012.

86. Tronstad, C.; Staal, O.M.; Sælid, S.; Martinsen, Ø.G. Model-based filtering for artifact and noise suppression with state estimation for electrodermal activity measurements in real time. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 2750–2753.

87. Ahmad, Z.; Tabassum, A.; Guan, L.; Khan, N. Ecg heart-beat classification using multimodal image fusion. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1330–1334.

88. Ahmad, Z.; Tabassum, A.; Guan, L.; Khan, N.M. ECG heartbeat classification using multimodal fusion. *IEEE Access* **2021**, *9*, 100615–100626. [CrossRef]

89. Rahim, A.; Sagheer, A.; Nadeem, K.; Dar, M.N.; Rahim, A.; Akram, U. Emotion Charting Using Real-time Monitoring of Physiological Signals. In Proceedings of the 2019 International Conference on Robotics and Automation in Industry (ICRAI), Rawalpindi, Pakistan, 21–22 October 2019; pp. 1–5.

90. Elalamy, R.; Fanourakis, M.; Chanel, G. Multi-modal emotion recognition using recurrence plots and transfer learning on physiological signals. In Proceedings of the 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII), Nara, Japan, 28 September–1 October 2021; pp. 1–7.

91. Gupta, V.; Chopda, M.D.; Pachori, R.B. Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals. *IEEE Sens. J.* **2018**, *19*, 2266–2274. [CrossRef]

92. Yao, H.; He, H.; Wang, S.; Xie, Z. EEG-based emotion recognition using multi-scale window deep forest. In Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, China, 6–9 December 2019; pp. 381–386.

93. Wickramasuriya, D.S.; Tessmer, M.K.; Faghih, R.T. Facial expression-based emotion classification using electrocardiogram and respiration signals. In Proceedings of the 2019 IEEE Healthcare Innovations and Point of Care Technologies,(HI-POCT), Bethesda, MD, USA, 20–22 November 2019; pp. 9–12.

94. Kim, S.; Yang, H.J.; Nguyen, N.A.T.; Prabhakar, S.K.; Lee, S.W. Wedea: A new eeg-based framework for emotion recognition. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 264–275. [CrossRef] [PubMed]

95. Schmidt, P.; Reiss, A.; Duerichen, R.; Marberger, C.; Van Laerhoven, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM International Conference on Multimodal Interaction, Boulder, CO, USA, 16–20 October 2018; pp. 400–408.

96. Ahmad, Z.; Khan, N.M. Multi-level stress assessment using multi-domain fusion of ECG signal. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 4518–4521.

97. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

98. Santamaria-Granados, L.; Munoz-Organero, M.; Ramirez-Gonzalez, G.; Abdulhay, E.; Arunkumar, N. Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS). *IEEE Access* **2018**, *7*, 57–67. [CrossRef]

99. Chen, M.; Wang, G.; Ding, Z.; Li, J.; Yang, H. Unsupervised domain adaptation for ECG arrhythmia classification. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 304–307.

100. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]

101. Zhang, X.; Liu, J.; Shen, J.; Li, S.; Hou, K.; Hu, B.; Gao, J.; Zhang, T. Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine. *IEEE Trans. Cybern.* **2020**, *51*, 4386–4399. [CrossRef]

102. Ullah, H.; Uzair, M.; Mahmood, A.; Ullah, M.; Khan, S.D.; Cheikh, F.A. Internal emotion classification using EEG signal with sparse discriminative ensemble. *IEEE Access* **2019**, *7*, 40144–40153. [CrossRef]

103. Chang, E.J.; Rahimi, A.; Benini, L.; Wu, A.Y.A. Hyperdimensional computing-based multimodality emotion recognition with physiological signals. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), Hsinchu, Taiwan, 18–20 March 2019; pp. 137–141.

104. Yang, C.J.; Fahier, N.; He, C.Y.; Li, W.C.; Fang, W.C. An ai-edge platform with multimodal wearable physiological signals monitoring sensors for affective computing applications. In Proceedings of the 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Seville, Spain, 12–14 October 2020; pp. 1–5.

105. Vijayakumar, S.; Flynn, R.; Murray, N. A comparative study of machine learning techniques for emotion recognition from peripheral physiological signals. In Proceedings of the 2020 31st Irish Signals and Systems Conference (ISSC), Letterkenny, Ireland, 11–12 June 2020; pp. 1–6.
106. Chao, H.; Dong, L. Emotion recognition using three-dimensional feature and convolutional neural network from multichannel EEG signals. *IEEE Sens. J.* **2020**, *21*, 2024–2034. [CrossRef]
107. Zhao, W.; Zhao, Z.; Li, C. Discriminative-CCA promoted By EEG signals for physiological-based emotion recognition. In Proceedings of the 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), Beijing, China, 20–22 May 2018; pp. 1–6.
108. Fabiano, D.; Canavan, S. Emotion recognition using fused physiological signals. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, UK, 3–6 September 2019; pp. 42–48.
109. Li, X.; Zhang, X.; Yang, H.; Duan, W.; Dai, W.; Yin, L. An EEG-based multi-modal emotion database with both posed and authentic facial actions for emotion analysis. In Proceedings of the 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Buenos Aires, Argentina, 16–20 November 2020; pp. 336–343.
110. Cimtay, Y.; Ekmekcioglu, E.; Caglar-Ozhan, S. Cross-subject multimodal emotion recognition based on hybrid fusion. *IEEE Access* **2020**, *8*, 168865–168878. [CrossRef]
111. Xie, J.; Xu, X.; Shu, L. WT feature based emotion recognition from multi-channel physiological signals with decision fusion. In Proceedings of the 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), Beijing, China, 20–22 May 2018; pp. 1–6.
112. Khateeb, M.; Anwar, S.M.; Alnowami, M. Multi-domain feature fusion for emotion classification using DEAP dataset. *IEEE Access* **2021**, *9*, 12134–12142. [CrossRef]
113. Pinto, J.; Fred, A.; da Silva, H.P. Biosignal-based multimodal emotion recognition in a valence-arousal affective framework applied to immersive video visualization. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 3577–3583.
114. Xing, B.; Zhang, H.; Zhang, K.; Zhang, L.; Wu, X.; Shi, X.; Yu, S.; Zhang, S. Exploiting EEG signals and audiovisual feature fusion for video emotion recognition. *IEEE Access* **2019**, *7*, 59844–59861. [CrossRef]
115. Yasemin, M.; Sarıkaya, M.A.; Ince, G. Emotional state estimation using sensor fusion of EEG and EDA. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 5609–5612.
116. Hssayeni, M.D.; Ghoraani, B. Multi-modal physiological data fusion for affect estimation using deep learning. *IEEE Access* **2021**, *9*, 21642–21652. [CrossRef]